Genetic mapping of the rat *agu* gene.

by

María Beatriz Durán Alonso.

Thesis submitted to the Faculty of Science, University of Glasgow, for the degree of Doctor of Philosophy.

Department of Molecular Genetics.

September 1997.

The research reported in this thesis is my own original work except were otherwise stated and has not been submitted for any other degree.

María Beatriz Durán Alonso, September 1997.

To my grandfather,

the most special person in the world.

Contents.

CHAPTER 2: Materials and Methods.

CHAPTER 3: Establishment of a rat genetic genome scan.

CHAPTER 4: Localisation of the *agu* locus to a single chromosomal region.

CHAPTER 5: Development of new genetic markers.

CHAPTER 6: Precise mapping of the *agu* locus.

CHAPTER 7: Reassignment of the microsatellite locus *D1Mgh18* to rat chromosome 3.

CHAPTER 8: General discussion.

List of Figures and Tables.

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

Chapter 5.

Chapter 6.

Chapter 7.

Chapter 8.

Appendix A.

Abbreviations.

ATP	adenosine triphosphate.
BAC	bacterial artificial chromosome.
bp	base pair.
BSA	bovine serum albumin.
^0C	degrees centigrade.
CEPH	Centre Etude Polymorphisme Humain.
Ci	curie.
cM	centimorgan.
cpm	counts per minute.
DGGE	denaturing gradient gel electrophoresis.
dH_2O	distilled water.
dNTP	deoxyribonucleotide triphosphate.
DTT	dithiothreitol.
EDTA	ethylene diamine tetra acetate.
g	gramme.
GABA	gamma amino butyric acid.
GDRDA	genetically directed representational difference analysis.
GPi	globus pallidus internal segment.
GPe	globus pallidus external segment.
GSH	reduced glutathione.
IPTG	isopropyl β-D-thiogalactopyranoside.
kb	kilo base pairs.
MAO-B	monoamine oxidase B.
Mb	megabase.
mg	milligramme.
µCi	microcurie.
µl	microlitre.
mmol	millimol.
MPP^+	N-methyl-4-phenylpyridinium.
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine.
ng	nanogramme.
6-OHDA	6-hydroxydopamine.

PAC	P1-derived artificial chromosome.
PCR	polymerase chain reaction.
PD	Parkinson's disease.
PEG	polyethylene glycol.
pmol	picomol.
RAPD	random amplified polymorphic DNA.
RDA	representational difference analysis.
RFLP	restriction fragment length polymorphism.
RNase	ribonuclease.
SAP	shrimp alkaline phosphatase.
SD	standard deviation.
SDP	strain distribution pattern.
SDS	sodium dodecyl sulphate.
SN	substantia nigra.
SNc	substantia nigra *pars compacta.*
SNr	substantia nigra *pars reticulata.*
SOD	superoxide dismutase.
SSC	standard saline citrate.
SSCP	single-strand conformation polymorphism.
SSLP	simple sequence length polymorphism.
STN	subthalamic nucleus.
Taq	*Thermus aquaticus.*
TBE	Tris-borate/EDTA buffer.
TE	Tris-EDTA.
TEMED	N,N,N',N'-tetramethylethylene diamine.
Tris	Tris (hydroxymethyl) aminomethane.
U	units of enzyme activity.
UV	ultraviolet light.
V	volts.
VTA	ventral tegmental area.
W	Watts.
X-gal	5-bromo-4-chloro-3-indolyl-β-galactopyranoside.
YAC	yeast artificial chromosome.

Abbreviations.

AS	Albino Swiss.
AS/AGU	Albino Swiss/Anatomy Glasgow University.
ATP	adenosine triphosphate.
BAC	bacterial artificial chromosome.
BN	Brown Norway.
bp	base pair.
BSA	bovine serum albumin.
^0C	degrees centigrade.
CEPH	Centre Etude Polymorphisme Humain.
Ci	curie.
cM	centimorgan.
cpm	counts per minute.
DGGE	denaturing gradient gel electrophoresis.
dH_2O	distilled water.
dNTP	deoxyribonucleotide triphosphate.
DTT	dithiothreitol.
EDTA	ethylene diamine tetra acetate.
F344	Fischer 344.
g	gramme.
GABA	gamma amino butyric acid.
GDRDA	genetically directed representational difference analysis.
GPi	globus pallidus internal segment.
GPe	globus pallidus external segment.
GSH	reduced glutathione.
IPTG	isopropyl β-D-thiogalactopyranoside.
kb	kilo base pairs.
MAO-B	monoamine oxidase B.
Mb	megabase.
mg	milligramme.
µCi	microcurie.
µl	microlitre.
mmol	millimol.

MPP^+	N-methyl-4-phenylpyridinium.
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine.
ng	nanogramme.
6-OHDA	6-hydroxydopamine.
PAC	P1-derived artificial chromosome.
PCR	polymerase chain reaction.
PD	Parkinson's disease.
PEG	polyethylene glycol.
pmol	picomol.
PVG	black hooded strain of rat.
RAPD	random amplified polymorphic DNA.
RDA	representational difference analysis.
RFLP	restriction fragment length polymorphism.
RNase	ribonuclease.
SAP	shrimp alkaline phosphatase.
SD	standard deviation.
SDP	strain distribution pattern.
SDS	sodium dodecyl sulphate.
SN	substantia nigra.
SNc	substantia nigra *pars compacta.*
SNr	substantia nigra *pars reticulata.*
SOD	superoxide dismutase.
SSC	standard saline citrate.
SSCP	single-strand conformation polymorphism.
SSLP	simple sequence length polymorphism.
STN	subthalamic nucleus.
Taq	*Thermus aquaticus.*
TBE	Tris-borate/EDTA buffer.
TE	Tris-EDTA.
TEMED	N,N,N',N'-tetramethylethylene diamine.
Tris	Tris (hydroxymethyl) aminomethane.
U	units of enzyme activity.
UV	ultraviolet light.

V	volts.
VTA	ventral tegmental area.
W	Watts.
X-gal	5-bromo-4-chloro-3-indolyl-β-galactopyranoside.
YAC	yeast artificial chromosome.

Acknowledgements.

I would like to thank my supervisors, Prof. Wayne Davies and Dr. Roger Sutcliffe for their advise and comments on the writing of this thesis. I acknowledge Prof. Martin J. Brodie, from the Western Infirmary of Glasgow, for his support. Thanks also to my assessors, Dr. Thora Glencorse and Dr. John Clutterback.

In the lab, I would like to thank Dr. David Livingstone, Dr. Paul Shiels, and Dr. Denise Donald for their advise, and Lynn Loughlin and Irene Houghton, for their invaluable assistance. Thanks to Dr. Greg Stewart for taking time to show me how to carry out automated sequencing. Many thanks to Mary Gardiner; it has been a real pleasure to work next to her. Thanks to Robert Heeley for his encouragement and for so many useful tips. A special thanks to Eunice-Georgia Stefanou, who has helped me get this thesis printed, spending extra time that she did not have.

Thanks also to the people in the Anatomy department for their help, especially to Prof. Tony Payne, Dr. Jackie Campbell and Dr. David Russell, for being so kind and helpful, and to Neil Bennett for doing such a great job.

My gratitude to the ladies in the prep-room, the wash-room, and the store, for making life so much easier and always brightening the day. Many thanks to Mary Burke and to Robert MacNab, for their kindness and for coming to the rescue in more than one occassion.

A special thank-you to Jenny Robertson and Pierre Schembri for their frienship. Thanks also to my "fellow brothers", Maurice, Tam, Ed, David and Hong for letting me share some very good moments with them.

Most of all, I want to thank all my family for their love and support. Very especially, I want to thank my mom, for her love, her understanding, her patience, and my brother, my best friend. Thanks to my father, for 'daring' me not to give up when everything looked black. A special thanks also to Hrvoje, who has given me the love and the strength I needed to overcome very painful and difficult moments in the last two years.

I thank my grandfather, the best model a person could ever have. I thank him for always being there for me, for giving me his strength.

I acknowledge the receipt of a CASE studentship from the BBSRC, held in conjunction between the Division of Molecular Genetics and Western Infirmary of Glasgow.

SUMMARY.

In 1993, a mutant strain, AS/AGU arose spontaneously in an enclosed colony of the Albino Swiss (AS) strain of rat, kept at the Anatomy department in Glasgow University. AS/AGU animals exhibit a set of locomotor abnormalities. They display a general instability and whole body tremor, are slow at initiating movement, show reductions in purposeful action, and perform poorly at locomotor tests such as mid-air righting. Moreover, there is evidence that these abnormalities are associated with a decrease in the levels of dopamine in the midbrain, and it has been demonstrated that L-dopa administration or fetal midbrain transplants reverse the majority of the symptoms, resembling the observations made on Parkinson's disease patients. These features make the AS/AGU strain a useful model for movement disorders due in significant part to failure of the dopaminergic transmission system.

Crosses of AS/AGU to other laboratory rat strains point to a single recessive mutation with essentially complete penetrance (*agu/agu*) as the cause of the abnormal phenotype. There is no evidence of sex linkage or maternal inheritance. In the absence of any evidence of the function of the *agu* gene product, positional cloning of this locus was begun.

The first step was the establishment of a genetic map location for the *agu* locus. A large series of microsatellite markers were analysed and used to identify which of the strains PVG, BN, and F344 differed to a greater extent from AS/AGU. Differences at 43%, 62% and 47% of the loci were recorded, respectively. BN and F344 were therefore selected as the reference strains in backcrosses to AS/AGU, in an attempt to maximise the number of informative markers which could be used to type the progeny. Linkage analysis studies on the backcross offspring identified linkage of one of the markers to *agu*, thus revealing the chromosomal location of this gene. Utilisation of a further set of microsatellite loci which had been previously described on the particular chromosome led to the establishment of two microsatellite-based linkage maps of the *agu*-containing region, defining a 25-30 cM interval. A more precise localisation of the gene was not possible at this point as other microsatellite loci already mapped to this region did not show allelic differences in our backcrosses. New markers were consequently developed.

The corresponding region of synteny in mouse was identified, and mouse microsatellite primers used to amplify rat genomic DNA. Approximately 30% of these primer sets yielded a single amplification product. However, no size difference was observed between the products obtained from each of the parental strains in the backcrosses and further analyses were not carried out.

New microsatellite markers were isolated from P1 clones. Selection of these clones was carried out by screening for the presence of PCR products from sequenced loci known to map or putatively mapping to the *agu*-containing region. These latter consisted of rat homologues of mouse sequences which had been mapped in the mouse region of synteny.

P1 clones were grown in the presence of sucrose, in order to select for molecules which had not lost the insert. Since the P1 plasmid is kept at a single copy, the copy number was induced by adding IPTG a few hours prior to harvesting the cells. Several methods were used to isolate P1 DNA, with purification through a CsCl gradient yielding the best results, in terms of quality and quantity of plasmid DNA isolated. In addition, the effect of the bacterial host on the plamid DNA was assayed, by comparing the yields obtained from the original host, NS3529, and those obtained from DH10B cultures. Prior to these experiments, P1 DNA was transferred, via γδ-transposition, from the former to the latter strain. Clearly higher yields and better DNA quality were observed with the DH10B strain. Moreover, no alterations were observed when growing DH10B cells carrying a recombinant P1 which had been previously shown to be unstable in the NS3529 host.

A series of new markers were established. Three of these were shown to be informative between AS/AGU and F344. Consequently, the *agu*-region was narrowed down to an interval of approximately 11 cM. A fourth marker which was also informative between the aforementioned strains was shown to map to a different chromosome than expected. The corresponding chromosomal location was identified by screening a panel of somatic cell hybrids with the primer set. Two uninformative markers had been isolated from the same P1 clone. Their chromosomal location was therefore analysed. One of them was demonstrated to map on the newly identified chromosome. The other could not be mapped, as both the mouse and the rat DNAs

yielded PCR products of the same size. Further analyses of the corresponding P1 clone demonstrated that this was chimaeric. This was the same clone that had been shown to be unstable during the growth of NS3529 cultures.

Further refinement of the *agu* region resulted ultimately from the release of new rat genetic maps in 1996, which contributed several new markers to the project. One of these microsatellites was shown to be informative in both the BN and the F344 backcrosses and mapped at less than 0.5 cM from *agu*. This marker was thus selected as the starting point for the physical mapping of this region of the rat genome.

Thus the goal of this work, to obtain a genetic map location for *agu*, was fully achieved, and enabled a rapid entry into the physical mapping and gene identification phase of positional cloning of the *agu* gene.

CHAPTER 1

General Introduction.

1.0 Introduction.

Pathological states such as Duchenne muscular distrophy (Hoffman *et al.*, 1987), hemophilia A (Lawn, R.M., 1985; Antonarakis and Kazazian, 1988) and Wilson disease (Bull *et al.*, 1993; Tanzi *et al.*, 1993) result from alterations in a single gene and show a simple pattern of inheritance. In contrast, many common disorders such as diabetes (Davies *et al.*, 1994a; Hashimoto *et al.*, 1994), hypertension (Sever, 1995) and Alzheimer's disease (Farrer *et al.*, 1991; Rao *et al.*, 1996), involve the interaction of genetic and environmental factors. In many of these cases, no clear inheritance pattern is observed, as the presence of individual factors may not be sufficient to lead to disease. Moreover, different genetic and/or environmental factors may each cause apparently the same disease phenotype (e.g. Alzheimer's disease). These features make pathogenesis more difficult to investigate in multifactorial diseases. A genetic component can usually be deduced from the incidence of disease in first degree relatives to probands, compared with the incidence in the general population. Comparisons are also established, whenever this is possible, between the risk in monozygotic and in dizygotic co-twins. Similarly, identification of regions of identity by descent (IBD) between individuals who are not closely related is facilitated in isolated populations, who are highly inbred (Nystuen *et al.*, 1996). Environmental and genetic factors can be implicated in observed discontinuities in disease incidence in the different populations.

Extensive searches for the genes involved in either monogenic or multigenic diseases in man could not be attempted until recently, except in the instances where a single locus was responsible for the mutant phenotype and this appeared associated to a particular allele at a locus of known position in the genome. Initial studies to identify disease-loci were based on available biochemical markers and on genetic markers (section 1.3) known as minisatellites (Jeffreys *et al.*, 1985) and restriction fragment length polymorphisms (RFLPs) (Botstein *et al.*, 1980). The more recent development of highly polymorphic (i.e. different alleles of the particular locus exist in the population), PCR-typable markers known as microsatellites (Weber and May, 1989; Litt and Luty, 1989), has allowed the establishment of high-resolution human genetic maps (Weissenbach *et al.*, 1992; Gyapay *et al.*, 1994; Murray *et al.*, 1994). In turn, the availability of an array of markers distributed all throughout the genome permits systematic genome-wide screens for genetic linkage of mutant genes. Yet, analysis of the components in the development of disease states in man can be very complex.

Important complications arise from (i) the genetic heterogeneity in man; (ii) the impossibility to control the environmental agents the individuals are exposed to; and (iii) incomplete pedigrees, as a result of the long time span that separates each generation. It is thus very difficult for studies of inheritance in man to be based on the physical examination of more than three generations. Many diseases are not manifest early in life (e.g. Parkinson's disease), further hindering the analyses, as older relatives will be deceased by the time the illness becomes symptomatic in an individual. The generally small size of the families adds to the problem of getting enough samples to study segregation patterns. Considerable effort has gone into eliminating, at least partly, some of these impediments. Extensive computer-based statistical tests are now available that deal with multifactorial diseases (Weeks and Lathrop, 1995). Human lymphoblastoid cell lines have been established at the CEPH (Centre d'Etude du Polymorphisme Humain) which can be indefinitely propagated in the laboratory. Therefore, data obtained with different genetic markers in different projects can be pooled together, on the basis of having used these same reference families (Dausset *et al.*, 1990).

Animal models have been traditionally used to aid the identification of disease genes in man. The advantages that these models offer are multiple. Inbred strains are most frequently employed, which carry a single allele at the majority of the homologous loci; these genotypes greatly facilitate the analysis of inheritance in the progeny obtained from crosses between strains. Moreover, segregation in laboratory animals can be followed up throughout many generations and on a larger scale than in human families. Genetic background and the environmental conditions can be actively controlled, facilitating the dissection of the various genetic and environmental components in a disease, and the study of their contribution to the observed phenotype. In this way, loci can also be identified which do not exert a strong effect on the disease phenotype (i.e. modifier genes). A possible involvement of any of the identified factors in specific human diseases can then be assessed. Comparative mapping plays a crucial role in these studies (Eppig and Nadeau, 1995); human homologues of various genes found to be responsible for pathological states in research animals have already been shown to be the cause of similar alterations in man. An example of this is the gene

Pax-3, associated to both the mutant 'splotch' phenotype in mouse and the human Waardenburg syndrome type 1 (WS1) (Tassabehji *et al.*, 1992).

Recently, new approaches have been developed which are based on the creation of transgenic animals. These genetically engineered animals carry a mutated gene or set of genes, thus allowing the study of the effect that modifications of the gene(s) have on the living organism. Nevertheless, modelling of the human genetic lesion in research animals may not lead to the same pathology. Differences in the phenotypic expression of a particular mutation can correlate to differences in the biochemical pathways. A clear example is found in the murine models of the Lesch-Nyhan syndrome. In these animals, the gene coding for the enzyme hypoxanthine phosphoribosyltransferase (HPRT) has been disrupted, in an attempt to mimic the human disorder. However, not until a second pathway (adenine phosphoribosyltransferase-dependent) is disrupted, do these animals exhibit some aspects of the Lesch-Nyhan disease (Wu *et al.*, 1993). Moreover, different alterations in the same gene can result in different phenotypes. An example is observed in models of β-thalassaemia. Insertional disruption of the β-major gene causes severe thalassaemia in mice, in contrast to the milder form which results when this gene is deleted (Shehee *et al.*, 1993). Furthermore, development of a disease phenotype may not only depend on the occurrence of a particular alteration of the genotype, but also on the presence of other genomic sequences. In this way, transgenic mice which carry the full-length cDNA found in Huntington's disease patients show no behavioural abnormalities, the CAG repeat being remarkably stable. This suggests the effect of other genomic sequences on repeat instability in man (Goldberg *et al.*, 1996).

Animal models can be classified into two major groups: (a) induced models (Kumar *et al.*, 1995; Sengstock *et al.*, 1994), where the pathological symptoms exhibited by the animal have been caused by the researcher through surgical interventions, controlled diets, treatment with chemicals, and (b) spontaneous models (Li *et al.*, 1991; Muramatsu *et al.*, 1994; Löscher *et al.*, 1996), which arise due to alterations in the genotype or to intercrossing of standard strains which are selected for phenotypes of interest. The former allow the establishment of correlations between structure and disease. Important observations can be made during these analyses which lead to the development of future ways to improve or cure similar pathologies in man. Nevertheless, these models are induced and thus do not necessarily relate to the mechanisms involved in pathogenesis in man. Identification of the factors underlying

the mutant phenotype in spontaneous animal models can provide more reliable clues towards the establishment of causative agents in human disease.

In 1993, a mutant sub-strain, AS/AGU (*agu/agu*), arose spontaneously in an enclosed colony of the Albino Swiss (AS) strain of rat, kept at the Anatomy department in Glasgow University. These animals exhibit locomotor abnormalities which have been correlated to alterations in the basal ganglia, resembling various aspects of Parkinson's disease in man (Clarke and Payne, 1994; Campbell *et al.*, 1996a). Identification of the gene responsible for the mutant phenotype in these rats is expected to provide new information on mechanisms of degeneration of dopaminergic systems in the midbrain. It could also offer a new candidate gene for studies on PD families. A positive correlation of allelic variation at this locus and the occurrence of Parkinson's disease would validate the AS/AGU mutant sub-strain as a true model for the disease. Studies on these animals could reveal particular biochemical pathways of possible involvement in PD and also aid the development of treatments which at least ameliorate some of the symptoms observed in the patients. A brief description of the AS/AGU sub-strain of rat is presented below (section 1.1). Various aspects of Parkinson's disease in humans are then discussed (section 1.2); laboratory animals of current use in the study of PD are also presented. An overview of the mapping tools and techniques presently available to the researcher is given in section 1.3.

1.1 The AS/AGU sub-strain of rat: A new model of movement disorder.

The AS/AGU sub-strain of rat exhibits a series of neurochemical and behavioural abnormalities which appear to involve the basal ganglia. Locomotor abnormalities of AS/AGU rats are first observed at around postnatal day 10, and reliable assessment of the phenotype is possible at 4-6 weeks. While impairment can be detected from an early age, the disorder is progressive. A major characteristic of the fully developed condition is a marked difficulty in initiating movement. The rats show rigidity of the hindlimbs resulting in a staggering gait and a tendency to fall over every few steps; a slight whole body tremor is detectable (Clarke and Payne, 1994). A battery of non-invasive neurological locomotor tests allows the confirmation of the mutant phenotype. These tests are concerned with a) movement frequency; b) latency of

initiation of movement; c) stability when going down a slanted plank; d) capability to right itself in mid-air during a 50 cm drop from an upside-down position. AS/AGU animals are slow at initiating movement, they fail to show purposeful movement, and they perform very poorly in the two latter tests.

The sub-strain is true breeding and crosses between these animals result in offspring which are all affected. Conversely, crosses between this sub-strain and others (AS, Sprague-Dawley) yield F1 animals which are all normal. Backcrosses of the F1 progeny to the AS/AGU parent give both normal and affected litters in a ratio of 1:1. These observations point to a single recessive mutation of full penetrance as the cause of the abnormal phenotype. The AS/AGU karyotype appears normal. Reciprocal crosses yield equivalent phenotypic ratios, arguing against either sex-linked or maternal inheritance (Campbell *et al.*, 1996a).

Analyses of the CNS structures in AS and AS/AGU animals do not reveal any gross abnormalities. In contrast to the weaver mouse, which exhibits a severe reduction in striatal dopamine levels and has been proposed as a model of Parkinson's disease (section *1.2.3*), AS/AGU animals do not present any abnormality in the gross histology of the cerebellum. It has been observed that the SN *pars compacta* (SNc) of AS/AGU rats is thinner than normal, correlating to a decrease in the number of tyrosine hydroxylase (TH)-immunoreactive (TH-ir) neurons in this area (Clarke and Payne, 1994). Tyrosine hydroxylase is the enzyme that catalyses the synthesis of catecholamines and is a marker for dopaminergic cells.

Loss of TH-ir neurons is accompanied by an overall reduction in the levels of dopamine in the neostriatum of the AS/AGU animals, compared to AS rats (Payne *et al.*, 1994; Campbell *et al.*, 1996a). Analyses by micropunch followed by high performance liquid chromatography with electrochemical detection (HPLC-ECD) indicate that the dorsal and the middle regions of the caudate-putamen are most affected, with reductions of around 30-35% and 25% in dopamine levels, respectively. Ventral and anterior regions are less affected (Campbell *et al.*, 1996a). This topographic specificity is also observed in the weaver mouse (Roffler-Tarlov and Graybiel, 1984), and resembles that seen in PD patients (Kish *et al.*, 1988). Assessment of the dopamine levels in animals of 3, 4, 6, 9, and 12 months of age has demonstrated that significant differences between AS and AS/AGU rats are only observed from 6 months onwards (Campbell *et al.*, submitted). Interestingly, the reduction observed in dopamine levels in AS/AGU animals by post mortem micropunch is much smaller than that recorded when

dopamine is sampled from the extracellular fluid of the striatum by microdialysis (reduction of the order of -80%) (J.M. Campbell, A.P. Payne and T.W. Stone, personal communication). This suggests that the *agu* mutation is associated with a defect in the functional release of dopamine in the striatum. In addition to dopamine depletion, AS/AGU animals show deficits in the noradrenalin levels, resulting from neuronal loss at the locus coeruleus (Scott *et al.*, 1994). Depletion of serotonin levels is also recorded (Stewart *et al.*, 1994). Studies of [^{14}C]2-deoxyglucose uptake on 43 separate regions of the brain of AS/AGU and AS control animals have indicated a significant reduction of 20-25% in metabolic activity in the SNc, the subthalamic nucleus and the ventrolateral thalamus of AS/AGU rats. Smaller, but not statistically significant reductions in 2-deoxyglucose uptake have been observed in other brain areas (A. Lam and J. McCulloch, personal communication).

Both acute and chronic administration of L-dopa reverse some of the locomotor abnormalities observed in AS/AGU animals and restore dopamine levels in the neostriatum to those observed in AS animals. Although the rats do not lose their disordered hindlimb gait, they show greatly increased frequency of initiation of movement, more extensive periods of movement, and ability to execute a righting reflex and walk down an inclined ramp. Essentially the same response is observed in animals which have been treated with a high dose of L-dopa for 5 days, and in rats that have been given a much lower dose of 25mg/kg/day for 4 weeks (Campbell *et al.*, 1997). Reversal of the locomotor dysfunction is also observed following fetal midbrain transplants into the striatum. These responses resemble those observed in PD patients (section *1.2.4*).

1.2 Parkinson's Disease.

Parkinson's disease is a motor system disorder, firstly recognised by James Parkinson in 1817. It shows a world-wide distribution and its incidence rate increases steeply with advancing age. The symptoms usually appear between the ages of 50 and 70, although approximately 5 to 10% of the patients are under the age of 40.

Parkinson's disease demonstrates a great deal of clinical heterogeneity (Poewe and Wenning, 1996); different patients can show different combinations and severity of symptoms. Primary symptoms are a characteristic "pill-rolling" tremor, muscular

rigidity, a significant impairment in movement initiation (akinesia), a reduction in the amplitude and velocity of voluntary movements (bradykinesia), and impaired balance and co-ordination. Other symptoms may accompany the former, such as visual deficits, speech and handwriting alterations, olfactory dysfunction, difficulty in chewing and swallowing, memory loss, skin problems; dementia is also shown by a proportion of the patients.

The motor disturbances are known to result mainly from the destruction of a very high proportion of dopamine-producing cells, primarily located in an area of the ventral midbrain known as the *substantia nigra pars compacta* (SNc). At least 70% of these dopamine-producing cells either die or become impaired in Parkinson's disease sufferers. This is in clear contrast to the loss of approximately 4% of the original set of dopamine-producing neurones observed during normal ageing. Loss of dopaminergic cells results in greatly decreased levels of dopamine in the striatum, the main receptor area for the SNc. This region of the brain controls movement, balance, and walking. In addition to the dopamine deficiency in the striatum, the dopamine levels are also reduced in other parts of the brain related to the corticolimbic system. A deficiency in noradrenaline, another chemical transmitter, has also been observed. Noradrenaline is found primarily in the *locus coeruleus* and the lateral tegmental area, both involved in governing the involuntary autonomic nervous system. Likewise, the levels of serotonin and gamma amino butyric acid (GABA) can also be lower in PD patients than in non-affected individuals. These deficiencies may account for some of the secondary symptoms in the disease. A brief description of the neurochemical pathways affected in Parkinson's disease is given in section *1.2.1.* Characteristic in Parkinson's disease is the presence of eosinophilic inclusions in the degenerating neurones. These formations, known as Lewy bodies (LBs), are also observed in other neurodegenerative processes such as Alzheimer's disease, as well as in the brains of normal aged individuals, and could be just markers of neurodegeneration (Langston, 1996).

1.2.1 Parkinson's disease and the basal ganglia.

The human basal ganglia [Figure 1-1] include the neostriatum, formed by the putamen and the caudate, and the globus pallidus, subdivided into an internal segment (GPi) and an external segment (GPe). The substantia nigra (SN), which comprises the substantia nigra *pars reticulata* (SNr) and the substantia nigra *pars compacta* (SNc),

and the subthalamic nucleus (STN) are also normally included in the basal ganglia, due to their close anatomical and functional association to the neostriatum and the globus pallidus. Together, these structures form a forebrain system which receives signals from various parts of the brain and integrates the information, directing then the execution of appropriate responses (Parent, 1990) by selecting patterns of motor activity. The neostriatum is the main 'receptor' area. Different pieces of information arrive here through separate pathways. The overall division observed in the neostriatal input is maintained in its output through the globus pallidus and the SN (Alexander and Crutcher, 1990). Moreover, a higher level of compartmentalization has been observed within the target nuclei. In this way, for example, the neostriatum appears divided in striosomes and matrix; these areas differ in their biochemical constituency and establish different connections with other neural structures (Graybiel, 1990).

Controlled motor activities are the result of complex interactions between the cortex, the basal ganglia, and other areas of the brain such as the thalamus (Graybiel, 1990; Parent, 1990; Smith and Bolam, 1990; Alexander and Crutcher, 1990) [Figure 1-2]. Information is sent from the cortex to the neostriatum by neurons that use excitatory amino acids (EAAs), mainly glutamate, as a transmitter. The neostriatum processes incoming sensorimotor information with messages from the SNc which seem to respond to signalling from the limbic system. Two striatal output pathways have been identified: a 'direct' pathway, which involves the GPi and the SNr, and an 'indirect' pathway, which involves the GPe and the STN. Inhibitory GABA-ergic signals sent from the neostriatum to the GPi and to the SNr result in the disinhibition of thalamo-cortical neurons and thus favours motor activity. Inhibitory GABA/enkephalin signals from the neostriatum to the GPe leads to the disinhibition of the STN, which then sends excitatory glutamatergic signals to the GPi and to the SNr, therefore reinforcing the inhibitory action of these latter on the thalamus. Disinhibition of the thalamo-cortical neurons results in excitatory signals to various premotor cortical areas. This is, however, an oversimplified description of the interactions involved in the control of motor and other related activities. Different subpopulations of neurons can be present within the same structure (e.g. in the neostriatum, in the GPe) which are involved in opposite responses. Additional complexity results from the existence of other connections between the different structures (e.g. the STN also innervates the

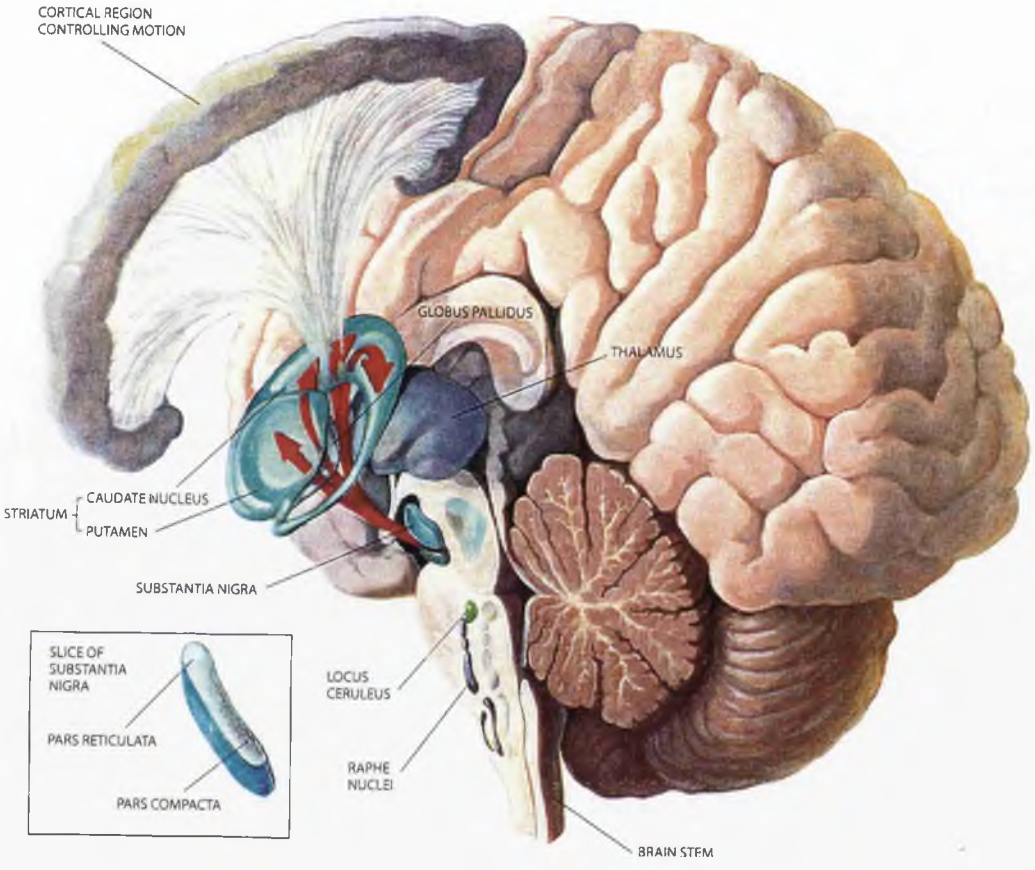


FIGURE 1-1. Diagram of the brain regions which are affected in Parkinson's disease, taken from Youdim and Riederer (1997).

The subthalamic nucleus, located below the thalamus, is not shown.
Loss of dopamine-producing neurons, particularly marked in the substantia nigra *pars compacta* (SNc), leads to imbalances in neurochemical transmission, which ultimately affect motion.

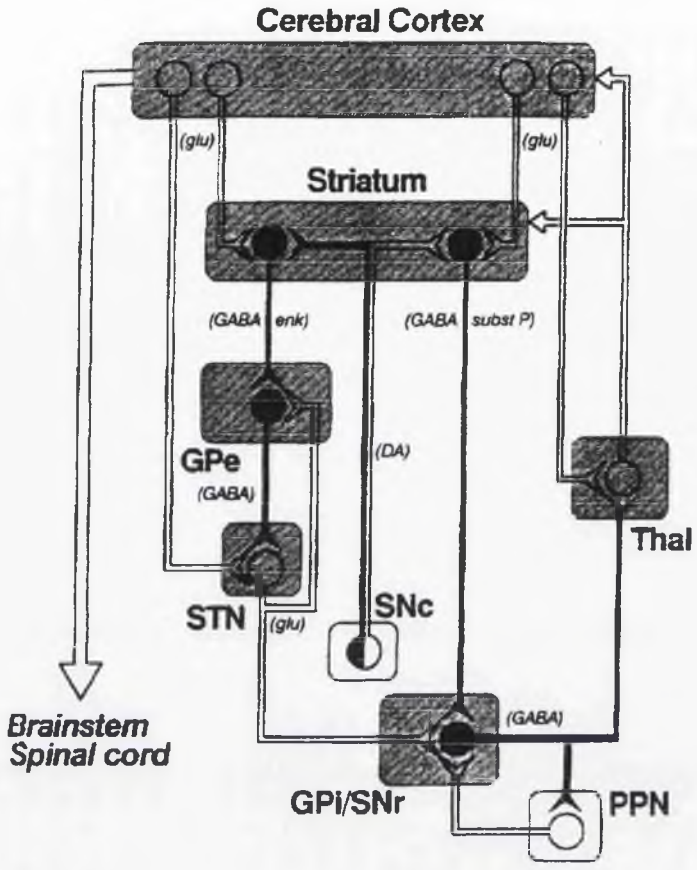


FIGURE 1-2. Simplified overview of the basal ganglia-thalamo-cortical circuitry, taken from Alexander and Crutcher (1990).

Open symbols indicate excitatory neurons and filled symbols inhibitory neurons. Striatal cholinergic interneurons, thought to receive signals from both the cortex and dopaminergic neurons (Carlsson and Carlsson, 1990), are not shown. Intrinsic feedback loops and other connections between the nuclei exist which are not presented in this diagram. GABA-ergic projections also carry: enkephalin (enk), to the GPe; substance P (subst P) and dynorphin, to the GPi; either enkephalin or substance P and/or dynorphin to the SNr (Fernandez and Dujovny, 1997).

Other abbreviations not shown in the text:
glu: glutamate.
DA: dopamine.
PPN: pedunculopontine nucleus.
Thal: thalamus.

neostriatum), as well as from intrinsic feedback loops within each nucleus; these are not shown in Figure 1-2.

Fine regulation of both the innervation of target areas and the response to received signals seems to correlate to complex interactions between the various neurochemicals in the brain. Disruption of the normal balance between them can lead to movement disorders (Calabresi *et al.*, 1996; Chesselet and Delfs, 1996). As mentioned above, dopamine depletion is the pathological hallmark of Parkinson's disease. Dopamine is the main carrier of signals between the SNc and the neostriatum and plays a major role in neurotransmitter expression in the striatal cells (Graybiel, 1990) and in the control of purposeful muscle activity through several mechanisms. The activity of dopaminergic neurons regulates glutamatergic signalling from the cortex (Calabresi *et al.*, 1996). The mechanism is not clear yet, although an interaction of both glutamatergic and dopaminergic messages at cholinergic interneurons in the neostriatum is suspected (Carlsson and Carlsson, 1990; Fernandez and Dujovny, 1997). Dopaminergic transmission also plays a crucial role in the integration of information received from sensorimotor areas and from limbic centres, associated to emotion and motivation. Messages sent from the latter to the neostriatum are translated into specific signals from the SNc to the striatal areas involved in motor functions. Dopamine exerts different actions on different subpopulations of striatal neurons. It activates the 'direct' pathway, therefore decreasing the phasic inhibition of thalamo-cortical neurons by the GPi and the SNr, and it inhibits the striatal output to the 'indirect' pathway, increasing the inhibition of the STN by the GPe. Both responses favour motor activity. Dopamine depletion in Parkinson's disease results in a decreased striatal output through the 'direct' pathway and in increased activity of the STN in the 'indirect' pathway. Overactivity of the STN in PD does not correlate to an expected inhibition of the GPe and it is thought to result from other interactions (Chesselet and Delfs, 1996); these could correspond to signals from the cortex or from the SNc, which also innervate this area.

Other dopaminergic areas are also affected in parkinsonian brains, although to a lesser extent. This is the case of the ventral tegmental area (VTA), associated to the limbic circuitry. Serotonin levels, which are high in the normal basal ganglia, are reduced in both the striatum and the globus pallidus in PD patients. A 50-85% loss of

noradrenergic neurons in the locus coeruleus has also been reported, associated to a 40-70% decrease in the noradrenaline concentrations in the cortex and limbic receptor areas.

The fact that Parkinson's disease does not become clinically manifest until the striatal dopamine levels are decreased to about 20% of the normal levels seems to indicate the action of some compensatory mechanism to maintain striatal dopamine transmission. Increased dopamine biosynthesis by the remaining neurons, hypersensitivity of the dopamine receptors in the striatal neurons, and dopamine synthesis in striatal cells which do not normally produce this neurochemical are some of the possible explanations for the delayed appearance of PD symptoms (Zigmond *et al.*, 1990). Decreased levels of dopamine-transporter, involved in the reuptake of dopamine from the synapse and thus in the termination of neurotransmission, have been observed in the remaining nigral cells in the brain of PD patients (Harrington *et al.*, 1996). Although this reduction could represent an attempt to increase synaptic dopamine concentrations, Harrington *et al.* (1996) consider it to be more likely the result of decreased activity of these neurons.

1.2.2 Cause of Parkinson's disease. Possible mechanisms of cell damage.

Despite its world-wide incidence, the cause of most of the cases of Parkinson's disease is still unknown. Initial studies on monozygotic and dizygotic twins argued against an important role of genetic factors in the pathogenesis of PD, as the concordance rates observed were very low (Duvoisin *et al.*, 1981). Moreover, most cases of PD are sporadic; only a few families present multiple affected individuals. The occurrence of several cases of PD in a family does not necessarily imply the involvement of genetic factors, since, given the high incidence of this neuropathology, this could be merely coincidental. These facts, together with the observation that 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a chemical found in impure drug preparations, causes neuronal damage similar to that seen in PD, suggested an environmental origin of this disease (Rajput, 1993). In agreement with this hypothesis were also the reports on a higher risk of PD in rural settings, where exposure to herbicides, pesticides and drinking well water were put forward as risk factors. However, no geographic clustering of PD cases is observed. This is one of the reasons that have led to what is currently the most accepted hypothesis for the cause of PD,

which proposes the action of environmental factors that determine the development of Parkinson's disease in genetically susceptible individuals (Poirier *et al.*, 1991; Burn *et al.*, 1992; Langston, 1996). It seems clear that ageing is also a contributing factor. Supportive of a genetic involvement in PD are the results obtained when applying newer techniques such as positron-emission tomography (PET) scanning of the brain of patients and normal co-twins or other relatives. This method allows one to detect the levels of striatal ^{18}F-6-fluorodopa (^{18}F-dopa) uptake and thus the assessment of the ability of the striatum to convert exogenous L-dopa to dopamine. A reduction of at least 35% of putamen ^{18}F-dopa uptake is observed in PD patients, whilst the caudate is relatively spared (Brooks, 1991). Brooks (1991) also reports a subnormal putamen ^{18}F-dopa uptake in some of the unaffected twins in monozygotic and dizygotic pairs, as well as in some relatives of Parkinson's disease sufferers. These and other subsequent studies (Burn *et al.*, 1992) show that the concordance rates between twins are higher than was previously considered. Ambiguous diagnostic criteria and, possibly, the classification of subclinically affected co-twins as unaffected in the initial studies, probably account for the discrepancy in the results. Given that PD becomes usually manifest in the later stages of life, absence of parkinsonian symptomatology in the co-twin of a PD case does not exclude the possibility that the former will also become affected later on. Genetic factors have already been established in some of the familial cases of PD (Polymeropoulos *et al.*, 1996; Polymeropoulos *et al.*, 1997). Harris and coworkers (1995) reported that a mutation in the ceruloplasmin gene leads to abnormal iron metabolism and, ultimately, to basal ganglia degeneration, as seen in PD. Genetic anticipation has been observed in some cases of Parkinson's disease, and the involvement of an unstable trinucleotide repeat proposed (Payami et al., 1995). Since both the familial and the sporadic cases of Parkinson's disease appear identical, it seems probable that genetic factors are also involved in the latter. The occurrence of PD is considered to result from the action of a single major gene, with low penetrance, in interaction with other loci which exert a smaller effect, or from the interplay of several loci (multifactorial model) (Lazzarini *et al.*, 1994); the genetic factors are thought to interact with environmental agents (Johnson, 1991; Burn *et al.*, 1992). Low penetrance could explain the absence of various cases in a family, as these are normally too small for a second case to occur. The fact that symptoms appear only when the levels of striatal dopamine are greatly reduced seems to point at a threshold operating in PD.

Deficient DNA repair mechanisms, the absence of a dopaminergic specific growth factor provided by the normal target striatal cells to the SN neurones, excitotoxicity, disturbances of calcium homeostasis, or a viral infection that, like the poliomyelitis virus, selectively destroys specific neuronal populations, are some of the factors which have been proposed as causative agents of the neurodegeneration observed in Parkinson's disease (Poirier *et al.*, 1991; Koller, 1997). A viral aetiology seems, however, improbable, as injections of parkinsonian brain tissue extracts into monkeys have failed to induce PD in these animals; nevertheless, this could be explained by different responses to viral infections in different primates. There is some evidence supporting a role of the immune system in PD. McGeer and colleagues (1988) reported activated microglia in the basal ganglia of parkinsonian brains. Furthermore, both the serum and the cerebrospinal fluid of a large proportion of PD patients carry an antibody that reacts with the dopamine neurons. Production of this antibody seems to follow release of an encrypted antigen from degenerating dopamine cells. Therefore, it may contribute to disease progression, rather than being the cause of neurodegeneration (Carvey *et al.*, 1991).

The close resemblance between PD and MPTP-induced toxicity strongly suggests the action of an MPTP-like environmental toxin in Parkinson's disease. Susceptibility to the toxin(s) could be dictated by metabolic errors such as deficient detoxification by enzymes like the hepatic cytochrome P-450 system or like those involved in sulphur metabolism (Williams *et al.*, 1991; Tanner, 1991; Langston, 1996), as these enzymes control the type and the quantity of chemical toxins that enter the brain. Were this the case, exposure of the individual to otherwise inoffensive doses of a certain compound could lead to PD. The pattern of neuronal loss characteristic of this pathology could result from the presence of specific uptake systems for the toxin in these populations of cells.

It has also been hypothesised that neuromelanin, a product of the oxidation of dopamine, plays a role in toxicity (Poirier *et al.*, 1991; Furtado and Mazurek, 1991). This concept is based on the observations that the presence of neuromelanin-bearing cells correlates mostly to the areas of cell loss in PD, that this pigment accumulates with age, and that aged animals are more susceptible to MPTP-induced parkinsonism than young ones. It has been proposed that neuromelanin could be permitting a long-

term trapping of toxic compounds taken up by the cells. On the other hand, not all the areas affected in PD are melanized.

Studies on the mechanism of MPTP-toxicity suggest that abnormalities in the mitochondria could lead to the development of Parkinson's disease. The absence of maternal inheritance in PD does not rule out this possibility, since several of the mitochondrial proteins are in fact encoded by the nuclear DNA. Furthermore, heteroplasmic mitochondrial mutations could lead to a non-Mendelian pattern of inheritance, as observed in Leber's hereditary optic neuropathy (Holt *et al.*, 1989; Brown *et al.*, 1992). A significant reduction of the activity of Complex I at the SN and the striatum has indeed been demonstrated in several PD patients. Complex I-deficiency is not specific to this pathology, and it can be found in other disorders such as Huntington's disease (Parker, 1991). The mitochondrial abnormalities seen in PD could thus be a secondary effect rather than the cause of neurodegeneration, although it must be kept in mind that the same phenotypic alteration can be the result of different biochemical defects.

The hypothesis of a "free radical toxicity" proposes that the neuronal death observed in PD results from oxidative damage to the cell by the formation of free radicals (Olanow, 1992; Jenner and Olanow, 1996). Contrary to early reports, it is now believed that many of the antioxidant defenses appear intact in PD patients [Figure 1-3]. Interestingly, an increase in the activity of superoxide dismutase (SOD) is observed. The levels of reduced glutathione (GSH), involved in the clearance of H_2O_2 by the enzyme glutathione peroxidase, are lowered in parkinsonian brains (Perry *et al.*, 1982). Studies on animal models suggest that, rather than leading to nigral pathology on its own, GSH depletion renders the dopaminergic system susceptible to toxins (Owen *et al.*, 1996). These alterations in antioxidant mechanisms run parallel to higher than normal levels of iron, particularly in the SN and striatal areas (Sofic *et al.*, 1988). Free iron is considered a potential catalyst of free radical reactions. The high concentrations of iron measured in PD brains do not correlate to increased levels of ferritin, which binds iron. This suggests that the excess iron exists in its free form. The high iron levels do not correlate either to increased numbers of transferrin receptors on the surviving melanized neurons in PD. In contrast, the levels of lactoferrin, which strongly binds iron, and lactoferrin receptors, are increased in the SNc of PD patients. This could lead to increased intracellular iron levels (Faucheux *et al.*, 1997). An increase in the iron

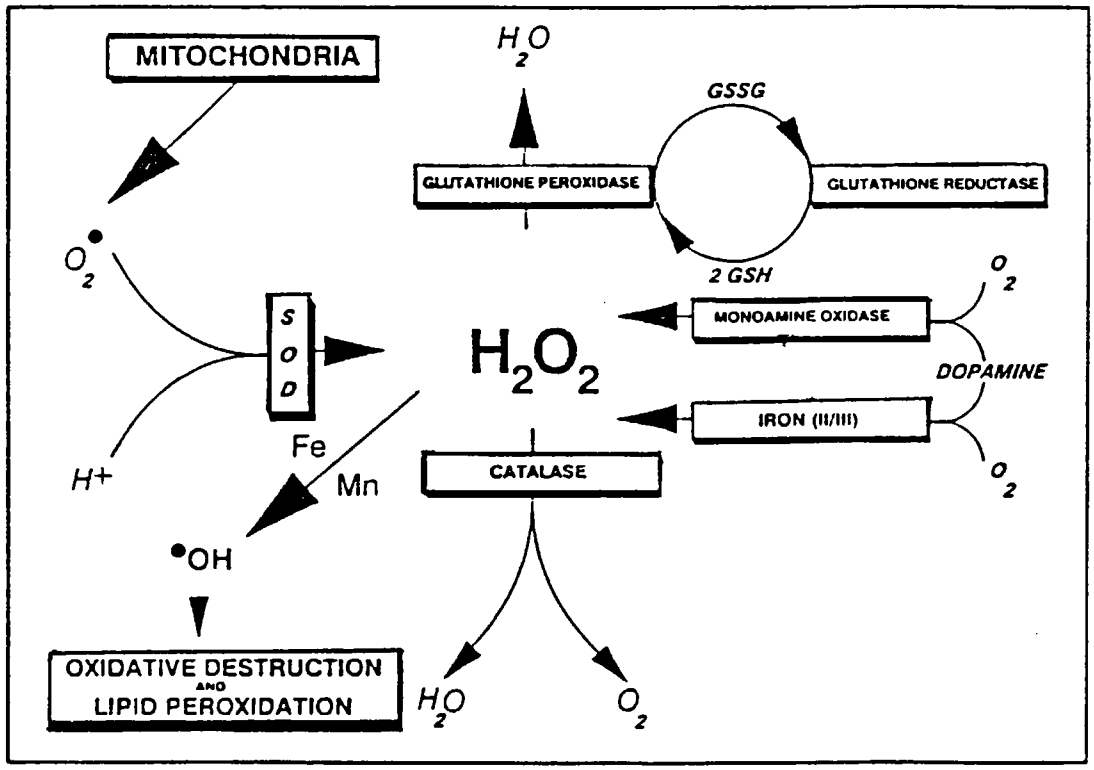


FIGURE 1-3. Antioxidant systems in dopaminergic neurons.

Diagram taken from Poirier *et al.* (1991). GSSG: oxidized glutathione. GSH: reduced glutathione. SOD: superoxide dismutase. $O_2^{\bullet}$: superoxide radical. •OH: hydroxyl radical.

levels is not specific for PD, and it is observed in other neurodegenerative disorders. It could thus be a secondary effect in the disease. Nonetheless, it probably is a contributing factor to the pathology. These observations on PD brains could well relate to increased oxidative stress, with the accumulation of free radicals ($OH^{\cdot}$, $NO^{\cdot}$, $O_2^{\cdot-}$) and concomitant damage to cellular structures. There are reports on oxidative damage to lipids, DNA, and proteins in the brain of Parkinson's disease patients, although the contribution of drug treatments to this damage has not been clearly evaluated yet.

1.2.3 Animal models.

No suitable genetic animal models have been available to date for the study of Parkinson's disease. The weaver mouse shows a loss of dopamine neurons in the SNc and a severe decrease in striatal dopamine levels (Roffler-Tarlov and Graybiel, 1984). However, other areas affected in PD patients, such as the locus coeruleus, are normal in these animals. Moreover, important cerebellar abnormalities, which are not seen in PD, are associated with the weaver phenotype. Weaver mice carry a mutation (*wv*) in a gene encoding a G-protein coupled inward rectifier potassium channel (GIRK2) (Patil *et al.*, 1995). This mutation results in the replacement of a glycine for a serine in a Gly-Tyr-Gly motif in the H5 pore region, leading to loss of selectivity for K^+. Mutant *wv*GIRK2 channels conduct Na^+ as well as K^+ (Tong *et al.*, 1996). Moreover, these channels show constitutive activation, which could be a secondary consequence of intracellular Na^+ accumulation (Silverman *et al.*, 1996). The Na^+ influx must, in turn, enhance the channel activity. In addition, *wv*GIRK2 channels are also permeable to Ca^{2+}, providing another mechanism for cellular degeneration in the weaver mouse (Silverman *et al.*, 1996). Sequencing of the human homologue of the *weaver* gene in fifty PD patients demonstrated no abnormalities. These results suggest a different etiology for the nigral degeneration observed in weaver mice and in PD in man (Bandmann *et al.*, 1996). Nevertheless, the failure to establish a correlation between the human homologue of the *weaver* gene and PD in the patients studied could merely reflect that allelic variations at this locus are not responsible for a large proportion of the PD cases in the general population.

Damage models are normally employed which mimic the parkinsonian symptoms. By studying the mechanisms involved in the abnormal phenotype, clues may be obtained to the possible pathways leading to PD in man (Langston, 1996). As

indicated above (section *1.2.2*), successful identification of mitochondrial abnormalities and deficiencies in detoxifying enzymes has been carried out in human patients following initial observations on experimental animals. These models also play a crucial role in the development of possible treatments for the parkinsonian symptoms. The benefits of L-dopa, selegiline, and other drugs, as well as those of surgical interventions (transplantation of fetal tissue, pallidotomy) are some examples (section *1.2.4*).

Two damage models of PD are mainly employed, obtained by treatment with either MPTP, or with 6-hydroxydopamine (6-OHDA). The mechanism of MPTP-induced neurodegeneration has been clearly elucidated (Furtado and Mazurek, 1991). Toxicity occurs when MPTP crosses the blood brain barrier before it is converted by MAO-B to its active form, N-methyl-4-phenylpyridinium ion (MPP^+). MPP^+ formation in the brain takes place mainly in astrocytes. MPP^+ is then released into the extracellular fluid and taken up by the dopaminergic cells, into their mitochondria. Here, MPP^+ inhibits the enzyme system Complex I, first stage of the mitochondrial electron transport chain. This can lead to a rapid depletion of ATP levels, gradually decreasing energy levels, and generation of free radicals.

MPTP-models closely resemble the clinical and biochemical features of PD in man (DeLong, 1990; Furtado and Mazurek, 1991). Although not all the animals develop the characteristic resting tremor, they show muscular rigidity, akinesia and bradykinesia. Damage to the dopaminergic neurons in the SNc and decreased dopamine levels in both the SNc and the striatum are observed. There is also some degree of cell loss in the locus coeruleus and the raphe, resembling that in PD patients. Interestingly, this damage is associated with eosinophilic intraneuronal inclusion bodies, resembling Lewy bodies. Large variations in the vulnerability to MPTP-toxicity are observed between different species. A correlation between neuromelanin levels and MPTP toxicity has been put forward, in an attempt to explain the greater susceptibility shown by primates and humans when compared to rodents. The observation that older animals are more sensitive to MPTP toxicity than young ones supports this concept. On the other hand, rats show a natural resistance to the toxic action of MPTP. Comparative studies between Brown Norway rats and C57BL mice (Zuddas *et al.*, 1994) suggest that the different vulnerability to MPTP observed in these animals may result from differences in the intraneuronal distribution of MPP^+. Differences in the vesicular storage of MPP^+ and its release could mediate susceptibility to MPTP toxicity.

Interestingly, Harrington and coworkers (1996) observed decreased levels of a synaptic vesicle amine transporter, VMAT2, in the SNc of PD sufferers. This transporter is responsible for the repackaging of dopamine into synaptic vesicles and its expression has been associated to resistance to MPP^+ in *in vitro* cell lines (Liu *et al.*, 1994). Nevertheless, Harrington *et al.* (1996) reported that some of the remaining nigral cells in PD appeared shrunken and atrophied, despite a considerably strong VMAT2 signal.

Neurotoxicity of the dopamine analogue 6-OHDA requires direct administration of the drug into the brain, as it can not cross the blood-brain barrier. This neurotoxin is, like MPTP, selectively transported into catecholaminergic neurons via high-affinity catecholamine uptake systems. Increased free radical levels and a decreased antioxidant capability in the target neurons have been observed following administration of 6-OHDA. Changes in membrane fluidity result, possibly due to lipid peroxidation. These changes seem to lead to a marked increase in the concentration of cytosolic calcium. This may, in turn, cause alterations in cellular processes such as signal transduction (Kumar *et al.*, 1995).

Evidence supporting a key role of oxidative stress in the pathology of PD, together with the observation of abnormal levels of iron in the SN of parkinsonian patients have led to the establishment of another model of PD. Intranigral infusion of moderate amounts of an iron citrate solution in rats leads to a dose-dependent SN neurodegeneration and a concomitant decrease in the levels of dopamine and its metabolites in the striatum. Very importantly, single exposure induces persistent and progressive changes, which occur through a number of months following infusion (Sengstock *et al.*, 1994), thus resembling the progressive degeneration observed in most PD patients. This feature is not found in the MPTP- or the 6-OHDA- models, where the changes observed following damage do not progress, and tend instead to recover with time, following single dose administration.

1.2.4 Treatment of Parkinson's disease.

A wide array of clinical approaches are being applied in the treatment of Parkinson's disease. These measures diminish or eliminate some of the parkinsonian symptoms. The most effective approach followed to date is the administration of L-dopa or levodopa, a precursor of dopamine. Administration of dopamine faces the problem that dopamine does not cross the blood-brain barrier; in contrast, L-dopa can

readily enter the brain. Although it does not stop the progression of the disease, L-dopa treatment delays this progression. Maximum benefit is achieved when this drug is administered early after disease onset; in contrast, life expectancy is not altered in those cases when treatment commences later in the disease. L-dopa is used alone or in combination with other drugs such as anticholinergics, which help partially restore the normal balance between acetylcholine and dopamine in the brain, the antiviral drug amantadine, which is thought to enhance the synthesis, release or reuptake of dopamine from surviving nigral cells, or dopamine agonists like bromocriptine and pergolide. These accompanying drugs can complement the action of L-dopa. This is especially important when the response to this compound starts to wear-off, usually during the progression of the disease. Moreover, around 20% of PD patients do not respond to L-dopa treatment. Important side-effects to this treatment are dyskinesias, or involuntary movements, and on-off effects, when the patient intermittently changes from normal to parkinsonian (Poewe and Wenning, 1996). These adverse side-effects are most commonly encountered following long-term administration of the drug. These alterations are a serious problem, as the initial dose of L-dopa must often be gradually increased in order to continue providing the same beneficial effects. In some cases, cessation of the treatment for a short period of time ('drug holiday') re-establishes the adequate response to L-dopa once treatment starts again. Co-administration of L-dopa with carbidopa, a peripheral dopa decarboxylase blocker, helps decrease some of the adverse side-effects and allows a reduction of the amounts of levodopa needed. In some cases, destruction of the subthalamic nucleus has been carried out to treat serious adverse effects caused by long-term administration of L-dopa.

Studies on animal models have demonstrated the benefits of transplanting fetal SN grafts into parkinsonian brains, as well as those of surgical interventions where the globus pallidus is lesioned (pallidotomy), with the concomitant suppression of the pallidal over-inhibition on the thalamus (Fernandez and Dujovny, 1997). Different outcomes have been obtained following transplantation of dopamine neurons in the brain of PD patients. Studies on research animals indicate the action of free radical toxicity during or following grafting, and demonstrate that manipulation of the tissue to be grafted, aimed towards the enhacement of antioxidative properties in these cells, leads to a greater survival of the dopamine neurons and to a faster and more extensive functional recovery (Nakao *et al.*, 1995).

Administration of trophic factors, such as brain-derived neurotrophic factor (BDNF) and glial cell-line derived neurotrophic factor (GDNF), has led to significant improvements in animal models of PD (Lapchak *et al.*, 1997; Fahn, 1997). Various strategies can be applied to deliver these factors to the brain. Viral vectors are being currently developed which allow both direct transfer of genetic material into the brain, or transformation of cell lines prior to transplantation (Horellou *et al.*, 1997).

Treatment of research animals with the drug known as selegiline or deprenyl prior to exposure to MPTP protects these animals from neurodegeneration. This drug has beneficial effects on PD patients. Although its mechanisms of action still remain controversial, it is currently considered that inhibition of MAO-B by selegiline may allow the accumulation of dopamine in the surviving cells in the SN. Together with an improvement in the symptomatology, a putative neuroprotective action of selegiline has been postulated (Jenner and Olanow, 1996; Tatton and Chalmers-Redman, 1996). This latter seems to be independent from the inhibition of MAO-B and exerted, instead, through transcriptional effects which lead to increased synthesis of trophic, antioxidant, and antiapoptotic factors, such as SOD, GSH, BCL-2. Selegiline appears to prevent damage to mitochondria which may be associated to the initiation of apoptosis. Were apoptosis involved in the neurodegeneration observed in disorders such as Parkinson's disease, this could well explain the neuroprotective effects of selegiline (Tatton and Chalmers-Redman, 1996).

1.2.5 Preclinical detection of Parkinson's disease.

The possibility of neuroprotective approaches makes early detection of Parkinson's disease crucial. This early diagnosis could facilitate the identification of the causative agents in PD. The cure for this pathology will not be available until the factors and the mechanisms involved are known. Identification of susceptibility genes in this disorder is, undoubtedly, a key element to the establishment of the environmental agents which precipitate the occurrence of PD, since these genes will point at specific biochemical pathways which can lead to the disease (Johnson, 1991). Moreover, knowledge of the susceptibility genes involved in Parkinson's disease could allow the identification of populations at a higher risk of developing this disorder; preventive therapy could then be attempted in these groups.

Therefore, early diagnosis of Parkinson's disease is of great importance for the understanding and efficient treatment of the disorder (Langston and Koller, 1991). In general, the alterations in the patients are all subtle at first, and worsen with time. Clear symptomatology of Parkinson's disease normally correlates with advanced stages of neurodegeneration. Detection of the early manifestations of the disease is therefore crucial to identify potential new cases when the neuronal damage is not extensive yet. Moreover, PET studies (Burn *et al.*, 1992) and reports on a premorbid personality of PD point towards a possible identification of putative cases even before any subtle alterations are shown. Biomarkers are being sought among metabolic enzymes which have shown reduced activity levels in some groups of patients, as well as in the levels in brain and in the cerebrospinal fluid of the metabolites produced in biochemical pathways which may be involved in Parkinson's disease (Carlsson and Fornstedt, 1991; Ellenberg, 1991). Although still controversial, many authors report the existence of a set of personality features which are commonly observed in individuals who later on develop parkinsonian symptoms (Paulson and Dadmehr, 1991). Comparison of twin pairs often shows personality differences which date back to adolescence or early adulthood, the affected twin being in general more serious, introverted and nervous than the unaffected co-twin (Duvoisin *et al.*, 1981). A lower incidence of alcoholism and cigarette smoking is encountered in PD patients. It has been suggested that the premorbid personality observed in Parkinson's disease patients may reflect innate intrinsic levels of dopamine. Studies on mouse have shown differences in midbrain TH activity and in the numbers of dopaminergic cells between different inbred strains (Baker *et al.*, 1980; Baker *et al.*, 1983). Interestingly, these variations seem to lead to differences in drug-induced and spontaneous behaviours which are mediated by dopaminergic neurons, such as motor activity and exploratory behaviour (Reis *et al.*, 1983).

The possibility of a premorbid personality in PD, together with the observation that other disorders which are commonly found in the patients, such as loss of interest in activities, olfactory dysfunction, subjective sensory abnormalities, are often present before the disease becomes manifest, indicate that alterations in brain function start several to many years before PD is diagnosed (Koller *et al.*, 1991). Long latency is observed in other neurological disorders (Koller *et al.*, 1991), such as Huntington's disease, where the alteration is present throughout the individual's life but the disease is

not manifest until mid-life. A slow progression of the symptoms in PD cases over many years supports a slow rate of cell death. In disagreement with this hypothesis is the observation by McGeer and colleagues (1988) of an active neuropathological process at the time of death in PD patients, where the number of cells undergoing phagocytosis is up to six times greater than that seen in normal brains. McGeer and coworkers estimated a loss of 12000 to 43000 SNc neurons per year in the diseased brains. Considering a normal cell complement of around 450000 neurons in the SNc, these authors predicted a complete depletion of cells in the SNc in not many years, were this a long-term process. Instead, they argued that the marked neurodegeneration observed in PD is the result of an active process commencing late in life, and that suppression of the ongoing process at these later stages could still interfere with the progression of the disease. On the other hand, the process observed by McGeer and his group could have coincided with the first years following disease onset in the individuals, as a fast progression of the symptoms has been reported to occur at these early stages; in general, this progression is markedly slower in later stages.

1.3 Identification of disease-causing genes.

The isolation of genes which determine susceptibility to or occurrence of disease states is of prime importance. Extensive work has been undertaken to date to identify disease genes in man. As mentioned earlier, these efforts often encounter serious problems which result from the complex inheritance patterns of polygenic traits and the impossibility to actively control the genetic background. These obstacles can sometimes be overcome through the study of highly inbred populations or, very importantly, by carrying out research on suitable animal models.

Two main approaches have been followed to identify disease-associated genes. One, 'Functional Cloning', requires knowledge of the biochemical defect leading to the disease process, and therefore of the function and/or the protein product of the sought-for gene. The other is known as 'Positional Cloning' since identification of the gene relies solely on its position in the genome.

1.3.1 Positional cloning approaches.

Positional cloning methods rely on finding the region containing the disease-causing gene through the establishment of linkage of genetic markers to the gene of interest. These analyses are greatly facilitated when a cytogenetic rearrangement, such as a deletion, has taken place. In the absence of such event, a genome-wide search is required to establish genetic linkage of at least one map element to the target locus. A requisite for any marker to be useful in these studies is that it is informative, that is, that different allelic forms of the marker exist in the population; these variations are in general phenotypically neutral.

Recently, new approaches have been developed which allow the isolation of sequences in the vicinity of a disease gene(s) by scanning the whole genome in a single pass, instead of sequentially assessing segregation patterns of individual loci in relation to the disease phenotype (Brown, 1994). These are methods of choice when a 'functional cloning' approach is not possible and a dense genetic map of the organism under study is not available, as identification of nearby sequences does not depend on genetic markers having been established within the region (Jonsson and Weissman, 1995). These approaches may yield better results than other available techniques when several loci are involved in the phenotype. Representational difference analysis (RDA) (Lisitsyn *et al.*, 1993; Lisitsyn, 1995) is based on the subtraction of common sequences between two DNA samples, and the selective amplification by PCR of restriction fragments that are present in one of them ('tester') but differ or do not exist in the other ('driver'). Since a high genetic variation is oberved in most populations, this method normally results in the identification of differences throughout the genome. In order to find variations between the genomes in the region carrying the disease gene, both DNA samples must contain the same combination of alleles except in the region of interest (Genetically Directed Representational Difference Analysis, GDRDA). This is achieved by employing congenic strains of laboratory animals or by making use of transmission genetics. The rationale behind the latter is that a pool of DNAs extracted from affected individuals will present the same allelic variation which is found in the normal population all throughout the genome except in a small region where the disease gene maps. This technique has already been shown to give excellent results when searching for sequences which are specific to regions carrying recessively acting genes (Lisitsyn *et al.*, 1994). An important disadvantage of this technique is, however, that it can only

detect changes in the genome which cause gross differences in the size of restriction fragments. Therefore, very small mutations outside restriction site sequences can not be detected with this method. A second technique is known as genomic mismatch scanning (GMS) (Nelson *et al.*, 1993; Brown, 1994). This is based on the ability of mismatch repair proteins to recognise single base pair mismatches in heteroduplexes formed between allelic sequences derived from each of two genomes. Both genomes are digested with a restriction enzyme which yields DNA fragments of a large enough size as to ensure the presence of base pair differences between those sequences which were not inherited from the same source. The combined action of several enzymes results in the tagging of homoduplexes (i.e. both strands derive from the same genome) and of heteroduplexes which contain mismatches, so that only mismatch-free hybrid molecules survive treatment with the enzyme *Exo*III. These hybrids are subsequently purified, labelled, and used to probe ordered arrays of DNAs which represent the entire genome. Regions inherited from the same source are thus identified. Excellent results have been obtained by Nelson and colleagues (1993) when applying this technique to yeast. Application of this method to man faces the greater complexity and the abundance of repetitive elements in this genome. Success of the GMS technique on the human genome would permit the detection of sequences inherited in common by the affected individuals in one or various pedigrees. A considerable fraction of these sequences would then most likely correlate to disease gene(s)-regions.

1.3.1.1 Genetic markers.

Different types of markers can be used when analysing inheritance patterns in populations. Variations at *biochemical* and *immunological markers* are most commonly studied by means of electrophoretic and histochemical staining procedures and immunoassays (Adams *et al.*, 1984; Adams and van Zutphen, 1990; Günther, 1990a). A disadvantage which is often encountered by these tests is the difficulty to correctly characterise some allelic variants, limiting their usefulness for genetic typing. In some cases, the results obtained following an electrophoretic screen can be affected by the quality of the tissue. Difficulties can also arise when selecting electrophoretic conditions and the support medium, or when attempting to interpretate complicated banding patterns (e.g. those observed in gels stained for aldehyde dehydrogenase (AHD) activity in vertebrates). Moreover, there are cases where heterozygotes may be difficult

or impossible to distinguish from either of the parental or the sibling homozygotes. In contrast to these, the allelic forms of many *morphological* and *physiological markers* can be easily determined. This is the case of some *colour* and *behaviour markers*, where the presence or absence of a given allele in an organism can be established by simply attending to some particular phenotypic features exhibited by the individual (Hedrich, 1990a; Serikawa *et al.*, 1992).

Identification of allelic forms corresponding to other types of genetic markers relies on the direct analysis of the DNA sequence. Variation among individuals can result from a) differences at restriction site sequences; b) differences in the number of repeat units which constitute a repetitive sequence; or c) other changes in the DNA sequence which are mainly found in non-coding regions.

Some differences in the DNA sequence can result in either the creation or the disruption of a restriction site. Thus, probes that bind to this region of the genome demonstrate a difference in the size of the respective hybridising fragment when the genomic DNAs are digested with the particular restriction endonuclease. These markers are known as *Restriction Fragment Length Polymorphisms* (*RFLPs*) (Botstein *et al.*, 1980; Günther, 1990b). RFLP maps have been constructed for man (Donis-Keller *et al.*, 1987), and other organisms (Nam *et al.*, 1989; Burr and Burr, 1991). A disadvantage of RFLPs is that only a few allelic forms are normally observed in each case. In the instances where one of these alleles is present in the majority of the individuals, the utility of the marker in genetic typing is further reduced. Moreover, digestion with 20-30 restriction endonuclesases may be required to detect a difference between two genomes; this conveys a considerable workload since RFLP analyses involve blotting and hybridisation protocols.

Variable number of tandem repeats (*VNTRs*) are, as the name indicates, repetitive sequences in the genome which may consist of a different number of repeat units in different individuals (Charlesworth *et al.*, 1994). This variation leads to differences in the length of these genomic regions. These markers are also known as *minisatellites* (Jeffreys *et al.*, 1985; Pravenec *et al.*, 1996), and are more informative than RFLPs. In the cases where the minisatellite repeat unit is of a considerable size (some can be up to 50 bp) and/or the number of repeats is large, analysis of the VNTR is carried out as in RFLP assays. Minisatellites spanning smaller DNA regions can be studied by using the polymerase chain reaction (PCR). *Microsatellites* (Litt and Luty,

1989; Weber and May, 1989; Love *et al.*, 1990) are a second type of tandem repeat sequences where the unit size is 1 to 6 nucleotides long. Therefore, allelic forms of these markers can be easily typed by amplification with PCR primers which bind to unique sequences flanking the repeat. These markers have also been called *sequence tagged microsatellite sites* (*STMSs*; Serikawa *et al.*, 1992), *simple sequence length polymorphisms* (*SSLPs*; Jacob *et al.*, 1995), *simple sequence repeats* (*SSR*; Jacob *et al.*, 1991; Goldmuntz *et al.*, 1993), or *simple tandem repeats* (*STRs*; Smith, 1995; Edwards *et al.*, 1991).

Different approaches can be followed in order to detect other variations in the DNA sequence which do not result in divergent restriction patterns between the genomes (e.g. deletions, insertions, nucleotide mutations at restriction sites) and which do not correlate to changes in the number of repeat units in VNTRs. Very efficient methods are based on the amplification of the genomic region under study and the analysis of the PCR products obtained. Small changes in the length of the products can be detected following electrophoresis on denaturing sequencing gels. More subtle changes such as single nucleotide substitutions require highly sensitive methods such as *denaturing gradient gel electrophoresis* (*DGGE*) and *single-strand conformation polymorphism* (*SSCP*) analysis. The former is based on the melting characteristics of DNA (Fischer and Lerman, 1983; Sheffield *et al.*, 1989; Traystman *et al.*, 1990); SSCP analysis (Orita *et al.*, 1989; Levitt, 1991; Russell, 1994) is based on the folded conformations that single DNA strands adopt when running through non-denaturing polyacrylamide gels. Both features depend on the DNA sequence.

Noncoding sequences in the genome are an excellent source of informative markers. Variations in the DNA sequence in these regions are in general phenotypically neutral and are thought to occur frequently (approximately once in every few hundred base pairs). The transcribed 3' untranslated regions of eukaryotic genes are a very good target for the establishment of genetic markers. These sequences often span several hundreds of base pairs; SSCP analyses have shown that multiple polymorphisms can be identified within a single region (Levitt, 1991). Development of these markers requires knowledge of the transcribed 3' untranslated sequence, and this in turn implies the availability of the cDNA sequence of the corresponding gene. Therefore, markers established within these regions can be used to locate the respective genes on the genetic maps. This is not always the case with microsatellite markers, since many of

these are not transcribed (Moore *et al.*, 1991; Levitt, 1991); assignment of a microsatellite to a gene requires then the availability of the genomic sequence.

Microsatellite markers.

This type of polymorphisms are considered to result from unequal meiotic exchanges or slippage of the DNA polymerase during replication (Litt and Luty, 1989; Levinson and Gutman, 1987; Schloetterer and Tautz, 1992). The presence of microsatellites in all the eukaryotes tested to date and their absence in bacteria suggests that these sequences possess some function. Microsatellites of the type $(CA)_n$ and $(TC)_n$ ($n \geq 14$) have only been found in nuclear DNA, which strongly supports the action of one or more mechanisms unique to eukaryotic nuclear DNA to retain these repeats during evolution (Gross and Garrard, 1986). It has been suggested that microsatellite sequences could act as hotspots for recombination (Slightom *et al.*, 1980; Treco *et al.*, 1985). Other reports do not support such a role (Stallings *et al.*, 1991); it could happen that presence of the repeat sequence is necessary but not sufficient for a recombinational hotspot to occur. Naylor and Clark (1990) demonstrated a down-regulatory role for a $(GT)_n$ repeat upstream of the rat prolactin gene, and Hamada and colleagues (1984) showed that a similar repeat could enhance transcription when inserted into plasmids carrying the choramphenicol acetyltransferase (CAT) reporter gene. A year later Thomas and coworkers (1985) reported the presence of microsatellite regions in nontranscribed sequences in rDNA; these repeats appeared to be absent from the transcription unit. The group established a model postulating that some repetitive sequences could be involved in transcription enhancer activities through conformational and other structural influences. It has been hypothesized that these repeats could play a role in packaging and condensing DNA into chromosomes in eukaryotes. $(GT)_n$ repeats are much more frequent in euchromatin than in heterochromatin and might be providing a DNA conformation that can be condensed and decondensed during the different phases of the cell cycle (Stallings *et al.*, 1991). Trinucleotide repeats coding for polyglutamine or polyproline tracts have been identified in protein coding regions (Wharton *et al.*, 1985; Lindahl *et al.*, 1990). These microsatellites seem to play a role in the normal function of the protein. Evidence is being found that supports that variation in the number of repeats at SSLP loci can cause quantitative changes in gene expression and function, and that these changes may have an effect on the phenotype (Lindahl *et*

al., 1990). Consequently, it has been proposed that microsatellites may be a major source of the extensive quantitative genetic variation observed in the populations (Kashi *et al.*, 1997).

Microsatellites have rapidly become the basis in the construction of dense genetic maps (Weissenbach *et al.*, 1992; Gyapay *et al.*, 1994; Dietrich *et al.*, 1994; Serikawa *et al.*, 1992; Jacob *et al.*, 1995) since they offer important advantages in comparison to other markers. Some of these advantages are presented below:

a) SSLPs are abundant and widely spread throughout the eukaryotic genomes. The $(CA)_n$ motif appears as the most common repeat, with approximately 10^5 copies in the genomes of man, mouse and rat (Beckmann and Weber, 1992; Love *et al.*, 1990; Serikawa *et al.*, 1992). In addition, microsatellites have shown very little tendency to cluster (Luty *et al.*, 1990), in contrast to some hypervariable minisatellite loci which cluster at the proterminal ends of chromosomes (Royle *et al.*, 1988).

b) Great variability has been observed at microsatellite loci. In 1990, Weber reported that the number of alleles identified at a given $(CA)_n$-microsatellite locus in man correlates to the total length of the repeat. The same trend was observed with tri- and tetranucleotide repeats (Edwards *et al.*, 1991). Similar results were obtained with mouse (Cornall *et al.*, 1991) and cattle (Moore *et al.*, 1991). Studies by Serikawa and colleagues (1992) on several rat strains indicated that di- to tetranucleotide repeats which consist of 10 or more repeat units have a high probability of showing allelic variants.

c) Numerous studies have shown these markers to be stable through many generations (Weber and May, 1989; Love *et al.*, 1990; Serikawa *et al.*, 1992; Pravenec *et al.*, 1996), and not subject to frequent mutations as has been reported for certain minisatellite polymorphisms (Jeffreys *et al.*, 1987; Jeffreys *et al.*, 1988).

d) Newly developed markers can be made readily available to other labs by simply releasing the sequences of the PCR primers.

e) Given the extreme sensitivity of PCR (Saiki *et al.*, 1988), typing of microsatellite loci requires much smaller amounts of DNA than those needed to carry out blotting and hybridisation procedures.

f) Labelling of the PCR primers allows the multiplexing of different PCR reactions. Adequate sequencers and computer software are now available which permit

automation of the typing process, resulting in great amounts of good quality typing data being rapidly obtained and processed (Edwards *et al.*, 1991; Levitt *et al.*, 1994; Smith, 1995).

1.3.1.2 Physical mapping.

Identification of the gene of interest requires that the genetic map which has been established for that region of the genome be developed into a physical map. This translates into the isolation of genomic fragments which cover the region. These sequences are normally identified following screening of genomic DNA libraries. Initially, screening is carried out with the available genetic markers known to map closest to the sought-for gene. End-fragment sequences are then obtained from the positive clones and used to screen the library for new genomic sequences, and so forth. Analysis of restriction patterns and of the presence or absence of unique sequences identified within each of the isolated genomic fragments allows the alignment of overlapping clones into larger units or *contigs*. This process is known as *chromosome walking* (Feil *et al.*, 1990a; Albertsen et al., 1994; Nehls *et al.*, 1995). Some important considerations in the establishment of a physical map are discussed below.

Genomic DNA libraries.

Different cloning systems can be used in the construction of DNA libraries (Monaco and Larin, 1994):

-*Cosmid vectors*, based on the *cos* site-mediated packaging of recombinant molecules into bacteriophage lambda particles.

-*Yeast artificial chromosomes (YACs)*, linear vectors consisting of a telomeric sequence at each end, a centromere, an origin of replication, and yeast selectable markers to allow propagation in yeast.

-*P1 vector*, based on bacteriophage P1 elements.

-*Bacterial artificial chromosomes (BACs)*, based on the F-factor plasmid.

-*P1-derived artificial chromosomes (PACs)*, based on both the P1 and the F-factor systems.

Each cloning system presents an array of advantages and disadvantages. Thus the establishment of contigs across specific regions of the genome is frequently

favoured by the use of more than one type of genomic DNA library. YACs (Burke *et al.*, 1987) offer the greatest advantage in terms of insert length, since these chromosomes can carry DNA fragments of several hundred kilobase pairs in size. In comparison, cosmids allow for a maximum insert size of 40-45 kb; P1 vectors (Sternberg, 1990a, 1990b), BACs (Shizuya *et al.*, 1992) and PACs (Ioannou *et al.*, 1994) can carry DNA fragments of 70-100 kb, up to 300 kb, and 100-300 kb, respectively. A greater insert size results in a smaller number of clones needed to cover a genomic region. However, the YAC system poses some important problems such as low transformation efficiencies when preparing libraries, a high rate of chimaerism and instability of the genomic insert (Nehls *et al.*, 1995; Ochman and Buckholtz, 1995; Harshman *et al.*, 1995), and difficulties in purifying large amounts of YAC DNA from the 15 Mb of yeast chromosomal background. Modifications have been introduced to at least partially overcome some of these problems. YAC libraries have been constructed in recombination-deficient hosts (Haldi *et al.*, 1994), as chimaerism is considered to arise mainly from recombination events in the host cell (Nehls *et al.*, 1995). Also, some host strains have been identified which yield higher transformation rates than others (Burke *et al.*, 1987).

The main improvement has been, however, the use of other vectors such as P1 (Sternberg, 1990a; Pierce and Sternberg, 1992) and BACs (Shizuya *et al.*, 1992) in the construction of new libraries. These systems, while still enabling the cloning of large genomic fragments, present a greater insert stability, mainly due to their maintenance in the host as low copy number plasmids; chimaerism does not appear to be a frequent phenomenon either (Nehls *et al.*, 1995; Harshman *et al.*, 1995). It has been observed that YAC instability often correlates to the genomic region in particular (Nehls *et al.*, 1995) and that some sequences are frequently underrepresented or even absent in this type of libraries. In contrast, some of these fragments have been readily identified in equivalent P1 libraries (Albertsen *et al.*, 1994). Also, higher transformation efficiencies are achieved (Pierce *et al.*, 1992). Construction of P1 libraries involves the preparation of different packaging extracts to be used at different stages in the formation of infectious particles. These steps are eliminated when using BACs, since efficient transformation is then carried out by electroporation (Sheng *et al.*, 1995). However, this latter system lacks the positive selection for recombinant clones offered by the P1 vector p*Ad*10*sac*BII (Pierce and Sternberg, 1992). Also, the DNA yields obtained are

very low. More recently, the PAC system was developed, which combines the advantages from both the P1 and the BAC systems (Ioannou *et al.*, 1994).

Detection of unstable and/or chimaeric clones.

Insert instability and the occurrence of chimaerism are a serious problem when carrying out chromosome walking. It is therefore crucial that these events are detected as soon as possible. Changes such as deletions can be identified through comparisons of the marker content and the restriction pattern obtained from the insert with those observed in the genomic region of interest as well as through correlations to other clones (cosmid clones, for example) which have already been confirmed and characterised (Feil *et al.*, 1990b; Wada *et al.*, 1990). Chimaerism can be assessed by confirming the chromosomal location of unique sequences identified in the genomic insert. This can be done by screening *somatic cell hybrids.* These are clones which have retained a different set of chromosomes or chromosomal regions from the species under study on a genetic background derived from another species. Any sequence can be used to screen a panel of these clones as long as the products obtained from both species can be distinguished from each other. Correlation of the presence or absence of the product from the species of interest with the presence or absence of chromosomes (or chromosomal regions) from this species in each hybrid clone allows the mapping of this sequence. A second approach allows one to map whole inserts; this method is known as *fluorescent in situ hybridisation* or *FISH.* It consists on the labelling of the genomic insert (or of the whole clone) with a fluorescent dye and *in situ* hybridisation on metaphase spreads. Furthermore, more than one fragment can be labelled and used at a time by employing different fluorescent dyes; this allows one to confirm the order of the different clones in relation to each other (Albertsen *et al.*, 1994).

Correlation of the genetic map to the physical map.

Genetic maps are constructed on the basis of the recombination frequencies observed between loci. Mathematical functions have been designed which allow more accurate calculations, taking into account factors such as the simultaneous occurrence of a cross-over in each of two contiguous intervals in the chromosome (Green, 1981).

Genetic distances are normally used to estimate the corresponding physical distances. The relation between both parameters varies according to the species (e.g. 1 cM is considered to be the equivalent to 1 Mb in humans, whereas 2.7 kb is considered

the equivalent to 1 cM in yeast). However, these are general estimates; there are regions in the genome where a small physical interval correlates to a region of relatively high meiotic recombination frequency, and *vice versa*. The former can be seen at the telomeric regions of many chromosomes; in contrast, centromeric regions often present recombination frequencies which are much lower than those anticipated.

Physical maps are constructed once a genetic map has been established which is not larger than a few cM in length (Albertsen *et al.*, 1994; Pawar *et al.*, 1995). Yet, this normally represents too great a distance to start searching for individual genes in the region, as a 1 cM region can contain several tens of genes. Therefore, further reduction of this region is in general attempted by searching for recombination events between newly identified informative marker loci and the gene of interest. Identification of recombinants leads to a higher resolution of both the genetic and the physical maps. New flanking markers are thus defined, with the concomitant elimination of outside sequences from any further analysis (Zhou *et al.*, 1995). The observations made by Harshman and colleagues (1995) underline the importance of establishing fine genetic maps prior to searching for candidate genes in the region. Quite a large number of techniques are currently available to carry out these searches. Numerous groups have reported the isolation of disease-associated genes, normally following a combination of these methods (Rommens *et al.*, 1989; Miki *et al.*, 1994; Wang *et al.*, 1996). These approaches are discussed in Chapter 8. An overview of the steps involved in the positional cloning of disease genes is shown in Figure 1-4.

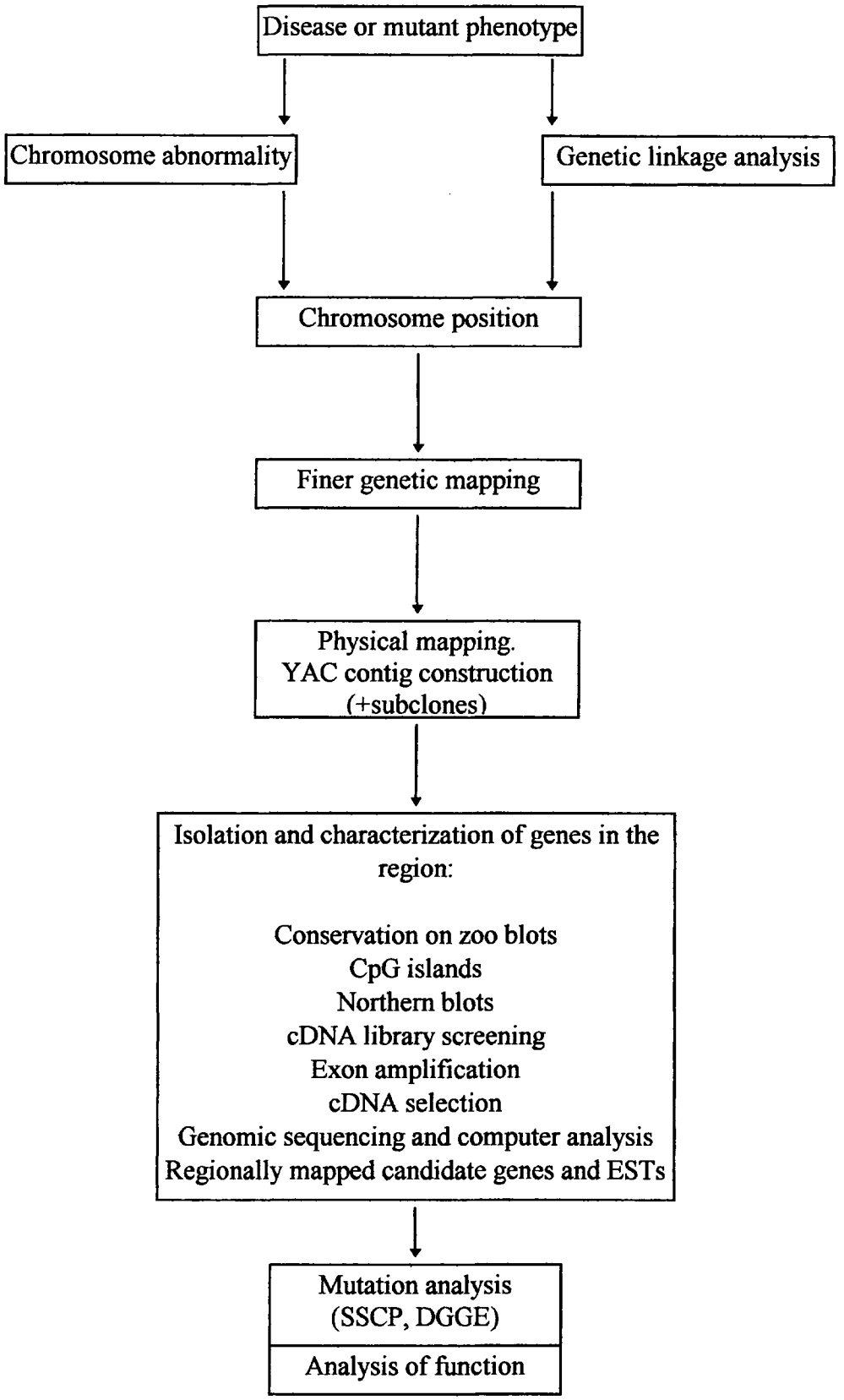


FIGURE 1-4. Diagram showing the steps commonly involved in positional cloning. Modified from Bailey *et al.* (1997) and Monaco (1994).

1.4 Aims of the project.

The ultimate aim of this project is the positional cloning of the *agu* gene. This will be the basis for subsequent studies on the mechanism of action of the mutation carried by AS/AGU animals and on a possible correlation of PD in man to mutations in the human homologue of the *agu* gene or in other genes involved in the same biochemical pathway. Likewise, a putative role in neurodegeneration might then be assigned to the environmental factors which act within this pathway.

My work was directed towards obtaining a sufficiently exact genetic location of the *agu* gene to allow the initiation of physical mapping approaches. Several steps were followed:

1._ Identification of informative markers between AS/AGU and other strains of rat. Development and optimisation of an efficient, reliable system to screen the rat genome.

2._ Selection of the most informative crosses to the AS/AGU sub-strain.

3._ Genetic scan of the rat genome. Identification of the chromosomal location of the *agu* gene.

4._ Mapping of the *agu* gene to a genetic interval of 1 cM or less.

CHAPTER 2

Materials and Methods.

2.1 Chemicals and biochemicals.

Agarose MP (multipurpose), chloramphenicol, lysozyme and 5-bromo-4-chloro-3-indolyl-β-galactopyranoside (X-gal) were obtained from Boehringer Mannheim, Lewes, UK. MetaPhor agarose was obtained from FMC Bioproducts, Rockland, USA. The DNA molecular weight markers were obtained from BRL, Gibco Ltd., Paisley, UK. Lambda phage DNA was obtained from Sigma Chemical Co., Poole, UK.

Kanamycin, ampicillin, streptomycin sulfate, isopropyl β-D-thiogalacto-pyranoside (IPTG), polyethylene glycol (PEG), ethidium bromide, blue dextran, phenol red, N,N,N',N'-tetramethylethylene diamine (TEMED), N,N-dimethylformamide, and urea were obtained from Sigma Chemical Co., Poole, UK.

Formamide was obtained from Aldrich Chemical Co., Gillingham, UK. Anachem, Luton, UK supplied 6%(w/v) sequencing solution (19:1) and 40% (w/v) acrylamide/bis-acrylamide stock solution (19:1).

Radiochemicals were bought from DuPont/ NEN Research products, Herts, UK.

All other chemicals were of analytical reagent grade and were mainly obtained from either BDH Chemical Ltd., Poole, UK, or from Fisher Scientific, Loughborough, UK.

2.2 Oligonucleotides, enzymes, kits and other materials.

All the enzymes were obtained from BRL, Gibco Ltd., Paisley, UK, except : *Taq* DNA polymerase and T4 polynucleotide kinase, obtained from Promega Ltd., Southampton, UK; RNAse , obtained from Sigma Chemical Co., Poole, UK; lysozyme, bought from Boehringer Mannheim, Lewes, UK; shrimp alkaline phosphatase (SAP), obtained from USB, Amersham International p/c, Little Chalfont, UK.

Oligonucleotides were either synthesized in the department, ordered from BRL, Gibco Ltd., Paisley, UK, or bought from Research Genetics Inc. (Huntsville, Alabama, USA). Oligonucleotides prepared in the department were synthesized on either an ABI 391 or an ABI 392 DNA/RNA synthesizers (Applied Biosystems Inc., Hertford, UK). Some rat DNAs were obtained from Harlan Olac (Oxon, UK).

The Puregene DNA Isolation kit was bought from Flowgen Instruments Ltd., Lichfield, UK. The panel of somatic cell hybrids was kindly provided by Professor Claude Szpirer, Laboratoire de Genetique, Universite Libre de Bruxelles, Belgium.

The P1 library was a gift from R. J. MacDonald, Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas.

The pDT Blue cloning system was bought from ams Biotechnology, Witney Oxon, UK.

2.3 Nucleic acid isolation and manipulation.

2.3.1 Commonly used buffers.

a) *TE buffer*.

10 mM Tris·HCl, pH 8.0; 1 mM EDTA.

b) *TBE buffer (5x)*.

54 g Tris, 27.5 g boric acid, 20 ml 0.5 M EDTA (pH 8.0), made up to 1 litre in distilled water. The pH should be c. 8.3.

c) *TAE buffer (50x)*.

242 g Tris, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA (pH 8.0), made up to 1litre in distilled water.

d) *Agarose gel loading buffer (6x)*.

4 g of sucrose dissolved in 6 ml TE; 2.5 mg bromophenol blue and 2.5 mg or less of xylene cyanol FF added to this solution. Volume made up to 10 ml with distilled water.

e) *Formamide/EDTA/XC/BPB gel-loading buffer*.

10 ml of formamide, 10 mg of xylene cyanol FF, 10 mg of bromophenol blue and 200 µl of 0.5 M EDTA (pH 8.0).

These and other standard solutions were all prepared according to Sambrook *et al.* (1989).

2.3.2 DNA isolation.

2.3.2.1 Genomic DNA isolation.

Most of the genomic DNAs handled during the course of this project were used as templates in PCR reactions. Initially, a few DNA samples were obtained from Harlan Olac (Oxon, UK). Thenceforth, all the DNAs were prepared in house, the first samples being extracted from frozen spleens with the use of liquid nitrogen (Method A). The majority of the DNAs were prepared using the Puregene DNA isolation kit from Flowgen Instruments Ltd. (Method B). Once in solution, the DNAs were all kept at 4°C.

a) *Method A.*

The frozen tissue was ground with continuous addition of liquid nitrogen. It was then transferred to a precooled 50 ml Falcon tube and the liquid nitrogen let to evaporate completely.

9.5 ml of NET buffer (100 mM NaCl, 10 mM Tris-HCl (pH8.0), 25 mM EDTA (pH8.0)) was gently added onto one side of the tube, resulting in a highly viscous solution. 0.5 ml of 10% SDS and 50 µl of 20 $mg.ml^{-1}$ proteinase K was then added and the mixture incubated at 50°C for some time.

The DNA was extracted by adding an equal volume of phenol:chloroform (ratio 1:1) to the sample, leaving the tube on a roller for 15-20 minutes. The sample was then centrifuged at 2000 rpm for 10 minutes, and the upper aqueous phase transferred to a new tube by using a pasteur pipette cut at its end to avoid shearing the DNA. The phenol/chloroform extraction was repeated. The DNA was finally precipitated with 2 volumes of absolute ethanol in the presence of 0.3 M sodium acetate, pH 5.2. The DNA was extracted by swirling it round the end of a pasteur pipette bent at its tip. After being immersed in 70% ethanol, the pipette was left in 1 ml of TE and the DNA gently resuspended. 2 µl of RNase (10 $mg.ml^{-1}$) was added to the sample and this was incubated at 37°C for 30 minutes. Another phenol/chloroform extraction step followed. The DNA was again precipitated and resuspended in 0.5-1 ml of TE. If required, the tube was left at 65°C for 15-30 minutes to help the DNA get into solution.

b) *Method B.*

The DNAs were extracted using the Puregene DNA isolation kit, following the vendor's instructions.

10-20 mg of frozen tissue (normally spleen) was homogenized in 600 µl of cell lysis solution, with the help of disposable pestles. Lysates were incubated at 65°C for 15-60 minutes. Following this step, 2µl of RNase A solution (4 $mg.ml^{-1}$) was added to each lysate; the tubes were inverted several times and then left at 37°C for 15-30 minutes. It followed a protein precipitation step where, having let the samples cool to room temperature, 200 µl of protein precipitation solution was added to each lysate. The mixtures were vortexed vigorously for 20 seconds and then centrifuged at 14000xg for 3 minutes. A tight pellet was expected; in those instances in which the protein pellet appeared as loose, the samples were again vortexed, left on ice for five minutes, and centrifuged at 14000xg for 5 minutes. The pellet consisted of the precipitated proteins; the supernatant, containing the DNA, was transferred into a clean 1.5 ml eppendorf tube, and the DNA precipitated with 600 µl of isopropanol. The tubes were inverted gently until white threads of DNA were clearly visible. The samples were then centrifuged at 14000xg for 2 minutes, with the DNA appearing at the side of the tube as a small white pellet. The supernatant was discarded and the DNA pellet washed with 200 µl of 70% ethanol. The tubes were centrifuged again at 14000xg for 2 minutes, the ethanol carefully discarded, as the pellets might be loose, and the samples dried for 5-10 minutes. Finally, 100 µl of TE was added to each tube, and the DNA allowed to get into solution by leaving the samples either at room temperature overnight or at 65°C for 1 hour. Most DNAs were resuspended at room temperature but the quality of the samples did not differ from that of the DNAs resuspended at 65°C.

2.3.2.2 Plasmid DNA isolation.

3 ml LB-cultures were set up and grown overnight, as explained below. The cultures were then spun down in a microcentrifuge at 14000 rpm for one minute. The supernatant was removed by aspiration and the bacterial pellet thoroughly resuspended in 200 µl of GTE buffer (50 mM glucose; 25 mM Tris, pH8.0; 10 mM EDTA, pH8.0). 300 µl of freshly prepared 0.2 N NaOH/1%SDS was added to the tube; the contents were gently mixed by inversion and incubated on ice for five minutes. The solution was neutralized by adding 300 µl of 3 M potassium acetate, pH4.8; the tube was again

gently inverted and incubated on ice for five minutes. The cell debris and most of the chromosomal DNA was removed by centrifugation at 14000 rpm for 10 minutes and the supernantant transferred to a clean tube, to which 2 µl of RNase A (10 $mg.ml^{-1}$) was added. The sample was incubated at 37°C for 30 minutes. Two consecutive extractions with 400 µl of chloroform followed. The nucleic acid was precipitated by adding an equal volume of 100% isopropanol and immediately centrifuging the tube for 10 minutes at room temperature. The DNA pellet was washed with 500 µl of 70% ethanol and air dried. The DNA was normally resuspended in 40-50 µl dH_2O at 37°C for 30 minutes.

2.3.2.3 P1 DNA isolation.

As for small scale plasmid DNA isolations, P1 DNA was prepared according to protocols based on the alkaline lysis procedure (Birnboim and Doly, 1979). Several methods were used:

I._ Birnboim and Doly method.

Applied to 25 ml and 50 ml culture volumes.

The cells were pelleted in a centrifuge at 8000 rpm for 10 minutes and resuspended in 25 mM Tris-HCl (pH8.0), 50 mM glucose, 10 mM EDTA; 2 $mg.ml^{-1}$ of lysozyme was also added in most isolations. The resuspended cells were incubated at room temperature for 10 minutes. 0.2 N NaOH/1% SDS was added to the sample and the mixture was incubated on ice for 10 minutes. Finally, potassium acetate, pH4.8, was added and the reaction incubated on ice for an additional 10 minutes. The volumes used in each case were: 250 µl of lysis solution, 450 µl of 0.2 N NaOH/1%SDS and 350 µl of 3 M potassium acetate (pH4.8), for 25 ml culture volumes. When working with culture volumes of 50 ml, 2 ml of lysis solution, 4 ml of 0.2 N NaOH/1%SDS and 3 ml of 3 M potassium acetate (pH4.8) was used. Following the neutralisation step, the lysate was centrifuged at 12000-15000 rpm for 15 minutes and the supernatant transferred into a clean tube and extracted with an equal volume of phenol/chloroform. The P1 DNA was precipitated by adding 2 volumes of 100% ethanol. The DNA pellet was washed with 1-5 ml of 70% ethanol, air dried, resuspended in TE and treated with RNase at 37°C for 1 hour. DNA pellets obtained from 25 ml culture volumes were resuspended in 24-40 µl of TE and treated with 0.4 µg of RNase. DNA pellets from 50 ml cultures were resuspended in 500 µl of TE and treated with 4 µg of RNase. Following RNase

treatment, the DNA was re-precipitated with 0.3 M sodium acetate (pH5.2) and 2 volumes of 100% ethanol. The pellet was washed with 1 ml of 70% ethanol, air dried, and resuspended in a final volume of 50-150 µl of TE.

II._ DNA purification of P1 DNA using Qiagen-tip 100.

Culture volumes of 500 ml were used when applying this method.

The Qiagen *Midi* purification protocol was followed with the recommended volumes of P1, P2, P3 and QC buffers being doubled. The DNA pellet was resuspended in 500 µl of TE.

III._ Modified Qiagen Midi protocol I.- Phenol/chloroform extraction.

This was a variant of the protocol mentioned above. The supernatant obtained after removing the cell debris and most of the chromosomal DNA following neutralisation with 3 M potassium acetate (pH4.8), was extracted with an equal volume of phenol/chloroform instead of being left to flow through a Qiagen-tip 100. After extraction, the DNA was precipitated with an equal volume of 100% isopropanol and centrifuged at 15000 rpm for 30 minutes at 4°C. The DNA pellet was washed with 2 ml of 70% ethanol, air dried and resuspended in 400 µl of TE.

IV._ Modified Qiagen Midi protocol II.- CsCl gradient purification.

Culture volumes of 250 ml were used when applying this method of P1 DNA purification.

The Qiagen Midi protocol was followed to the point where the cell debris and most of the chromosomal DNA had been removed by centrifugation, following the neutralisation step. The clear supernatant was then transferred to a clean tube and the nucleic acid precipitated with 1 volume of 100% isopropanol. The pellet was washed with 70% ethanol, air dried, and gently resuspended in 3.5 ml of TE. The DNA was purified through a CsCl gradient, as explained in Kimmerly *et al.* (1994). The pellet was resuspended in 100-200 µl of TE.

In some instances where the P1 DNA had been isolated through a typical Birnboim and Doly method (I and III), this DNA was further purified through a *proteinase K treatment* (Southard-Smith *et al.*, 1994). A final concentration of 0.05 $\mu g.ml^{-1}$ RNase was added to the DNA solution, and the sample incubated at 37°C for 30

minutes. EDTA, SDS, and proteinase K were then added to final concentrations of 5 mM, 0.1%, and 0.2 $mg.ml^{-1}$, respectively. This mixture was incubated at 55°C for 1 hour. The plasmid DNA was extracted once with phenol and once with chloroform. It was then precipitated with 2 volumes of ethanol, in the presence of 0.3 M sodium acetate, pH 7.0, and finally resuspended in TE.

Wide-bore tips were employed in all the P1 DNA manipulations, to avoid shearing the DNA. The DNAs were resuspended at 37°C for 10-30 minutes and not by pipetting the solution up and down. All the P1 DNAs were stably kept at 4°C.

2.3.3 Oligonucleotide synthesis.

2.3.3.1 Design of primers for PCR and sequencing reactions.

Oligonucleotides which were 17 to 25 nucleotides long and had melting temperatures (Tm ≈ {(A+T) x 2°C + (G+C) x 4°C}; Suggs *et al.*, 1981) between 50°C and 65°C were selected. Both PCR primers in a set were designed to have similar Tm, and not to be complementary at their 3' ends. The programs GeneJockey II and MacVector were occasionally used to aid the design of the primer sets.

2.3.3.2 Synthesis and deprotection of oligonucleotides.

The oligonucleotides prepared in house were synthesized on either an ABI 391 or on an ABI 392 DNA/RNA synthesizers.

The beads were extracted from the synthesis columns and then thoroughly resuspended in 1 ml of fresh liquid ammonium. The samples were left at room temperature for 1 to 2 hours and then centrifuged at 14000 rpm for 15-20 minutes, to remove the beads. The supernatants, carrying the DNA, were transferred to clean Nunc tubes; 1 ml of liquid ammonium was added. The tubes were left at 50°C for 8-17 hours. An aliquot of each sample was transferred to another tube and the DNA precipitated with 2 volumes of 100% ethanol, in the presence of 0.3 M sodium acetate, pH 5.2. Precipitation was carried out at -70°C for 2 or more hours. The DNA was initially resuspended in 100 µl of dH_2O. These were the concentrated stocks; the DNA concentrations were calculated in each case and primer working stocks prepared by making the appropriate dilutions. Both the concentrated and the diluted primer stocks were kept at -20°C. The remaining ammonium stocks were kept at -70°C.

2.3.4 Quantitation of nucleic acid.

In general, nucleic acid concentrations were determined spectrophotometrically at A=260 nm. In a 1 cm path length an absorbance value of 1.0 corresponds to 50 μg/ml for double stranded DNA and 20 μg/ml for oligonucleotides.

Spectrophotometric quantitation of plasmid DNA (pGEM-3Zf(+), pBluescript II KS+, some P1 DNAs) was not accurate, since the presence of contaminating chromosomal DNA in the sample would result in higher OD readings. Therefore, approximate concentrations were calculated by running an aliquot of the plasmid DNA against a λ/*Hin*dIII marker, loaded at a known concentration, and comparing the intensities of the bands on the gel. Approximate concentrations of P1 DNAs were calculated by adding the estimated concentrations of each of the bands obtained following single or double digestion of the DNA with restriction endonucleases.

2.3.5 P1 DNA Library.

A brief summary on the construction and some of the characteristics of P1 libraries follows. A more detailed description is found in Pierce *et al.*, 1992; Pierce and Sternberg, 1992; Southard-Smith and MacDonald, 1993.

2.3.5.1 The p*Ad*10*sac*BII *vector.*

The P1 vector p*Ad*10*sac*BII differs from its precursor, pNS582*tet*14*Ad*10 (p*Ad*10, described by Sternberg (1990a)) in that the former contains an additional cassette (*sac*BII) that provides a positive selection for cloned inserts [Figure 2-1].

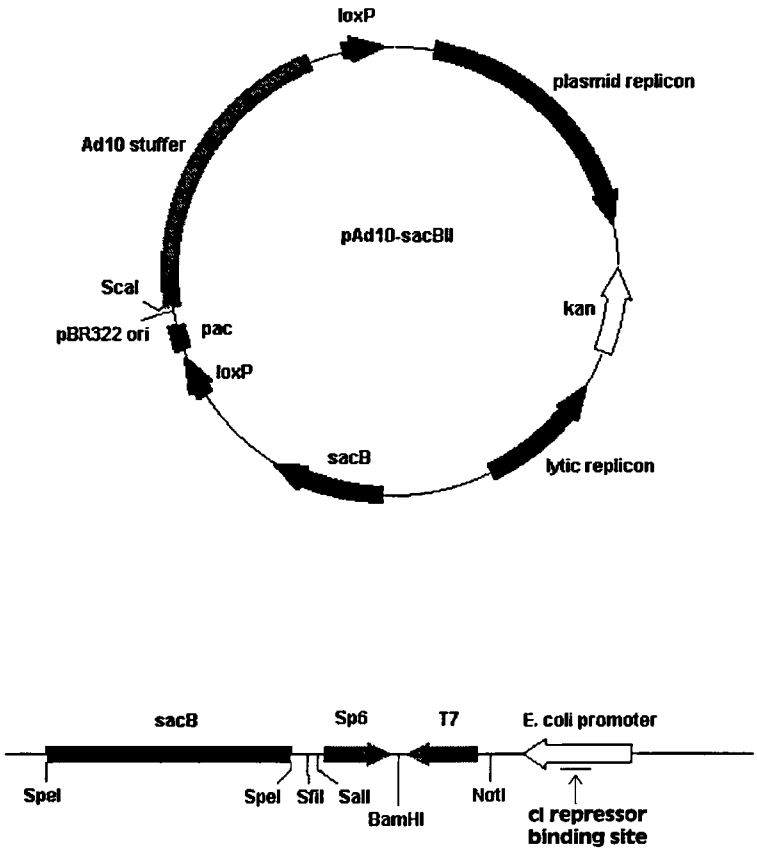


FIGURE 2-1. Map of the p*Ad*10*sac*BII vector and cloning region (Pierce and Sternberg, 1992; Pierce *et al.*, 1992; Sternberg, 1992).

The *sac*BII cassette contains the *sac*B gene from *Bacillus amyloliquifaciens*, which codes for the protein levansucrase. This enzyme converts sucrose to levan. Expression of this gene when sucrose is present in the medium kills the bacterium, since accumulation of levan is lethal in *E.coli*. In the P1 vector system, expression of the *sac*B gene is regulated by a synthetic near-consensus *E.coli* promoter. The cloning site is a unique *Bam*HI located between this gene and its promoter. Therefore, those P1 clones carrying an insert will grow in the presence of sucrose; in contrast, cells where the expression of the *sac*B gene has not been interrupted, will die.

Digestion of the plasmid DNA at the unique *Sca*I and *Bam*HI restriction sites generates a small (~4.4 kb) arm containing a *lox*P site, a phage P1 packaging site, *pac*, and the *sac*BII cassette. It also generates a larger arm (~27 kb) which contains the remaining elements of the P1 plasmid: an 11 kb adenovirus DNA fragment, which has disrupted an ampicillin-resistance (*amp*^{r}) gene in the P1 vector; a plasmid replicon which stably maintains the plasmid DNA at a copy number of 1 P1 molecule per *E.coli* cell; a second *lox*P site; a kanamycin- resistance (*kan*^{r}) gene; a lytic replicon, regulated by the *lac* operon promoter, which is repressed in cells containing the *lacI*^{q} repressor. Addition of the *lac* inducer isopropyl-β-D-thiogalactopyranoside (IPTG) to the growth medium results in the activation of this lytic replicon and allows the copy number to be increased 10- to 20-fold.

2.3.5.2 Construction of the rat genomic DNA P1 library.

DNA fragments in the approximate size range of 70 to 100 kb were selected following partial digestion with *Sau*3AI of high molecular weight genomic DNA extracted from Sprague-Dawley rats. The ends of both P1 arms were treated with alkaline phosphatase, prior to ligation to the genomic digestion products.

Appropriate ligation events resulted in a foreign DNA fragment sandwiched between both arms such that the *lox*P sites were oriented in the same direction. Infectious P1 phage particles were generated through a two-stage *in vitro* packaging reaction of these ligation products. In the P1 system, packaging is initiated at the *pac* site. Encapsidation terminates in a sequence-independent manner by a "headful" cutting reaction. Therefore, the amount of DNA packaged is determined by the size of the P1 head. This meant that very large DNA inserts resulted in the packaging of only one of the *lox*P sites. Recombinant molecules which carried very small inserts were not successfully packaged. Infectious particles were finally obtained by the addition of phage tails. These were then used to infect the *E.coli* strain NS3529, containing the *lacI*^{q} and the *cre* genes on two separate integrated λ prophages. The *cre* gene codes for the P1 Cre recombinase protein, needed for the cyclization of the P1 DNA molecules after their injection into the host. This process involves recombination at the two *lox*P sites. Thus, only those molecules containing both of these sequences (i.e. mostly carrying 70-95 kb inserts) were cyclized. Approximately 105,300 independent clones were obtained. Characterisation of the size of the genomic insert was allowed by the

unique restriction sites *Not*I and *Sfi*I, located at each side of the *Bam*HI cloning site. Digestion with these enzymes resulted in the extraction of the insert fom the P1 molecule. The average insert size was 75 kb. Therefore, greater than 2.5 genome equivalents were represented in this library. Colonies were collected and arrayed in 209 frozen glycerol pools. These primary pools contained on average 550 clones/pool (in a range of 230 to 1200 clones/pool).

2.3.5.3 Screening of the P1 library.

Identification of a single P1 clone of interest was achieved in 3 stages:

a) Identification of a positive primary pool.

A small aliquot was removed from each glycerol pool and resuspended in 100 µl of dH_2O. The suspensions were thoroughly vortexed and heated at 100°C for 10 or 20 minutes to prepare lysates. An aliquot of the lysate (1 µl) was used for PCR amplification.

b) Identification of a positive secondary pool.

Once a positive primary pool was identified, an aliquot was taken from this and resuspended in 3 ml of LB medium. The cell density was calculated by measuring the absorbance at 600 nm in a spectrophotometer. Appropriate dilutions were made and plated which allowed to obtain single colonies on the plates. These dilutions were prepared considering that an OD_{600} value of 1 corresponded to a density of 10^8 cells.ml^{-1}. Individual colonies were picked and gridded on new plates. Once these colonies grew on the new plates, each row of 10 colonies was taken as a secondary pool. A small volume of cells was removed from each of the colonies in a secondary pool and these were resuspended in a volume of 100 µl of dH_2O. Lysates were prepared as before and again screened by PCR.

c) Identification of a single positive P1 clone.

A small fraction was taken from each of the colonies which constituted a positive secondary pool. Each of these was resuspended in a separate volume of 100 µl of dH_2O. Individual lysates were then prepared and, again, screened by PCR.

All the lysates were kept at -20°C.

2.3.6 Polymerase chain reaction (PCR).

In general, 50-100 ng of template DNA and 40-50 ng of each the forward and the reverse primers was used in each reaction. Most amplifications were carried out in a volume of 10 µl, in the presence of 1x magnesium free-*Taq* polymerase buffer (50 mM KCl, 10 mM Tris-HCl, 0.1% Triton X-100), 1-3 mM $MgCl_2$, and 0.5 mM dNTPs. 1.5 U of *Taq* polymerase was added to each sample, normally following a hot start.

Amplifications were carried out on a GeneAmp 9600 System. The standard cycling conditions were one cycle at 99°C for 10 minutes, preceding the addition of *Taq* polymerase to the samples, at 80°C; one cycle at 94°C for 3 minutes; 25-35 cycles at 94°C for 30 seconds, Tann. (annealing temperature; in the range 50°-65°C) for 30 seconds, and 72°C for 30 seconds; a final extension cycle followed, at 72°C for 3 minutes.

In a few of the reactions where mouse primers were being used to amplify heterologous rat DNA, formamide was added to a final concentration of 5%.

Multiplexed amplifications with the primers R191 and D1Gu10 were carried out in a final volume of 15 µl. 100 ng of DNA template, 50 ng of each oligonucleotide primer, 1 mM $MgCl_2$, 1x *Taq* polymerase buffer, and 0.7 mM dNTPs was added to each tube. Reactions were run following a hot start, with the addition of 1.5 U of *Taq* polymerase.

Amplifications with radioactively labelled primers were run on a Perkin Elmer Cetus apparatus. A 1:50 ratio of labelled to unlabelled forward primer was used in each reaction. A hot start was not used in these amplifications and *Taq* polymerase was directly added with all the other reagents. Reactions were carried out in a final volume of 10 µl, overlaid with light mineral oil. Most primer sets worked optimally at identical conditions in both the GeneAmp 9600 system and the PE Cetus thermal cycler. A cycle at 95°C for 4 minutes preceded the amplification cycles, at 94°C for 30 seconds, Tann. for 30 seconds, and 72°C for 30 seconds; this cycling was followed by an additional cycle at 72°C for 3 minutes.

2.3.7 Labelling of DNA with ^{32}P.

T4 polynucleotide kinase (Promega) was used in these reactions. This enzyme exhibits a 5' polynucleotide kinase and a 3' phosphatase activity. It catalyzes the transfer of the γ-phosphate of ATP to a 5' hydroxyl group of DNA or RNA. Two different reactions were carried out with this enzyme:

- A forward labelling reaction, where a phosphate group (labelled with γ-^{32}P) was added to nucleic acids with 5' OH groups (*2.3.7.1* and *2.3.7.2*).

- An exchange labelling reaction, where the 5' phosphate group from DNA was removed, exchanged to ADP, and replaced using [γ-^{32}P]ATP as the donor (*2.3.7.3*).

2.3.7.1 Labelling of PCR primers.

50 ng of forward primer (6-8 pmoles) was mixed with 1 µl of 10x kinase buffer (Promega; 700 mM Tris-HCl, pH 7.6; 100 mM $MgCl_2$, and 50 mM DTT), 1 µl of [γ-^{32}P]ATP (10 µCi/µl, 4500 Ci/mmol), and 1 µl of enzyme (10 U), in a final volume of 10 µl. The sample was incubated at 37°C for 1 hour. The reaction was then placed at 70°C for 10 minutes and either directly used in PCR amplifications or stored at -20°C for future use.

2.3.7.2 Labelling of oligonucleotides.

$(CA)_{10}$, $(GC)_{10}$, $(CT)_{10}$, and the forward and reverse D1Mco1 primers were labelled with [γ-^{32}P] in 30 µl reactions containing approximately 50 ng of the oligonucleotide (5-8 pmoles), 3 µl of 10x kinase buffer (Promega), 5 µl of [γ-^{32}P]ATP (10 µCi/µl, 4500 Ci/mmol) and 2 µl (20 U) of T4 polynucleotide kinase. The enzyme was added last and the reaction carried out at 37°C for 1 hour. The sample was then heated up at 70°C for 10 minutes, prior to removal of the unincorporated radioactivity (*2.3.11.1*).

2.3.7.3 Labelling of a 1 kb DNA ladder marker.

5 µg of 1 kb DNA ladder marker (BRL) was labelled in a 25 µl-reaction, with the addition of 5 µl of 5x exchange reaction buffer (BRL; 250 mM Imidazole-HCl, pH 6.4, 60 mM $MgCl_2$, 5 mM 2-mercaptoethanol, and 350 µM ADP), 3 µl of [γ-^{32}P]ATP (10 µCi/µl, 4500 Ci/mmol), and 1 µl of T4 polynucleotide kinase (10 U). The reaction was carried out at 37°C for 30 minutes and then stopped by adding EDTA to a final

concentration of 5 mM or by heating the sample at 70°C for 10 minutes. An aliquot was diluted 1:50 to prepare the working stock.

2.3.8 Digestion of DNA with restriction enzymes.

Restriction digests of most plasmids were carried out in 10-20 µl reactions for 1-1 1/2 hours. In some instances, restriction digests of P1 DNA were set up in final volumes of 30-40 µl and left at 37°C for 2-4 hours. All the digestions were carried out at 37°C. The BRL React buffers provided with each enzyme were appropriately used at a final concentration of 1x React buffer; 1 µl (1µl = 10 U) of restriction enzyme was used per 10 µl reaction. When DNA was digested with two enzymes these were selected so that both yielded 100% activity in the same buffer (e.g. *Eco*RI and *Not*I both show 100% activity in 1x React buffer 3).

2.3.9 Gel electrophoresis.

2.3.9.1 Agarose gel electrophoresis.

Horizon™ 58 and Horizon® 11.14 horizontal gel electrophoresis systems (BRL) were most commonly used. Eight to 14 samples could be loaded on the former, and up to 28 on the latter. In some instances, when large sets of backcross animals were being typed, the Kodak Biomax HR2025 system was used; 40 to 80 samples were run on these gels.

When studying strain differences at any new microsatellite locus, 4% MetaPhor agarose gels were always used. In the cases where a strain difference was large, any further characterisation of this SSLP and/or typing of the corresponding backcross progeny was often carried out on standard 4% agarose gels.

Analysis of the products obtained following amplification of P1 lysates or somatic cell hybrid DNA template was commonly carried out on standard 1-2% agarose gels, with the exception of those instances when the PCR product was very small. In these latter cases, 4% agarose gels were used.

Most gels were cast and run on 1x TBE. A few were prepared and run in 1x TAE; in these latter, changes in the running buffer were required when running the samples for several hours. All the gels contained ethidium bromide at a concentration of 0.1-0.2 µg/ml. Running times of 45 minutes to 4 hours were needed, depending on the

agarose concentration and the level of resolution required. In general, 95-140 V were applied.

Plasmid DNAs and their restriction fragments were always run on 0.8-1% agarose gels, cast and run on 1x TBE. No ethidium bromide was added to these gels. Instead, these were stained after running the samples, in distilled water containing 0.4-0.6 μg/ml of ethidium bromide. Appropriate resolution of the bands was normally achieved in 45-60 minutes, at 85-90V.

2.3.9.2 Polyacrylamide gel electrophoresis.

Denaturing 6%-8% polyacrylamide gels were used to analyse microsatellite alleles of similar sizes. In general, 8% gels were employed in those cases where the products being analysed were smaller than 150 bp. Products of 200 bp or more were normally studied on 6% polyacrylamide gels, prepared by mixing 80 ml of 6% sequencing solution (19:1) (acrylamide 6% (w/v), Bis-acrylamide, ratio 19:1; 7 M urea; 1x TBE buffer, pH 8.3) (Anachem, Luton, UK) with 0.8 ml of 10% ammonium persulfate (APS) and 35 μl of N,N,N',N'-tetramethylethylene diamine (TEMED). In order to prepare 8% polyacrylamide gels, 42 g of urea was dissolved in 20 ml of 5xTBE buffer and 10 ml of dH_2O; the solution was left to cool; 20 ml of 40% (w/v) acrylamide/bisacrylamide stock solution (19:1) was added, and the volume made up to 100 ml with dH_2O. One ml of 10% APS and 32 μl of TEMED was added and the mix immediately poured between two 20.3 cm x 60 cm plates. In all cases, the 10% APS solution had been either freshly made or kept at 4°C for no longer than a week.

Gels (0.4 mm in thickness) were run on IBI Base Runners (International Biotechnologies Inc.). Prior to loading the samples on the gel, this was pre-run at 60 W for 20-40 minutes. Approximately 10 μl of loading buffer was added to each PCR sample. Loading buffer was also added to an aliquot of the diluted labelled ladder. The samples were heated at 100°C for 10 minutes and immediately placed on ice. 5 μl of each sample was loaded on the gel; the remaining volumes were kept at -20°C.

6% and 8% polyacrylamide gels were run on 1x TBE, at 60 W for 3 hours and at 15-20 W for 8-10 hours, respectively. Then the gels were lifted on Whatman 3MM paper, covered in Saran wrap, and dried under vacuum at 80°C for 30-45 minutes. The gels were placed in cassettes, together with X-ray film, and placed at -70°C for 10 hours to 2 days.

2.3.9.3 *Molecular weight markers.*

A 1 kb DNA ladder marker (BRL; 12 bands each containing from 1 to 12 repeats of a 1018 bp DNA fragment, and 11 additional bands of 75, 134, 154, 201, 220, 298, 344, 396, 506, 517 and 1636 bp, respectively) was normally used as the size standard when running PCR products on agarose gels. In some cases, a 100 bp ladder marker (BRL; 15 fragments ranging in length from 100 bp to 1500 bp, at 100 bp-increments, and an additional 2072 bp-fragment) was employed. Both ladders were diluted in dH_2O and loading buffer to yield a final concentration of 100 $ng.ml^{-1}$. A total of 500 ng was loaded per well.

An aliquot of the 1 kb ladder marker was radioactively labelled (*2.3.7.3*) to be run on polyacrylamide gels. Approximately 5 ng of marker was then loaded per well.

λ/*Hin*dIII fragments (fragment sizes, in kb: 0.56, 2.0, 2.3, 4.4, 6.6, 9.4, 23.1) were used as size markers when running plasmid DNA; the 1 kb ladder was also often run on these gels. The former was prepared by digesting λ phage DNA with the enzyme *Hin*dIII, at 37°C for 1 1/2 - 2 hours, and then diluting the reaction to a final concentration of 25 ng of λ DNA per µl. Following digestion with *Hin*dIII and subsequent dilution, the sample was incubated at 65°C for 10 minutes and immediately put on ice for another 10 minutes. A total of 250 ng of λ/*Hin*dIII DNA was loaded per well.

All the markers, with the exception of the radioactively labelled 1 kb marker, were kept at 4°C. The labelled marker was kept at -20°C.

2.3.10 Southern blotting.

Following electrophoresis and resolution of the fragments obtained from the digestion of plasmid DNA with restriction endonucleases, the DNA was tranferred to nylon membranes (Magna nylon transfer membrane; Micron Separations Inc.).

The gel was left in 0.25 M HCl for 10 minutes. This solution was discarded and the gel washed in distilled water. The gel was kept immersed in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 15 minutes. Passed this time, this solution was replaced with fresh 1.5 M NaCl, 0.5 M NaOH and the gel soaked for an additional 15 minutes. Gentle agitation was kept throughout the washes. The gel was finally removed from the denaturing solution and placed on a flat surface (normally a glass plate). A piece of membrane cut according to the measures of the gel was briefly immersed in denaturing solution and carefully placed on the gel. The DNA was transferred to the nylon membrane by capillary action. After blotting for at least 8 hours, the membrane was

washed in 2x SSC (0.3 M NaCl, 30 mM tri-sodium citrate) to remove any adhering agarose, and exposed to UV radiation. Membranes were kept at -20°C in sealed polythene bags.

2.3.11 Hybridisation of oligonucleotides to filter bound nucleic acid.

2.3.11.1 Removal of unincorporated radionucleotide.

30 µl of blue dextran solution and 10 µl of phenol red solution was added to the 30 µl-labelling reaction after this was heated up at 70°C for 10 minutes (section *2.3.7.2*). The sample was let to flow through a Sephadex G25 column (NAP™ S columns; Pharmacia Biotech.), which had been previously equilibrated in TE buffer. Flow was maintained through two 0.5 ml TE-washes. The blue fraction, containing the labelled oligonucleotide, was collected. The specific activity of the probe (SA=cpm per µg of DNA) was estimated by measuring on a scintillation counter the counts given by an aliquot of the collected fraction, and then applying the following calculation:

$$SA = \frac{\text{cpm in 1 µl x total volume of the blue fraction (µl) x } 10^3\text{(ng)}}{\text{Amount of DNA in the labelling reaction (~50 ng)}}$$

In general, specific activities in the 10^8-10^9 range were obtained with all the probes.

2.3.11.2 Prehybridisation.

Filters were placed in cylinders containing a hybridisation solution. This was always prepared fresh, and contained:

- 2.5 ml of 20x SSC (3 M NaCl, 0.3 M tri-sodium citrate).
- 100 µl of 0.1 M sodium pyrophosphate.
- 1 ml of 50x Denhard't solution (2.5 g Ficoll (type 400), 2.5 g of polyvinyl-pyrrolidone, 2.5 g BSA(fraction V; Sigma); volume made up to 250 ml with dH_2O. Filtered and kept at -20°C).
- 0.5 ml of 1 M sodium phosphate buffer, pH 6.8.
- 2 ml of formamide.
- 170 µl of 10 $mg.ml^{-1}$ salm sperm DNA.

Volume made up to 10 ml with dH_2O.

Volumes of 3-10 ml were used in the prehybridisation step, depending on the size and the number of filters in each cylinder. Prehybridisation was carried out at 42°C for 1-2 hours.

2.3.11.3 Hybridisation and washing.

The prehybridisation solution was discarded and the probe was added to a volume of fresh hybridisation solution that was approximately half of that used for prehybridisation. Enough probe was added in order to obtain about 10^6 cpm.ml^{-1}. Hybridisation was carried out for 10-20 hours at 42°C.

Following hybridisation, the membranes were washed twice in 2x SSC/0.1% SDS, at room temperature for 15 minutes. A last wash followed, in 1x SSC/0.1%SDS, at 50°C, for an additional 15 minutes. The filters were then removed and placed in sealed polythene bags. These were placed in a cassette, together with X-ray film that was exposed at -70°C for 10 to 72 hours.

2.3.12 Photography/Autoradiography.

Gels stained with ethidium bromide were viewed on an UV transilluminator and photographed using either a Mitsubishi image and video copy processor or a Polaroid® MP-4 Land camera.

Autoradiography was performed in metal cassettes (medical chest X-ray type) using Fuji medical X-ray films. Enhancement of the intensity of the autoradiographic images was obtained by exposing the film at -70°C. The X-ray films were developed using an X-OGRAPH Compact X2 automatic processor.

2.3.13 QIAEX II- DNA agarose gel extraction and purification in aqueous solutions.

QIAEX II resin (Qiagen Ltd., Dorking, UK) was used to extract DNA from agarose gels, and also to purify and concentrate DNA from aqueous solutions without the need for phenol extraction or ethanol precipitation steps. The vendor's instructions were followed. In general, samples were eluted only once, in 20 µl of dH2O. Only in those cases when large amounts of DNA were being purified were 2 elution steps (i.e. 2x 20 µl) carried out.

I._ DNA purification in P1 lysates.

In some instances, the confirmation of the presence of a particular locus in P1 primary pools required purification of the DNA in the lysates. The lysates were then prepared in a volume of 30 µl of dH_2O, instead of 100 µl, and the DNA extracted with QIAEX II resin.

II._ DNA extraction from agarose gels.

DNA fragments obtained from the digestion of P1 and other plasmid clones were excised from 0.8-1% agarose gels after these had been stained in ethidium bromide for a very short time. All the non-nucleic acid impurities present, including the agarose, were then removed. The same approach was used when preparing vector DNA for cloning experiments. In this case, vector DNA was run on a low percentage agarose gel (0.8-1%), following linearisation and dephosphorylation steps. The band corresponding to the linearised plasmid was extracted and purified, with the consequent removal of any remaining uncut plasmid.

PCR products were extracted from 2% agarose gels except when these products were of a considerable small size. In these cases, 4% low melting point agarose gels were run, the product band extracted, and the DNA purified as before.

2.3.14 Removal of the 5' phosphate from linearised DNA.

In order to avoid self ligation of the vector when carrying out cloning experiments, the 5' terminal phosphate groups on linearised plasmid DNA were removed prior to any ligation stages. The enzyme shrimp alkaline phosphatase (SAP) was used for this purpose. SAP works in most restriction buffers; therefore, an aliquot (1 U) of this enzyme was added to 20 µl of vector DNA restriction digests. The sample was then left at 37°C for 30 minutes. Removal of the enzyme was achieved by running the dephosphorylated sample on a low percentage gel and then extracting the band corresponding to the linearised vector (section *2.3.13 (II)*). Prior to this step, the sample was heated at 65°C for 10-15 minutes.

2.3.15 Shotgun cloning of P1 DNA fragments.

Approximately 0.5 to 1 μg of P1 DNA was digested with *Sau*3AI enzyme in 10-20 μl reactions. A phenol/chloroform extraction followed; the aqueous phase was transferred to a clean tube and further extracted with an equal volume of chloroform. The DNA was precipitated with 2.5 volumes of 100% ethanol, in the presence of 0.3 M sodium acetate, pH 5.2. The DNA was resuspended in 10 μl of dH_2O and ligated to 50-60 ng of vector DNA.

2.3.16 Ligation of DNA fragments.

The vector was dephosphorylated when it had been cut with only one enzyme. When cloning into pGEM-3Zf(+) vector which had been cut with both *Kpn*I and *Hin*cII, no phosphatase treatment was required, since the two ends were non-cohesive. When cloning specific band-purified DNA fragments, concentrations of both the insert and the cloning vector were estimated prior to ligation by running respective aliquots on a low percentage gel, against 250 ng of a λ/*Hin*dIII marker. The molar ratio of insert to vector was approximately 3:1. This ratio was estimated to be higher in the shotgun cloning experiments (section *2.3.15*).

All the ligations were performed in 20 μl of 1x ligation buffer (50 mM Tris-HCl, pH 7.6; 10 mM $MgCl_2$; 1 mM ATP; 1 mM DTT, 5% (w/v) PEG 8000), containing 1.5 U of T4 DNA Ligase (BRL). The reactions were incubated for 4 hours at room temperature, or overnight at 16°C.

Cloning of the *D3Gu1* PCR product was carried out using the pDT Blue Cloning System (ams Biotechnology). The yield of amplification product was considerably low (10-15 ng in a 10 μl reaction); 5.5 μl of a fresh PCR reaction was mixed with 2 μl pDT vector DNA (100 ng), 1 μl of 10x ligase buffer, 1 μl of 10 mM ATP, and 0.5 μl of T4 DNA ligase. The reaction was incubated at 16°C overnight.

2.3.17 Colony hybridisation.

Colony lifts were prepared with nylon membranes cut to the appropriate sizes. A first set of membranes were placed on the plates and removed after 30 seconds. The

second set were left on the plates for 1 minute and then removed. Four Petri dishes ((a) to (d)) were used as trays containing Whatman 3MM paper soaked in:

(a) 10% SDS,

(b) 0.5 M NaOH; 1.5 M NaCl,

(c) 1.5 M NaCl; 0.5 M Tris, pH 7.4,

(d) 2x SSC,

respectively.

The filters were sequentially placed on each of these plates, the side of the membrane carrying the colonies up. The filters were left on the plates (a), (b), (c), and (d), for 3, 5, 5, and 5 minutes, respectively. The DNA was then fixed to the membranes by UV-crosslinking. The membranes were washed in 0.1x SSC/0.1% SDS at 80°C for 1 hour, and the cell debris removed by thorough rubbing of the surface of these filters whislt still immersed in solution. The membranes were then prehybridised and hybridised at 42°C, in the same solution than that used to probe blots of plasmid DNA digests (section *2.3.11*). A larger volume of hybridisation mix was employed both in the prehybridisation and in the hybridisation steps. Approximately 10^6 to 10^7 cpm per ml of hybridisation solution was used. Washes were also carried out under the same conditions than when working with Southern blots. X-ray film was exposed at -70°C for 2-3 days.

2.3.18 Sequencing.

Sequencing was carried out on an ABI 373 sequencer apparatus (Perkin Elmer), using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit. The manufacturer's instructions were followed.

2.3.18.1 Preparation of DNA templates.

Plasmid clone DNA and PCR products were sequenced. Good results were mainly dependent on the purity and the amount of DNA used in each sequencing reaction. Different approaches were followed to obtain the required yields of clean DNA.

I._ P1 DNA.

P1 DNA was isolated from the *E.coli* NS3529 and the DH10B strains, following either the CsCl purification protocol (section *2.3.2.3*) or another modification of the alkaline lysis protocol. This latter is described below.

A single colony of the P1 clone was seeded into 15 ml of TB containing 25μg.ml^{-1} kanamycin (when working with NS3529; 50 μg.ml^{-1} kanamycin when using DH10B). The culture was grown at 37°C overnight. The cells were pelleted at 6000 rpm for 5 minutes and then resuspended in 3 ml of dH_2O. Another centrifugation step followed, as before; this time the pellet was thoroughly resuspended in 2 ml of GTE (1% glucose/ 50 mM Tris, pH8.0/ 10 mM EDTA) and left on ice for a few minutes. 3 ml of 1% SDS/ 0.2 N NaOH was added and the tube inverted 3-5 times and left on ice for 10 minutes. 3 ml of 3 M sodium acetate (pH 4.8) was then added, the tube gently inverted 3-5 times, and left on ice for 20 minutes. The sample was centrifuged at 15000 rpm for 10 minutes to remove the cell debris and most of the chromosomal DNA. The supernatant was transferred to a new tube and the DNA precipitated with an equal volume of isopropanol. The DNA was recovered as a clear pellet at the side of the tube after centrifuging at 15000 rpm for 10 minutes. This pellet was dissolved in 0.4 ml of 0.3 M sodium acetate, pH4.8. An equal volume of phenol/chloroform (ratio 1:1) was added and the sample vigorously vortexed and centrifuged at 14000 rpm for 5 minutes. The phenol/chloroform extraction was repeated 1 or 2 more times. 1 ml of 100% ethanol was finally added to the recovered aqueous phase and the tube centrifuged at 14000 rpm for 10 minutes. The DNA was washed with 0.4 ml of cold 70% ethanol, air dried, and dissolved in 0.4 ml of TE; 5 μl of RNase A was added and the tube incubated at 37°C for 2 hours. 4 μl of 1 M $MgCl_2$ and 200 μl of 40% PEG_{8000} was then added and the sample vortexed and centrifuged at 14000 rpm for 15 minutes. The pellet was rinsed with cold 70% ethanol and air dried; it was then dissolved in 0.4 ml of 0.5 M ammonium acetate. One ml of 100% ethanol was added and the sample centrifuged at 14000 rpm for 10 minutes. The pellet was rinsed twice with 0.4 ml of cold 70% ethanol, air dried, and finally dissolved in 55 μl of dH_2O.

All the isolated P1 DNAs were dialysed for 6-14 hours prior to sequencing.

II._ Other plasmid DNAs.

pGEM-3Zf(+), pBluescript II KS+, and pDT Blue clone DNAs were isolated from 3 ml LB cultures, as previously described (section *2.3.2.2*). The DNA was resuspended in 32 µl of dH_2O and re-precipitated by adding 8 µl of 4 M NaCl and 40 µl of 13% PEG_{8000}. The sample was thoroughly mixed, left on ice for 20 minutes, and centrifuged at 14000 rpm for 15 minutes, at 4°C. The DNA pellet was washed with 70% ethanol, air-dried, and finally resuspended in 20 µl of dH_2O.

III._ PCR products.

PCR reactions were carried out under the conditions established as optimal, in a volume of 25 µl. Purification of the amplification product was rapidly achieved using MicroSpin™ S-400 HR columns (Pharmacia Biotech.); the manufacturer's instructions were followed. In a few instances, the yield of purified product was too low to be used in sequencing reactions. In these cases, the PCR reaction was repeated and the product run on an agarose gel and purified with QIAEX II resin. In the instances where the yield was still too low, two modifications were introduced:

a) The number of cycles in the PCR reaction was increased. The appearance of background bands did not pose a problem, since only the specific PCR product was extracted from the gel, purified, and sequenced.

b) In the cases when increasing the number of PCR cycles did not result in sufficient amounts of product being obtained, a first PCR amplification was carried out and the PCR product run on a low melting point agarose gel. A small sample was taken from the product band and used as the template in a second amplification reaction. The final product was run on an agarose gel and purified with QIAEX II resin.

2.3.18.2 Sequencing of DNA templates.

Approximately 3 to 5 µg of P1 DNA was used as template in each sequencing reaction. Aliquots (1 or 2 µl) of the purified plasmid DNAs and the PCR products were run on agarose gels to estimate the concentration of these DNAs. Approximately 30-100 ng of PCR product was added to each sequencing reaction. The amount of plasmid DNA used as template for sequencing normally ranged between 0.3 and 0.5 µg.

All the sequencing reactions were carried out in a volume of 20 µl. Eight microlitres of the Terminator Reaction Mix (available with the sequencing kit) and

3.2 pmoles of the appropriate sequencing primer was added to the template. Sequencing was carried out on a GeneAmp PCR System 9600. The cycling conditions were 1 cycle at 96°C for 1 minute, followed by 25 cycles at 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes.

When sequencing PCR products, the corresponding PCR primers were used. Sequencing of pGEM-3zf(+) and pBluescript II KS+ clones was carried out with the -21 M13 Forward and the M13 Reverse primers:

-21 M13 FORWARD PRIMER: 5'- TGTAAAACGACGGCCAGT -3'
M13 REVERSE PRIMER: 5'- CAGGAAACAGCTATGACC -3'

Sequencing of pDT Blue clones was carried out with the -21 M13 Forward primer and a T7 primer:

T7 PRIMER: 5'- CTAATACGACTCACTATAGG -3'

P1 clone DNAs were sequenced with either the SP6 or the T7 sequencing primers described by Kimmerly *et al.* (1994):

SP6 PRIMER: 5'- GGCCGTCGACATTTAGGTGACAC -3'
T7 PRIMER: 5'- CCGCTAATACGACTCACTATAGGG -3'

Once the cycling had finished, the DNA was precipitated with 2.5 volumes of 100% ethanol, in the presence of 0.3 M sodium acetate, pH 5.2. Each product was finally resuspended in 4 µl of *loading buffer* (deionized formamide and 25 mM EDTA (pH 8.0) containing 50 mg/ml blue dextran, in a ratio of 5:1 formamide to EDTA/Blue dextran). The samples were heated at 90°C for 2 minutes before loading.

Sequencing gels (0.4 mm in thickness) were prepared by mixing 50 ml of 6% sequencing solution (19:1) (acrylamide 6% (w/v), Bis-acrylamide, ratio 19:1; 7M urea; 1x TBE buffer, pH8.3) (Anachem, Luton, UK) with 250 µl of 10% APS and 25 µl of TEMED. The samples were run at 30 W (2500 V; 40 mA) for 14 hours, in 1x TBE. Sequencing data were collected and analysed using Model 373 Software (Applied Biosystems, Perkin Elmer).

2.3.19 Computer-assisted sequence analysis.

Search for specific sequences in the database was carried out by using the "GCG Sequence Analysis Software Package, Version 7" from the Genetics Computer Group at the University of Wisconsin (UWGCG). The programs Stringsearch and Fetch were used to identify sequences which corresponded to specific character patterns in the database and to copy those sequences or data files into appropriate directories, respectively.

Primer design was aided by the MacVector™ 4.1.4 and the GeneJockeyII Sequence Processor (Biosoft, Cambridge, UK) programs. The latter was used to analyse the sequences obtained following sequencing in the ABI 373 system. Often, DNAs corresponding to the same clone (i.e. carrying the same insert) were extracted from 2 different cultures (i.e. 2 separate colonies) and sequenced in parallel. These two sequences were aligned (Gap) to identify possible misreadings. When PCR or sequencing primers were to be design, only the regions of complete identity were taken into account. This alignment function was also used when searching for gene-specific regions in the rat muscle creatine kinase gene.

Obtained sequences were compared to non-redundant sequences in the GenBank+EMBL+DDBJ+PDB database through BLAST Sequence Similarity Searching, based on the BLAST algorithm (Altschul *et al.*, 1990).

2.4 Statistics.

χ^2 values were calculated as:

$$\chi^2 = \sum_{i}^{k} [(x_i - m_i)^2 / m_i]$$

where:

k is the number of classes of animals (two parental classes, $agu^+ x^r/agu^- x^a$ and $agu^- x^a/agu^- x^a$; and two recombinant classes, $agu^+ x^a/agu^- x^a$ and $agu^- x^r/agu^- x^a$, as indicated in Figure 4-1).

x_i is the number of animals observed in each class.

m_i is the number of animals expected in each class, assuming random assortment of *agu* and the SSLP locus *x* alleles.

The degrees of freedom (df) are calculated as k-1.

The standard deviation (Green, 1981), SD, was calculated as:

$$SD = \sqrt{\frac{r(1-r)}{N}}$$

where r is the recombination frequency and N the number of animals typed.

The 95% lower (L_1) and upper (L_2) limits were calculated as:

$$L_1 = r - x \,.\, SD$$

$$L_2 = r + x \,.\, SD$$

where:
r= recombination fraction
SD=standard deviation
x is a coefficient obtained with Gaussian curves. For a 95% confidence interval, both the lowest and the highest values for which the probability is 2.5% are rejected. The x value corresponding to this probability is x= 1.96. Calculation of the 95% lower limit for the interval *C - D1Mit3* (Chapter 4) resulted in a negative value. Therefore, the highest values for which the probability was 5% were rejected ($>L_2$). This correlated to an x value= 1.645.

Conversion of the recombination frequencies (r) to genetic distances (d, in cM) was carried out by applying the Kosambi's mapping function (Green, 1981):

$$d = 1/4 \,[\log_e (1+2r) - \log_e (1-2r)]$$

2.5 *E. coli* strains.

The *E.coli* strains which were used in this work are shown below.

STRAIN	GENOTYPE	REFERENCE/SOURCE
DH5α	*sup*E44 Δ*lac*U169(ϕ80*lac*ZΔM15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1	Hanahan, 1983.
DH10B	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lac*ZΔM15 Δ*lac*X74 *end*A1 *rec*A1 *deo*R Δ(*ara-leu*)7697 *ara*D139 *gal*UK *nup*G *rps*L	Grant *et al.*, 1990.
F3	DH10B [F::Tn10d-Cam]	Kimmerly *et al.*, 1994.
INVαF'	F' *end*A1 *rec*A1 *hsd*R17 (r_k^-, m_k^+) *sup*E44 *thi*-1 *gyr*A96 *rel*A1 ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169λ⁻	TA cloning® kit. (Invitrogen BV, Leek, The Netherlands)
NS3529	*rec*A1 *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) λimmλLP1 λimm434*nin*5X1-*cre*	Pierce *et al.*, 1992.
TG1	*sup*E *hsd*Δ5 *thi* Δ(*lac-pro*AB) F' [*tra*D36 *pro*AB⁺ *lac*I^q *lac*ZΔM15]	Gibson, 1984.

2.6 Plasmid vectors.

The pDT Blue vector was obtained with the pDT Blue cloning system (ams Biotechnology). Plasmid pBluescript II KS+ was obtained from Stratagene (La Jolla, CA, USA). Plasmid pGEM-3Zf(+) was obtained from Promega.

2.7 Media for propagation of *E.coli.*

a) <u>*L-broth (LB)*</u>.

10 g tryptone, 5 g yeast extract, 5 g NaCl, 1 g glucose, 20 mg thymine, made up to 1 litre in distilled water and adjusted to pH 7.0 with NaOH.

b) *2x YT-broth*.

16 g tryptone, 10 g yeast extract, 5 g NaCl, made up to 1 litre in distilled water and adjusted to pH 7.0 with NaOH.

c) *T-broth (TB)*.

12 g tryptone, 24 g yeast extract, 4 ml glycerol, made up to 900 ml in distilled water. Autoclaved and let to cool to 60^0C or less. Then, 100 ml of a sterile solution of 0.17 M KH_2PO_4, 0.72 M K_2HPO_4 added.

d) *SOB medium*.

20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 10 ml 250mM KCl, made up to 990ml with distilled water and adjusted to pH 7.0 with NaOH. Autoclaved and let to cool down to room temperature. 10 ml of sterile 1 M $MgCl_2$ added before use.

e) *SOC medium*.

1 litre of SOB medium to which 20 ml of sterile 1 M glucose has been added.

f) *L-Agar*.

As L-Broth without glucose and the addition of 15 $g.l^{-1}$ agar.

2.8 Antibiotics and Indicators.

The antibiotic concentrations used for both liquid and plate selection were as follows:

Antibiotic	Stock solution	Working concentration
Ampicillin	100 $mg.ml^{-1}$, prepared in dH_2O.	100 $\mu g.ml^{-1}$
Chloramphenicol	100 $mg.ml^{-1}$, prepared in ethanol.	10 $\mu g.ml^{-1}$
Kanamycin	100 $mg.ml^{-1}$, prepared in dH_2O.	25 or 50 $\mu g.ml^{-1}$
Streptomycin	100 $mg.ml^{-1}$, prepared in dH_2O.	100 $\mu g.ml^{-1}$

The indicators X-gal (5-bromo-4-chloro-3-indolyl-β-galactopyranoside) and IPTG (isopropyl β-D-thiogalactopyranoside) were both prepared at a concentration of 100 $mg.ml^{-1}$, and used at the final concentrations of 40 and 50 $\mu g.ml^{-1}$, respectively. IPTG was dissolved in distilled water; X-gal was prepared in N,N-dimethylformamide. When being employed to induce the P1 copy number in *E. coli* host cells, a 1 M IPTG solution was prepared. This was added to the cultures to a final concentration of 1 mM.

All the antibiotics and the indicators above were stored at -20°C. When preparing plates, these solutions were added to the molten agar once this had been cooled to approximately 50°C. In general, the plasmids which did not carry an insert gave rise to blue colonies whilst those which did resulted in the growth of white colonies.

2.9 Sterilisation.

All growth media were sterilised by autoclaving at 121°C for 15 minutes at a pressure of 15 lbs. Thermolabile substances such as antibiotics were sterilised by filtration through 0.22-micron filters. Those antibiotic solutions which had been prepared in ethanol need not be sterilised. Sucrose and glucose solutions were either filtered through 0.22-micron filters or autoclaved at 104°C for 10 minutes at a pressure of 5 lbs.

2.10 *E.coli* growth conditions.

Liquid cultures of *E.coli* strains from which pGEM-3Zf(+), pBluescript II KS+ and pDT Blue clone DNAs were to be isolated were grown in L-broth with the corresponding ampicillin selection. Volumes of 3 ml of broth were normally seeded and the cultures incubated at 37°C for 9-17 hours in an orbital shaker at ca. 250 rpm. In the case of the subclone 37B, 5 ml cultures were also prepared and grown at 30°C for approximately 15-20 hours. P1 clones were grown in volumes of 50, 100, 250 and 500 millilitres; most cultures were prepared in L-broth and a few in 2x YT- or T-broth. When large cultures (100 ml or more) were to be prepared, 3-5 ml cultures were first grown overnight and aliquots taken to seed the larger volumes. A final concentration of

5% sucrose was added to all the P1 cultures; the antibiotic kanamycin was added to a final concentration of 25 μg.ml^{-1} when using NS3529 host cells and to a concentration of 50 μg.ml^{-1} when working with DH10B cells. Most P1 cultures were grown at 30°C. Normally 1 mM IPTG was added to the P1 cultures once these had reached an $AB_{590\text{-}595}$= 0.15-0.2; these were then grown for an additional 4-20 hours.

E.coli strains were also propagated on L-agar plates with the appropriate selection. Plates contained ca. 25-50 ml of agar. Except the P1 clones, all the plates were incubated overnight at 37°C. P1 colonies were grown in the presence of 5% sucrose at 30°C for 20-36 hours.

2.11 Preservation of *E.coli* strains.

A 0.75-1 ml aliquot of each culture was mixed with 1 ml of 40% (v/v) glycerol, 2% peptone (w/v), and stored at -70°C. The strains were revived by scraping the surface of this frozen stock with a toothpick and streaking onto agar. This allowed the isolation of single colonies on the plate.

2.12 P1 DNA transfer into DH10B by an F' episome intermediate.

The transfer into DH10B cells was done via γδ-transposition and it involved several steps.

1. A single colony of NS3529 containing the P1 of interest was seeded into 1 ml of LB containing 50 μg.ml^{-1} kanamycin, and grown at 37°C overnight. At the same time, an F3 colony, grown in a plate containing 10 μg.ml^{-1} chloramphenicol, was picked to inoculate a 5 ml LB culture which was also grown at 37°C overnight.

2. 1.8 ml of LB was added to a sterile glass test tube; 100 μl from each overnight culture was also added and the mix (M1) incubated on a roller (30 rpm) at 37°C for 3 hours. A 10^{-5} dilution of mating mix M1 was plated on L-agar containing 50 μg.ml^{-1} kanamycin/ 10 μg.ml^{-1} chloramphenicol, and grown at 37°C overnight. A control was prepared where 5 μl of the F3 and 5 μl of the NS3529 cultures was separately spotted

on a plate which also contained 50 μg.ml^{-1} kanamycin and 10 μg.ml^{-1} chloramphenicol. This control plate was incubated at 37°C overnight.

3. As expected, no colonies grew on the control plate, since these cells did not carry both kanamycin and chloramphenicol resistance genes.

Between 400 and 600 colonies grew on the M1 plate; these must be NS3529 cells which had received the F::Tn10d-Cam factor from the F3 donor and therefore exhibited resistance to both antibiotics.

A single colony was picked into 1ml of LB containing 50 μg.ml^{-1} kanamycin and 10 μg.ml^{-1} chloramphenicol, and grown at 37°C overnight. At the same time, a DH10B colony was picked from an L-agar/ 100 μg.ml^{-1} streptomycin plate, to inoculate 5 ml of LB. This culture was grown at 37°C overnight.

4. Again, 100 μl from each overnight culture was added to 1.8 ml of LB in a sterile glass test tube and the mix incubated on a roller (30 rpm) at 37°C for 3 hours. 100 μl of this mating mix (M2) was plated on L-agar containing 50 μg.ml^{-1} kanamycin/ 10 μg.ml^{-1} chloramphenicol/ 100 μg.ml^{-1} streptomycin. A control was prepared, as before, with 5 μl of M1 and 5 μl of DH10B cultures being separately spotted on a plate containing these three antibiotics. Both plates were incubated at 37°C overnight.

5. No colonies were observed on the control plate. Approximately 400-500 colonies grew on the M2 plate; these must be the DH10B cells which had received both the F factor and the P1 clone from the donor parent.

Some modifications were introduced to the original method when transferring P1*(D1Mco1)* DNA:

A considerable fraction of the gridded P1(*D1Mco1*) bacterial mass was removed with a toothpick and resuspended in 100 μl of LB. 90 μl was used to inoculate 1 ml of L-broth containing 50 μg.ml^{-1} kanamycin and 5% sucrose. This culture was grown at 37°C for nearly 2 hours. PCR amplification showed that the *D1Mco1* sequence was still present.

The P1 culture was added to 1.8 ml of LB, together with 100 μl of an overnight F3 culture, and the mix incubated at 37°C for 3 hours. 100 μl of this mix (M1) was plated and grown at 30°C for 36 hours, until very small colonies were visible on the plate.

Six M1 colonies were picked and each resuspended in 100 μl of LB. PCR confirmed that all of them contained the *D1Mco1* sequence. 90 μl of each suspension

was employed to seed each of six 600 µl LB cultures (containing 50 µg.ml^{-1} kanamycin and 10 µg.ml^{-1} chloramphenicol) and these grown at 37°C for 2 1/2 hours. Again, the *D1Mco1* PCR product was obtained from each of these cultures. Two of them were pooled, pelleted, resuspended in 100 µl of LB, and mated to an overnight DH10B culture. 1 ml from this mix (M2) was centrifuged and the pellet resuspended in 200 µl of LB. 100 µl was plated on LB medium containing 50 µg.ml^{-1} kanamycin/ 10 µg.ml^{-1} chloramphenicol/ 100 µg.ml^{-1} streptomycin, and grown at 30°C until a few colonies were visible.

2.13 Introduction of plasmid DNA into *E. coli.*

2.13.1 Preparation of competent cells.

The cells were kept at 4°C throughout all the manipulations when preparing competent cells. INVαF' cells, available with the TA cloning® kit (Invitrogen BV., Leek, The Netherlands), were also used.

2.13.1.1 Electrocompetent cells.

These cells yielded the highest transformation efficiencies (10^8-10^{10} transformants per µg of DNA). TG1 and DH5α cells were made electrocompetent following the steps below.

1._ A volume of 400 ml of L-broth was inoculated with 4 ml of a fresh overnight culture. The cells were grown at 37°C with vigorous shaking, to an approximate A_{600} of 0.5-0.6.

2._ The cells were kept on ice for 15-30 minutes and then pelleted by centrifuging at 5000 rpm for 15 minutes at 4°C.

3._ The supernatant was discarded and the cells thoroughly resuspended in 400 ml of ice cold dH_2O. The centrifugation step was repeated as before.

4._ The supernatant was removed and the cells resuspended in 250 ml of ice cold dH_2O. Another centrifugation step followed.

5._ The pellet was resuspended in 20 ml of ice cold 10% glycerol and centrifuged as before.

6._ The cell pellet was finally resuspended in approximately 0.7 ml of ice cold 10% glycerol. Aliquots of 40 µl were prepared and left on dry ice for 15 minutes. These aliquots were stored at -70°C for up to ~6 months.

2.13.1.2 $CaCl_2$ method.

Competent INVαF' and TG1 cells were prepared following the protocol described below. A transformation efficiency of ~10^6 was normally obtained. The cells were kept at 4°C for a maximum of 2-4 days.

1._ A volume of 100 ml of L-broth was inoculated with 1 ml of a fresh 3 ml overnight culture, and the cells grown at 37°C until an A_{595} of about 0.6 was reached.

2._ The cells were centrifuged at 3000 rpm at 4°C for 5 minutes and the supernatant discarded.

3._ The cells were resuspended in 25 ml of ice cold 50 mM $CaCl_2$ and left on ice for 20 minutes. The centrifugation step was repeated.

4._ The cells were finally resuspended in 5 ml of ice cold 50 mM $CaCl_2$ and kept at 4°C.

2.13.2 Transformation procedures.

When transforming INVαF' cells which had been obtained with the TA cloning® kit, the manufacturer's instructions were followed.

2.13.2.1 Electro-transformation.

Electroporation was carried out on a Gene Pulser apparatus (Bio-Rad Laboratories), set at 25 µF, 200 Ω, and 2.50 kV. 0.2 cm cuvettes were used.

The DNA used in these transformations must be in a low ionic strength buffer. In the cases where a ligation product was being used in the transformation, this DNA was first precipitated with 2-2.5 volumes of 100% ethanol, in the presence of 0.3 M sodium acetate (pH 5.2), and then resuspended in 7-8 µl of dH2O.

1._ A 40 µl aliquot of electrocompetent cells was gently thawed. The suspension was mixed with 1-4 µl of DNA (depending on the DNA concentration).

2._ The mixture was transferred to the bottom of a cold electroporation cuvette and left on ice for 0.5-1 minute. The cuvette was then placed in a chilled safety chamber slide. This was pushed into the chamber until it was seated between the contacts in its base.

3._ Electroporation was carried out at the settings mentioned above.

4._ After the pulse the cuvette was rapidly removed from the chamber and 1 ml of SOC medium added to resuspend the cells.

5._ The cells were left on ice for 1 minute and then transferred to a sterile universal. The suspension was left on an orbital incubator at 37°C for 1 hour, prior to plating.

2.13.2.2 Transformation of $CaCl_2$-competent cells.

In this case, ligation products were directly used in the transformation, without the need for a precipitation step to remove salts. Aliquots of 5 µl of ligation reaction were used in these occasions.

1._ The DNA was mixed with a 100 µl aliquot of the cell suspension and left on ice for 30 minutes.

2._ The sample was heated at 42°C for 90 seconds and incubated on ice for an additional 2 minutes.

3._ 300 µl of 2x YT-broth was added to the cell suspension. The sample was left on an orbital incubator at 37°C for 1 hour, prior to plating.

CHAPTER 3

Establishment of a rat genetic genome scan.

3.1 Introduction.

As described in Chapter 1 (section 1.1), the *agu* mutation did not correlate with any detectable alteration in the karyotype of the AS/AGU sub-strain. The function of the product of the *agu* gene at the molecular level could not be predicted and selection of candidate genes was therefore not a viable approach to gene identification at this stage. In order to determine the position of the *agu* locus, linkage of the gene to at least one genetic marker was sought. Detection of linkage was to involve the analysis of the cosegregation pattern of the *agu* gene with each of a large series of informative SSLPs in the offspring obtained from crosses of standard rat lines to the AS/AGU sub-strain. Since *agu* is a recessive mutation, the F1 population is phenotypically normal. Therefore, backcrosses of the F1 offspring to the affected parent were carried out, with the concomitant segregation of the mutant and the normal *agu* alleles in the resulting progeny.

Central to these linkage studies was the availability of a wide selection of microsatellites for which the PCR conditions had been optimised. Unambiguous identification of each particular SSLP allele in each animal required obtaining primarily the expected PCR product in the reactions. Very extensive work was thus directed towards this aim.

3.1.1 Development of the rat genetic map.

A wide selection of microsatellites was needed to diminish the possibility of regions of the genome not being covered by any marker. This latter situation could hinder the identification of the *agu* locus, should this gene map to any of these regions.

Robinson (1972) estimated the length of the rat genome to be approximately 2400 cM, based on counts of chiasmata in meiosis. According to Jacob *et al.* (1991) about 200 markers would be required to ensure a 99% chance of each marker being linked to a map of this size, assuming a random distribution of the markers. At the time the project commenced, in 1993, this aim presented itself as a difficult task. The reason was the absence of an available detailed genetic map of the rat, with microsatellite markers having just started to be used in this species.

Although the start of rat genetics (Lindsey, 1979) was parallel to that of the mouse, the rat soon became the model of choice for physiologists and other biomedical

researchers whereas the mouse became the preferred model for geneticists. This was translated into very different rates at which the two genetic maps developed. By 1990, over 1300 loci had been identified in the mouse, with more than 960 of these having been mapped on specific chromosomes. In contrast, about 330 genetic loci had been described in the rat and approximately 200 had been mapped (Hedrich, 1990b). With the discovery of microsatellites in some eukaryotic species (Litt and Luty, 1989; Weber and May, 1989) and the observation that these were also present in the mouse, the genetic map of this rodent rapidly progressed (Love *et al.*, 1990; Cornall *et al.*, 1991). Three-hundred and seventeen SSLPs had been typed in the mouse by 1992. These markers were detectably linked to approximately 99% of the mouse genome, at an average spacing of 4.3 cM (Dietrich et *al.*, 1992). In comparison, important contributions towards the development of the rat genetic map had just started to be made.

Some of the first large collections of data arrived from Levan *et al.* (1990) and Hedrich (1990b). Thirteen linkage groups were defined, with only six of them being assigned to specific chromosomes. Markers such as minisatellites were not developed in the rat yet. Levan and coworkers carried on with the assignment of 214 loci to 20 of the 22 rat chromosomes in 1991. This number was raised to 370 loci in the following year (Levan *et al.*, 1992). Nevertheless, very limited linkage information was available, with gene orders and distances not being established yet.

It was the groups of Hilbert and coworkers (1991) and Jacob *et al.* (1991) who first demonstrated that rat microsatellites showed similar properties to those found in man and mouse and then proceeded to generate a large collection of these genetic markers. Jacob and colleagues (1991) developed 290 SSLPs. Of these, 112 were studied in F2 populations. Linkage analyses defined 23 linkage groups, each containing 2 or more markers, and 10 unlinked microsatellites. More than half of the markers were unambiguously assigned to specific chromosomes by means of somatic cell hybrid analyses. Nonetheless, the bulk of these data was not published till later on (Jacob *et al.*, 1995). Meanwhile, some of the markers became commercially available through Research Genetics (Huntsville, Alabama, US).

A major contribution to the rat map was that made by Serikawa and coworkers in 1992. Searches through published rat sequences and screening of a rat genomic library led this group to the identification of 134 SSLPs, corresponding to 118 loci. Sixty-three of these were newly assigned to chromosomes. Linkage analyses with most

of the SSLPs and 29 other markers were carried out. One-hundred and ten markers were mapped into 24 linkage groups. Twelve others did not show linkage. A final stage was reached where all the chromosomes except 15 and Y had been assigned one or more loci. All these markers taken together seemed to offer a 50-70% coverage of the rat genome. The same group of researchers extended the number of rat SSLPs to 141, in 1993 (Kuramoto *et al.*, 1993). Some of these were assigned to discrete chromosomal subregions by fluorescent *in situ* hybridisation techniques.

The work carried out by Remmers and colleagues was also very important, with frequent additions to the maps of the different chromosomes (Remmers *et al.*, 1992; Goldmuntz *et al.*, 1993; Zha *et al.*, 1993; Remmers *et al.*, 1993a; Remmers *et al.*, 1993b). Some of these SSLPs were added to the set of markers being offered by Research Genetics.

The work described above constituted the first serious attempts to develop a detailed linkage map of the rat genome, correlating with the increasing importance of this rodent as a genetic model. All the mapping data published up to the middle of 1993 was collected by Yamada *et al.* (1994) [Appendix A; Figure A-1]. Five-hundred and thirty-nine rat loci, of which 167 were typed by microsatellite markers, were assigned to specific chromosomes. Another 8 still remained in the linkage groups originally defined by Hedrich in 1990. Two-hundred and seventy-six of these loci were located on linkage maps. Only those loci for which microsatellite markers were available were practically of use in this project.

Parallel to the progress in the generation of genetic markers and mapping of loci in the rat, cytological and linkage studies were being carried out to compare the emerging rat genetic map to those of mouse and man (Yoshida, 1978; Levan *et al.*, 1991; Serikawa *et al.*, 1992; Yamada *et al.*, 1994). Extensive chromosome homology was established between the mouse and the rat. This observation was of great importance since Stallings and coworkers (1991) had shown the conservation of a considerable proportion of microsatellite sequences between closely related species. Moore and colleagues (1991) had further demonstrated that this conservation could be close enough to allow the use of PCR primers designed in one species to analyse microsatellite allelic differences in another. These initial experiments were followed by other groups (Jacob *et al.*, 1991; Kondo *et al.*, 1993; Fredholm and Winter, 1995). In contrast to the high success rate (56%) achieved by Moore and colleagues when

amplifying ovine DNA with bovine primers, and that obtained by Fredholm and Winter (1995) on the *Canidae* family, only 12-16% of the mouse/rat PCR primers used by Kondo and coworkers (1993) efficiently amplified the corresponding heterologous DNAs. Nevertheless, these results were of special interest since it offered the possibility of using mouse primers to screen for SSLPs in regions of the rat genome where this type of markers were still scarce.

A major breakthrough was the publication in 1995 of a rat genetic linkage map which for the first time appeared to cover the vast majority of the genome (Jacob *et al.*, 1995). Jacob and colleagues developed 1170 SSLPs and mapped 432 of them in a single F2 population. With the exception of two markers on chromosome 9, all the others mapped into any of 21 large linkage groups, at an average spacing of 3.7 cM. Linkage groups were assigned to 21 of the rat chromosomes [Appendix A; Figure A-2]. The sequences of the PCR primers were deposited with Research Genetics, becoming commercially available. Parallel to this work were the studies by Maihara and coworkers (1995). This group obtained 125 PCR primers from the aforementioned company, proceeding to their analysis on different inbred strains. Fifty-six of these markers were mapped onto the existing linkage map published by Yamada *et al.* (1994). This provided a means of integrating the data collected up to the middle of 1993 with the new markers developed by Jacob and colleagues (1995).

Quite a few loci putatively involved in disease phenotypes have been identified using the rat as a genetic model (Brown *et al.*, 1996; Gauguier *et al.*, 1996). Yet the rat map remains incomplete in comparison with those of man (Murray *et al.*, 1994) and the mouse (Dietrich *et al.*, 1994). Further work is being carried out to expand and increase detailed coverage of this map. Some groups of researchers are characterising markers throughout the whole genome (Yokoi *et al.*, 1996; Pravenec *et al.*, 1996). Others are leading the development of linkage maps which correspond to specific chromosomes (Goldmuntz *et al.*, 1995; Du *et al.*, 1995; Du *et al.*, 1996; Ding *et al.*, 1996). This is often the case when more detailed maps of a specific region(s) already shown to carry disease-associated loci are required (Gu *et al.*, 1996; Galli *et al.*, 1996). More complete maps are being constructed containing loci detected by a variety of procedures. A good example is found in the linkage map reported by Pravenec *et al.* (1996) where a variety of 500 morphological, biochemical, immunogenetic and molecular genetic markers have been integrated. Szpirer and colleagues have recently collected the information on all the genes and anonymous sequences mapped to date in the rat (Szpirer *et al.*, 1996).

Approximately 600 genes have been mapped to specific chromosomes. A linkage map containing more than 650 markers has been generated.

3.1.2 Considerations in the establishment of informative crosses to the AS/AGU sub-strain.

A large linkage panel was required to ensure identification of the *agu* region. It was therefore crucial to maximise the number of SSLPs that are informative in the backcrosses. Selection of reference parents from strains placed at the highest phylogenetic distances from AS/AGU should satisfy this requirement. However, two immediate disadvantages were encountered in the rat when compared to the situation in the mouse. Firstly, the phylogenetics of inbred mouse strains were well documented. In contrast, very little information was available about the relations between laboratory inbred strains of rats. The trees of genetic relatedness of rat strains that were available at the time (Arenas *et al.*, 1981; Festing and Bender, 1984) reflected a small subset of loci involved in specific functions. These functions could restrict the degree to which those loci diverged between different rat strains. Therefore these were not a good representation of the overall changes in the whole genome and could be misleading, as later claimed by Canzian and coworkers in 1995. Secondly, whereas interspecific crosses could be exploited in the mouse, this approach was not feasible in rats. This difference was critical in the establishment of dense genetic maps. It had been demonstrated that the wild mouse subspecies *Mus spretus* differed from any of the laboratory mouse *M. musculus* strains at 72-92% of the SSLPs analysed whilst these inbred strains only differed at about 50% or less of the markers when compared to each other (Love *et al.*, 1990; Cornall *et al.*, 1991). In contrast, the Norway rat, *Rattus norvegicus*, and the black rat, *Rattus rattus* (Hayashi *et al.*, 1979; Baverstock *et al.*, 1983), can not be intercrossed and interspecific mapping is therefore not available in rat.

Over 140 inbred strains of rats were available by 1990 (Greenhouse *et al.*, 1990). Three of these were selected to be analysed, together with the AS/AGU sub-strain, with each of a series of 75 microsatellites. These were the *Brown Norway* (BN), the F344 and the PVG strains. Phylogenetic studies had presented the former as the most distant to any of the other inbred strains (Festing and Bender, 1984). These observations had been further supported by results obtained by Serikawa and colleagues

when typing several rat strains with microsatellite primers (Serikawa *et al.*, 1992; Kuramoto *et al.*, 1993).

The greatest divergence from AS/AGU was demonstrated by the BN and the F344 strains (Shiels *et al.*, 1995). Therefore, these two strains were used to establish backcrosses to the former.

3.2 Results.

3.2.1 Initial set of SSLPs: Optimisation of the PCR conditions.

This part of the project commenced towards the end of 1993 and continued for most of the following year. An initial set of 75 microsatellites were studied in duplicate by both Dr. P. Shiels and myself. Some sequences were obtained from some of the data published by Serikawa *et al.* (1992) and Kuramoto *et al.* (1993) and synthesized in house. Other primers were bought from Research Genetics (Huntsville, Alabama, US); no sequence or linkage data were at the time available for these latter SSLPs, merely the chromosomal locations. The PCR conditions were optimised so that a single product resulted from each reaction; this was accomplished by adjusting the concentration of $MgCl_2$ and changing the annealing temperature and the number of PCR cycles. The annealing temperature was usually identical or 1°-5°C lower than the melting temperature (Tm) of each primer pair, this being estimated for practical purposes as Tm= {(G+C) x 4°C} + {(A+T) x 2°C} (Suggs *et al.*, 1981). Those cases where the Tm of each of the oligonucleotides in the primer set differed by 5° to 10°C or more proved to be more difficult to optimise. In general, the annealing temperatures ranged between 50°-65°C. Most primer sets worked optimally in the presence of 1mM $MgCl_2$, with higher concentrations resulting in accompanying background. Many yielded clear products after 25 PCR cycles; a few others required running 30-35 cycles. Our PCR conditions differed, on many occasions, from those previously published. The conditions for each of the 75 SSLPs are presented in Appendix B, as part of the publication by Shiels *et al.* (1995). The table shows primer set MIT-R1380 as mapping to chromosome 19, as this was the information given by Research Genetics. However, this primer has been subsequently mapped to rat chromosome 6 (Maihara *et al.*, 1995; Jacob *et al.*, 1995).

3.2.2 Selection of parental strains.

DNAs from PVG, BN, F344 and AS/AGU rats were either obtained from Harlan Olac (Oxon, UK) or prepared from the rat spleens. Two different methods were followed in the extraction of DNA (Chapter 2; section 2.3.2.1), the first involving homogenisation of rat tissue with liquid nitrogen. This was soon replaced by the use of a DNA isolation kit (Flowgen Instruments Ltd., UK). Aliquots of each DNA sample were used as template in PCR reactions which were run according to the conditions previously determined as optimal. The products were run on 4% MetaPhor agarose gels which are of higher resolution than the standard agarose gels [Figure 3-1]. The results of this work are presented in Appendix B (Shiels *et al.*, 1995). BN and F344 were shown to diverge from AS/AGU to a greater extent (at 62% and 47% of the loci, respectively) than the PVG strain (at 43% of the SSLPs).

Two separate backcrosses were set up with the BN and the F344 strains as reference parents. The reason behind setting up the F344 backcross in addition to the BN backcross was to maximise the number of microsatellites that were informative, as some of these had only shown allelic differences between the F344 and the AS/AGU strains and were thus not useful in the cross to the BN strain.

3.2.3 Optimisation of further SSLPs.

During the time needed to produce the first batches of backcross animals, a further set of primers were bought from Reseach Genetics, optimised, and tested on the four rat strains. Two new members of the group, Mr. M.Canham and Mr. A. McCallion, and myself carried out this part of the work. The main problem was the coverage of the larger chromosomes 1-5, especially 1, 3 and 5. The size of these chromosomes was estimated to be in the range of 134 cM (chromosome 5) to 221 cM (chromosome 1). A single informative marker had been identified in each of the chromosomes 11, 15, 16, 17 and Y. These were medium-sized chromosomes, with lengths of 70-85 cM. More problematic were chromosomes 9 and 20, for which markers which were informative for any of the backcrosses had not been identified yet. No informative markers had been found on the X chromosome either. However, coverage of the autosomes was the most immediate concern since inheritance of the *agu* mutation did not seem to demonstrate any linkage to sex.

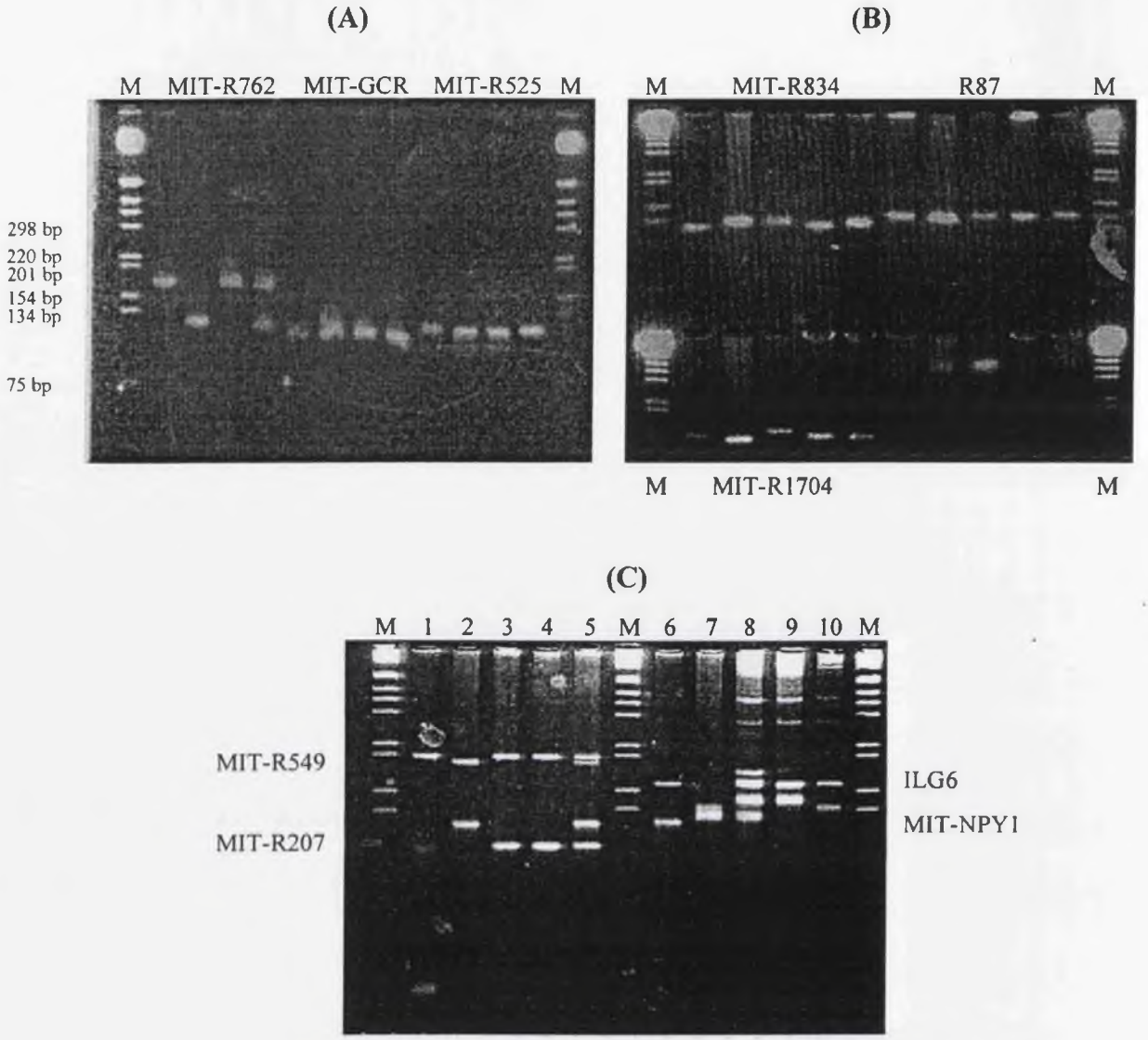


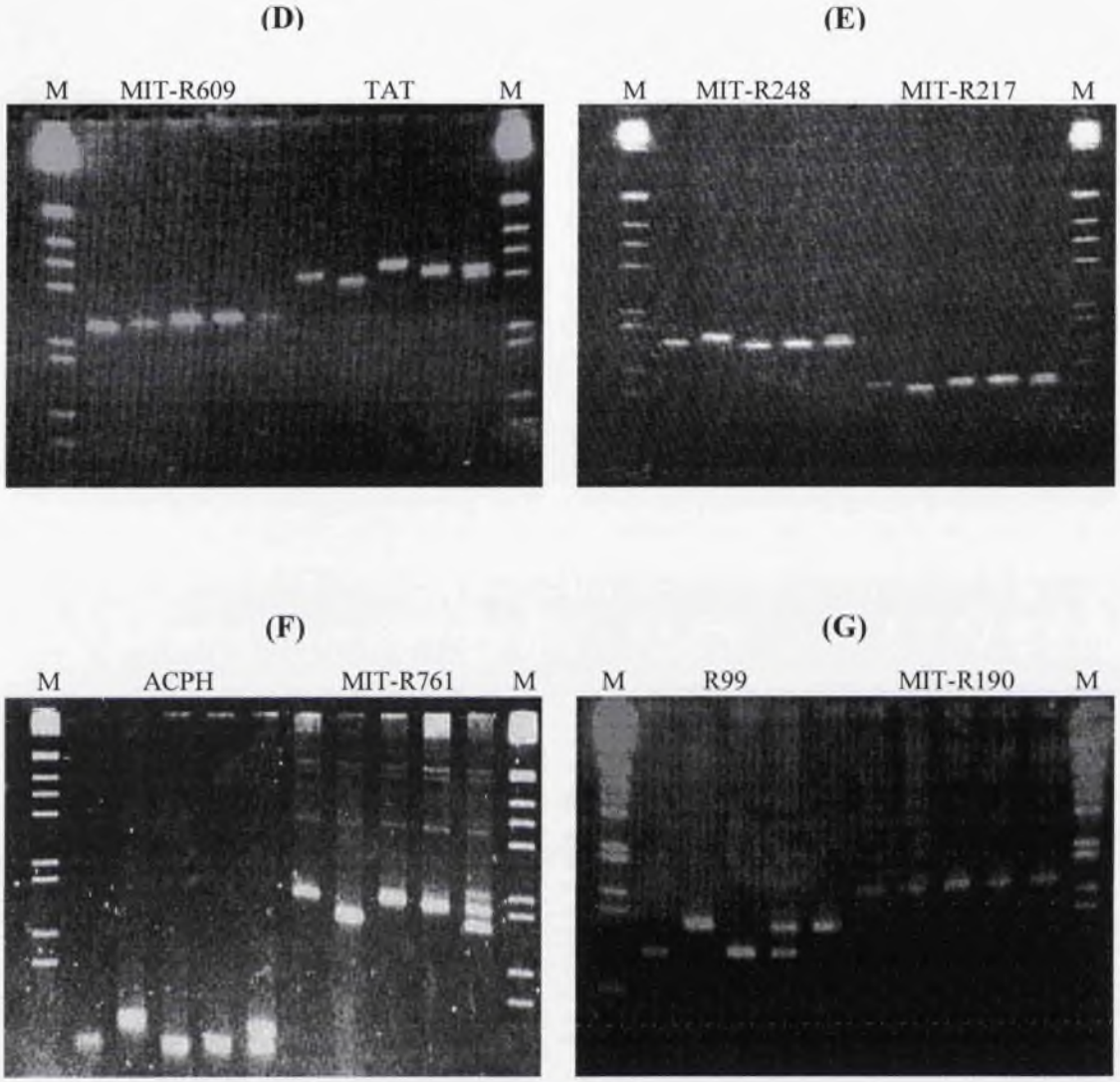
FIGURE 3-1. Resolution of PCR products on 4% MetaPhor agarose gels.

A 1kb DNA ladder (M) was loaded as the size marker on all the gels.

(A) From left to right: products resulting from the amplification of PVG, BN, AS/AGU and BN:AS/AGU (mixed in equal volumes) DNA templates with each of the primers MIT-R762, MIT-GCR, and MIT-R525, respectively.

(B) Top half of the gel, from left to right: products resulting from the amplification of PVG, BN, F344, AS/AGU and BN:AS/AGU (mixed in equal volumes) DNA templates with th primers MIT-R834 and R87, respectively. Bottom half: amplification products obtained with th primer set MIT-R1704; the order of the samples is as indicated above.

(C) Resolution of multiplexed PCR products obtained in separate reactions with the primer MIT-R549 and MIT-R207 on PVG, BN, F344, AS/AGU and BN:AS/AGU (mixed in equa volumes) DNA templates (lanes 1-5), and of the products obtained in separate reactions with th primers ILG6 and MIT-NPY1 on PVG, BN, BN:AS/AGU (mixed in equal volumes), AS/AGU and F344 DNA templates (lanes 6-10).



(D) From left to right: Amplification products obtained with the primer sets MIT-R609 and TAT on PVG, BN, F344, AS/AGU and BN:AS/AGU (mixed in equal volumes) DNA templates.

(E) From left to right: Amplification products obtained with the primer sets MIT-R248 and MIT-R217 on PVG, BN, F344, AS/AGU and BN:AS/AGU (mixed in equal volumes) DNA templates.

(F) From left to right: Amplification products obtained with the primer sets ACPH and MIT-R761 on PVG, BN, F344, AS/AGU and BN:AS/AGU (mixed in equal volumes) DNA templates.

(G) From left to right: Amplification products obtained with the primer sets R99 and MIT-R190 on PVG, BN, AS/AGU, BN:AS/AGU (mixed in equal volumes), and F344 DNA templates.

The results obtained with some of the new SSLPs are presented in Table 3-1. As estimated from the data, the total number of loci shown to differ between AS/AGU and each of the other rat strains after the new additions closely followed the values already obtained with the initial set of 75 microsatellites. The PVG and F344 strains were now exhibiting a very similar degree of divergence from the mutant strain (approximately 45% and 42% of the loci, respectively). BN and AS/AGU differed at 63% of the analysed SSLPs.

The backcross animals started to arrive and extensive typing work was quickly commenced by Dr. P. Shiels and myself. As described in Chapter 4, a locus was found in these initial screenings which appeared to be linked to the *agu* gene. Thenceforth, work was directed towards the optimisation and testing of SSLPs which mapped to the particular chromosome, any further analysis of other primer sets coming to an end. Consequently, analyses with heterologous mouse primers were not needed at this point.

3.2.4 Clarification of an ambiguous result.

The release in 1995 of the primer sequences which had already for some time been commercially available from Research Genetics made possible the comparison of the sequences of the primer sets previously bought from this company with others which had to that date been analysed in the project. It was then observed that the forward and the reverse sequences of the primer sets R101 (Serikawa *et al.*, 1992) and CASAG1 (Hilbert *et al.*, 1991) were identical and thus amplified the same microsatellite in the α-casein (*Csn1*) locus. Since our data recorded some differences in the results obtained with these two primers (Shiels *et al.*, 1995), these must be artifactual. Therefore the analyses had to be repeated to eliminate the discrepancies. These discrepancies are listed below.

a) The annealing temperature determined as optimal for 'R101' had been 52°C, whereas that for 'CASAG1' had been 53°C, both PCRs running 25 cycles, in the presence of 1 mM $MgCl_2$. Single bands had indeed been obtained with 'R101' at 52°C; identical results were observed when this primer was used in PCRs run at an annealing temperature of 53°C.

b) A small difference had been recorded between the sizes of the products obtained with 'R101' (120-140 bp) and those yielded by 'CASAG1' (115-130 bp). Nonetheless, these results could be easily reconciled as, in general, the sizes of the

Chromosome	Primer name	Annealing Temperature (0^C)	PCR Product size (bp)	Allele differences
1	IGF2	53	135 -160	AGU=F344=PVG>BN
1	LEUKOS1 [b]	70	130	AGU=BN=F344=PVG
1	D1N64 [a]	66	300	AGU=BN=F344=PVG
2	MIT-PKLG	55	110 - 120	PVG>F344>AGU=BN
3	MIT-R593 [a]	57	170 - 190	BN>AGU=PVG>F344
3	MIT-R139 [a]	56	170 - 200	AGU=F344=PVG>BN
4	MIT-R514	57	270 - 290	PVG>BN>AGU=F344
4	SPRD [a]	57	120	AGU=BN=F344=PVG
5	MIT-R1698	53	135	AGU=BN=F344=PVG
5	MIT-R252	53	195 - 220	F344>BN=PVG>AGU
5	MIT-R589 [a]	53	220	AGU=F344=PVG>BN
5	MIT-R1138 [a]	53	100 - 140	AGU=F344>BN>PVG
5	MIT-R735 [b]	55	125 - 135	PVG>AGU=F344>BN
5	MIT-R980 [b]	53	220	AGU=F344>BN=PVG
7	MIT-R513	53	275	AGU=BN=F344=PVG
7	MIT-R1071	53	230 - 250	PVG>BN>AGU=F344
7	MIT-R605 [a]	53	90 - 100	F344=PVG>AGU=BN
10	MIT-MHCG	55	170 - 190	AGU>F344=PVG>BN
12	MIT-R534	54	135	AGU=BN=F344=PVG
12	MIT-R354	53	270 - 300	AGU=F344>PVG>BN
13	MIT-R120 [b]	51	250	AGU=BN=F344=PVG
13	MIT-R982	53	125 - 160	BN>AGU=F344=PVG
14	MIT-R1692 [a]	54	135 - 150	BN>AGU=F344=PVG
14	MIT-R1041 [b]	64	165 - 190	BN=PVG>F344>AGU
14	MIT-ALBZA	53	230 - 250	AGU=F344=PVG>BN
14	MIT-AFPGA	53	170 - 190	AGU>BN>PVG*
16	MIT-R220	57	170 - 190	AGU=F344>PVG>BN
16	MIT-R1254	53	150 - 160	AGU=PVG>BN=F344
18	MIT-ALBA	53	230 - 290	PVG>AGU=F344>BN

TABLE 3-1. Strain differences observed at each of a series of 29 SSLP loci.

All the primer sets worked optimally in the presence of 1mM $MgCl_2$; 25 PCR cycles were run in all the reactions.

[a] and [b] Primer sets for which the PCR conditions were optimised by Mr. A. McCallion and Mr. M. Canham, respectively.

* As previously reported for the primer R99 (Shiels *et al.*, 1995), two alleles were obtained when typing F344 DNA with the primer MIT-AFPGA. Only single products were produced when amplifying any of the other DNAs.

PCR products were not completely accurate but estimates calculated in reference to a 1 kb DNA ladder marker. The set of PCR products obtained with each of the primers were run side by side; no differences in the size range were observed.

c) The rat strains BN and PVG differed at the *Csn1* locus when analysed with 'CASAG1', the BN allele being larger than that present in the PVG strain. However, this variation was not found when using the primer 'R101'.

To that date, the analysis of all the PCR products had always been carried out on 4% MetaPhor agarose gels, prepared with 1xTBE; these gels displayed a considerable capability for high resolution. However, in the case of 'R101' / 'CASAG1', the ambiguous results obtained with the BN and the PVG strains indicated the need for a higher resolution of these PCR products. The difference observed on Metaphor with the 'CASAG1' primer pair had been barely detectable and therefore confirmation was needed. The corresponding forward primer was radioactively labelled and used in PCR reactions; the products were run on an 8% polyacrylamide gel [Figure 3-2]. Clear resolution of each of the products was achieved in this gel system, confirming that a difference existed between the BN and the PVG strains, the former carrying a larger *Csn1* allele than the latter. The 'R101' results could not be confirmed on polyacrylamide gels as the stocks which were being used run out.

The fact that no difference was observed between the BN and the PVG amplification products obtained with 'R101' when these were run on MetaPhor agarose gels was probably due to some variations in the casting and/or the running of the gels, or to differences between the two primer stocks since these were obtained from different sources. Another explanation would be that the 'R101' stocks did not correspond to such primer set. This seems unlikely, given the identical annealing temperatures and the approximate sizes of the PCR products. The 'CASAG1' primer set had been bought from Research Genetics; linkage of the amplified locus to other markers on chromosome 14 (work carried out by Mr. M. Canham) supported that these stocks corresponded to the expected primer sequences.

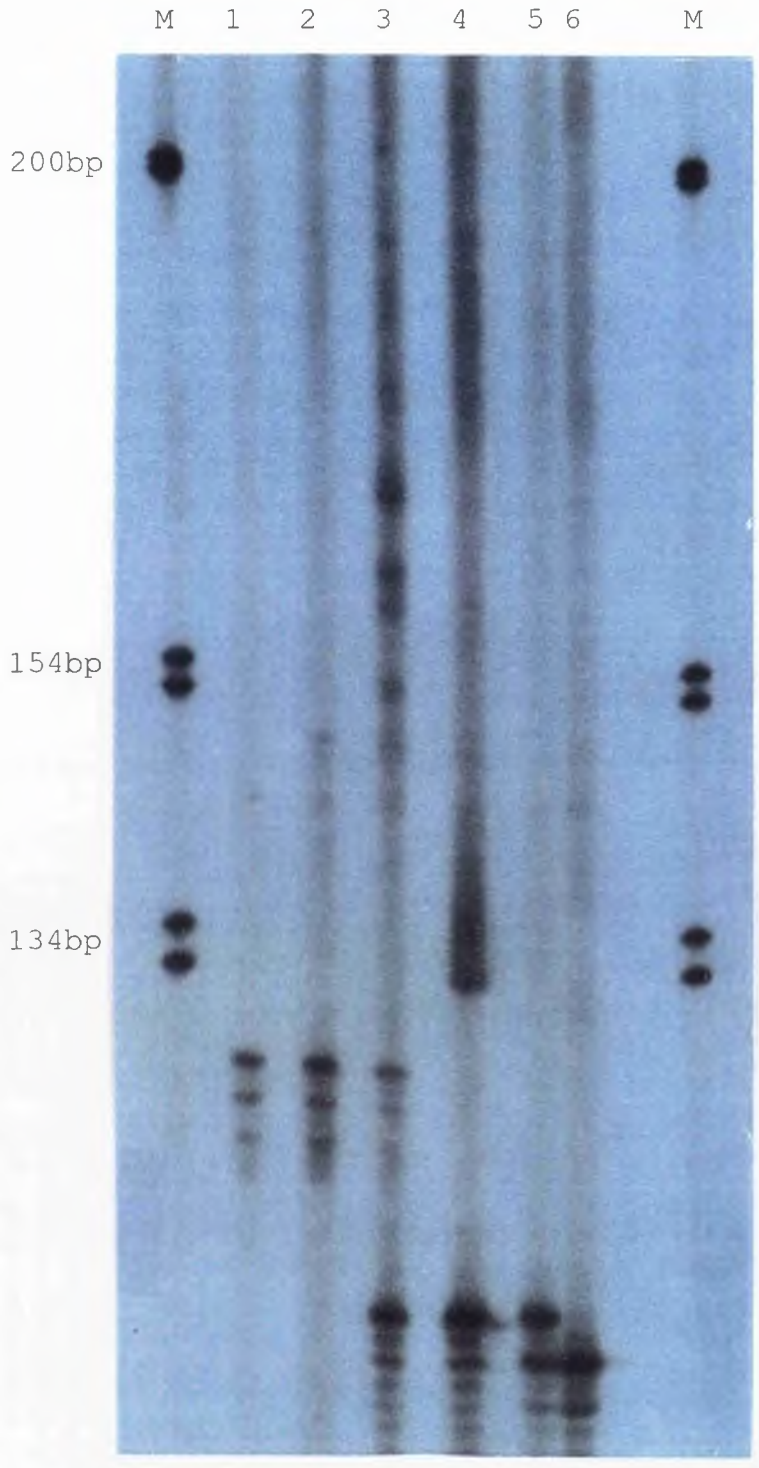


FIGURE 3-2. Resolution of 'CASAG1'-PCR products on an 8% polyacrylamide gel.

Amplification products from: F344 (1), AS/AGU (2), (AS/AGU x BN) F1 (3), BN (4), BN and PVG DNAs mixed in equal volumes (5), and PVG (6) DNA templates. A radioactively labelled 1 kb DNA ladder marker was used as the size standard (M).
The observed laddering pattern is a common feature of dinucleotide repeats and it is considered to result from slipped mispairing in the repeat during PCR amplification (Luty *et al.*, 1990; Hughes, 1993).

3.3 Discussion.

3.3.1 Factors affecting the reliable interpretation of typing data.

A linkage analysis approach to the localisation of the *agu* gene in the genome was based on the assumption that each parental strain was homozygous at every single locus and that no mutations occurred at the marker loci when transmitted from generation to generation. Consistent co-amplification of two DNA fragments under already optimised PCR conditions may result from either the presence of two different alleles at the particular locus (i.e. heterozygosity) or from amplification of two different loci of very similar sequences (e.g. gene families, pseudogenes). In two instances, amplification of F344 DNA extracted from two different animals yielded divergent results where either 1 or 2 bands were consistently amplified, one being always of the same size as the AS/AGU PCR product. This was observed with the primer sets R99 (Serikawa *et al.*, 1992) and MIT-PLANH (Research Genetics). Therefore, these two loci must have been homozygous in one animal and heterozygous in the other. Residual heterozygosity (Hedrich, 1990c) of the parental strains can, in these cases, lead to erroneous interpretation of the genotyping results and may result in failure to detect existing linkage. Another primer, MIT-AFPGA (Research Genetics) also resulted in the co-amplification of two discrete bands from the F344 DNA template. This primer was only tested on one animal; however, a single allele was reported by Jacob and colleagues (1995) and it is likely that this locus was heterozygous in the F344 animal studied.

Mutation rates at the genetic markers are considered to be very low (Weber and May, 1989; Serikawa *et al.*, 1992). Most of the microsatellite loci analysed in this project consist of dinucleotide repeats and the mutation rate at these loci has been observed to be even lower than at some tri- and tetranucleotide repeat sites (Weber and Wong, 1993; Ellegren, 1995). Moreover, only those mutations which affect the allelic length and yet do not disrupt codominant inheritance patterns can cause problems in the linkage analyses. Otherwise, changes in allele size are readily detectable.

3.3.2 Establishment of an efficient large scale typing system.

Typing of the progeny was a very labour-intensive effort. A genome-wide search approach required the analysis of a very large number of informative loci. A number of

800 animals was planned to be initially handled until the specific chromosomal region carrying the *agu* gene was clearly established. Once markers were identified in proximity to this locus, finer genetic map distances would be obtained through the analysis of a larger number of offspring. It was therefore of prime importance to develop a typing procedure where a maximal output was obtained, by reducing the time, cost and effort inputs. This was achieved through improvements on the processing of the DNA samples and other parameters affecting the typing of those DNAs.

The first DNAs were prepared according to a long protocol involving the homogenisation of the rat tissue with liquid nitrogen. This method was soon replaced by a DNA isolation kit. As demonstrated once the first series of backcross animals started to arrive, more than 50 DNAs could be extracted per person per day with this kit. Practically all the DNAs were shown to be of a high quality. High DNA yields were consistently obtained from small amounts of spleen, presenting this tissue as a very large source of DNA. Consistent yields made spectrophotometer recordings for each of the samples unnecessary. A standard dilution factor was applied when preparing the working stocks. This was a remarkable improvement since taking the OD readings constituted one of the most time-consuming steps in the preparation of DNA samples.

A great number of PCR reactions were to be carried out. The possibility of multiplexing different amplification reactions, of current application in typing work, was considered. However, the vast majority of the studies where this approach is followed use fluorescently-labelled primers which allow the unambiguous identification of the products amplified by each of the individual primers. Labelled primers were not being used in this project. Therefore, a first requisite for the multiplexing of several PCR reactions to be feasible was a clear difference in the sizes of the products amplified by each primer pair. Similar product sizes could give rise to problems when trying to identify the different amplification fragments on the gel. This factor greatly restricted the number of reactions which could be multiplexed in future typings of the backcross animals. Further restrictions were imposed by the fact that the primers must be informative in the same backcross (i.e. the primers typed the same set of animals). Moreover, multiplexing of different amplifications in the same PCR reaction involved the use of primers which worked under the same conditions. Should the primers require different PCR conditions, multiplexing could only be feasible when loading the samples on the gel [Figure 3-1(C)] but not in the PCR reaction itself. Multiplexing of some PCR reactions was initially tested on the four strains of rat under study. However, a higher

failure rate was obtained in comparison to single locus amplifications. These observations are in agreement with the data reported by Lord and coworkers (1995). Therefore, the great majority of the typings were carried out in independent reactions.

Differently from minisatellites, the repeat units found in microsatellites can be as small as 1-2 bp. Unambiguous detection of an allelic difference at a single repeat unit requires running polyacrylamide gels. This is expensive and time-consuming. Instead, 4% MetaPhor agarose gels were solely employed in the establishment of this initial genetic scan of the rat genome. This gel system offers high resolution of the products in running times of 1-3 hours. According to the manufacturer, 4% MetaPhor agarose gels prepared and run in TBE buffer give comparable resolution to 8% polyacrylamide gels when analysing fragments in the 70-300 bp range. Concentrations of 5% may be required to get very fine resolution of fragments less than 100 bp. These gels are however more difficult to prepare and need longer running times. Experience with the 'R101'/'CASAG1' primer showed the resolution of 4% MetaPhor gels is in fact not so high. These particular PCR products were in the range 115-130 bp, very close to the threshold size which can be resolved on these gels. In general, polyacrylamide gels seem to be the most reliable system to study strain differences in these instances, although new alternative methods are now available, such as the use of 6% NuSieve agarose gels (Plaha *et al.*, 1995). Advantages of this latter system is the high resolution achieved and the possibility to run each gel at least five times without repouring. This feature is in contrast to 4% MetaPhor gels, since our attempts to reuse the same gel, with or without repouring, failed.

3.3.3 Strain distribution pattern of alleles at 103 microsatellite loci.

In agreement with the results obtained by other groups (Serikawa *et al.*, 1992; Maihara *et al.*, 1995), the BN strain displayed the greatest number of divergent loci when compared to the AS/AGU, the F344 and the PVG strains, differing at 63%, 68% and 65% of the tested SSLPs, respectively. These observations are all consistent with the phylogenetic tree for inbred strains recently constructed by Canzian and colleagues (1995), where BN is presented as the most distant of 13 inbred rat strains.

Given the frequency of informative SSLP loci, BN was chosen to be the reference parental strain in a first backcross to AS/AGU. F344 and PVG diverged at 42% and 45% of the loci, respectively, when compared to the mutant strain. Both F344

and PVG exhibited a proportion of approximately 9% of the loci which differed between either of them and AS/AGU but did not show allelic differences between this latter and the BN strain. F344 differed at a greater proportion of SSLP loci from BN and was selected to be the reference strain in a second backcross.

The selection of BN and F344 to act as the reference parental strains translated into 75% of the studied microsatellites (i.e. 77 SSLPs) being informative in the backcrosses. According to the calculations by Jacob *et al.* (1991), the proportion of the rat genome expected to lie within *d* cM of a marker, assuming random distribution of the genetic markers being employed, is approximately $1\text{-}e^{-2Nd/D}$, where N is the number of markers and D is the size of the rat genome. Given an estimated genetic length of the rat genome D=2,400 cM* (Robinson, 1972), and a number N=77 markers, and in general assuming a random distribution of such markers, 96% of the genome should lie within 50 cM of a marker and 85% within 30 cM of a marker. Complete coverage of the genome had thus not been achieved. Moreover, assumption of an even distribution of the particular markers throughout the genome was already proven wrong. No informative markers had been found on the chromosomes 9, 20 or X. Medium-sized chromosomes such as 11 (ca. 81 cM), 15 (ca. 81 cM) and 17 (ca. 73 cM) had been assigned single markers, whereas others such as chromosomes 10 (ca. 94 cM) and 14 (ca. 85 cM) had been assigned 9 and 8 informative microsatellite loci, respectively. Despite this lack of coverage of the rat genome, linkage of the *agu* gene to one of the characterised SSLPs was detected during the initial typings of backcross offspring. The study of rat or mouse microsatellites which mapped outside the identified chromosome was not pursued. Further analyses focused instead on SSLPs which mapped to this chromosome.

* The length of the genetic map, D, varies according to different estimates (Robinson, 1972; Jacob *et al.*, 1995; Toyota *et al.*, 1996). Consequently, calculation of the degree of coverage obtained with the available markers depends on the value given to D. More recently released rat genetic maps (Toyota *et al.*, 1996; Bihoreau *et al.*, 1997) estimate the length of the rat genome as approximately 2000 cM. Calculation of the degree of coverage provided by the 77 informative markers which were available at the time this part of the project was carried out, based on D= 2000 cM, would suggest that 98% of the genome should lie within 50 cM of a marker and 90% within 30 cM of a marker.

CHAPTER 4

Localisation of the *agu* locus to a single chromosomal region.

4.1 Introduction.

A series of 103 SSLPs were studied in the AS/AGU, the BN, and the F344 strains of rat. Allelic differences between these latter and AS/AGU were identified at 63% and 42% of the loci, respectively. These two strains were selected as reference parents in backcrosses to AS/AGU. The initial goal was to obtain eight-hundred offspring, 400 from each cross. Towards the end of 1994, typings with the optimised primer sets started on the batches of litters, as they arrived.

Linkage of any given microsatellite marker to the *agu* locus was assessed by correlating the amplification pattern with the phenotype exhibited by each animal [Figure 4-1]. One of two possible PCR results was obtained from each DNA sample: a single band, of the same size as that corresponding to the AS/AGU homozygote, or two bands, of the same size as the products obtained from the AS/AGU and the reference strain, respectively.

A microsatellite marker was found which demonstrated a highly significant deviation from random reassortment with the *agu* locus, allowing the chromosomal localisation of this gene. Other SSLPs which mapped to the same chromosome were studied. Those which were informative between AS/AGU and one of the reference strains were used for further linkage analyses, resulting in a more precise location of *agu*. Other markers were available which were not tested, as they could not be typed by PCR (Hilbert *et al.*, 1991; Goldmuntz *et al.*, 1993; Pravenec *et al.*, 1996).

4.2 Results.

Each backcross animal was individually identified and its phenotype studied (work done by Dr. R. Sutcliffe and Mr. N. Bennett) at weekly intervals from 5 weeks after birth onwards, by applying a series of tests already mentioned in Chapter 1 (section 1.1). Once the presence or the absence of the mutant phenotype was unambiguously demonstrated, the animals were killed and their spleens extracted and stored at -70^{0}C (Mr. N. Bennet, Dr. R. Sutcliffe, Dr. P. Shiels, Mr. M. Canham, Mr. A. McCallion, Dr. D. Donald and Mrs. M. Gardiner). DNA was prepared from each tissue sample, as described in Chapter 2 (section *2.3.2.1*).

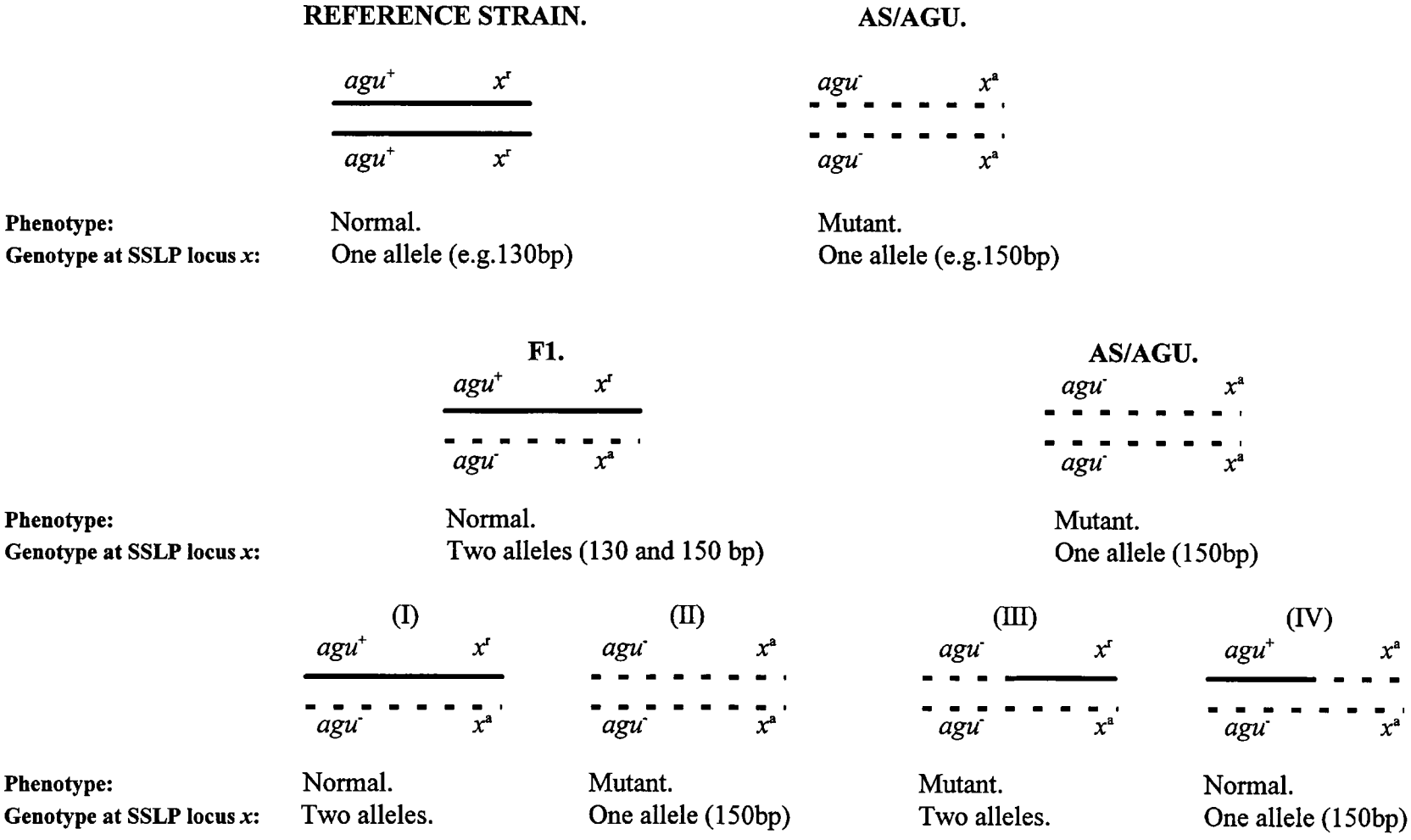


FIGURE 4-1. Diagram showing the possible combinations of phenotype and the amplification pattern for a given SSLP locus in the backcross progeny. The model assumes none or 1 cross-over between the loci. In the instances where 2 or more recombinations have occurred between *agu* and the SSLP locus, these can pass undetected, falling into categories (I) or (II). Internal markers are then required to identify these events.

4.2.1 Identification of the chromosome which carries the agu gene.

Typings with each primer set were carried out by amplifying ca.100 ng of each DNA sample in PCR reactions for which the optimal conditions had already been determined. The products were then run on 4% MetaPhor agarose gels. In some instances, variations in the DNA concentration resulted in the presence of background bands.

On average, 20-50 backcross progeny were initially typed with each of 26 SSLPs (work done by Dr. P. Shiels and myself) [Figure 4-2]. The significance of any deviation from a 1:1 segregation from the *agu* locus was statistically assessed by means of χ^2 analyses. These data are presented in Table 4-1. One of the primer sets, R33 (Serikawa *et al.*, 1992), demonstrated clear linkage to the *agu* locus (χ^2 (df=3)= 33.38, $p<0.001$). This microsatellite mapped to the kallikrein (renal) gene (*Kal*). Typing of 118 [AS/AGU x (AS/AGU x BN) F1] animals (by Dr. P. Shiels) yielded a recombination frequency with respect to the *agu* locus of 24.6% ± 4.0%. This finding pointed to chromosome 1 as the chromosome to which the *agu* gene mapped.

4.2.2 Narrowing the chromosomal region to which the agu gene maps.

The next step following the identification of the chromosome which putatively carried the *agu* gene was to confirm this result as well as to define a more precise location of the gene. To do this a selection of SSLPs known to map to chromosome 1 were obtained and tested for being informative in our backcrosses. The PCR conditions and strain distribution pattern (SDP) of this new set of SSLPs are shown in Table 4-2. Linkage estimates of the informative loci on chromosome 1 to *agu* are shown in Table 4-3. Clear linkage of *agu* to the loci *D1Mit1*, *D1Mgh7*, *D1Mit9*, *D1Mco2* and *D1Cep4* confirmed that *agu* mapped to rat chromosome 1 [Figure 4-3].

Two maps were established, one for each backcross. These are shown in Tables 4-4 and 4-5. The locus *D1Cep4* (initially referred to as *2B1* by Serikawa and colleagues (1992) and Goldmuntz *et al.* (1993)) was identified as the closest to *agu* (4.3 ± 1.3 cM). A precise mapping function for the rat genome has not been established yet. In general, either the Haldane or the Kosambi's functions are employed to transform recombination frequencies into genetic distances (Green, 1981; Jacob *et al.*, 1995; Gu *et al.*, 1996). When the Kosambi's function was applied to our data (Chapter 2; section 2.4), estimated distances and obtained recombination frequencies were identical when these

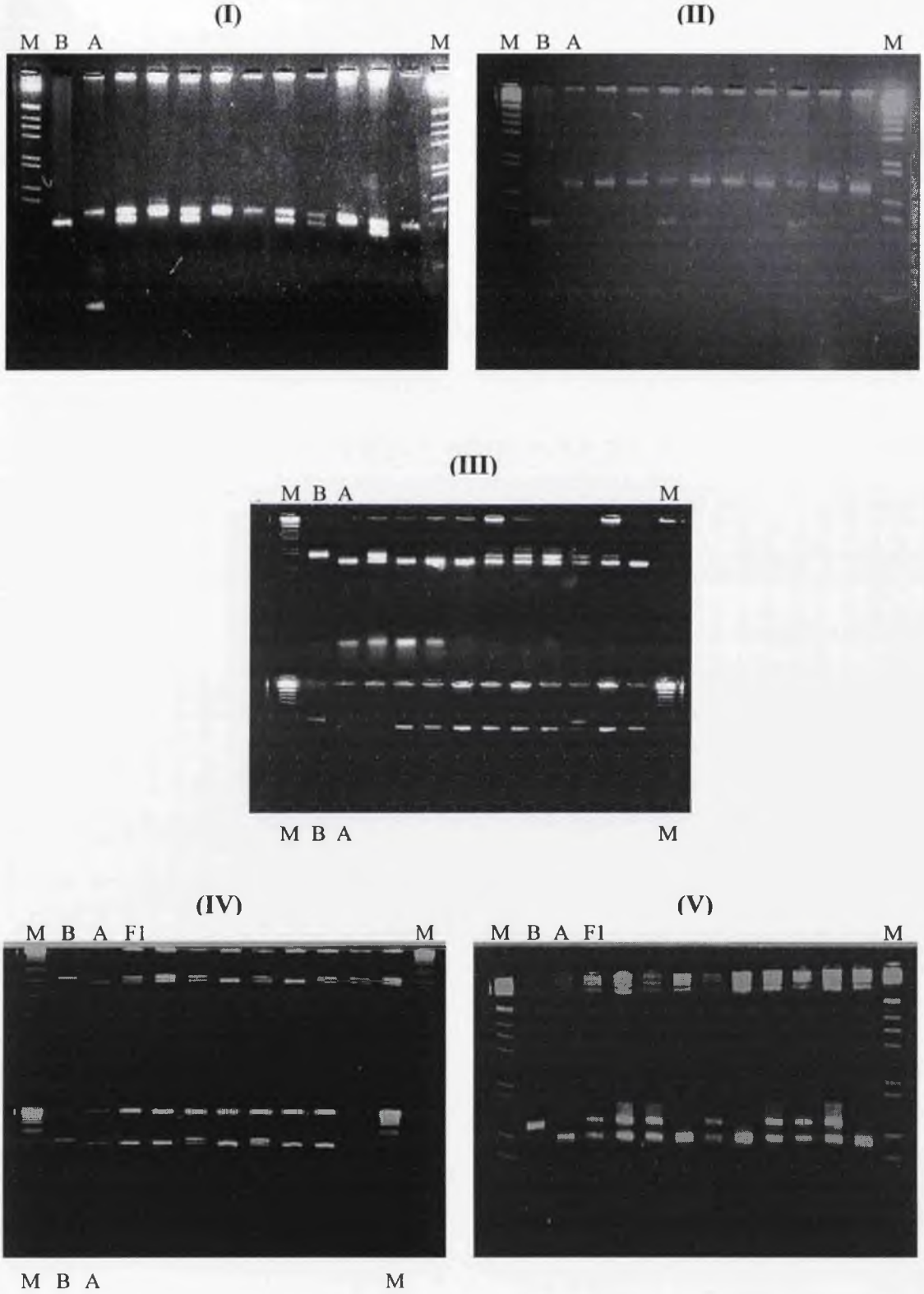


FIGURE 4-2. Typing of [AS/AGU x (AS/AGU x BN) F1] progeny with SSLP markers.

Resolution of the PCR products obtained following amplification of DNA template from BN (B), AS/AGU (A), (AS/AGU x BN) F1 (F1), and [AS/AGU x (AS/AGU x BN) F1] animals, with each of the primer sets R30 (gel (I)), MIT-R762 (gel (II)), R256 (gel (III)), and MIT-R844 (gels (IV) and (V)). A 1 kb DNA ladder was used as the size marker (M).

SSLP locus (Primer)	Chromosome	No. of animals typed	χ^2 value (df=3)
D1Mgh22 (IGF2)[a]	1	50	2.32 (0.9>p>0.5)
Lsn (R161)	1	34	4.29 0.5>p>0.1)
Kal (R33)[a]	1	118	33.38 (p<0.001)
Pkl (R34)[a]	2	35	0.6 (0.9>p>0.5)
Cpb (CARBO8b)[a]	2	18	2.4 (0.5>p>0.1)
D2Mit9 (R479)[a]	2	17	0.64 (0.9>p>0.5)
D2Mit14 (R360)[a]	2	17	4.5 (0.5>p>0.1)
D2Mit17 (R248)[a]	2	20	2.6 (0.5>p>0.1)
D3Mit13 (R244)[a]	3	20	2.6 (0.5>p>0.1)
A2m (R76)[a]	4	36	1.5 (0.9>p>0.5)
D4Mit19 (R1209)[a]	4	44	2.72 (0.5>p>0.1)
D6Mit4 (R1126)[a]	6	14	1.42 (0.9>p>0.5)
D6Mit10 (R1380)	6	25	2.36 (0.9>p>0.5)
Myh3 (R6)[a]	10	20	8 (0.05>p>0.025)
D13Mit3 (R578)	13	32	2.5 (0.5>p>0.1)
Ren (R256)	13	50	3.92 (0.5>p>0.1)
Csna (CASAG1)	14	36	6.9 (0.1>p>0.05)
Igfbp3 (R113)	14	47	0.74 (0.9>p>0.5)
D15kyo1 (KR62)	15	21	4.71 (0.5>p>0.1)
D16Mit2 (R762)	16	32	2.5 (0.5>p>0.1)
D16Mit3 (R1254)	16	37	2.03 (0.9>p>0.5)
D16Mit5 (R220)	16	34	5.8 (0.5>p>0.1)
D17Mit2 (R834)	17	25	0.76 (0.9>p>0.5)
D18Mit8 (R844)	18	20	2.0 (0.9>p>0.5)
Adrb2 (R30)	18	36	2.9 (0.5>p>0.1)
Tat (TAT)	19	20	2.8 (0.5>p>0.1)

TABLE 4-1. Segregation of *agu* and SSLP alleles in backcross progeny.

[a] typings carried out by Dr. P. Shiels.

χ^2 values were calculated as described (Chapter 2; section 2.4).

LOCUS	PRIMER NAME	ANNEALING TEMPERATURE (^{0}C)	PCR PRODUCT (bp)	ALLELE DIFFERENCES	SOURCE.
*Bsis**	R151	62	140 - 150	PVG>AGU=BN=F344	Serikawa *et al.*, 1992.
D1Mco2	D1Mco2	54	155 -175	AGU=BN>F344=PVG	Gu *et al.*, 1996.
*D1Mgh2**	MSA	57	145 - 175	F344>AGU=BN	Jacob *et al.*, 1995.
*D1Mgh7**	R5583	50	135 - 155	F344>AGU=PVG>BN	"
D1Mgh17	R5731	62	125 - 150	AGU>F344>BN	"
D1Mit1	R1673	55	130 - 170	PVG>AGU=F344>BN	"
D1Mit3	R1049	55	130 - 140	BN>AGU=F344	"
D1Mit9	R234	60	130 - 145	AGU=BN>PVG>F344	"
Dat1	KR214	50	195	AGU=BN=F344=PVG	Yokoi *et al.*, 1996.

TABLE 4-2. PCR conditions and SDP of alleles corresponding to nine SSLP markers on chromosome 1.

All the reactions, except those with the primer KR214, were run for 25 cycles and were carried out in the presence of 1 mM $MgCl_2$. Efficient amplification of the *Dat1* locus required 30 PCR cycles.

* SSLP loci analysed by Dr. P. Shiels.

BACKCROSS	Locus (Primer name)	Number of animals typed	Recombination frequency (±SD)
[AS/AGU x (AS/AGU x BN) F1]	*D1Mgh17* (R5731)[c]	96	48.9 ± 5.1
	D1Mit1 (R1673)	400	5.5 ± 1.1
	D1Mgh7 (R5583)[b]	296	19.9 ± 2.3
	Kal (R33)[a]	118	24.6 ± 4.0
	C[c]	123	41.46 ± 4.4
	D1Mit3 (R1049)	23	47.8 ± 10.4
	D1Mgh22 (IGF2)[a]	50	54.0 ± 7.0
[AS/AGU x (AS/AGU x F344) F1]	*D1Mit9* (R234)	50	28.0 ± 6.3
	D1Mco2	258	7.0 ± 2.9
	D1Cep4 (R191)[d]	258	4.3 ± 1.3
	Lsn (R161)	34	44.1 ± 8.5

TABLE 4-3. Linkage estimates between informative SSLP markers on chromosome 1 and the *agu* locus.

These are based on the typing of animals from the BN and the F344 backcrosses.

Although informative for both backcrosses, *Kal*, *D1Mgh17* and *D1Mgh7* were only used to type [AS/AGU x (AS/AGU x BN) F1] offspring, as the presence of two alleles in the [AS/AGU x (AS/AGU x F344) F1] progeny could not be clearly established on 4% MetaPhor gels.

Work carried out by: [a] Dr. P. Shiels; [b] Miss N. Craig; [c] Dr. P. Shiels and myself; [d] Mr. A. McCallion and myself.

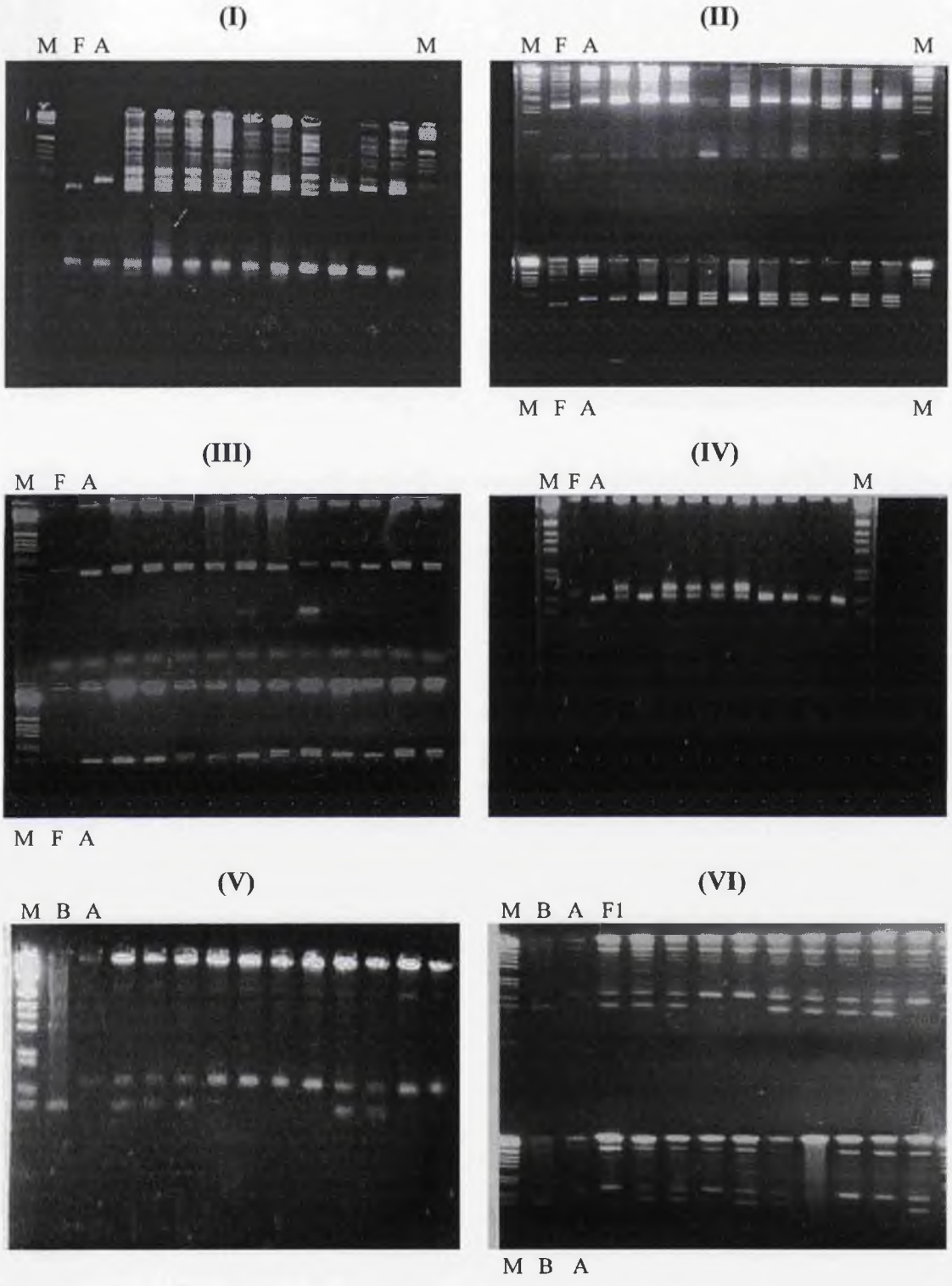


FIGURE 4-3. Typing of backcross animals with SSLP markers which are linked to the *agu* gene.

Gels (I) to (IV): Resolution of the PCR products obtained following amplification of DNA template from F344 (F), AS/AGU (A), and [AS/AGU x (AS/AGU x F344) F1] animals with the primer sets D1Mco2 (gels (I) and (II)), and R191 (gels (III) and (IV)).
Gels (V) and (VI): Resolution of *D1Mit1* amplification products obtained with DNA template from BN (B), AS/AGU (A), (AS/AGU x BN) F1 (F1) and [AS/AGU x (AS/AGU x BN) F1] animals.
A 1 kb DNA ladder was used as the size marker (M).

[AS/AGU x (AS/AGU x BN) F1]

LOCUS NAME	DISTANCE BETWEEN LOCI (cM)	95% LOWER AND UPPER LIMITS (cM)
D1Mgh17		
	46.9	36.9 - 56.9
D1Mit1		
	5.5	3.4 - 7.6
agu		
	41.5	32.9 - 50.1
C		
	4.3	0 - 11.2
D1Mit3		
	31.2	8.5 - 53.9
D1Mgh22		

TABLE 4-4. Order of loci on rat chromosome 1, based on the typing of [AS/AGU x (AS/AGU x BN) F1] offspring.

[AS/AGU x (AS/AGU x F344) F1]

LOCUS NAME	DISTANCE BETWEEN LOCI (cM)	95% LOWER AND UPPER LIMITS (cM)
D1Mit9		
	20.8	4.5 - 37.1
D1Mco2		
	3.5	1.4 - 5.6
D1Cep4		
	4.3	1.8 - 6.8
agu		
	44.1	27.4 - 60.8
Lsn (R161)		

TABLE 4-5. Order of loci on rat chromosome 1, based on the typing of [AS/AGU x (AS/AGU x F344) F1] offspring.

The order of the loci was established by multiple 3 point cross analyses, and conveys the smallest possible number of double crossovers in each region.

The loci *Kal* and *D1Mgh7* have been omitted in Table 4-4, as typing of [AS/AGU x (AS/AGU x BN) F1] progeny yielded two different orders for these SSLP markers.

Recombination frequencies have been used as direct estimates of the distance between adjacent loci (section *4.2.2*). Calculation of the 95% lower and upper limits was carried out as described (Chapter 2; section 2.4).

latter were lower than 6%. When applied to recombination frequencies in the range 20-32%, the distances obtained were not much different from the recombination values. All recombination frequencies above 40% resulted in distances of over 50 cM. Because it is not clear how adequate the Kosambi's function is when applied to very large intervals (Donis-Keller *et al.*, 1987; Jacob *et al.*, 1995), and given the good approximation of the recombinantion frequencies *per se* in the smaller intervals, the recombination rate was initially used as a direct estimate of the distances between loci.

D1Mgh7 and *Kal* could not be mapped in relation to the flanking loci, *agu* and *C*, since two different orders were obtained. *agu* - (20.3 ± 3.7 cM) - *D1Mgh7* - (21.2 ± 3.8 cM) - *Kal* was the order observed when studying these three markers on 118 [AS/AGU x (AS/AGU x BN) F1] progeny. When the same 118 animals were typed at *Kal*, *D1Mgh7* and *C*, the order obtained was *Kal* - (21.2 ± 3.8cM) - *D1Mgh7* - (22.9 ± 3.9 cM) - *C*. These two orders were mutually exclusive since the coat colour locus *C* had been found to map near the *D1Mit3* locus and at 41.5 ± 4.4 cM from the *agu* gene.

D1Mit1 and *D1Mgh7/Kal* were taken as the flanking markers of the region of interest. New SSLPs which mapped within this area were identified and analysed on the different strains, following optimisation of the PCR conditions [Table 4-6]. These loci were *Cype*, *Cear*, *D1Mgh5*, *D1Mgh6*, *D1Mgh18*, *D1Mco1*, *Calm3* and *Atp1a3*. The primer sets D1Kyo1 (locus *Ckmm*; mouse M105) and R93 (locus *Cebpa*), although not assigned to any linkage group on rat chromosome 1, were also tested, as these loci had already been mapped to this chromosome as well as to the chromosomal region of synteny in mouse. Other microsatellites mapping to cytochromes P450 and members of the *CEA* (carcinoembryonic antigen) gene family were also tested (this work being done by Dr. P. Shiels and Mr. M. Canham, respectively). None of these markers, with the exception of *D1Mgh18*, showed allelic differences between any of the rat strains under study, even when analysed on 6-8% polyacrylamide gels.

D1Mgh18 showed strain differences between AS/AGU and the F344 strain. However, an unexpected result was obtained with this SSLP since typing of [AS/AGU x (AS/AGU x F344) F1] animals indicated that it was not linked to the gene of interest nor to other two markers in the region which were informative for the same cross, *D1Cep4* and *D1Mco2*. This is discussed in Chapter 7.

At this point it became necessary to develop new microsatellite markers from the region of interest. This work is discussed in Chapter 5. In addition, it was later

LOCUS	PRIMER NAME	ANNEALING TEMPERATURE (°C)	PCR PRODUCT (bp)	ALLELE DIFFERENCES	SOURCE.
Atp1a3	ATP1A3	64	150	AGU=BN=F344=PVG	Gu *et al.*, 1996.
Calm3	CALM3	60	280	AGU=BN=F344=PVG	Gu *et al.*, 1996.
Cear (Cgm4)	R100	60	165	AGU=BN=F344=PVG	Serikawa *et al.*, 1992.
*Ckmm**	D1Kyo1	55	110	AGU=BN=F344=PVG	Kondo *et al.*, 1993.
Cype	CY45E1	58	125	AGU=BN=F344=PVG	Jacob *et al.*, 1995.
D1Mco1	D1Mco1	51	145	AGU=BN=F344=PVG	Gu *et al.*, 1996.
D1Mgh5	CEAC	57	140	AGU=BN=F344=PVG	Jacob *et al.*, 1995.
D1Mgh6	R5272	56	125	AGU=BN=F344=PVG	Jacob *et al.*, 1995.
D1Mgh18	CYSS	56	110 - 135	AGU=BN=PVG>F344	Jacob *et al.*, 1995.
*Cebpa**	R93	55	115	AGU=BN=F344=PVG	Serikawa *et al.*, 1992.

TABLE 4-6. SSLP-containing loci known to map or putatively mapping (*; see text, section *4.2.2*) to the region flanked by *D1Mit1* and *Kal/D1Mgh7* on rat chromosome 1.

All the PCR reactions, except those carried out with the CALM3 primer pair, were run for 25 cycles. Efficient amplification of the *Calm3* locus required running 30 PCR cycles. All the PCR reactions worked optimally in the presence of 1 mM $MgCl_2$. The SDP of the alleles at all the loci apart from *D1Mgh18* were studied on both 4% MetaPhor agarose gels and on 6-8% polyacrylamide gels.

observed that, contrary to the earlier maps (Serikawa *et al.*, 1992; Kuramoto *et al.*, 1993; Yamada *et al.*, 1994), which placed the loci *Pkc* (protein kinase C, type I) and *Cgm4* (*Cear*) at a considerable distance from the *D1Cep4* locus (at 12 and 23 cM, respectively) and distal to *Cype*, the maps released by Gu *et al.*, Yokoi *et al.*, and Ding *et al.*, all in 1996, indicated that these genes mapped between the two latter loci. The protein kinase C, type I gene was mapped by Gu *et al.* (1996) at 2.5 cM from the *D1Cep4* locus, following typing of over 150 animals. Therefore, given its location, this locus was analysed again in the AS/AGU and the reference strains (by Dr. P. Shiels). This time the PCR products were run on a 6% polyacrylamide gel, following radioactive labelling of the R158 forward primer (Serikawa *et al.*, 1992). A difference was observed in the size of the products obtained with F344 and BN DNA templates and that obtained with AS/AGU DNA. Although this primer was informative for both backcrosses, it was initially used to type [AS/AGU x (AS/AGU x F344) F1] offspring (work done by Dr. P. Shiels). No recombinants were observed following typing of 264 animals, rendering *Pkc* the closest locus to the *agu* gene.

4.3 Discussion.

Linkage of the kallikrein locus to the *agu* gene (24.6 ± 4.0 cM) was observed following typing of backcross offspring with a series of 26 SSLPs, indicating that *agu* is located on rat chromosome 1. Chromosome 1 is the largest in rat, considered to be about 10% of the female haploid genome (*Committee for a Standardized Karyotype of Rattus norvegicus* (1973)). A single linkage group of 130 cM and consisting of 18 microsatellite loci was described for chromosome 1 by Yamada and colleagues in 1994 [Appendix A; Figure A-1]. The map constructed by Jacob and coworkers in 1995 contained 40 SSLPs and covered 155.2 cM [Appendix A; Figure A-2]. Further microsatellite markers were subsequently developed by other groups (Gauguier *et al.*, 1996; Yokoi *et al.*, 1996). Important contributions were made by Gu *et al.* (1996), with the establishment of a linkage map of chromosome 1 which extended through 193 cM [Appendix A; Figure A-3]. Also, Ding *et al.* (1996) identified 2 linkage groups on this chromosome, altogether covering about 191 cM [Appendix A; Figure A-4]. By this time

the length of the female haploid genome was estimated as 3000 Mb or 2250 cM, which indicated a length of 226 cM (301 Mb) for chromosome 1.

Following the observations with the kallikrein locus, a further set of 16 microsatellite markers were tested to refine the location of *agu* on chromosome 1. Of these, 5 were not informative for either of the backcrosses, 3 were informative in the BN backcross and 6 in the F344 backcross. The other 2 were informative in both backcrosses, although they were only used to type BN backcross progeny, given the small size difference between the F344 and the AS/AGU amplification products.

The informative SSLPs and a coat colour marker were used to type backcross progeny. Clear linkage of six of these markers to the *agu* locus confirmed its location on rat chromosome 1 and established two linkage maps of this chromosome, one for each backcross [Figure 4-4]. Other microsatellites were also analysed which resulted not to be informative for any backcross. In addition, analysis on polyacrylamide gels showed a difference in the size of the AS/AGU and the BN and the F344 products at the *Pkc* locus. Typing of [AS/AGU x (AS/AGU x F344) F1] offspring placed this marker in close proximity to the *agu* gene. Unambiguous location of the loci *Kal* and *D1Mgh7* could not be determined since two discordant orders were established. These typings were carried out once; therefore confirmation of each individual genotype should be the first step towards identifying the correct order. This was not judged to be a high priority in this project, since these loci are distant from *agu*. The order *agu - Kal - D1Mgh7 - C* [Figure 4-5] is in agreement with the map published by Gu *et al.* (1996), where the order *D1Cep4 - Pkc - Kal - D1Mgh7* is presented. However, whereas this group shows a distance of 7.8 cM between *Kal* and *D1Mgh7*, the recombination rate obtained in this work is greater than 21%. A close similarity in the order and distances between the loci has been otherwise observed between the work published by Gu *et al.* (1996) and the results obtained in this project. This strongly suggests that the order *agu - Kal - D1Mgh7 - C* is correct and that the contradictory data and the large distance established between *Kal* and *D1Mgh7* are the result of typing errors.

The order of all the other mapped loci is in agreement with previous linkage maps (Jacob *et al.*, 1995; Gu *et al.*, 1996; Ding *et al.*, 1996). Contrary to the map released by Yamada *et al.* (1994), *Pkc* maps at the other side of the *D1Cep4* locus, near the *Cype* gene. This is the order found in the maps of Gu *et al.* (1996) and Ding *et al.* (1996). The distances between the loci are very similar to most of the published data.

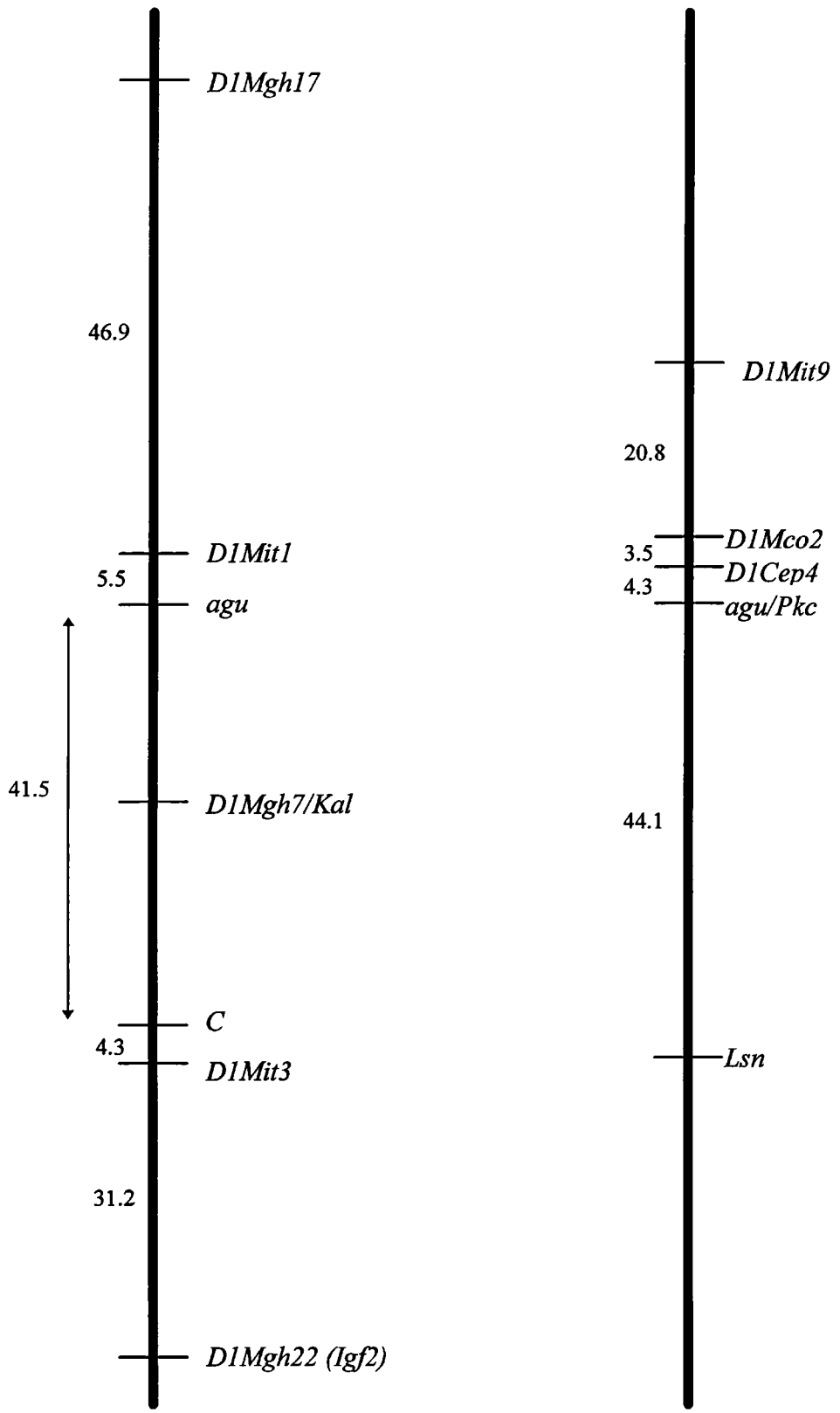


FIGURE 4-4. Linkage maps for rat chromosome 1.

The left panel was obtained following the typing of [AS/AGU x (AS/AGU x BN) F1] offspring. Data from typing [AS/AGU x (AS/AGU x F344) F1] animals led to the establishment of the map on the right.

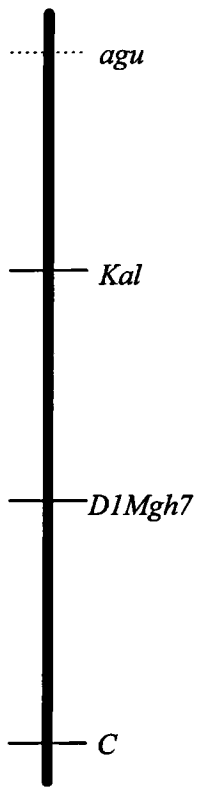


FIGURE 4-5. Probable order for *D1Mgh7* and *Kal* on rat chromosome 1.

Order obtained following the typing of 118 [AS/AGU x (AS/AGU x BN) F1] animals at the loci *Kal*, *D1Mgh7* and *C*. A different order was obtained when the same set of animals were typed at the loci *agu*, *Kal* and *D1Mgh7* (section *4.2.2*). Evidence that the order shown is correct comes from the data published by Gu *et al.* (1996) and from the genetic map recently released by Bihoreau *et al.* (1997).

CHAPTER 5

Development of new genetic markers.

5.1 Introduction.

A high-resolution genetic map is essential for positional cloning experiments. Genetic markers provide the means to identify the chromosomal position of the locus in question and to subsequently narrow down the location to a distance of 1 cM or less. A physical map can then be obtained for further analysis (Rommens *et al.*, 1989; Harshman *et al.*, 1995; Wang *et al.*, 1996).

In the case of this project, by the time a chromosomal fragment of approximately 25-30 cM in length (ca. 35-45 Mb) was established as containing the *agu* gene, no other informative SSLPs were available which mapped to this region. The linkage to *Pkc* had not yet been identified. It thus became necessary to search for new microsatellites within that chromosomal area. Two main approaches were followed. One approach was based on the observations by Stallings and colleagues (1991) and Moore and coworkers (1991) that some PCR primers defined in one species can be used to amplify conserved repetitive sequences in related species. This possibility has been exploited by other groups (Jacob *et al.*, 1991; Kondo *et al.*, 1993; Fredholm and Winter, 1995). As previously done by Kondo and colleagues (1993), mouse primers were used in this work to amplify rat genomic DNA. The other method was based on the screening of a rat genomic P1 library to identify clones carrying sequences which mapped or putatively mapped to the region under study. New SSLPs were then identified within the insert by probing the P1 clones with labelled oligonucleotide repeats and sequencing the hybridising DNA fragments.

5.1.1 Development of new SSLP markers.

Comparative mapping between different species can sometimes be a reliable means towards a preliminary characterisation of chromosomal regions which are not covered in one of the species. The robustness of this approach is closely dependent on the number of genetic markers which are present in the linkage map of each of the species being compared; these sequences are the basis to the accurate identification of conserved regions between the genomes.

The mouse and human regions of synteny with the rat chromosomal fragment of interest were sought. Not surprisingly, a closer homology was observed between the rodent maps than between these and the human map. Levan and colleagues (1991) and

Yamada *et al.* (1994) reported, respectively, that 15 out of 16 and 18 out of 22 genes mapped to rat chromosome 1 were located on mouse chromosome 7. The human homologues of these rat genes were mainly distributed on chromosomes 11 and 19. The same correlation to the human chromosomes was observed with mouse genes on chromosome 7 (Copeland *et al.*, 1993). Comparison of our results with the work done by Saunders and Seldin in 1990 indicated that the *agu*-containing region was likely to correlate to the proximal region of mouse chromosome 7 and to human chromosome 19q.

5.1.1.1 Amplification of rat genomic DNA with heterologous mouse SSLP primers.

A set of 10 mouse SSLP primer sets were selected (by Dr. P. Shiels) that mapped to the region in mouse chromosome 7 which appeared to be syntenic to the area of interest in rat chromosome 1. These were used (by Dr. P. Shiels and myself) to amplify rat genomic DNA, aiming to identify homologous rat markers. Some of the primer pairs yielded a very high background of amplified bands that impeded the identification of a single main product, even when 5% formamide was added to the PCR reaction. Others did not yield any product. Four amplifications resulted in a defined band; however, when the AS/AGU product was compared to those from the BN and the F344 genomic DNAs, no strain difference was observed. Therefore, no further work was done on these SSLPs. Other mouse primer sets were used to amplify DNAs from a panel of mouse x rat somatic cell hybrid clones (work done by Miss N. Craig) to confirm that the rat PCR product mapped to the rat chromosome of interest. However, the attempts to differentiate the mouse PCR product from the rat product failed, rendering this assay not useful in practice.

5.1.1.2 Isolation of rat P1 genomic clones for SSLP analysis.

Development of new rat microsatellite markers was achieved following the completion of a whole series of experiments on a rat genomic P1 library. This library is described in Chapter 2 (section *2.3.5*). P1 clones can carry large inserts (70-95 kb on average) (Sternberg, 1990b; Smoller *et al.*, 1991; Sternberg, 1992), and this system has been reported to be much more stable than, for example, YACs (Albertsen *et al.*, 1994; Nehls *et al.*, 1995). It was therefore considered that any given P1 clone containing a selected marker was likely to carry other sequences which are adjacent to this marker in

the chromosome. These inserts were therefore ideal substrates for the development of new microsatellite markers in the region.

Comparison of the rat and mouse genetic maps allowed the selection of a set of genes which, being placed in both the rat region of interest and the mouse chromosomal fragment of synteny, were considered to be putatively close to the *agu* gene. These loci were used to identify chromosomal regions containing SSLP markers which could be useful in refining the existing genetic map.

Most of the chosen loci contained already established SSLPs which had not shown any strain difference when the AS/AGU, the BN and the F344 strains of rat had been tested (Chapter 4). The primer pairs which amplified these microsatellite regions were now selected for future screening of the P1 library. ATPase, Na+K+ transporting, alpha 3 polypeptide (*Atp1a3*; primer set ATP1A3), calmodulin III (*Calm3*; primer pair CALM3) and *D1Mco1* (primer D1Mco1) were some of these loci; some members of the cytochrome P450 and of the *CEA* families were also selected by Dr. P. Shiels and Mr. M. Canham, respectively, to identify P1 clones of interest. The primer set R158 was also used to screen the P1 library, following identification of *Pkc* as the closest marker to *agu*. The estimated position of these loci is shown in Appendix A.

Analysis of the genes present in the mouse region of synteny led to the selection of an additional number of rat homologues, even though these were not mapped in the rat yet. These were the creatine kinase, muscle form (*Ckmm*; locus *D1Kyo1*), the ATPase-like vacuolar proton channel (*Atpl*), the ionotropic glutamate receptor, kainate 5 (*Grik5*), the sodium channel beta-1 subunit (*Scnn1b*) and the prostaglandin I receptor (*Ptgir*) genes. No microsatellite had been identified in any of the latter four genes nor in some of the *CEA* genes. In these instances, a primer set was designed to the 3' untranslated region (3' UTR) of the gene. These primer pairs were initially used in SSCP analyses (by Dr. P. Shiels, Mr. M. Canham and Miss N. Craig); no differences were detected between the AS/AGU rat and any of the reference strains. These primers were then employed to screen the P1 library.

My work focused mainly on the *Scnn1b* gene (Miss N. Craig designed a corresponding PCR primer set, D1Gu102, and optimised the amplification conditions), and on the *D1Mco1* rat SSLP.

5.2 Results.

The amplification products obtained with the primer sets D1Gu102, D1Mco1, CALM3, and a *Ckmm*-specific primer pair which was designed following comparison of both the muscle and the brain creatine kinase (Benfield *et al.*, 1984; Bendfield *et al.*, 1985) [Figure 5-1], were sequenced. These sequences were each confirmed with the published sequence prior to the utilisation of the primers to screen the P1 library. All the amplification products were, as expected, specific to the corresponding loci [Figure 5-2]. The primer sets CALM3 and CKM were stored for a possible future use. Thenceforth my work was centered on the *Scnn1b* and the *D1Mco1* loci.

5.2.1 Identification of P1 clones containing DNA sequences from the region of interest.

The optimised primer sets for selected rat loci were used to screen the P1 library. Identification of single P1 clones carrying the respective PCR sequence was achieved in 3 stages. These are described in section *2.3.5.2.*

Around 12 primer sets were initially tested by the group. Most of the genes were found following screening of the first fifty primary P1 pools; a few of them required testing further sets of pools. In a small number of cases the PCR signals were not reproducibly obtained in consecutive amplifications. Confirmation of the presence of the PCR sequence was then achieved by purifying the DNA in the lysate with QIAEX II-resin (section *2.3.13*) and using this DNA as template in the PCR reactions.

5.2.2 Sequence confirmation of the PCR product obtained from a single P1 clone.

An aliquot (1 µl) from each of the P1 cell cultures was used as template in PCR reactions; the amplification products were purified and sequenced. P1 clones selected for the rat loci *Scnn1b* and *D1Mco1*, respectively, yielded PCR products corresponding to the published sequences of these two loci. These P1 clones were called P1(*Scnn1b*) and P1(*D1Mco1*)) [Figures 5-3 (A) and 5-3 (B)].

5.2.3 P1 DNA isolation from a positive single clone.

Cultures were prepared with each of the two P1 clones under study. Bacteria were always grown in the presence of 25 $\mu g.ml^{-1}$ kanamycin and 5% sucrose, since the

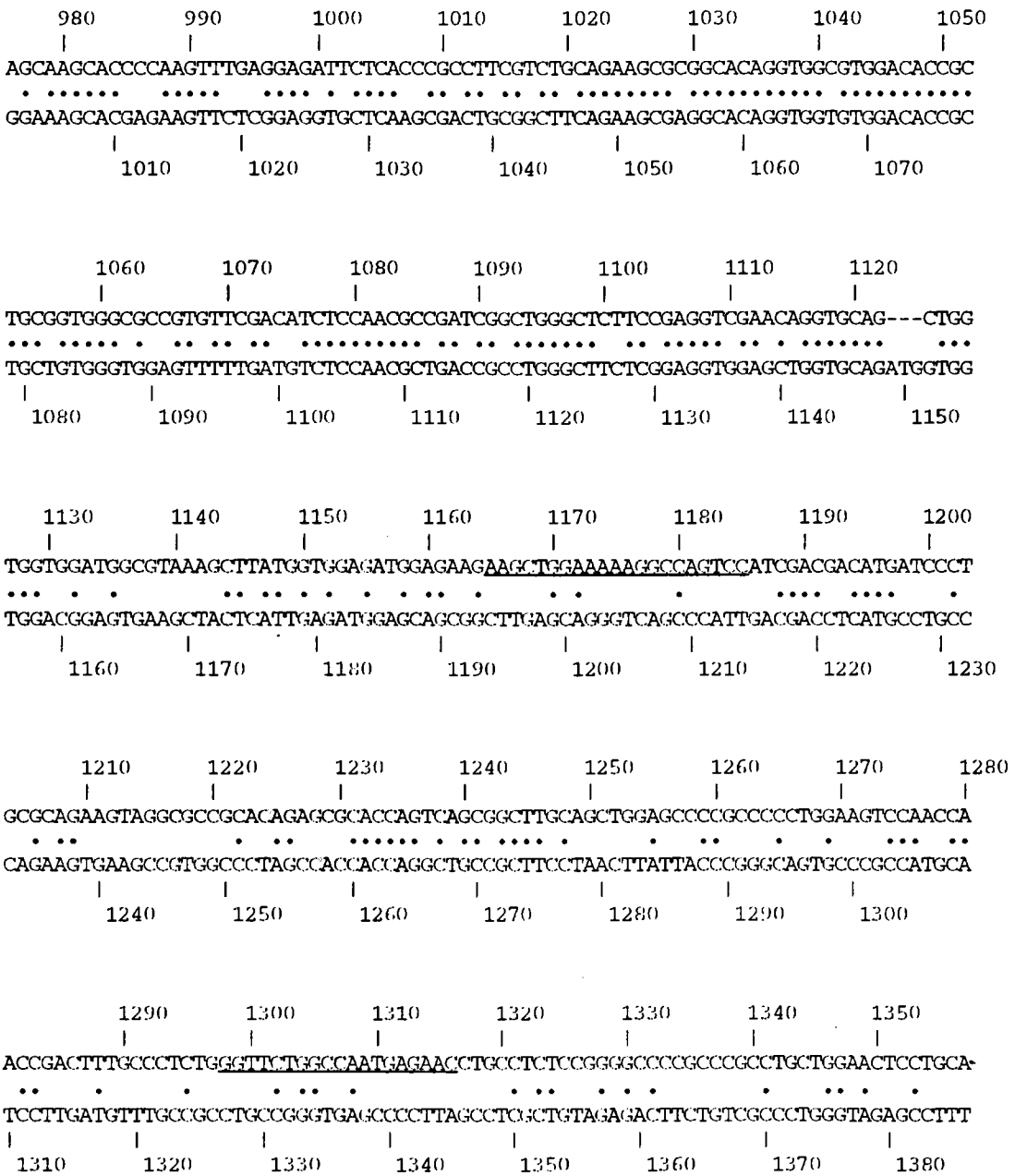


FIGURE 5-1. Design of a *Ckmm*-specific PCR primer set

The sequences of the rat skeletal muscle creatine kinase (*Ckmm*) composite mRNA (top sequence; Benfield *et al.*, 1984; Accession M10140) and the rat brain creatine kinase (*Ckb*) mRNA (bottom sequence; Benfield *et al.*, 1985; Accession M14400) were aligned and compared, in search for regions specific to the *Ckmm* gene. Gene-specific regions, shown above, were found at the 3' end of the sequences compared. A *Ckmm*-specific PCR primer was designed (underlined). Amplification with this primer set was carried out at an annealing temperature of 63°C for 25 cycles, in the presence of 1 mM $MgCl_2$.

(a) Rat skeletal muscle creatine kinase composite mRNA, complete cds.
(Length = 1410; Accession M10140)

```
Query:    22 GACTACATGATCCCTGCGCAGAAGTAGGCGCCGCACAGAGCGCACCAGTCAGCGGCTTGC 81
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  1189 GACGACATGATCCCTGCGCAGAAGTAGGCGCCGCACAGAGCGCACCAGTCAGCGGCTTGC 1248

Query:    82 AGCTGGAGCCCCGCCCCCTGGAANNCCAACCAACCGACTTTGCCCTCTGGGTTCTGGCCA 141
             ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Sbjct:  1249 AGCTGGAGCCCCGCCCCCTGGAAGTCCAACCAACCGACTTTGCCCTCTGGGTTCTGGCCA 1308

Query:   142 ATGANAAC 149
             |||| |||
Sbjct:  1309 ATGAGAAC 1316
```

(b) R.norvegicus CaMIII gene for calmodulin III
(Length = 7727; Accession X14265)

```
Query:   162 ACTGCCTGTGGGGGCGCGTNTGATGGGCCACAGGCGTTCCAGTTTTTCTGTTGTTGCCTG 103
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct:  5011 ACTGCCTGTGGGGGCGCGTCTGATGGGCCACAGGCGTTCCAGTTTTTCTGTTGTTGCCTG 5070

Query:   102 GCACCAGGAATCAGTTCNTAGAGCACAGGACTTGGTTTTGTTGATTGCTGCCATATCTTC 43
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct:  5071 GCACCAGGAATCAGTTCCTAGAGCACAGGACTTGGTTTTGTTGATTGCTGCCATATCTTC 5130

Query:    42 AATGCNTGGTGCATA 28
             ||||| |||||||||
Sbjct:  5131 AATGCCTGGTGCATA 5145


Query:   250 TCACTTGCTTCCTTCTACCAGAGNTCAGATACCCTGACCCCAACTCAAGCTGGCAG 195
             |||||||||||||||||||||| |||| ||||||||||||||||||||||||||||
Sbjct:  4911 TCACTTGCTTCCTTCTACCAGAGCTCAGCTACCCTGACCCCAACTCAAGCTGGCAG 4966
```

FIGURE 5-2. BLAST searches with the sequences obtained from amplification with (a) the CKM forward primer (CKM-F) and (b) the CALM3 reverse primer (CALM3-R).

As indicated, the two PCR products were confirmed as *Ckmm* (Benfield *et al.*, 1984) and *Calm3* (Nojima, 1989) fragments, respectively. A tract of nine (TCCC) is found in the *Calm3* gene, expanding from nucleotide 4972 to nucleotide 5008.

Rattus norvegicus chromosome 1 microsatellite sequence *D1Mco1*.
(Length = 342; Accession U19349)

```
Query:    59 TTTCAACACCTTTCAAGTAAACAAATGTGTTAATGAAATAGAAGAGGAGTTTTTGTTTT 1
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct:   120 TTTCAACACCTTTCAAGTAAACAAATGTGTTAATGAAATAGAAGAGAAGTTTTTGTTTT 178

Query:   115 CTTATCTGCGAGTGGATACCATA 93
             |||||| |||| ||||||||||
Sbjct:    62 CTTATCAGCGAATGGATACCATA 84
```

FIGURE 5-3 (A). Confirmation of the *D1Mco1* sequence obtained from a P1(*D1Mco1*) culture.

BLAST search with the sequence obtained following amplification of DNA from a P1(*D1Mco1*) culture with the reverse D1Mco1 primer. The sequence of the forward D1Mco1 primer is underlined. A $(TG)_{17}$ stretch expands from nucleotide 85 to nucleotide 118.

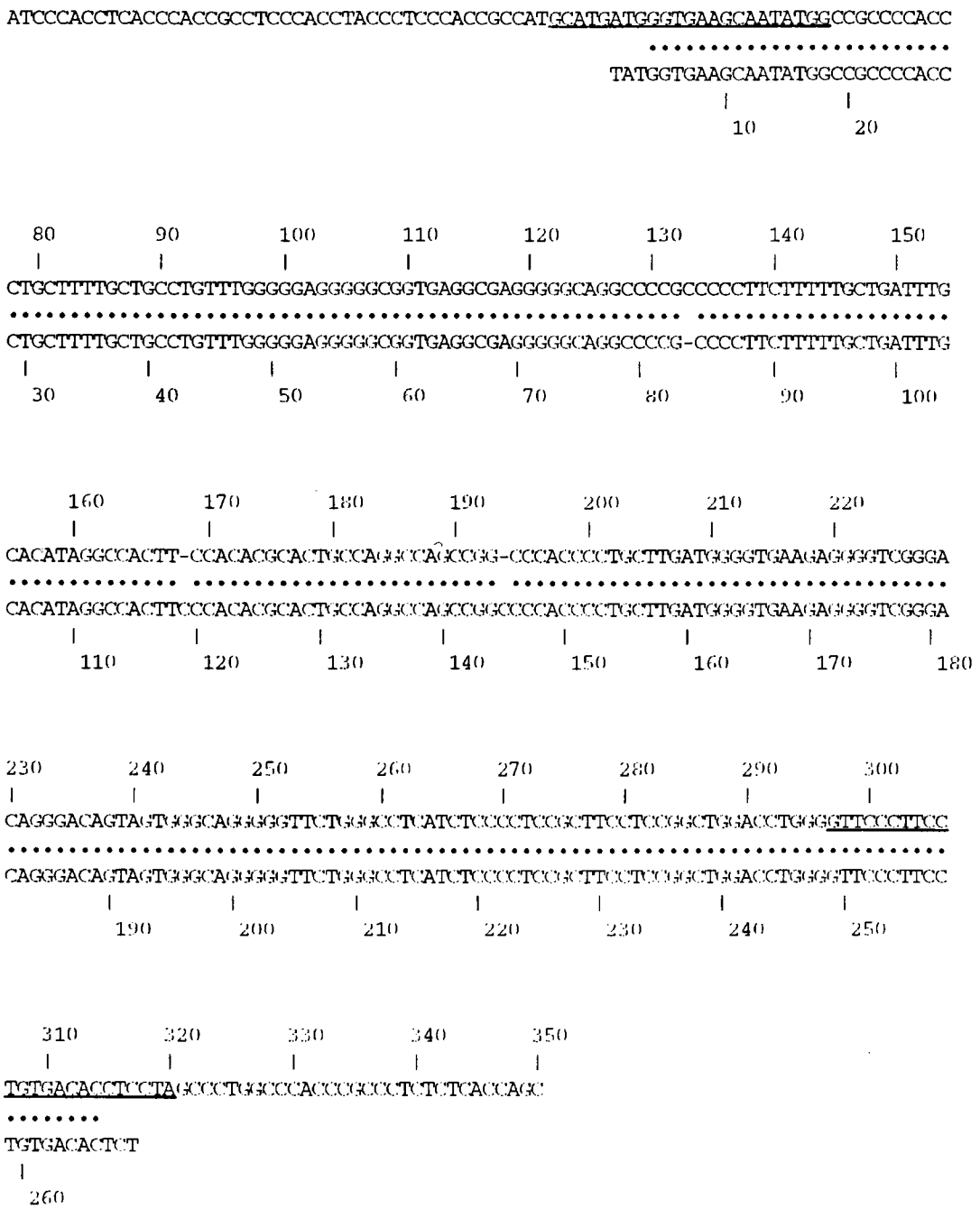


FIGURE 5-3 (B). Sequence alignment of the *D1Gu102*-PCR product obtained from the P1(*Scnn1b*) clone and the rat *Scnn1b* gene.

Comparison of a fragment of the rat *Scnn1b* gene (top; Isom *et al.*, 1992; Accession M91808) with the sequence obtained from amplifying DNA from a P1(*Scnn1b*) culture with the primer set D1Gu102 (bottom). Both the sequence of the forward primer and the sequence complementary to the reverse primer are underlined. Amplification was carried out at 55°C for 30 cycles, in the presence of 1 mM $MgCl_2$.

vector used in the construction of the library was p*Ad*10*sac*BII. IPTG was added to the cultures when the cells were in early log phase, resulting in the inactivation of the $lacI^q$ repressor and the concomitant increase in the copy number of the plasmid (section *2.3.5.1*).

Prior to isolation of the P1 DNA, PCR reactions were carried out to confirm the presence of the *Scnn1b* and the *D1Mco1* sequences in their respective cultures. Presence of the *Scnn1b* sequence was always confirmed; however, none of the corresponding cultures showed amplification of the *D1Mco1* locus. Confirmation of the presence of this sequence was only achieved when the cells were grown on solid media (section *5.2.5.1*).

Various methods, described in section *2.3.2.3*, were employed to isolate P1 DNA. The P1 clone carrying the *Scnn1b* fragment was isolated following the Birnboim and Doly procedure, and also the Qiagen 'Midi' method and the modified Qiagen 'Midi' protocol I [Figure 5-4]. Isolation of P1 DNA containing the *D1Mco1* sequence was done by applying either of the two modified Qiagen 'Midi' protocols (I and II).

The Birnboim and Doly procedure was the most rapid method of isolation, the P1 DNAs being in general very clean; however, this method yielded the lowest amounts of DNA (ca. 2-5 μg), since it employed small volumes of bacterial culture. Repeated isolations were therefore required to ensure continued availability of the particular DNA.

Isolation of P1 using Qiagen-tip 100 was unsatisfactory. The considerable yield of P1 DNA (ca.10-50 μg) was contaminated with a considerable amount of bacterial chromosomal DNA. The modified Qiagen 'Midi' protocol I was adopted as a compromise between these two methods. High yields of considerably clean P1 DNA (ca. 20-50 μg) were isolated.

In a few instances, the single phenol/chloroform extraction step was not efficient enough to purify the P1 DNA. This problem was more commonly encountered when 500 ml cultures were being used, and resulted in either complete absence of digestion or incomplete digestion when the DNAs were treated with restriction endonucleases; different enzymes exhibited different activity efficiencies in each case [Figure 5-5]. This problem was overcome by repeating the phenol/chloroform extraction 1 or 2 more times or, preferably, by introducing a proteinase K treatment step followed by a second phenol/chloroform extraction (section *2.3.2.3*).

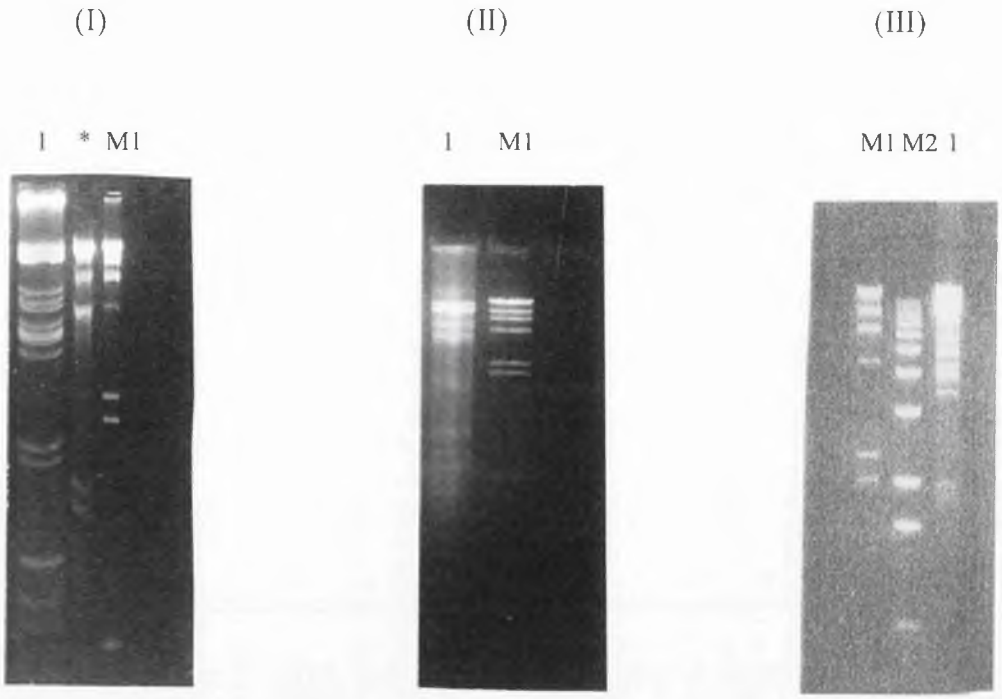


FIGURE 5-4. Comparison of the P1 DNA obtained following various methods of isolation.

The three gels show P1(*Scnn1b*) DNA following digestion with *Bgl*II.

(I) Lane 1: 24 µl of a total of 28 µl of DNA isolated from a 25 ml culture with the Birnboim and Doly method.

(II) Lane 1: 50 µl of a total of 500 µl of DNA isolated from a 500 ml culture, following the Qiagen 'Midi' procedure.

(III) Lane 1: 8 µl out of 400 µl of DNA isolated from a 500 ml culture following the modified Qiagen 'Midi' protocol I.

The size markers are λ DNA digested with *Hin*dIII (M1) and a 1 kb DNA ladder (M2). The asterisk (*) indicates a very old sample of a λ / *Hin*dIII marker.

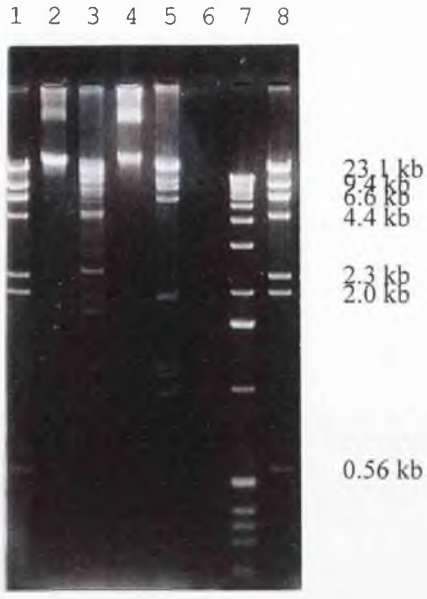


FIGURE 5-5. P1(*D1Mco1*) DNA that has lost the *D1Mco1* marker sequence.

Resolution on a 1% agarose gel of the restriction fragments obtained following digestion of the DNA extracted from a P1(*D1Mco1*) culture that has lost the reference *D1Mco1* sequence, with *Bgl*II (lane 2), *Eco*RI/*Hin*dIII (lane 3), *Kpn*I/*Hin*cII (lane 4), *Bam*HI (lane 5). Restriction with the enzymes *Bgl*II, *Kpn*I and *Hin*cII failed. Successful digestion with *Bgl*II was observed after the DNA was treated with proteinase K and phenol/chloroform-extracted.

The size standards are a λ/*Hin*dIII marker (lanes 1 and 8) and a 1 kb DNA ladder (lane 7).

The modified Qiagen 'Midi' protocol II, although the most cumbersome and time-demanding procedure, proved to be the most efficient method for P1 DNA isolation, in terms of both quality and quantity (ca. 50-150 μg). It therefore became the method of choice when large amounts of clean P1 DNA were to be prepared.

5.2.4 Transfer of P1 clones to DH10B.

The quantity and the quality of P1 DNA isolated from single clones was of central importance for the subcloning of SSLP-containing DNA fragments as well as for the direct sequencing of the ends of the P1 insert. Previous communications had reported the isolation of greater amounts of P1 DNA from *cre*$^-$ host strains than from *cre*$^+$ bacteria such as NS3529; this DNA also appeared to be of better quality (Kimmerly *et al.*, 1994). All the isolated P1 clones were therefore transferred to DH10B cells, which did not express the Cre recombinase enzyme. The transfer into DH10B cells was carried out via γδ-transposition. The steps followed are described in Chapter 2 (section 2.12).

Restriction digests of P1(*Scnn1b*) DNA isolated from NS3529 and from DH10B demonstrated an at least 5- to 10-fold increase in the amounts of DNA obtained from the latter host strain. This was also the case when working with the P1(*D1Mco1*) clone. The difference was accentuated when IPTG was added to the DH10B cultures. [Figure 5-6].

5.2.5 Observed instability of the P1(D1Mco1) clone.

5.2.5.1 Loss of the D1Mco1 sequence. Growth on solid media.

As mentioned earlier, problems arose when trying to grow the P1(*D1Mco1*) clone in liquid media. PCR reactions set up on different cultures demonstrated that the *D1Mco1* sequence had been lost in all of them. This result was consistently encountered with cultures of various volumes which had been grown at either 37°C or 30°C and with or without IPTG; only a few 3 ml cultures which had been grown at 30°C and which were not yet dense showed presence of the marker sequence, and further growth of these cultures resulted once again in loss of the PCR signal. To confirm that the absence of PCR product was due to the loss of the particular sequence in the P1 DNA and not to inhibition of the PCR reaction by contaminants such as excess cell debris, the DNA was

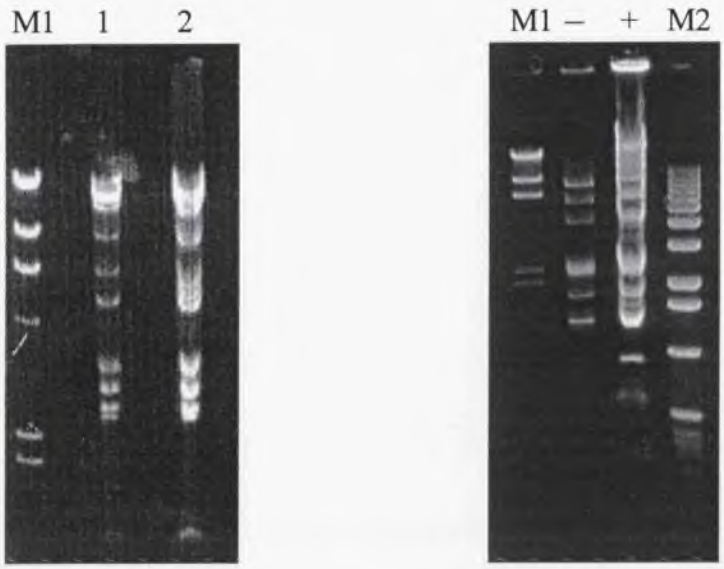


FIGURE 5-6. Effects of IPTG and the host strain on P1 DNA yields.

(I) *Not*I/*Bam*HI digests of P1(*Scnn1b*) DNA extracted from 500 ml NS3529 (lane 1) and DH10B (lane 2) cultures. In each case, 7µl of a total of 400µl of isolated DNA was digested.

(II) *Eco*RI digests of P1(*D1Mco1*) DNA isolated from 250 ml DH10B cultures which were (+) or were not (-) induced with IPTG prior to extraction of the DNA. Each DNA was resuspended in a final volume of 200 µl; digestion was carried out with 6 µl of DNA from the uninduced culture (-), and on 3 µl of DNA from the induced culture (+).

The size standards used were a λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2).

purified from a small volume of culture (30 µl) with QIAEX II-resin, and the PCR repeated; no product was obtained.

The P1 DNA was isolated and analysed in restriction reactions with *Eco*RI, *Bam*HI, *Bgl*II and *Kpn*I, showing an insert of at least 40-60 kb [Figure 5-5]. Thus, the insert had not been completely lost, although some rearrangement had occurred in the P1 DNA which must have been advantageous for the growth of this clone.

Maintenance of the *D1Mco1* sequence was attained by growing the clone on solid medium. Great part of the bacterial mass corresponding to the gridded P1(*D1Mco1*) clone was removed with a toothpick and resuspended in 180 µl of L-broth; 30 µl aliquots were plated on each of 6 LB plates containing 25 $\mu g.ml^{-1}$ kanamycin and 5% sucrose, and grown at 30°C for 22 hours, until a confluent lawn of cells was observed on each plate. The colonies were eluted with L-broth and pooled into a Nalgene tube, in an approximate final volume of 25 ml. An aliquot (1 µl) of this culture was used as template for amplification with the D1Mco1 primer set; this time, a strong signal was obtained. This DNA was isolated following the modified Qiagen 'Midi' protocol II.

When the restriction patterns of both the DNA which had lost the PCR sequence (P1 DNA(PCR^-)) and the newly isolated DNA (P1 DNA(PCR^+)) were compared following *Bam*HI and *Bgl*II digestions, no clear differences were observed. Therefore, the rearrangement(s) which had taken place during growth in liquid media did not seem to involve drastic alterations of the P1(*D1Mco1*) DNA.

5.2.5.2 Transfer of P1(D1Mco1) to DH10B cells.

This process involved several cycles of growth which had by now proved to be unfavourable for the maintenance of the *D1Mco1* sequence in the P1 DNA. This problem was conclusively demonstrated when transfer of the P1 clone to DH10B following the appropriate mating steps (section 2.12) resulted in DH10B colonies which had lost the *D1Mco1* fragment. On the other hand, it had been observed that small cultures maintained this sequence while still exhibiting a weak growth. Therefore, transfer to DH10B cells was again attempted, this time with the introduction of a few modifications to the original protocol, these concerning growth times and culture volumes. The procedure is described in Chapter 2 (section 2.12).

Several DH10B colonies were picked following transfer of the P1 molecule, and PCR reactions were carried out on the corresponding lysates. All the clones contained the *D1Mco1* sequence. Small (3 ml, 50 ml) and large (500 ml) cultures were prepared and grown at 37°C; this time, no loss of the PCR signal was observed.

5.2.5.3 Stability of the P1(D1Mco1) clone in the DH10B host strain.

Absence of the *D1Mco1* PCR product could be the result of either deletions involving one or both D1Mco1 primers, or of some rearrangement where both oligonucleotide sequences were maintained but at a different position than in the original fragment, in such a manner that amplification of the intervening sequence was not feasible (since no PCR product was obtained at all). Both the P1 DNA which had failed to give the *D1Mco1* PCR product and the DNA which showed amplification of this fragment were isolated from 50 ml DH10B cultures following the Birnboim and Doly procedure. Two sets of cultures were prepared in each case. One (culture (a)) was directly seeded with a single colony from the M2 plate, and the other (culture (b)) inoculated with 0.5 ml from a 3 ml culture which had been seeded with another M2 single colony and grown at 37°C overnight. Both 50 ml cultures were grown at 37°C for approximately 14 hours. Digestion of the isolated DNA with *Kpn*I followed and the fragments were run on 1% agarose gels and blotted onto nylon membranes.

Both the forward and the reverse D1Mco1 primers were radioactively labelled and used in separate experiments to probe the filters prepared. No signal was observed when probing P1 DNA(PCR^-) with either primer. This indicated deletion of both sequences from the original P1 DNA during growth. Probing of P1 DNA(PCR^+) prepared from each of the two 50 ml cultures ((a) and (b)) resulted in hybridisation signals with both labelled primers; both the forward and the reverse probe lit up the same *Kpn*I fragment. However, this hybridising fragment was shown to be different in each digest, corresponding to a ca. 7 kb band in DNA (a) and to a >20kb band in DNA (b); this was associated to a difference in the respective digestion patterns [Figure 5-7].

To test the possibility that the P1 DNA(PCR^+) was not stable and may be still prone to change into the P1 DNA(PCR^-) form as the number of growth cycles increased, a set of four 50 ml cultures were prepared where the second culture was inoculated with 0.5 ml from the first culture (this having been seeded with a single

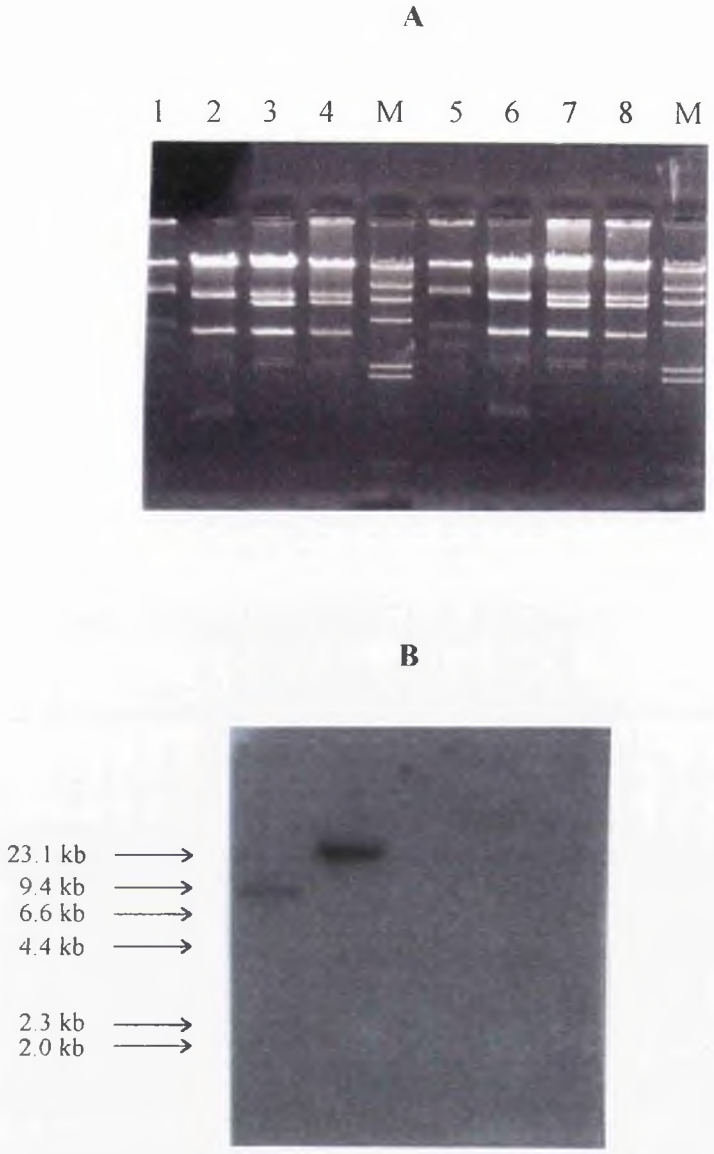


FIGURE 5-7. Instability of the P1(*D1Mco1*) clone. Loss of the *D1Mco1* sequence.

(A) *Kpn*I digests of DNA extracted from 50 ml P1(*D1Mco1*) cultures.
Lanes 1 and 5: P1 DNA (PCR^+) from culture (a) [P1 DNA (PCR^+)-(a)].
Lanes 2 and 6: P1 DNA (PCR^+) from culture (b) [P1 DNA (PCR^+)-(b)].
Lanes 3 and 7: P1 DNA (PCR^-) from culture (a) [P1 DNA (PCR^-)-(a)].
Lanes 4 and 8: P1 DNA(PCR^-) from culture (b) [P1 DNA (PCR^-)-(b)].
Details are given in the text. The size standard is a λ/*Hin*dIII marker (M).

(B) The two identical halves of the gel were blotted and probed separately with the forward and the reverse D1Mco1 primers, respectively. Identical results were obtained with both probes.

colony from the M2 plate) after an overnight growth at 37^0C, and so forth. PCRs on these cultures showed that all of them retained the *D1Mco1* sequence. Surprisingly, when *Kpn*I digests of these DNAs were prepared, a fourth digestion pattern (identical for all of them) was observed which was different from the two DNA(PCR$^+$) digests and from the unique pattern obtained with DNA(PCR$^-$) [Figure 5-8]. The D1Mco1 primers hybridised to the larger >20 kb fragment.

To ascertain that the differences in the digestion and hybridisation patterns were not due to contaminants in the DNA, impeding complete digestion with the restriction enzyme, two controls were added: a) some of the samples underwent treatment with proteinase K, followed by a phenol/chloroform extraction, and were subsequently digested and compared to the original patterns, and b) *Kpn*I digests of P1 DNA(PCR$^+$) which had been purified through a CsCl gradient were compared to the previous results. The DNAs which had undergone proteinase K treatment exhibited the same digestion and hybridisation patterns as the originals. DNA purified through the CsCl gradient gave the same results as those obtained with DNA(PCR$^+$)-(a) [Figure 5-9]. Growth in the absence or presence of IPTG did not have any effect on either the digestion pattern or the size of the hybridising fragment.

5.2.6 Identification of repeat sequences in P1 inserts.

The isolated DNAs were digested with various restriction endonucleases and the restriction patterns analysed on 0.8%-1% agarose gels. Both P1(*Scnn1b*) and P1(*D1Mco1*) were found to carry inserts of considerable size (at least 50-60 kb).

Southern blots of the digests were prepared and probed with radioactively labelled oligonucleotide repeats [$(AC)_{10}$, $(CT)_{10}$ and $(GC)_{10}$]. Analysis of both P1 clones demonstrated that probing with $(AC)_{10}$ resulted in several hybridisation signals of considerable strength [Figures 5-10 (A) and 5-10 (B)]. No clear signals were obtained with $(CT)_{10}$ or $(GC)_{10}$ probes. Thus (AC)-containing microsatellites were initially sought. Two approaches were used to subclone (AC)-containing fragments from the P1 insert: I) subcloning of specific gel-purified fragments; II) shotgun cloning of whole P1 digests.

The first approach was only followed with the P1(*Scnn1b*) clone, since several of the hybridisation signals corresponded to restriction fragments in the approximate

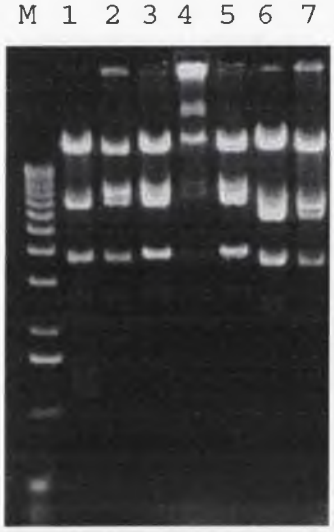
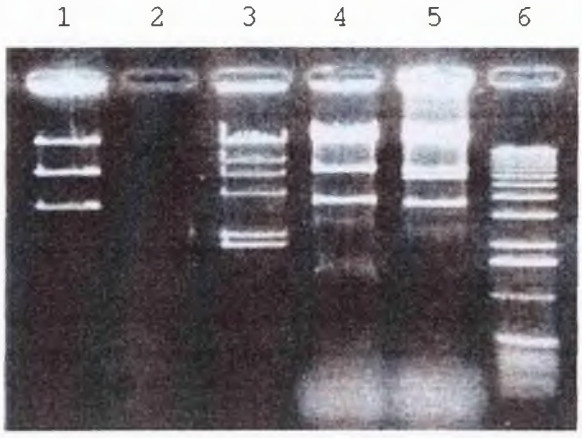


FIGURE 5-8. Apparent stability of the P1(*D1Mco1*) clone in the DH10B host through consecutive subcultures.

*Kpn*I digests of DNA extracted from P1(*D1Mco1*) cultures. Details are given in the text.
Lane 1: P1 DNA (PCR^+)-(b).
Lanes 2 to 5: set of four successive subcultures from an M2 (*D1Mco1*$^+$) colony.
Lane 6: P1 DNA (PCR^-)-(a).
Lane 7: P1 DNA (PCR^-)-(b).
A 1 kb DNA ladder (M) was used as the size standard.

(A)



(B)

FIGURE 5-9. Further confirmation of instability in the P1(*D1Mco1*) clone.

(A) *Kpn*I digests of DNA extracted from P1(*D1Mco1*) cultures. Lane 1: P1 DNA purified through a CsCl gradient. Lane 4: P1 DNA (PCR^+)-(b). Lane 5: P1 DNA (PCR^-)-(b). Details are given in the text. The size standards used are a λ/*Hin*dIII marker (lane 3) and a 1 kb DNA ladder (lane 6).

(B) The gel was blotted and probed with the reverse D1Mco1 primer.

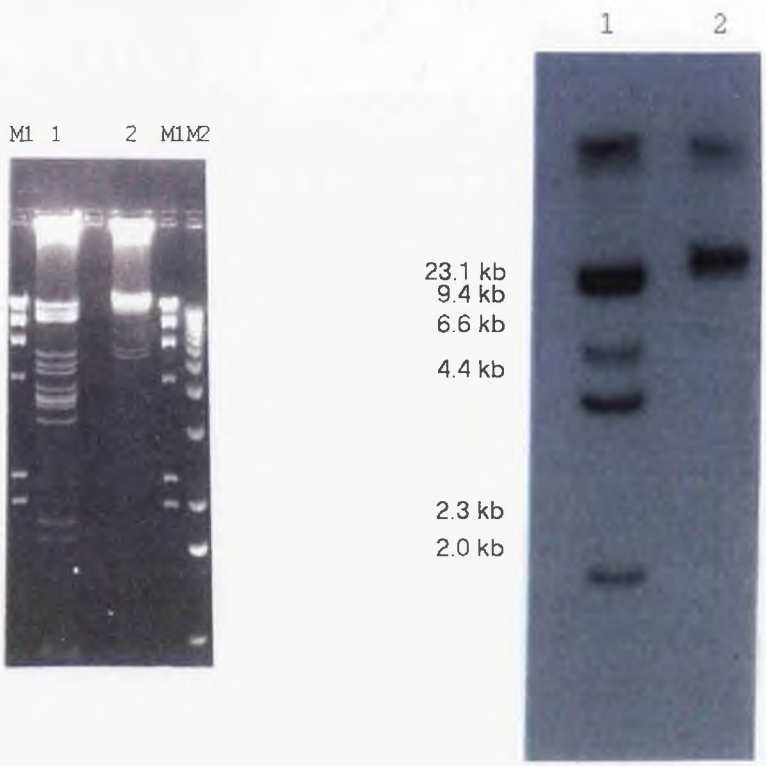


FIGURE 5-10 (A). P1(*Scnn1b*) DNA. Screen for repeat sequences.

(I) *Bgl*II (lane 1) and *Spe*I (lane 2) digests of P1(*Scnn1b*) DNA, run on a 1% agarose gel. The size standards are a λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2).

(II) The digested DNA was blotted onto a nylon membrane and probed with a radioactively labelled $(AC)_{10}$ oligonucleotide.

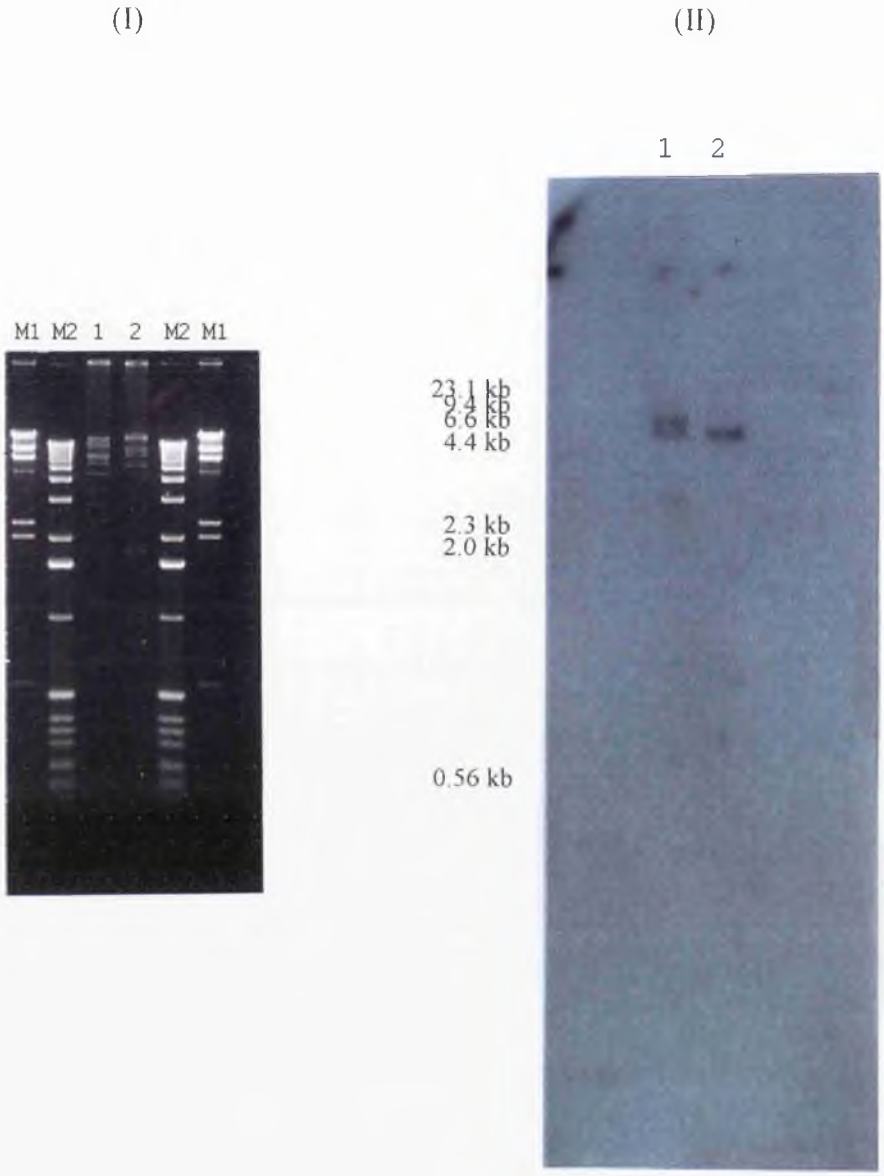


FIGURE 5-10 (B). P1(*D1Mco1*) DNA. Screen for repeat sequences.

(I) *Bgl*II (lane 1) and *Bam*HI (lane 2) digests of P1(*D1Mco1*) DNA, run on a 1% agarose gel. The size standards are a λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2).

(II) The digested DNA was blotted onto nylon membranes and probed with a radioactively labelled $(AC)_{10}$ oligonucleotide.

size range of 1.7-4.8 Kb which were clearly visible on the gel. This method was not applied to the P1(*D1Mco1*) clone as the fragments which hybridised to the $(AC)_{10}$-probe were either too faint on the gel and therefore the DNA prone to be lost during the purification process, or too large to ensure cloning. At the same time, shotgun cloning of the entire P1 digest represented a quicker approach where several microsatellite-containing clones could be identified in a single experiment. This second approach was applied to both P1 DNAs.

5.2.6.1 *Subcloning of gel-purified fragments.*

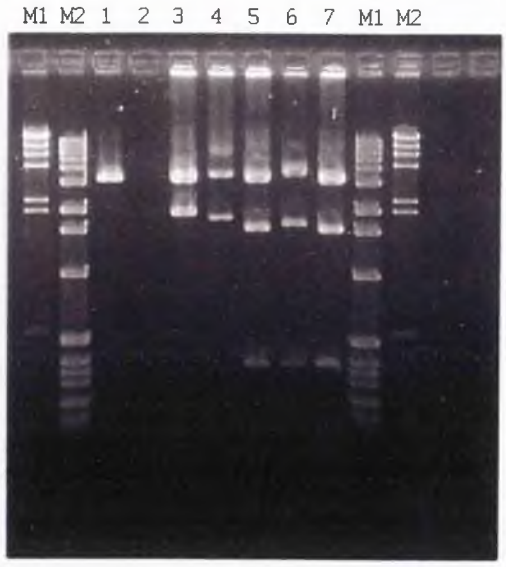
Four strong signals were observed when *Bgl*II digests of the P1(*Scnn1b*) DNA were probed with an $(AC)_{10}$ probe. Three of these signals corresponded to sets of bands in the approximate size ranges of 4.5-5 kb, 3.5-4 kb and 1.7-1.9 kb [Figure 5-10 (A)].

Each set of bands was extracted from the gel and purified with QIAEX II-resin (*2.3.13*). Ligation of each DNA fragment to a *Bam*HI-digested vector (either pGEM-3Zf(+) or pBluescript II KS+) was carried out, as described in Chapter 2. *Bam*HI was chosen as the site for cloning because no *Bgl*II sites were present in the multiple cloning site (MCS) of any of these vectors and both these enzymes generate ends which are compatible with each other. Cells were transformed and plated on medium containing ampicillin and the indicators X-Gal and IPTG. Both pGEM-3Zf(+) and pBluescript II KS+ carry a portion of the lacZ gene which provides α-complementation for blue/white selection of the recombinant clones; when grown in the presence of X-Gal and IPTG those recombinants carrying an insert are generally white whereas those lacking inserts are mostly blue.

White colonies resulting from each of the transformations were picked to inoculate 3 ml cultures which were grown at 37°C overnight. Plasmid DNAs were isolated from these cultures and analysed on 1% agarose gels. The presence of (AC)-repeats on those clones which carried an insert was confirmed by preparing Southern blots and probing them with the radioactively labelled $(AC)_{10}$ oligomer.

DNA extracted from pGEM-3Zf(+) clones which carried either of the two subcloned smaller P1 fragments resulted in the identification of a clone carrying an (AC)-containing 2 kb insert (clone 56A) [Figure 5-11]. The DNA from two of these colonies was prepared for sequencing.

(I)



(II)

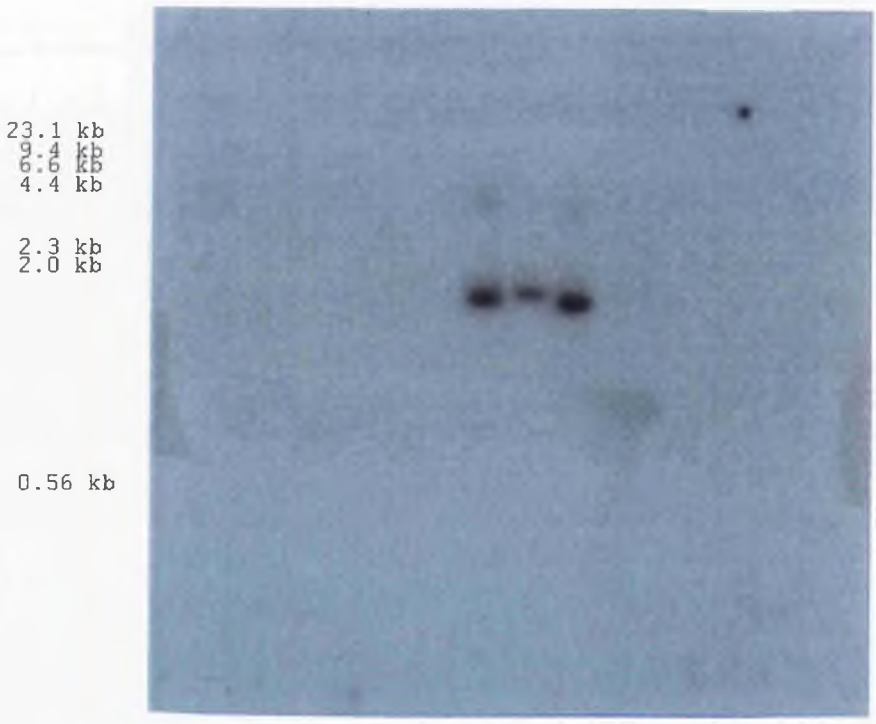


FIGURE 5-11. Subclones of the P1(*Scnn1b*) clone. Screen for repeat sequences.

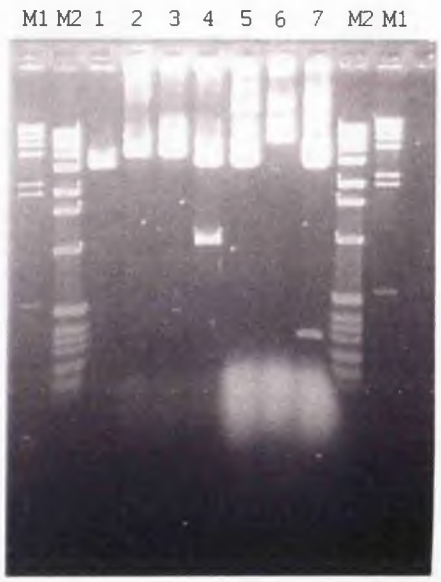
(I) Two different clones were obtained following subcloning of the P1 (*Scnn1b*) DNA *Bgl*II fragments in the size range of 1.7-2 kb. Lane 1: pGEM3Zf(+) plamid DNA; lanes 3 to 7: recombinant DNA obtained from five different white colonies. All the DNAs were digested with both *Eco*RI and *Hin*dIII. The size standards are a λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2).

(II) The digested DNA was blotted onto nylon membranes and hybridised to a radioactively labelled $(AC)_{10}$ oligonucleotide. A DNA fragment of approximately 1.6 kb lit up (clone 56A).

Three different insert sizes were observed in subclones of the set of P1 fragments in the 3.5-4 kb range. One of them, corresponding to an insert of approximately 3.6 kb in size was found to hybridise to the $(AC)_{10}$ probe [Figures 5-12 (I) and (II)]. Further subcloning of this DNA was carried out after a smaller *Kpn*I-*Hin*cII fragment was identified as containing the repeat [Figures 5-12 (III) and (IV)]. The corresponding band, of approximately 2.1 kb in size, was gel-purified and ligated to pGEM-3Zf(+), which had already undergone gel-purification following digestion with both enzymes *Kpn*I and *Hin*cII. Electrocompetent DH5α cells were transformed. A number of white colonies were picked from the plates, 3 ml cultures set up and grown, and the plasmid DNAs isolated. Several of these colonies were found to carry the *Kpn*I-*Hin*cII fragment (clone 56B); the presence of the (AC)-repeat was re-confirmed by probing a Southern blot of these DNAs. The DNAs from two of these positive colonies were prepared for sequencing.

Subcloning of the two DNA fragments in the 4.5-5 kb region gave rise to clones exhibiting two different digestion patterns and therefore carrying both fragments. The larger insert (ca. 4.6 kb) was found to carry the (AC)-repeat [Figure 5-13]. Digestion with different restriction endonucleases showed that this insert did not contain any *EcoRI*, *Not*I, *Kpn*I or *Hin*dIII sites, since double digestion of the plasmid DNA with the first two enzymes only resulted in the extraction of the insert from the pBluescript II KS+ vector and single digestions with *Kpn*I and *Hin*dIII led to linearization of the plasmid DNA. Digestion of this DNA with *Hin*cII gave rise to five fragments, including a band corresponding in size to the linearised vector on its own. Probing of a Southern blot resulted in a weak signal on a 2.5 kb fragment, which could indicate the presence of a very short (AC)-repeat, and two strong hybridisation signals, one corresponding to a 1.1 kb DNA band and the other to the vector fragment. There being only one *Hin*cII site in the pBluescript II KS+ vector, the larger hybridising signal must be the vector DNA together with a short stretch of insert DNA at its 3'-end, this fragment containing an (AC)-repeat. The proximity of the repeat to the 3'end of the vector DNA ensured its direct sequencing with the -21 M13 forward primer so that it was unnecessary to subclone this fragment. Therefore, this plasmid DNA was prepared for sequencing (clone 56C). Further subcloning of the other two possible hybridising fragments could have been carried out and at least another SSLP could have been found this way.

(I)



(II)

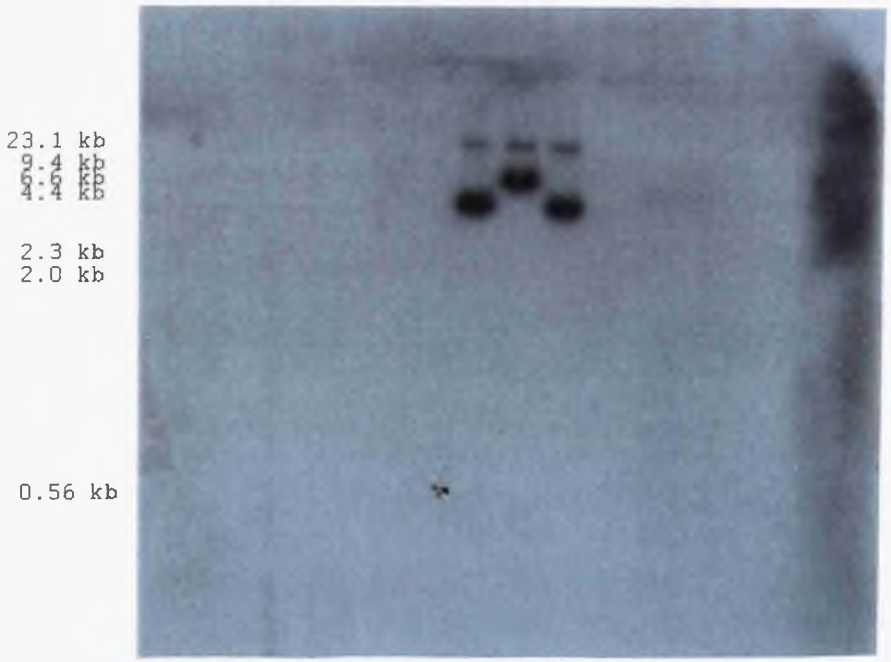
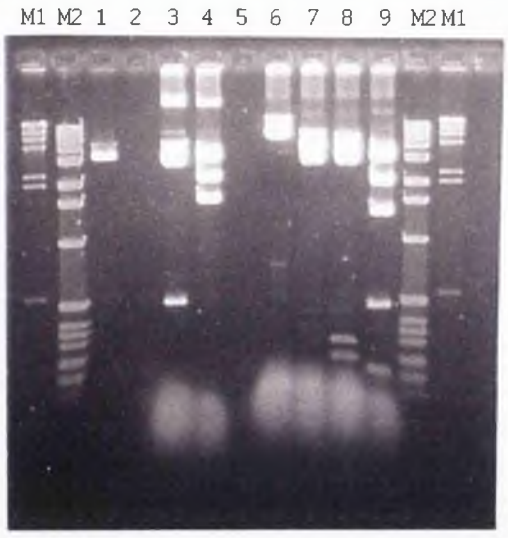


FIGURE 5.12. Subclones of the P1(*Scnn1b*) clone. Screen for repeat sequences.

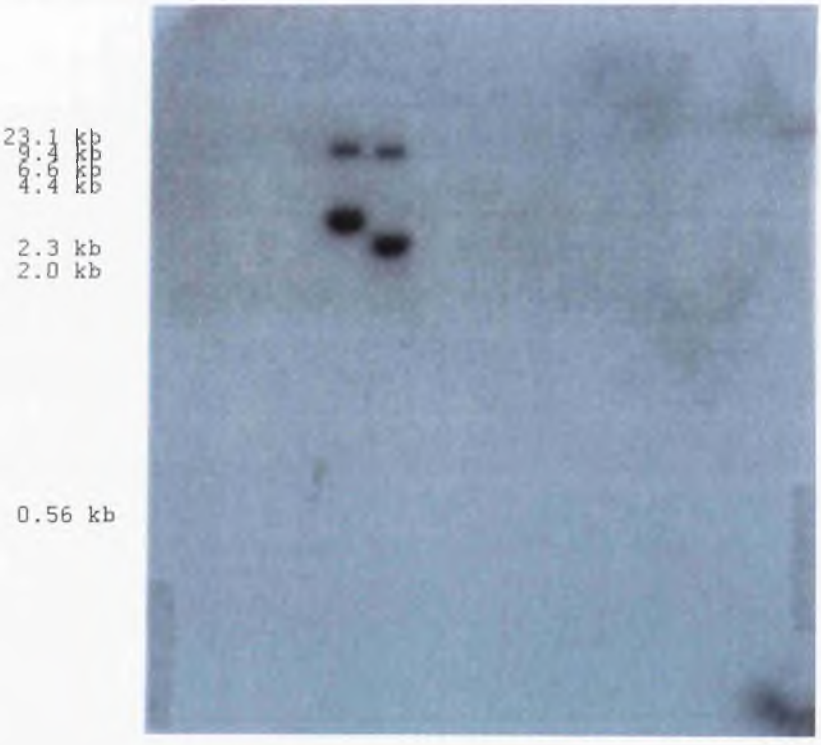
(I) Lane 1: Linearised pGEM-3Zf(+) plasmid. Lanes 2 to 4: *Eco*RI, *Hin*dIII, and *Eco*RI/*Hin*dIII digests of the DNA extracted from a single white colony (Q3-1) following transformation with a pGEM-3Zf(+) vector which carried either of the 3.5-4kb P1 (*Scnn1b*) DNA fragments. Lanes 5 to 7: restriction fragments of DNA extracted from another white colony (Q3-2); the same digestions as those above were carried out. The size standards are a λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2).

(II) The digested DNA was blotted onto nylon membranes and probed with a radioactively labelled $(AC)_{10}$ oligonucleotide. Clone Q3-2 was shown to carry an (AC) repeat.

(III)



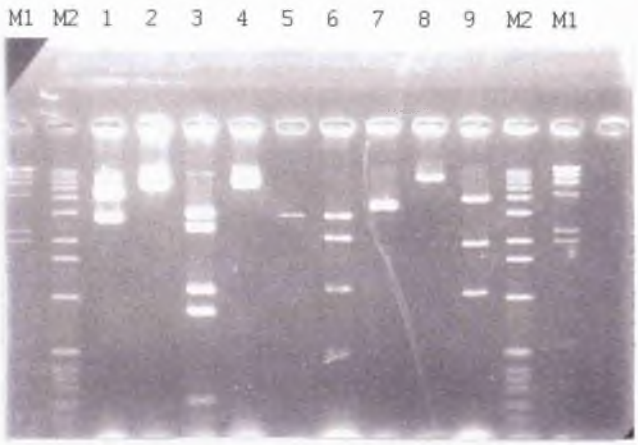
(IV)



(III) Lane 1: pGEM-3Zf(+) plasmid DNA digested with *Eco*RI. Lanes 3 and 4: *Ava*I/*Hin*dIII and *Kpn*I/*Hin*cII digests of clone Q3-2. Lanes 6 to 9: *Eco*RI, *Eco*RI/*Hin*dIII, *Ava*I/*Hin*dIII, and *Kpn*I/*Hin*cII digests of a third clone, Q3-3. The size standards are a λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2).

(IV) The digested DNA was blotted onto a nylon membrane and hybridised to a radioactively labelled $(AC)_{10}$ oligonucleotide. Clone Q3-2 was shown to carry an (AC) repeat in a ca. 2.1 kb *Kpn*I/*Hin*cII fragment. This fragment was subcloned into a pGEM3Zf(+) vector (clone 56B).

(I)



(II)

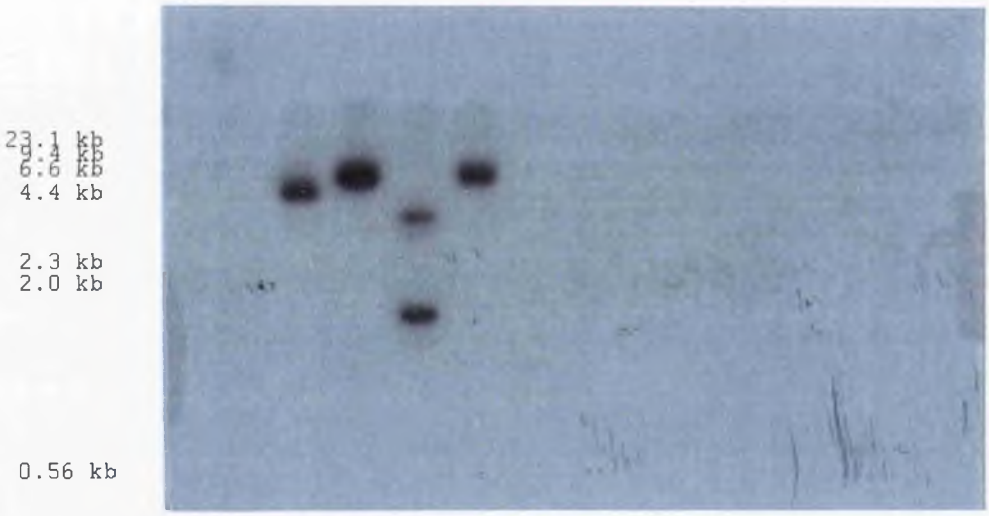


FIGURE 5.13. Subclones of the P1 (*Scnn1b*) clone. Screen for repeat sequences.

(I) Lanes 1 to 4 and lanes 6 to 9, respectively: *Eco*RI/*Not*I, *Kpn*I, *Hin*cII and *Hin*dIII fragments corresponding to two different pBlueScript subclones of the P1(*Scnn1b*) clone. Lane 5: pBlueScript II KS+ digested with *Eco*RI. The size standards are a λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2).

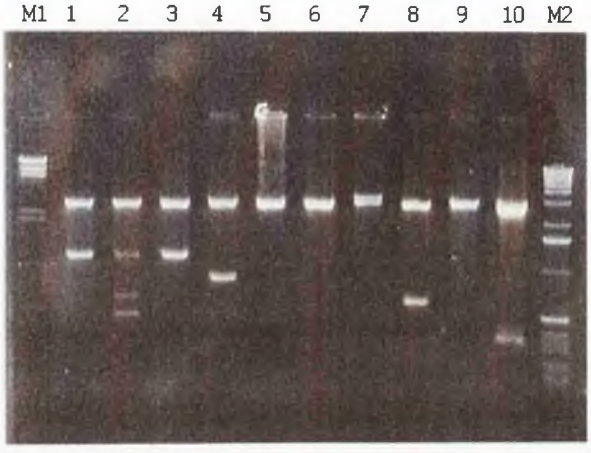
(B) The digested DNA was blotted onto a nylon menbrane and probed with a radioactively labelled $(AC)_{10}$ oligonucleotide. The results indicated that one of the clones carried at least two (AC) repeats (clone 56C).

However, the fact that three microsatellites had already been identified and shotgun cloning was going to be carried through led to the termination of any further work with these clones.

*5.2.6.2 Shotgun cloning of Sau*3AI-*P1 digests.*

Approximately 0.5-1 μg of both P1(*Scnn1b*) and P1(*D1Mco1*) DNAs, prepared by the modified Qiagen 'Midi' protocols I and II, respectively, was digested with *Sau*3AI. The DNA was then ligated to 50-60 ng of pBluescript II KS+ vector DNA which had been digested with *Bam*HI, SAP-treated and gel-purified with QIAEX II-resin, as previously described (Chapter 2). *Bam*HI was chosen as the cloning site in the vector since it cuts the vector DNA only once and it generates staggered ends which are compatible with those obtained from digestion with *Sau*3AI. TG1 cells made competent by the $CaCl_2$-method were transformed, plated, and grown at 37°C overnight. Two- to three-hundred white colonies were picked from each transformation, gridded on petri dishes, and grown at 37° for an additional 10 hours; duplicates of each plate were prepared. One set of duplicates were kept as master plates; the other set were used to prepare duplicate filters for colony hybridisation experiments, as explained in section *2.3.17*. The filters were probed with a radioactively labelled $(AC)_{10}$-oligomer. Four colonies from the P1(*D1Mco1*) clone and five from P1(*Scnn1b*) gave hybridisation signals in both duplicate filters; the signals observed on the set of filters firstly prepared were in most cases considerably stronger than those coming from the replicate filters prepared later. This was probably due to most of the bacterial mass having been taken up by the first set of filters. The signals on the filters were correlated to their respective colonies on the master plates and these were used to set up 3 ml LB cultures which were grown at 37°C overnight. Also, a white colony which did not give any hybridisation signal was randomly selected and a culture prepared. The plasmid DNAs were isolated from these 10 cultures and digested with both *Eco*RI and *Not*I, located at either side of the *Bam*HI site used for cloning. These inserts were shown to be approximately 1.3, 2.5, 1.3 and 0.8 kb in the four P1(*D1Mco1*)-subclones (37A, 37B, 37C and 37D), and 2.9, 3.1, 0.6, 2.9 and 0.35 kb in the P1(*Scnn1b*)-subclones (56D, 56E, 56F, 56G and 56H), respectively. The DNAs were run on a 1% agarose gel and blotted to be probed again with the (AC)-oligomer [Figure 5-14]. All the DNAs except the one prepared from the non-hybridising colony (thus being used as a negative

(I)



(II)

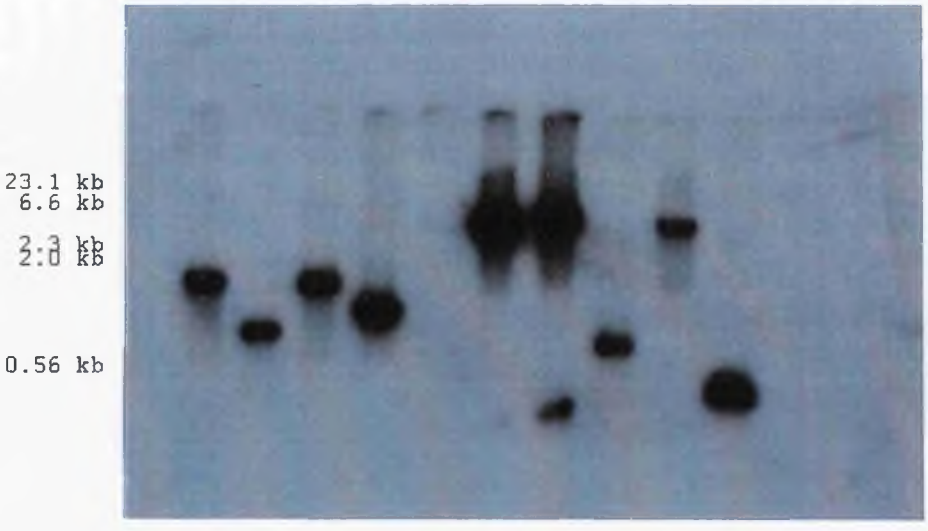


FIGURE 5.14. Subclones of *Sau3AI* DNA fragments from the P1(*Scnn1b*) and the P1(*D1Mco1*) clones.

(I) Lanes 1 to 4: P1 (*D1Mco1*)-subclones 37A, 37B, 37C, and 37D, respectively. Lane 5: pBlueScript II KS+ plasmid DNA digested with *Eco*RI. Lanes 6 to 10: P1 (*Scnn1b*)-subclones 56D, 56E, 56F, 56G, and 56H. The inserts were extracted from the pBlueScript vector DNA by digesting with *Eco*RI/*Hin*dIII. A λ/*Hin*dIII marker (M1) and a 1 kb DNA ladder (M2) were used as the size standards.

(II) The digested DNAs were blotted onto a nylon membrane and hybridised to a radioactively labelled $(AC)_{10}$ oligonucleotide. All the clones except that on lane 5 were shown to carry (AC) repeats.

control) showed hybridisation signals, confirming the results obtained with the colony hybridisation experiment. These DNAs were prepared for sequencing.

5.2.7 Establishment of new SSLPs.

Subcloning of P1(*Scnn1b*) and P1(*D1Mco1*) DNAs resulted in the identification of 8 and 4 (AC)-containing clones, respectively. A total of 14 DNAs were prepared (two of them, corresponding to clones 56A and 56B, were prepared in duplicate), and sequenced from each end with the -21 M13 Forward and Reverse primers.

Initially, (AC)-repeats were found within two of the three DNA inserts obtained from subcloning specific gel-purified fragments from P1(*Scnn1b*) (clones 56A and 56C) and also within the inserts present in the shotgun clones 56F and 56H. The sequences obtained from these latter two clones demonstrated to be the same; furthermore, this sequence was also found in the 56C clone. An (AC) repeat was identified within the ca. 0.8 kb insert carried by the P1(*D1Mco1*) subclone 37D. No repeats were established within any of the other DNAs (in the size range of 1.3 to 3.1 kb) following this first round of sequencing. The P1(*D1Mco1*) subclones 37A and 37C proved to be the same clone.

Internal primers were designed from the ends of the sequences obtained with both the forward and the reverse M13 primers to carry out further sequencing of the inserts. A (CA)-repeat was found in the 56B clone. No repeats were reached within any of the remaining P1(*Scnn1b*) shotgun subclones (2.9 and 3.1 kb in size). Since three microsatellites had already been identified within this P1 clone, no further work was carried out with these DNAs. An $(AC)_{25}$ repeat was identified in clone 37A. Certain anomalies were observed when studying the DNA from clone 37B; this case is discussed below (section *5.2.7.1*).

PCR primers were designed from the regions flanking each repeat. Two of the repeats were found near the 3' end of the sequences obtained; in these cases, primers could not be designed from the downstream flanking regions since multiple misreadings were already observed in these sequences. A primer was designed upstream of the repeat to be used for further sequencing of the insert. It corresponded to a sequence close to the repeat so that, once a second primer was designed from the other side of the (AC) region, both could be employed to amplify 150-300 bp fragments in PCR reactions. This was the case with clones 56B and 37D. Primers D1Gu15-F and D1Gu18-R were designed, respectively, from the sequences preceding each repeat and

used in a new set of sequencing reactions. D1Gu15-R and D1Gu18-F were designed from each of the new sequences, repectively. As well as yielding good sequences of the regions downstream the repeats, a $(TACA)_9$ motif was found towards the end of the new sequence obtained with the D1Gu15-F primer. The primer D1Gu19-F was designed from the upstream region but, again, the sequence downstream this new repeat was not good enough to design a reverse PCR primer. However, the 2.1 kb insert had also been sequenced from the other end with an internal primer, providing a sequence which overlapped to a certain degree with that containing the (TACA)-motif. This sequence was used to design the D1Gu19-R primer.

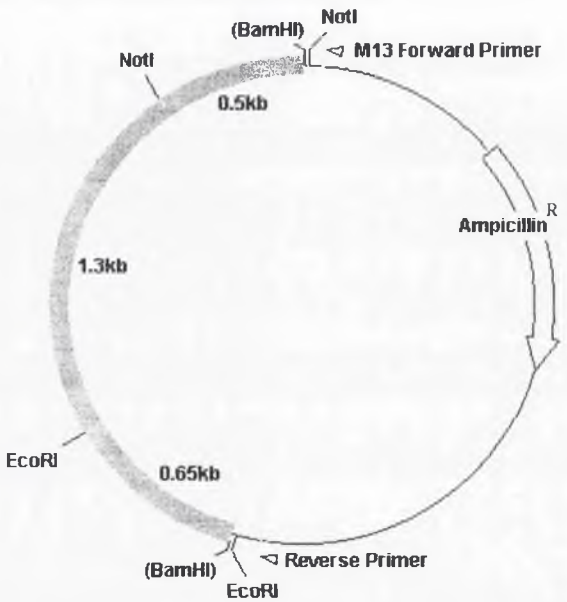
5.2.7.1 Observed instability in one of the P1(D1Mco1) subclones.

Digestion of the DNA isolated from the P1(*D1Mco1*)-subclone 37B with the enzymes *Eco*RI and *Not*I yielded 3 insert fragments, of approximately 0.5, 0.65, and 1.3 kb, respectively. However, the intensity of these bands was of a much lesser intensity than that of the vector DNA. Analysis of undigested plasmid DNA demonstrated the presence of at least two different plasmids in the sample. To test the possibility that these plasmids had been introduced in the same bacterial cell during transformation, an aliquot of this DNA was used to transform competent cells. A set of 22 white colonies were picked into and equal number of 3 ml L-broth cultures and these grown at 37°C overnight. Plasmid DNA was subsequently isolated from each culture and analysed on a low concentration agarose gel. Three different patterns were obtained this time, following digestion with both *Eco*RI and *Not*I: a) 3 fragments as before; b) 2 fragments, of 0.5 and 1.3 kb; c) a single 0.5 kb insert. The 0.5 kb fragment appeared as a strong band in all the digests, in contrast to the 1.3 kb and the 0.65 kb bands which were, when present, much weaker. Analysis of the undigested samples showed that only in the latter case (c) was a single plasmid present. At least two plasmids were present in the other cases. These observations indicated that the original plasmid DNA carried an insert of approximately 2.4 kb and this was undergoing deletions during growth, consequently leading to the observed mixture of different plasmid molecules in single bacterial cultures.

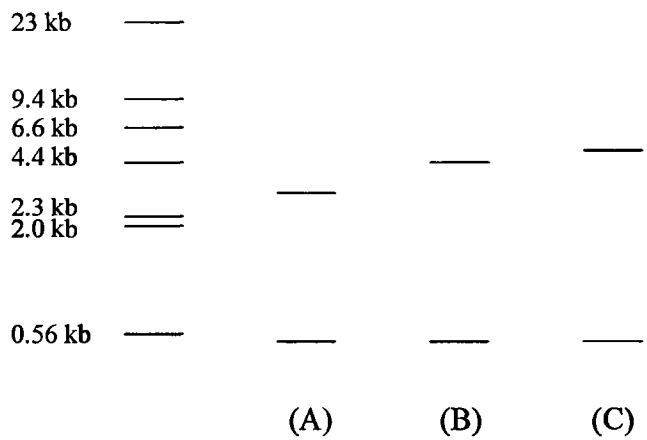
Sequencing of any of these DNA samples was not appropriate since it was the 0.65 kb fragment which had been shown to hybridise to the $(AC)_{10}$ probe. Nevertheless, sequencing of one of the samples where this fragment was present was attempted. As

expected, the low amounts of intact plasmid, together with the abundance of plasmid molecules which only carried a 0.5 kb insert, resulted in sequence only being obtained from the 0.5 kb fragment.

An aliquot of the original DNA was again used to transform electrocompetent TG1 cells. This time, the plates were incubated at 30°C. Cultures were seeded with each of 12 white colonies and the plasmid DNA isolated. The 3 patterns previously observed were obtained again. One of the samples was selected which still carried some intact plasmid; the DNA was digested with: i) *Eco*RI, ii) *Not*I, and iii) *Eco*RI/*Not*I. Analysis of the products obtained from each single digestion indicated that the 1.3, the 0.65 and the 0.5 kb bands corresponded to *Not*I/*Eco*RI, *Eco*RI/*Eco*RI, and *Not*I/*Not*I fragments, respectively. Therefore, the *Eco*RI and the *Not*I restriction sites in the insert must be distributed as follows:



The whole of the remaining DNA was then digested with *Not*I, in an attempt to separate plasmid DNA which carried the (AC)-containing fragment from deleted forms. The rationale behind this step was: a) The 0.65 kb *Eco*RI-*Eco*RI fragment (0.65 *E-E*) would remain attached to the vector DNA following digestion with *Not*I; and b) the 3 possible plasmid DNAs would give rise to different restriction patterns ((A), (B) and (C)). These are shown below, next to a λ/*Hin*dIII marker.



Pattern (A) results from the digestion of plasmid DNA carrying only a 0.5 kb insert; pattern (B) is obtained from those molecules carrying both the 1.3 and the 0.5 kb fragments; pattern (C) is only obtained following digestion of the intact recombinant molecule.

It was possible therefore to recognise the *Not*I restriction fragment which contained the 0.65 *E-E* insert sequence as a band of approximately 4.9 kb. This band was subsequently extracted from the gel and purified with QIAEX II-resin. An aliquot (2μl) of this DNA was digested with *Eco*RI. The presence of the 0.65 *E-E* fragment was thus confirmed. However, the yield of DNA was not high enough to ensure availability of the DNA were more than one sequencing run required to characterise the repeat. Consequently, the DNA was self-ligated and an aliquot of this DNA used to transform electrocompetent TG1 cells. A set of ten 3 ml L-broth cultures were prepared and the plasmid DNA isolated. The same single plasmid molecule was obtained from all of them; digestion of three of these DNAs with *Eco*RI/*Not*I yielded the expected 1.3 and 0.65 kb bands. This plasmid was thus stable. Two of the DNAs were prepared for sequencing.

Given the position of the cloned 0.65 *E-E* insert fragment (shown above), sequencing was carried out with the M13 Reverse primer. Two imperfect repeats were obtained. A PCR primer set, D1Gu13, was designed and the amplification conditions optimised.

5.2.8 Evidence of chimaerism in the P1(D1Mco1) clone.

Mapping studies with the newly developed SSLPs revealed that the microsatellite markers *D1Gu13* and *D1Gu18*, established within the P1(*D1Mco1*) clone, mapped to rat chromosome 8 and not to chromosome 1. The location of *D1Gu14* could not be confirmed in any of two mouse x rat somatic cell hybrid panels. However, *D1Mco1* was shown to map to chromosome 1, as reported by Gu *et al.* (1996). This work is described in Chapter 6.

Two possible explanations could account for these unexpected results:

a) The P1 clone originally gridded and identified as carrying the *D1Mco1* fragment was in fact a combination of two P1 clones, carrying inserts derived from chromosomes 1 and 8, respectively. These two DNAs had thus been simultaneously isolated (following growth on solid medium) and used in the shotgun cloning experiment.

b) The P1 DNA isolated was indeed extracted from a single clone and this clone was chimaeric, carrying a fragment of chromosome 1 and a fragment of chromosome 8.

Strong evidence in favour of this second possibility was obtained following amplification of P1 DNA with the four sets of primers (D1Mco1, {D1}Gu13, {D1}Gu14 and {D1}Gu18), following transfer of the P1 clone(s) to DH10B cells. Four different digestion patterns had been observed following *Kpn*I-digestion of P1 DNAs extracted from DH10B cells (section *5.2.5.3*). Three of these corresponded to clones which contained the *D1Mco1* fragment, in contrast to a fourth clone, which did not carry this fragment. PCR on each of these DNAs demonstrated that one of the clones carried all the four microsatellite loci; a second clone carried only the *D1Mco1* sequence; amplification of a third DNA template yielded a strong signal corresponding to *D1Mco1* and a very faint band from the *(D1)Gu14* locus. The DNA which did not carry the *D1Mco1* sequence carried both the *(D1)Gu14* and the *(D1)Gu18* fragments, but not the *(D1)Gu13* microsatellite [Figure 5-15].

5.2.9 <u>*Sequencing of the P1 insert ends.*</u>

Sequencing of the P1(*Scnn1b*) DNA insert was not carried out since typing with the D1Gu16 primer pair had shown this region of the chromosome to be quite distant from the *agu* locus. Prior to the characterisation of P1(*D1Mco1*) as a highly probable

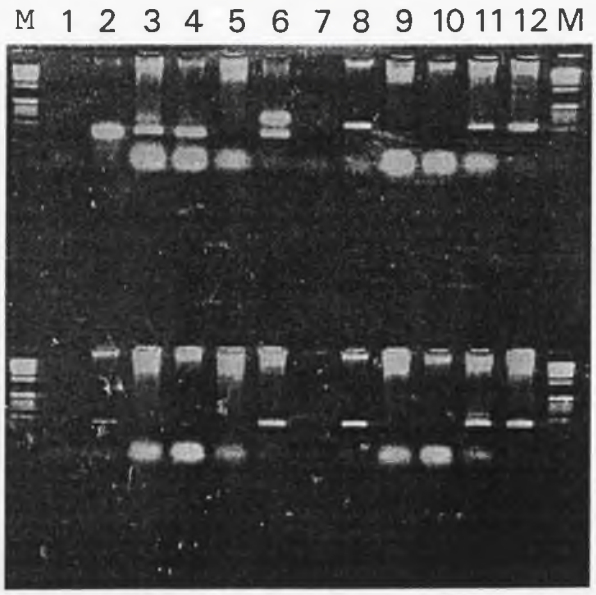


FIGURE 5-15. Evidence of chimaerism in the P1(*D1Mco1*) clone.

Screening of DNAs isolated from various DH10B cultures.

Top half of the gel.

Lanes 1 to 6: PCR products obtained with the D1Mco1 primer. Lane 1: no DNA control; lane 2: P1 DNA (PCR^+)-(a); lane 3: P1 DNA (PCR^+)-(b); lane 4: P1 DNA from the second subculture of an M2 ($D1Mco1^+$) colony; lane 5: P1 DNA (PCR^-)-(a); lane 6: rat genomic DNA.

Lanes 7 to 12: PCR products obtained with the (D1)Gu18 primer. The order of the DNA templates is as before.

Bottom half of the gel.

Amplification products obtained with the primer sets (D1)Gu13 (lanes 1 to 6) and (D1)Gu14 (lanes 7 to 12). The order of the templates is as before. Although not clear in the figure, a faint PCR signal was observed with (D1)Gu14 on lane 10 (*). Details are given in the text.

The marker used is a 1 kb DNA ladder (M).

chimaeric clone, both the P1(*D1Mco1*) DNA isolated from NS3529 cells and that obtained from the DH10B strain were used as templates in sequencing reactions. Two different methods were employed to prepare each of these DNAs:

a) The DNA was isolated from 250 ml cultures following the modified Qiagen 'Midi' protocol II and subsequently dialysed.

b) The P1 DNA was isolated and prepared following the protocol described in *2.3.18.1.*

Both P1(*D1Mco1*) DNA samples, each isolated from a different host (i.e. NS3529 and DH10B) were prepared following both methods above; each DNA was sequenced with the T7 and the SP6 primers.

Sequencing of the DNAs prepared according to the second protocol (b) failed. Sequencing of the P1(*D1Mco1*) DNA isolated from the NS3529 strain and purified through a CsCl gradient resulted in a very poor sequence when using the SP6 primer and failed when using the T7 primer. However, good sequences were obtained for each of the insert ends with the DNA isolated from the DH10B host which had been prepared according to the first protocol (a). It was also observed that the most successful sequencing of this P1 DNA was achieved when the sample had been freshly prepared; older samples failed to give sequences of comparable quality.

The sequences obtained for each end of the P1 insert were checked against the available database to be confirmed as unique sequences in the genome (i.e. non-repetitive) and, if applicable, correlate them to known sequences in rat and/or other organisms. The T7-end sequence was then shown to correspond to rat LINE sequences, which are highly repetitive. In contrast, the sequence obtained with the SP6 primer showed some homology to non-repetitive sequences in the database.

5.3 Discussion.

5.3.1 Comparative mapping.

Comparative mapping of the rat, the mouse, and the human genomes indicates that the rat chromosomal region of interest is syntenic to the proximal fragment of mouse chromosome 7 and to human chromosome 19q. Increasing numbers of mapped loci in each of the corresponding linkage maps (Gu *et al.*, 1996; Ding *et al.*, 1996; Brilliant *et al.*, 1996; Ashworth *et al.*, 1995) allow more detailed comparisons of these

regions (Stubbs *et al.*, 1996). Gene order and distances can then be more accurately correlated. Establishment of close homology at high resolution levels can play an important role in the identification of disease genes, as it permits the selection of 'candidate' sequences in the area (Eppig and Nadeau, 1995).

5.3.2 Amplification of rat genomic DNA with heterologous mouse primers.

Approximately 30% (7 out of 23) of the mouse PCR primers that were used to amplify rat genomic DNA in this project yielded single products. This is in contrast to the 12% success rate reported by Kondo *et al.* (1993). However, a very small number of mouse primer sets were used in this work, in comparison to the 166 primer pairs tested by Kondo and colleagues. It must also be kept in mind that, given that none of the amplified rat products showed differences between any of the parental strains in the backcrosses, these were not further analysed. Therefore, these fragments were not sequenced; it could be the case that some of these are not the expected rat homologues of the mouse SSLPs.

5.3.3 Rat bacteriophage P1 library.

The vector p*Ad*10*sac*BII was used to construct the library. This vector offers a positive selection system for clones carrying an insert. This feature represents a major improvement in comparison to its precursor vector, p*Ad*10 (Sternberg, 1990a; Sternberg, 1992). First, most of the clones which do not contain an insert are eliminated when grown in the presence of sucrose; thus, overgrowth of these clones in detriment of those which carry an insert does not normally occur. This permits pooling considerable numbers of individual clones when organising the library, without compromising the identification of single clones in the pool (Pierce *et al.*, 1992; Sternberg, 1992). Second, recombinant p*Ad*10*sac*BII molecules are directly selected following growth in the presence of kanamycin and sucrose. In contrast, screening for colonies which are both kanamycin-resistant and tetracycline-sensitive has to be carried out in order to identify p*Ad*10 recombinant molecules (as the cloning site in the P1 vectors maps within the tetracycline^R-gene).

5.3.3.1 Screening of the P1 library.

The library consists of 209 primary pools (550 clones/pool, on average). Most primer sets identified at least one positive pool following screening of 50-80 of these. The high rate of positive hits agrees with the claim that the library represents greater than 2.5 genome equivalents (Southard Smith and MacDonald, 1993).

During screening of these primary pools it was observed that some of the signals obtained with some primers were not reproducible. Two main factors were accountable for this: a) cell debris interfering with the PCR reaction, as demonstrated by the fact that purification of the DNA in the lysates with QIAEX II-resin resulted in consistent, clear signals; and b) residual nuclease activity in the sample. This latter was observed in a couple of instances when, correlating with a failure in the system to maintain the lysates frozen while stored, the particular samples did not give the expected PCR signals, even after purification of the DNA. When fresh lysates were prepared and immediately treated with QIAEX II, the expected PCR sequences were reproducibly obtained, confirming the original signals as true positives. When compared, the only noticeable difference between our protocol to prepare the lysates and that presented by Southard Smith and MacDonald (1993) was the 10 and 20 minutes employed, respectively, to boil the lysates. New lysates were prepared where minimal aliquots of each primary pool were removed and resuspended in 100 µl of dH_2O, and boiled for 20 minutes; thenceforth, clean, reproducible signals were obtained. Subsequent work (carried out by Dr. D. Donald and Ms. N. Craig) was based on the screening of large pools of clones picked following growth of each primary pool on plates, aiming to ensure the representation of each individual clone in the sample. Reproducible strong signals were obtained in these studies.

5.3.4 Comparison of P1 DNA extracted from two different host strains.

Progress in the manipulation of P1 clones was sometimes hindered, either because too small amounts of DNA had been isolated or because the quality of the sample was not as good as was desirable. A method had been described by Kimmerly and colleagues (1994) where P1 clones had been transferred to DH10B cells following a 2-step mating protocol which exploited the F factor transduction and γδ-transposition events. This group reported that the P1 DNA isolated from this new host was of a clearly higher quality. The same procedure was thus followed with our P1 clones, with

the same results. P1 DNA was isolated from both the NS3529 and the DH10B strains and used as template in sequencing reactions with the T7 and the SP6 primers; only the DNA extracted from the latter host yielded good quality sequences. Different methods were employed to prepare the DNA; although P1*(D1Mco1)* DNA only gave good sequences when purified through a CsCl gradient, other P1 DNAs have been successfully sequenced in the lab following a simple phenol/chloroform extraction or PEG_{8000} precipitation in the presence of $MgCl_2$. Also, a drastic decline in the success of the sequencing process was observed as the samples became older, with the freshly prepared DNAs always producing the best results. An increase in the amount of P1 DNA isolated from DH10B cells when compared to that obtained from the NS3529 strain was also clear, even in those instances where the NS3529 cultures had been induced and the DH10B cells had not.

Kimmerly and coworkers have proposed two arguments which could explain the superiority of the DNA extracted from DH10B: a) this strain lacks the Cre recombinase enzyme, impeding the formation of P1 multimers by lox-mediated recombination which might not be released as well as P1 monomers following cell lysis; and b) DH10B carries a mutation in the *endA* gene which, in its wild-type, encodes an endonuclease.

5.3.5 Identification of new rat SSLPs.

Radioactively labelled dinucleotide repeats were mainly used to screen for the presence of repetitive sequences in the inserts of the isolated P1 clones. This type of probe was preferentially used, since $(AC)_n$ and $(AG)_n$ are the most abundant repeats in the rat genome (Beckmann and Weber, 1992). Mononucleotide repeats were not sought since, contrary to the reported informativeness of these DNA stretches in mouse (Love *et al.*, 1990), analyses on the rat (Serikawa *et al.*, 1992) did not indicate that these motifs could be a rich source of informative markers in this species. Moreover, detection of strain differences at mononucleotide repeat loci may require resolution at the single base pair level. Tri- and tetranucleotide repeat probes were not widely used as these markers are less abundant than some dinucleotide repeats.

In agreement with numerous reports (Beckmann and Weber, 1992; Serikawa *et al.*, 1992), $(AC)_n$ was the most frequently identified motif. Screening of the P1 (*Scnn1b*) and the P1(*D1Mco1*) clones with an $(AC)_{10}$ probe resulted in the establishment of seven new SSLPs [Table 5-1]. The *Scnn1b* gene maps at approximately 8 cM from *Pkc* in the

P1 CLONE	SUBCLONE	SSLP LOCUS	PRIMER SEQUENCES(b)	REPEAT	LENGTH
P1 (*Scnn1b*)	56A	*D1Gu17*	AGCATGGGAATGGGAACCT AAGCTGACTTGGCATGTGC	$(AC)_{23}$	203 bp
	56B	*D1Gu15*	CTCCTGCTTCTCCTACC TTTGCATACATATGTAGATGC	$(AC)_{18}(AT)_6$	215 bp
		D1Gu19	CACAATACACATATAATATAC TTGGGCCACACATTCTC	$(TACA)_9$	239 bp
	56C 56F 56H	*D1Gu16*	TCGCATAAGCATGTGTGCAC ATACAACCGTGTGTTCCTGC	$(CA)_8TA(CA)_9(GA)_{24}$	208 bp
P1(*D1Mco1*)[a]	37A 37C	*D0Gu14* (*D1Gu14*)	TGTGGAAATAGAATCTCCCC TGAGCTCAGGGTGGACCT	$(AC)_{25}$	176 bp
	37B	*D8Gu1* (*D1Gu13*)	AACATCCAGTCTGTGACC TCAAAGAAAGAGAGGGGAA	$(CT)_7TT(CT)_{20}(CA)_{15}TA$ $(CA)_3(CT)_5$	171 bp
	37D	*D8Gu2* (*D1Gu18*)	CCAGTATGTGCAAGGTGG TTATGCTCAGCCACATGT	$(AC)_{16}$	230 bp

TABLE 5-1. Newly established rat SSLPs.

[a] The microsatellites *D1Gu13* and *D1Gu18*, identified in the P1 (*D1Mco1*) clone, were subsequently shown to map to rat chromosome 8 and were renamed *D8Gu1* and *D8Gu2*, respectively (Chapter 6). Chromosomal localisation of the marker *D1Gu14* could not be established and is currently called *D0Gu14*.

(b) Both the forward and the reverse primers are given in the 5' to 3' orientation.

mouse map. Selection of this and a few other genes (e.g. *Grik5*) as reference sequences for the isolation of P1 clones occurred at the initial stages of this part of the project, at a time in which the location of *Pkc* relative to the *agu* gene had not yet been established. Subsequent positioning of *Pkc* in our linkage map allowed a further refinement of the *agu*-containing region, and thus of the selection of "reference" loci to carry on the P1 library screening with. Therefore, a second set of marker *sequences (Pkc, Atpl, Ptgir, D1Mco1)* were used to screen the P1 library, in parallel to the on-going analyses of P1 clones carrying genes from the *CEA* family.

5.3.6 The P1(D1Mco1) clone.

5.3.6.1 Instability of the P1(D1Mco1) clone.

All the attempts to grow the P1(*D1Mco1*) clone in liquid media resulted in the loss of the *D1Mco1*-PCR product. This did not occur when the clone was grown in solid medium. Therefore, it seemed that P1(*D1Mco1*) was an unstable clone. Loss of some sequence(s) in the genomic insert must confer some advantage to the growth of the clone, with the new P1 variant overgrowing the original clone.

The subsequent observation that 2 of the SSLPs isolated from the P1(*D1Mco1*) clone mapped to rat chromosome 8 raised the possibility that the bacterial colony originally identified as containing this P1 consisted of 2 different clones, carrying chromosome 1 and chromosome 8 fragments, respectively. Were this the case, loss of the *D1Mco1* product in liquid cultures could result from the bacteria containing the P1 which carried a chromosome 8 fragment overgrowing the culture. Growth on solid medium would in turn lead to the simultaneous isolation of both P1 clones. However, this possibility was rejected, as probing of *Kpn*I-digests of P1 DNA isolated from various (*D1Mco1*$^+$)-DH10B cultures demonstrated differences in the restriction fragment which carried the SSLP under study. This could only result from instability in the P1(*D1Mco1*) clone.

No loss of the PCR product was observed when DH10B cells carrying the *D1Mco1* sequence were grown in liquid media. It thus seemed that the P1 was stable in this *cre*$^-$ strain. However, when DNA was isolated from several DH10B colonies and digested with *Kpn*I, three different patterns were obtained. On the other hand, consecutive culturing of one of the colonies did not result in any change in the digestion pattern. Therefore, it could be hypothesized that the P1(*D1Mco1*) clone is indeed stable

once inside DH10B hosts, and that the differences observed correspond to different mating events of individual DH10Bs to single NS3529 cells carrying each a P1 molecule which may or may have not undergone changes during the mating steps. The observation that P1 DNA which had lost the *D1Mco1* marker prior to mating yielded a unique digestion pattern (with each of *EcoRI* and *Kpn*I) when extracted from several DH10B colonies could indicate that complete stability of the P1 clone is achieved following the changes undergone by these particular P1 molecules, independently from the host strain.

5.3.6.2 Characterisation of P1(D1Mco1) as a chimaeric clone.

At least 2 of the 3 SSLPs isolated from P1 DNA which carried the *D1Mco1* sequence were shown to map to rat chromosome 8 ({D1}Gu13 and {D1}Gu18). These results could be explained by either the simultaneous isolation of DNA from 2 co-existing clones, or by the occurrence of chimaerism in the P1(*D1Mco1*) clone. Although the first possibility can explain the loss of the *D1Mco1*-PCR product in liquid cultures (5.3.6.1), several observations have been made which strongly argue against this hypothesis, in favour of the second explanation.

(a) A small number of restriction fragments were observed on low agarose gels following single *Kpn*I, *Bgl*II and *Bam*HI digestions of P1 DNA which carried the *D1Mco1* sequence. Moreover, summing of the estimated fragment lengths did not seem to indicate the presence of more than one P1 clone.

(b) Assuming that the (*D1Mco1*$^-$)-DNA correlates to the P1 carrying a fragment of chromosome 8, and that the (*D1Mco1*$^+$)-DNA was extracted from a mixed population of this latter and the real P1(*D1Mco1*) clone, presence of the P1 clone that carries a chromosome 8 fragment does not result in any clear difference between the *Bam*HI or the *Bgl*II digestion patterns of these two DNAs.

(c) *D1Mco1*-containing DNA was transferred into DH10B bacteria through the formation of an intermediate F' episome. Small modifications affecting culture volumes and growth times were added to the original protocol described by Kimmerly and colleagues (1994). Individual DH10B colonies were picked following transfer of the P1 DNA and tested for the presence of the *D1Mco1* sequence. All of them were positive. Accepting the co-existence of two P1 clones in the original NS3529 bacterial colony, the results indicate that only P1(*D1Mco1*) clones or heterologous colonies

consisting, again, of both P1 clones, were obtained following transfer of the P1 DNA. In contrast, colonies formed by P1 clones which only carried the chromosome 8 fragment were not observed. This seems unlikely, given the hypothesized faster growth of this latter clone in comparison to P1(*D1Mco1*).

(d) Analyses of the different P1 DNAs with all the D1Mco1, (D1)Gu14, (D1)Gu13 and (D1)Gu18 primer sets showed that the (*D1Mco1*$^-$)-DNA also lacked the *(D1)Gu13* sequence, whereas the other three types of (*D1Mco1*$^+$)-DNA represented a spectrum which ranged from the presence of all four SSLPs to the presence of *D1Mco1*, exclusively. The presence of SSLPs derived from different chromosomes in the DNA isolated from DH10B cells is the strongest evidence against the presence of two different P1 clones in the original bacterial colony, as it is extremely unlikely that these two clones would still co-exist in DH10B colonies following transfer of P1 DNA through two separate mating steps.

The observations made on the P1(*D1Mco1*) clone are unexpected, since most studies present P1 vectors as a very stable cloning system (Albertsen *et al.*, 1994; Nehls *et al.*, 1995; Harsman *et al.*, 1995), even when carrying repetitive sequences (Smoller *et al.*, 1991); also, reports on chimaeric P1 molecules are rarely encountered. Nehls and colleagues (1995) observed that YAC instability often relates to the nature of the DNA fragment being cloned. This could be the case with the P1(*D1Mco1*) clone. Chimaerism and instability could result from these regions in chromosomes 1 and 8 being highly unstable. The maintenance of different combinations of SSLPs in the P1 DNA as well as the difficulties to maintain the *(D1)Gu13*-containing fragment in standard pBluescript II KS+ subclones agree with this explanation.

CHAPTER 6

Precise mapping of the *agu* locus.

6.1 Introduction.

Allelic differences between the parental rat strains in the backcrosses were sought at each of the newly established SSLP loci. Informative markers were subsequently mapped by typing backcross progeny, and added to the previous genetic map. Typing of F344 backcross animals at the *D1Gu13* locus did not show linkage of this marker to the *agu* gene. Screening of somatic cell hybrid clones demonstrated that at least two of the microsatellites established within the P1(*D1Mco1*) clone, including *D1Gu13*, mapped instead to rat chromosome 8.

In parallel to this work, all the backcross progeny were typed with the closer flanking markers available at the time. Subsequent mapping of internal markers was carried out by only typing those animals that carried a crossover between any of the flanking loci and *agu*. This considerably reduced the number of animals to be typed, and allowed the screening of much larger numbers of backcross offspring. This approach ultimately led to the mapping of the *R158* microsatellite locus. This marker was found to map at less than 0.5 cM from *agu*.

6.2 Results.

6.2.1 Search for strain differences at the newly identified SSLP loci.

F344, AS/AGU and BN genomic DNAs were used as templates in PCR reactions with each of the new SSLP primers [Table 6-1]. When run on 4% MetaPhor agarose gels it was observed that PCR on F344 DNA with the D1Gu16 primer set resulted in two products (of approximately 205 and 220 bp), the smallest being the same size as that obtained when amplifying AS/AGU and BN DNAs. A third band, of approximately 370 bp in length, was also obtained from the AS/AGU template. [Figure 6-1]. Although weaker than the ca. 205 bp band, it was clear and, as confirmed later, it always accompanied the smaller AS/AGU product. Therefore, both this larger AS/AGU fragment and the ca. 220 bp F344 product were considered to result from the amplification of specific sequences present in the corresponding genomic DNAs, and not as PCR artifacts.

SSLP locus	Approximate PCR product size	Annealing temperature (°C)[b]	Allele differences
D1Gu17	203 bp	56	AS/AGU=BN=F344
D1Gu15	215 bp	54	AS/AGU=BN=F344(I)<F344(II)[c]
D1Gu19	239 bp	—	—
D1Gu16	208 bp	66	AS/AGU=BN=F344(I)<F344(II)[c]
D0Gu14[a] (*D1Gu14*)	176 bp	60	AS/AGU=BN=F344
D8Gu1[a] (*D1Gu13*)	171 bp	60	AS/AGU=BN<F344
D8Gu2[a] (*D1Gu18*)	230 bp	54	AS/AGU=BN=F344

TABLE 6-1. Allelic differences at newly developed SSLP loci.

[a] The microsatellites *D1Gu13* and *D1Gu18* were subsequently shown to map to rat chromosome 8 and were renamed *D8Gu1* and *D8Gu2*, respectively. Chromosomal localization of the marker *D1Gu14* could not be established and is currently called *D0Gu14*.

[b] All the PCR reactions were carried out for 25 cycles, in the presence of 1 mM $MgCl_2$. The primer set D1Gu19 was not tested.

[c] Two PCR products [(I) and (II)] were obtained from the amplification of F344 DNA template with each of the primer pairs D1Gu15 and D1Gu16.

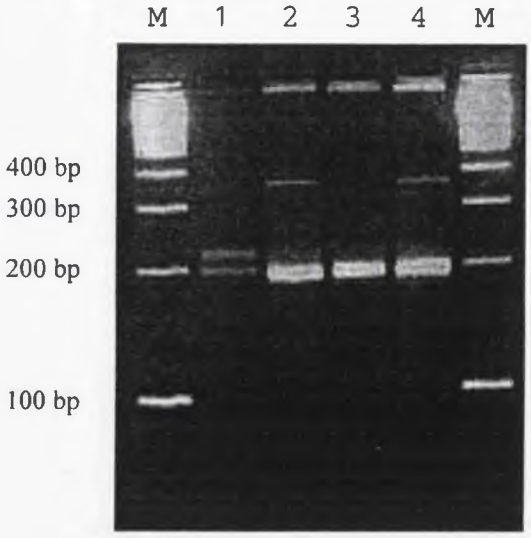


FIGURE 6-1. Resolution of the amplification products obtained with the primer set D1Gu16.

Resolution of the PCR products obtained from amplifying F344 (lane 1), AS/AGU (lane 2), BN (lane 3), and (AS/AGU x BN) F1 (lane 4) DNA templates with the primer set D1Gu16. A 100 bp DNA ladder was used as the size standard (M).

Linkage of the larger F344 product to the *agu* gene was assessed by typing a set of animals from the [AS/AGU x (AS/AGU x F344) F1] cross [Figure 6-2] and then searching for a correlation between the presence of this band and the display of a normal phenotype and *vice versa*. A χ^2 test showed this correlation to be in fact highly significant (χ^2 (df=3)= 106.4; $p<0.001$), pointing at *D1Gu16* as a new polymorphic marker between the F344 and the AS/AGU strains.

When the SSLP loci *D1Gu13*, *D1Gu14*, *D1Gu15*, *D1Gu17* and *D1Gu18* were analysed on 4% MetaPhor agarose gels none of them was found to be detectably polymorphic for any of the three rat strains under study. When resolved on 6% polyacrylamide gels, however, *D1Gu13* and *D1Gu15* were shown to be polymorphic between the F344 and the AS/AGU strains. The *D1Gu19* locus was not analysed.

6.2.2 Mapping of the new informative SSLPs.

One hundred and sixty-two [AS/AGU x (AS/AGU x F344) F1] animals were typed with the D1Gu16 primer, the results indicating an approximate distance of 9.3 ± 2.3 cM to the *agu* locus, distal to *D1Cep4*. Twenty-nine [AS/AGU x (AS/AGU x F344) F1] animals were typed with the D1Gu15 primer set [Figure 6-3]. Clear linkage to the *agu* locus was observed (χ^2 (df=3)= 27.41; $p<0.001$). A recombination frequency of 7% ± 4.7% was obtained. Another 9 litters where a crossover had been detected between *agu* and *D1Gu16* were also typed with this primer, yielding identical results to those observed with the D1Gu16 primer pair. These observations support the expected mapping of *D1Gu15* very close to *D1Gu16*.

Thirty-six [AS/AGU x (AS/AGU x F344) F1] animals were typed with the primer set D1Gu13. Contrary to the expected result, the corresponding locus did not show linkage to the *agu* gene (χ^2 (df=3)= 1.5; $0.5<p<0.9$). The location of *D1Gu13* was subsequently studied by screening a panel of mouse x rat somatic cell hybrids (Szpirer *et al.*, 1984). *D1Gu13* was then demonstrated to map to rat chromosome 8 [Figure 6-4], therefore explaining the lack of linkage to *agu*. The *D1Gu13* marker had been isolated from a P1 clone which carried the *D1Mco1* microsatellite (Chapter 5). This latter had been mapped by Gu and colleagues (1996) to the region of interest in rat chromosome 1. Consequently, the location of *D1Mco1* was assessed with the somatic cell hybrid panel. The results obtained agreed with those reported by Gu *et al.* (1996), confirming that *D1Mco1* maps to chromosome 1 in the rat. The same hybrid clones were also used

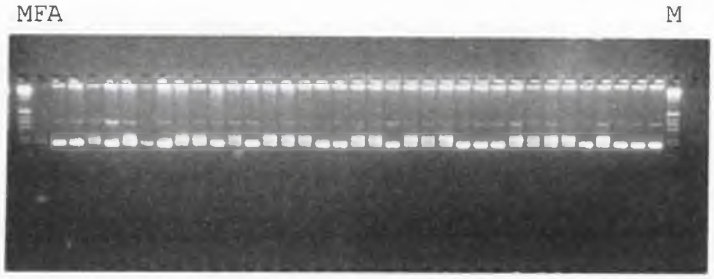


FIGURE 6-2. Typing of [AS/AGU x (AS/AGU x F344) F1] offspring with the primer set D1Gu16.

Resolution on a 4% agarose gel of the PCR products obtained from the amplification of F344 (F), AS/AGU (A), and [AS/AGU x (AS/AGU x F344) F1] DNA templates with D1Gu16. A 1 kb DNA ladder was used as the molecular weight standard (M).

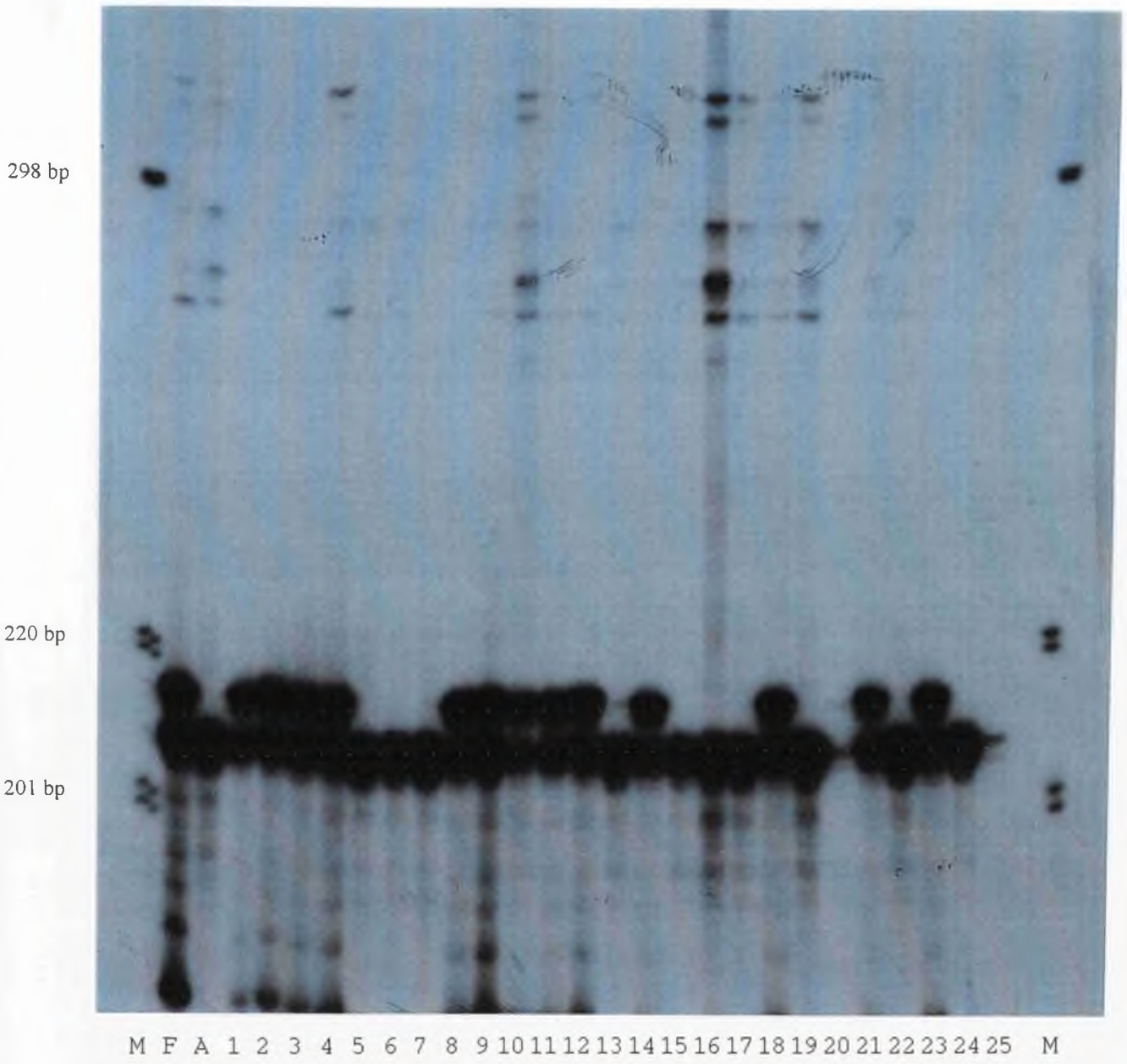


FIGURE 6-3. Resolution of *D1Gu15* PCR products on a 6% polyacrylamide gel.

The products obtained from amplification of F344 (F), AS/AGU (A), and [AS/AGU x (AS/AGU x F344) F1] (lanes 1-25) DNA templates with the primer set D1Gu15 were run on 6% polyacrylamide gels. The D1Gu15-forward primer was radioactively labelled prior to PCR amplification.
A radioactively labelled 1 kb DNA ladder was used as the size standard (M).

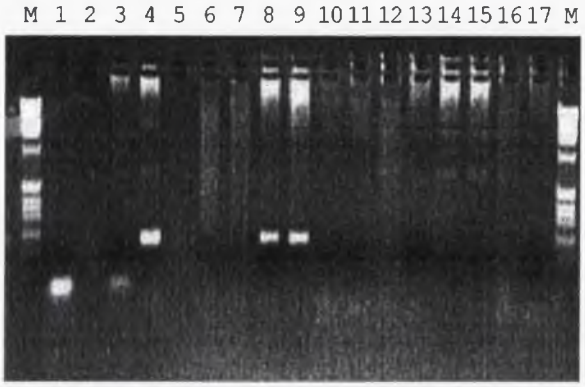


FIGURE 6-4. Screening of a panel of mouse x rat somatic cell hybrids with the primer set (D1)Gu13.

(D1)Gu13 was assigned to rat chromosome 8 following screening of a panel of mouse x rat somatic cell hybrid clones. The order of the DNA templates was as follows:

Lane 1: no DNA control.
Lane 2: empty lane.
Lane 3: BWTG3 mouse DNA.
Lane 4: Sprague-Dawley rat DNA.
Lane 5: Empty lane.
Lane 6: Clone LB251.
Lane 7: Clone LB600.
Lane 8: Clone BS140h1.
Lane 9: Clone BS511.
Lane 10: Clone LB510.6.
Lane 11: Clone LB330.
Lane 12: Clone LB860.
Lane 13: Clone LB780.
Lane 14: Clone LB1040.
Lane 15: Clone LB1040.TG3.
Lane 16: Clone LB780.6.
Lane 17: Clone LB330.TG3.

A 1 kb DNA ladder was used as the size marker (M).

to map the other two microsatellites which had been isolated from the same P1 clone, *D1Gu14* and *D1Gu18*. *D1Gu18* was shown to map to rat chromosome 8, whereas *D1Gu14* could not be mapped, as the rat PCR product could not be distinguished from that of the mouse. These data are presented in Table 6-2.

A third informative marker, *D1Gu10*, was identified in the *agu* region. This SSLP was isolated by Ms. N. Craig from a P1 clone carrying the *Grik5* gene. Typing of 203 [AS/AGU x (AS/AGU x F344) F1] animals by Ms. N. Craig showed that this locus mapped at 7 ± 1.8 cM from the *agu* locus, distal to *D1Cep4*.

6.2.3 Mapping of the R158 microsatellite locus.

The position of the rat *R158* microsatellite marker in relation to *agu* could not be defined following the typing of 264 backcross progeny, as no recombinants were identified. Therefore, unambiguous mapping of both genes demanded the typing of a larger set of animals. Moreover, analyses with this primer set required the running of radiolabelled PCR products on polyacrylamide gels. This method is more costly and time-consuming than the agarose gel system being used during most of this work. In order to increase the efficiency of the screening process at the *R158* locus as well as at any other new microsatellite marker close to *agu*, a panel was formed with the backcross animals where a recombination event had taken place between *agu* and any of the four closest flanking markers, *D1Mit1* and *D1Mgh7*, mapped on the [AS/AGU x (AS/AGU x BN) F1] cross, and *D1Cep4* and *D1Gu10*, mapped on the [AS/AGU x (AS/AGU x F344) F1] cross. Analysis with these markers was carried out on standard 4% agarose gels. Furthermore, a difference in the sizes of the *D1Cep4* and the *D1Gu10* PCR products allowed the multiplexing of both reactions, greatly improving the efficiency of these tests [Figure 6-5]. In addition to the typings already described above and in Chapter 4, a larger set of DNAs were prepared (by Mrs. M. Gardiner and myself) and I carried on with the screening for recombinants at either of the aforementioned loci. Altogether, a total of 400 BN backcross animals and over 500 F344 backcross progeny were typed. In parallel to this work, a third backcross was set up, [(AS/AGU x F344) F1 x AS/AGU). Nearly 400 offspring were typed at both *D1Cep4* and *D1Gu10*, in search for possible differences in the recombination frequencies in the (AS/AGU x F344) F1 males and the (AS/AGU x F344) F1 females. In contrast to a recombination frequency of 4.9% ± 0.96% in the interval *D1Cep4* - *agu* in F1 females, the

TABLE 6-2. Assignment of *(D1)Gu13* and *(D1)Gu18* to rat chromosome 8. Confirmation of the location of *D1Mco1* on rat chromosome 1.

Nine, twelve, and thirteen rat x mouse somatic cell hybrid clones were screened with the primer sets D1Mco1, (D1)Gu13, and (D1)Gu18, respectively. The results indicated that the two latter loci mapped to rat chromosome 8. *D1Mco1* was shown to map to rat chromosome 1, in agreement with the data published by Gu and colleagues (1996).

[a] Independent hybrid clones are those derived from distinct fusion events. Non-independent clones are indicated by asterisks.
[b] A + or a - indicates the presence or absence of the PCR product, respectively.
[c] A + indicates that the rat chromosome is present in more than 55% of the metaphases. A (+) indicates that the chromosome is present in 25-55% of the metaphases; (-) indicates that the chromosome is present in less than 25% of the metaphases ([d] in these two latter cases, the hybrid in question was not taken into account to establish the number of discordancies for that particular chromosome). A - indicates absence of the rat chromosome.
NT: The clone was not tested with that particular primer set.

Clones[a]	D1Mco1[b]	(D1)Gu13[b]	(D1)Gu18[b]	Rat chromosomes[c]																			
				X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
LB251	+	-	-	+	+	+	-	+	-	(+)	+	-	-	+	-	+	+	-	-	-	+	-	+	-
LB150.1	NT	-	NT	+	-	-	+	+	-	-	+	-	+	(+)	+	+	+	-	-	(+)	(+)	+	(+)	-
LB161	-	NT	-	+	-	+	+	+	+	+	+	-	+	+	-	(+)	+	+	+	+	+	+	+	(+)
LB780*	-	-	-	+	-	+	+	+	+	-	+	-	-	+	+	-	+	-	-	-	+	+	-	(+)
LB780.6*	NT	-	-	+	-	+	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	+	-	-
LB600	+	NT	-	+	+	+	+	+	+	(+)	+	-	(-)	+	+	+	+	+	+	+	-	+	+	-
LB860	-	-	-	+	-	+	+	+	-	-	+	-	+	-	+	+	+	-	+	+	+	+	-	(+)
LB330*	-	-	-	+	-	+	+	+	-	+	-	-	-	+	-	+	-	-	-	-	+	-	-	-
LB330.TG3*	NT	-	NT	-	-	-	-	+	-	-	-	-	-	-	-	(+)	-	-	-	-	-	-	-	-
LB510.6	-	-	-	+	-	+	+	+	-	-	+	-	-	-	-	+	+	+	+	+	+	+	-	-
BS140h1	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BS511	-	+	+	+	-	+	+	+	(-)	+	+	+	-	-	+	-	+	-	+	+	+	-	+	-
LB1040*	NT	-	-	+	-	-	+	+	(-)	+	+	-	-	+	+	+	-	-	+	+	-	+	-	+
LB1040.TG1*	NT	NT	-	-	-	-	-	+	-	+	-	-	-	-	+	(+)	-	-	+	+	-	+	-	-
LB1040.TG3*	NT	-	-	-	-	-	-	+	-	+	(-)	-	-	-	+	-	-	-	+	+	-	+	-	-
Number of discordant clones[d]:																								
D1Mco1				6	**0**	6	7	6	4	3	5	3	3	3	4	3	5	3	5	5	7	5	2	2
(D1)Gu13				8	2	6	6	9	2	3	6	**0**	3	6	6	7	6	2	4	4	5	8	1	2
(D1)Gu18				9	3	8	7	10	4	5	7	**0**	3	8	7	7	7	4	7	7	6	10	3	2

TABLE 6-2. Assignment of *(D1)Gu13* and *(D1)Gu18* to rat chromosome 8. Confirmation of the location of *D1Mco1* on rat chromosome 1.

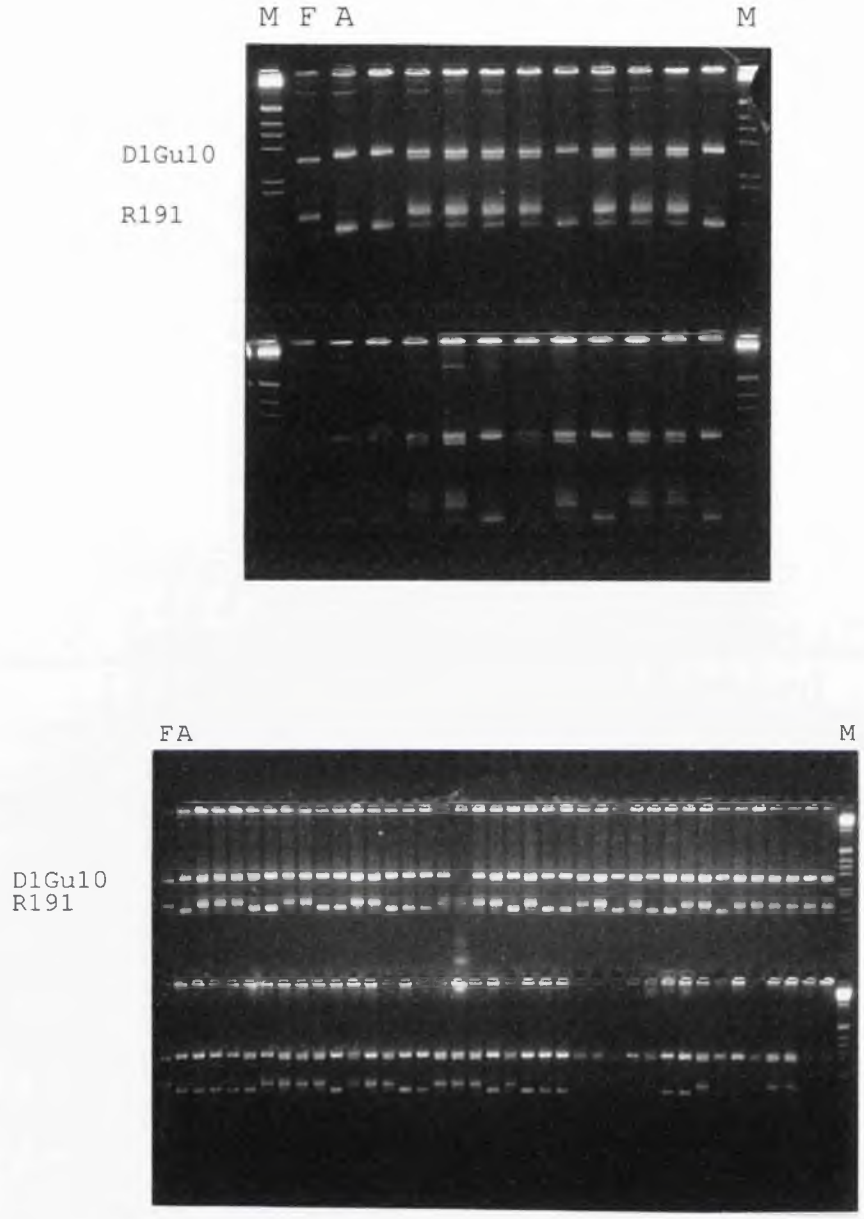


FIGURE 6-5. Resolution of amplification products obtained from multiplexed R191/D1Gu10-PCR reactions.

PCR reactions were carried out at 55°C for 27 cycles, in the presence of 1mM $MgCl_2$ and in a final volume of 15 µl. Both primer sets were added in each reaction. The amplification products were resolved on standard 4% agarose gels.

A: Product from the amplification of AS/AGU DNA template.

F: Product from the amplification of F344 DNA template.

M: 1 kb DNA ladder marker.

Remaining lanes: PCR products obtained from the amplification of DNA extracted from [AS/AGU x (AS/AGU x F344) F1] offspring.

recombination rate observed in this interval in F1 males was 2.5% ± 0.8%. Recombination frequencies in the interval *agu -D1Gu10* were similar in both sexes (i.e. 6.9% ± 1.1% in the F1 female, and 7.2% ± 1.3% in the F1 male).

Typing of recombinant animals with the R158 primer (by Ms. N. Craig) showed that this microsatellite locus maps at less than 0.5 cM from *agu* and distal to *D1Mit1*.

6.3 Discussion.

6.3.1 Mapping of new rat SSLPs.

The SSLP loci *D1Gu15* and *D1Gu16* showed differences between the AS/AGU and the F344 strains of rat. Amplification with the primer pair D1Gu16 yielded 2 specific bands with both the F344 and the AS/AGU DNAs. One of the products was of the same size in both strains, whereas the size of the larger fragments differed. Mapping of these strain-specific larger bands could not be carried out through the screening of somatic cell hybrid clones, as the parental rat strain used in their construction (Sprague-Dawley) only presented the lower band, common to all three strains, AS/AGU, BN, and F344. Two amplification products were also obtained with the D1Gu15 primer on F344 DNA. The lower band was of the same size as the product amplified from the AS/AGU and the BN DNAs. The larger fragment was specific to the F344 strain.

These observations could indicate the presence of residual heterozygosity at the *D1Gu15* and the *D1Gu16* loci, as previously observed in a few other cases (Chapter 3). On the other hand, the primer sets could be amplifying other loci in the genome which bear, specifically in the F344 strain, a very high homology to the sequence amplified in common in all the strains. These sequences can be, for example, pseudogenes or genes belonging to the same family, which have diverged to a lesser extent in the F344 strain than in the others. This explanation could also apply to the larger AS/AGU product obtained with the D1Gu16 primer set.

A clear correlation of the F344-specific larger band and the display of a normal phenotype was observed when typing [AS/AGU x (AS/AGU x F344) F1] litters with each of the primer sets D1Gu15 and D1Gu16 (i.e. 76 of 82 normal animals typed carried 2 *D1Gu16* alleles, and 18 of the 18 normal litters typed with the D1Gu15 primer carried 2 alleles at this SSLP locus). This eliminates the possibility that any of these two

loci is not linked to the *agu* gene. The same relationship was observed between presence of a single allele at each of these SSLP loci and the display of an affected phenotype. This strongly argues against the possibility that the different F344 animals used in the crosses to AS/AGU carried a different number of SSLP alleles (either 1 or 2) at *D1Gu15* or *D1Gu16*. Furthermore, comparison of the expected frequency of each genotype, assuming residual heterozygosity at either of these loci, with the observed data, clearly demonstrated that this was not the case. Residual heterozygosity at either *D1Gu15* or *D1Gu16* must result, in the absence of recombination between the SSLP and the *agu* loci, in a ratio of animals exhibiting a normal phenotype and carrying 2 alleles at the SSLP locus (agu^+ *D1Gu15(16)*^{r}/agu^- *D1Gu15(16)*^{a}), to animals exhibiting an affected phenotype and carrying 1 allele at the SSLP locus (agu^- *D1Gu15(16)*^{a}/agu^- *D1Gu15(16)*^{a}), of 1:2. In disagreement with this expectation, the data obtained indicate that both genotypes are found at approximately the same frequency. In addition, animals which carry a single SSLP allele and exhibit a normal phenotype (agu^+ *D1Gu15*^{a}/agu^- *D1Gu15*^{a}) should be observed at the same frequency than animals which exhibit a normal phenotype and carry two SSLP alleles (agu^+ *D1Gu15(16)*^{r}/agu^- *D1Gu15(16)*^{a}), and this is not the case. These results and the similar distances estimated between *D1Gu15*/*D1Gu16* and the *agu* loci (~7-9 cM), and the 8 cM distance reported between the *Scnn1b* and the *Pkc* genes in mouse (Brilliant *et al.*, 1996) strongly support the validity of using the *D1Gu15* and the *D1Gu16* F344-specific alleles as informative markers in studies of the segregation of loci in the *agu*-containing region. *D1Gu15*, *D1Gu16*, and a third microsatellite locus, *D1Gu10*, were added to our linkage map of rat chromosome 1, all distal to *D1Cep4*. The final map is shown in Chapter 8.

Three SSLPs were isolated from a P1 clone carrying the *D1Mco1* sequence. However, two of these, *D1Gu13* and *D1Gu18*, were shown to map to rat chromosome 8, following screening of a rat x mouse somatic cell hybrid panel (Szpirer *et al.*, 1984). Accordingly, these loci have been renamed *D8Gu1* and *D8Gu2*, respectively. The third one, *D1Gu14*, could not be mapped with this panel, as the rat product could not be distinguished from the mouse product. The same problem was encountered when the parental rat and mouse strains corresponding to a different panel of somatic cell hybrids were tested. Mapping of this marker could be attempted by probing blots of digested genomic DNA from the rat and the mouse parental strains and from the panel of

somatic cell hybrid clones with the (D1)Gu14 primers. Meantime, this locus is referred to as *D0Gu14*.

6.3.2 Mapping of closer markers.

Mapping of markers such as *R158*, located very proximal to *agu*, require the typing of very large numbers of animals, as increasingly smaller distances are associated to a reduction in the frequency of crossovers. An approach was followed where the backcross progeny were screened for the presence of recombination between *agu* and the closer flanking markers available. Given the proximity of these loci to *agu*, the failure to observe recombination in these intervals was considered as a strong indication that no crossovers had indeed occurred in the region. The alternative explanation, that this apparent lack of crossover events may result from more than one recombination event taking place in any of the intervals, was considered highly unlikely.

Two reciprocal F344 backcrosses were designed in order to study the frequency of recombination in the F1 male and in the F1 female, respectively. A considerable reduction in the recombination frequency was observed in the interval *D1Cep4* - *agu* in F1 males. This is in agreement with the numerous reports of an overall greater genetic length of the homogametic sex in various species (Haldane, 1922; Weissenbach *et al.*, 1992; Donis-Keller *et al.*, 1987). No marked differences were observed in the *agu* - *D1Gu10* segment. Non-uniformity of the differences in recombination between the male and the female genome has been previously described (Donis-Keller *et al.*, 1987). This heterogeneity suggests the existence of sex-specific points for recombination in the genomes.

Typing of the animals which carried a crossover in any of the intervals allowed a more efficient screening for the presence of recombination between *agu* and yet closer markers. In this way, approximately 200 animals, of a total of nearly 1300, were selected and typed with the primer set R158. This microsatellite marker was now shown to be separable from the *agu* mutation, and was mapped at less than 0.5 cM from the latter. This result allowed the initiation of physical mapping of the *agu* region, which is presently being carried out.

CHAPTER 7

Reassignment of the microsatellite locus *D1Mgh18* to rat chromosome 3.

7.1 Introduction.

A region containing the *agu* gene was initially defined, flanked by the loci *D1Mit1* and *D1Mgh7/Kal* (Chapter 4). Markers mapping within this part of rat chromosome 1 were optimised and tested for allelic differences between AS/AGU and either of the selected reference strains, BN and F344. One of these markers, *D1Mgh18* yielded then an unexpected result. This locus, mapping within the cystatin S gene (Cox and Shaw, 1992), demonstrated a strain difference between F344 and the mutant strain. According to the map released by Jacob and coworkers (1995) this SSLP mapped between the loci *D1Mit1* and *Cype*, close to *D1Mgh5*. However, when the corresponding primer set CYSS was used to type [AS/AGU x (AS/AGU x F344) F1] offspring, no linkage of *D1Mgh18* to the *agu* gene was detected. No linkage to *D1Mco2* or *D1Cep4* was identified either. These observations resulted in the location of *D1Mgh18* on the rat genome being called into question. This was firstly investigated by screening a panel of mouse x rat hepatocyte hybrids (Szpirer *et al.*, 1980).

Somatic cell hybrids have been extensively used in the establishment of genetic maps. This system was initially applied to the mapping of human and mouse loci (Minna and Coon, 1974; Kozak *et al.*, 1975). Subsequently this resource was made available for the mapping of rat loci. It was Weiss and Ephrussi who, in 1966, observed a preferential loss of rat chromosomes in rat x mouse somatic cell hybrids. Possible explanations for this phenomenon are differences in the replication times of the chromosomes in each species or adquisition of selective advantages by the clones where loss of rat chromosomes occurs.

Mouse x rat somatic cell hybrids were isolated by Szpirer and colleagues (1980) following fusion of mouse hepatoma cells and adult rat hepatocytes. The hybrid clones showed the expected random segregation of the rat chromosomes, thus being an excellent material to locate rat loci on their corresponding chromosome. The hybrids were subjected to detailed cytogenetic analyses (Szpirer *et al.*, 1984; Szpirer *et al.*, 1992). Presence or absence of each particular rat chromosome was firstly identified on high-quality G-banded metaphase spreads (Szpirer *et al.*, 1984). Unambiguous identification of rat chromosomes X, 1, 2, 3, 4, 7, 9, 11, 12, 13, 14 and 15 was readily achieved in most cases. Clear identification of a few other chromosomes was more difficult due to similarities to each other or to mouse chromosomes (e.g. rat chromosome 8, which is very similar to mouse chromosomes 9 and 13). Further

analyses, which included some small modifications to the fixation procedures, allowed a more accurate identification of the chromosomal constitution of the hybrids (Szpirer *et al.*, 1992). Two kinds of clones were obtained which contained either a single or a double set of mouse chromosomes. Both types segregated rat chromosomes and normally retained from 4 to 20 of these. In general, a larger number of mouse chromosomes correlated to decreased numbers of rat hepatocyte chromosomes. Observations by Weiss and Ephrussis (1966) seemed to indicate that the genomes of the two parental species are both active in the hybrid clones. On this basis, Szpirer and colleagues employed BWTG3, a mouse hepatoma cell line which is deficient in the enzyme hypoxanthine phosphoribosyltransferase (HPRT), to construct the hybrid panel. Selection of the clones was carried out in DHAT selective medium (containing hypoxanthine, aminopterin and thymidine). Therefore, the rat gene coding for this enzyme must be retained for the survival of the cells in such medium. This correlated with the presence of the rat X chromosome in practically all the clones, in agreement with observations in other mammals, where HPRT seems to map to this chromosome. No other selection was applied, and the rat autosomes which were retained varied among the different hybrids.

Numerous mapping studies using this and other panels of mouse x rat somatic cell hybrids have proved this system to be very suitable for mapping loci in the rat (Yoshida, 1978; Szpirer *et al.*, 1984; Szpirer *et al.*, 1992; Szpirer *et al.*, 1984; Levan *et al.*, 1991; Serikawa *et al.*, 1992). This was once again demonstrated in this work. Correlation of the rat chromosome constitution of each of the clones with the presence or absence of the *D1Mgh18* PCR product showed that, in fact, this locus did not map to rat chromosome 1. This result was consistent with previous observations when attempts to detect linkage of this SSLP to the *agu* gene and other markers in the region had failed. Confirmation of this result as well as a more precise location of *D1Mgh18* on the rat genetic map was achieved through linkage analysis studies.

7.2 Results.

7.2.1 Strain differences at the D1Mgh18 locus. Typing of [AS/AGU x (AS/AGU x F344) F1] offspring.

As mentioned in Chapter 4 (section *4.2.2*), allelic differences were observed at the *D1Mgh18* locus between the AS/AGU and the F344 strains of rat. Typing of backcross offspring was carried out as previously (i.e. about 100 ng of DNA was used as template in PCR reactions run at the conditions determined as optimal; the amplification products were then run on 4% MetaPhor gels).

The primer pair CYSS (locus *D1Mgh18*) was used to type 92 [AS/AGU x (AS/AGU x F344) F1] animals [Figure 7-1]. No linkage of the *D1Mgh18* locus to *agu* was detected, both exhibiting independent segregation in the offspring (χ^2 (df=3)= 1.13, 0.9>p>0.75). Lack of linkage to *D1Cep4* was also demonstrated, following analysis of 91 litters with the primers CYSS and R191 (locus *D1Cep4*), (χ^2 (df=3)= 0.82, 0.975>p>0.9). Similar results were obtained when typing 53 animals at both loci *D1Mgh18* and *D1Mco2* (χ^2 (df=3)= 7, 0.1>p>0.05).

Prior to these results it had been observed that the approximate sizes of the PCR products obtained with the F344 and the BN DNA templates were 110 and 135 bp, respectively [Figure 7-2]. This was in disagreement with the data presented by Jacob *et al.* (1995), where these products had been assigned sizes of 133 and 111 bp respectively.

7.2.2 Cloning of the D1Mgh18 PCR product. Sequence confirmation.

In order to confirm that no mistakes had been made during the synthesis of the primer set CYSS and that the discrepancies observed in relation to the original data were not the result of a variation in the primer sequences, the PCR product obtained from the F344 template was cloned into a pDT Blue vector (Chapter 2; section *2.3.16*). Part of the ligation product was used to transform electrocompetent TG1 cells and these plated on medium containing 100 $\mu g.ml^{-1}$ ampicillin as well as the indicators X-Gal and IPTG; the plates were incubated at 37^0C overnight. The yield of PCR product was quite low, and only a few colonies grew on the plates; most of these were white. Plasmid DNA was isolated from three cultures seeded with single white colonies and the presence of the PCR insert confirmed by amplification of 1 µl of each DNA sample

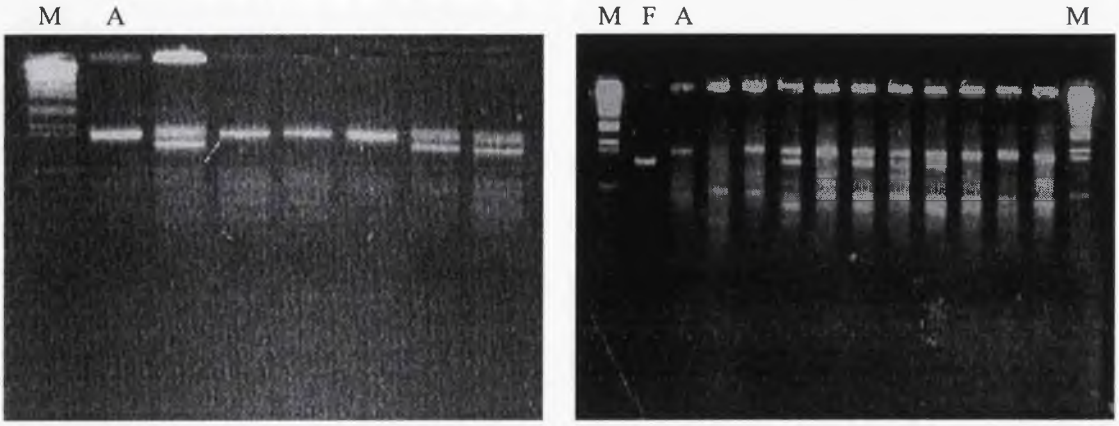


FIGURE 7-1. Typing of [AS/AGU x (AS/AGU x F344) F1] progeny with the primer pair CYSS.

Two examples of resolution of *D1Mgh18* PCR products on 4% MetaPhor gels. The DNA templates correspond to F344 (F), AS/AGU (A), and [AS/AGU x (AS/AGU x F344) F1] offspring. A 1 kb DNA ladder (M) was used as the size standard.

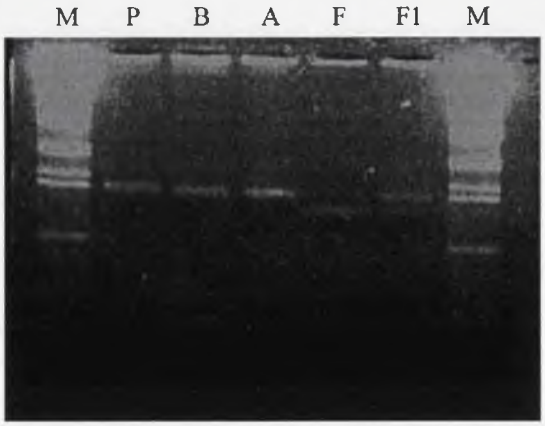


FIGURE 7-2. Strain distribution pattern of *D1Mgh18* alleles.

Resolution on a 4% MetaPhor gel of the PCR products obtained following amplification of PVG (P), BN (B), AS/AGU (A), F344 (F), and (AS/AGU x F344) F1 (F1) DNA templates with the primer set CYSS. A 1 kb DNA ladder (M) was used as the molecular weight standard.

with the primer set CYSS. This approach was an alternative to the analysis of the restriction pattern obtained following double digestion with, for example, the endonucleases *Hin*dIII and *Eco*RI, located at each side of the cloning site, and therefore useful to extract the insert from the pDT vector. The three DNAs isolated yielded amplification products of the right size (resolved in 4% MetaPhor agarose gels); no product was obtained from either a "no template" reaction or from a PCR reaction with pDT Blue DNA as the template.

Two of these three DNA samples were sequenced with the -21 M13 Forward primer and a T7 sequencing primer (Chapter 2; section *2.3.18.2*). This allowed the confirmation of the PCR primer sequences as those reported to amplify the *D1Mgh18* locus. A $(GT)_{24}$ motif interrupted twice by single (CT) and (TT) units constituted, together with the primer sequences, most part of the insert, with the CYSS-L primer already running into the (GT) stretch. The PCR product was found to consist of 115 bp; a BLAST search and direct comparison of both sequences [Figure 7-3] demonstrated the product to be identical to a fragment of the rat cystatin S (Cys S) gene (Cox and Shaw, 1992). These results thus confirmed the previous observations (i.e. lack of linkage to the *agu* gene; allele in the F344 strain of a smaller size than that in BN and AS/AGU rats) to indeed correspond to the *D1Mgh18* locus.

7.2.3 Mapping of the D1Mgh18 locus.

7.2.3.1 Screening of somatic cell hybrid clones.

The next step towards finding a possible interpretation of the data which had been obtained was to confirm the location of *D1Mgh18* on the rat map. This was done by screening a panel of mouse x rat somatic cell hybrids (Szpirer *et al.*, 1980) for the presence of the *D1Mgh18* locus, via PCR. Prior to screening, PCRs were carried out on both the mouse (BWTG3) and the rat (Sprague-Dawley) DNA templates to check that these yielded products of a different size. This was required to ensure that the rat signal could be unambiguously identified. No product was obtained with the mouse DNA template, presenting the panel of somatic cell hybrids as a valid assay to map the locus to its corresponding chromosome.

Fourteen hybrid clones were screened with the primer set CYSS [Figure 7-4]. Comparison of the segregation pattern of the different rat chromosomes with the presence or absence of the rat-specific PCR product in each clone demonstrated a

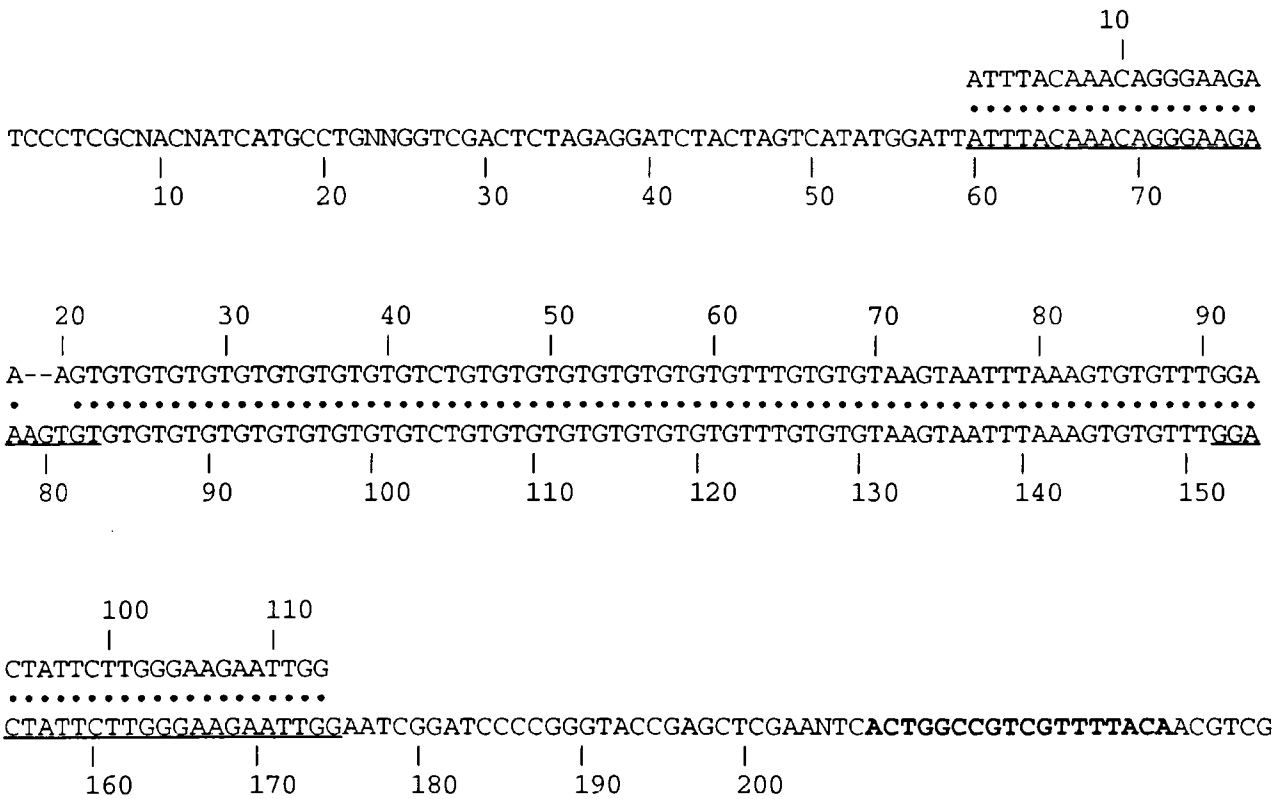


FIGURE 7-3. Sequence confirmation of the *D1Mgh18* PCR product with the rat *Cys* S gene.

The PCR product obtained following amplification of F344 DNA template with the primer set CYSS was cloned into a pDT Blue vector and sequenced with the -21 M13 forward primer and a T7 primer. The alignment of the sequence obtained with the T7 primer (bottom row) and a fragment of the rat *Cys* S gene (top row) is shown above. The sequences of the CYSS forward and reverse primers are underlined. The vector sequence which is complementary to the -21 M13 forward primer is indicated in bold.

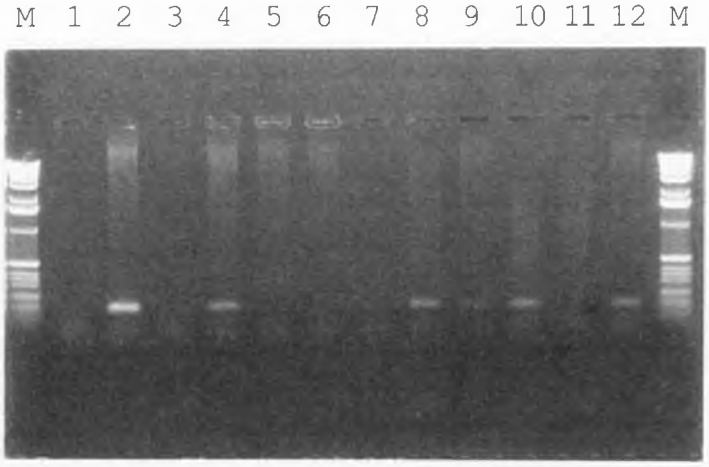


FIGURE 7-4. Screening of rat x mouse somatic cell hybrid clones with the primer set CYSS.

Amplification products obtained with the primer CYSS on DNA from BWTG3 mouse (lane 1), Sprague-Dawley rat (lane 2), and rat x mouse somatic cell hybrid clones. The products were run on a 1% agarose gel. A 1 kb DNA ladder (M) was used as the size standard.

Hybrid clones: LB251 (lane 3); LB600 (lane 4); LB780.8 (lane 5); LB1040.TG1 (lane 6); LB 210.I (lane 7); LB330 (lane 8); LB780 (lane 9); LB860 (lane 10); LB780.6 (lane 11); LB510.6 (lane 12).

perfect correlation of the presence of the amplification product with the maintenance of rat chromosome 3 in the hybrid clones, and *vice versa* [Table 7-1]. Different numbers of discordant clones were found for each of the other rat chromosomes, including chromosome 1. Thus, contrary to the published location of *D1Mgh18* on the latter, the data indicated that this microsatellite mapped to rat chromosome 3 (Durán Alonso *et al.*, 1997).

7.2.3.2 Linkage analysis.

Further mapping of the *(D1)Mgh18* locus was carried out by investigating linkage to other markers on chromosome 3 which were also informative for the [AS/AGU x (AS/AGU x F344) F1] cross. Only were 2 of these found among all the optimised SSLPs in the panel (Chapter 3): *D3Mit4* (primer set R593; Jacob *et al.*, 1995) and *R69* (locus *Scn2a*; Serikawa *et al.*, 1992). Sixty-four offspring were typed in common with the 3 primer sets. A clear linkage was observed between *(D1)Mgh18* and *D3Mit4* (χ^2 (df=3)=46.9, p<0.001). Linkage of the former to *Scn2a* was not significant (χ^2 (df=3)= 5.375, 0.5>p>0.1). Three point cross analysis indicated the order of the loci to be *(D1)Mgh18 - D3Mit4 - Scn2a* (Durán Alonso *et al.*, 1997). An additional number of animals had been typed for the individual adjacent pairs of loci. A total of 77 offspring were analysed at both the *(D1)Mgh18* and the *D3Mit4* loci; sixty-five animals were typed in common at the *D3Mit4* and the *Scn2a* loci. The distance between the first two markers was estimated as 10.4 ± 3.5 cM and that between *D3Mit4* and *Scn2a* as 41.5 ± 6.1 cM. These data translate into the following linkage map [Figure 7-5]:

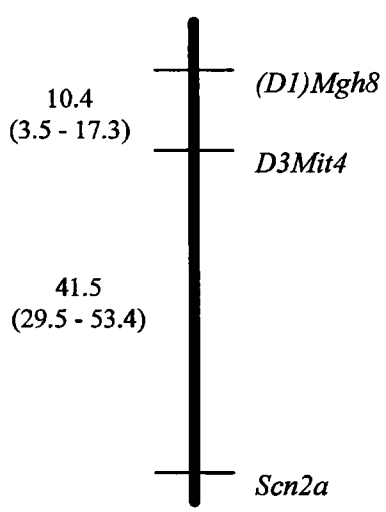


FIGURE 7-5. Order of the loci *(D1)Mgh18*, *D3Mit4* and *Scn2a* on rat chromosome 3.

Recombination frequencies are shown to the left of the chromosome. The 95% lower and upper confidence limits are indicated in parentheses.

Clones[a]	D1Mgh18[b]	Rat chromosomes[c]																				
		X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
LB251	-	+	+	+	-	+	-	(+)	+	-	-	+	-	+	+	-	-	-	+	-	+	-
LB150.1	+	+	-	-	+	+	-	-	+	-	+	(+)	+	+	+	-	-	(+)	(+)	+	(+)	-
LB161	+	+	-	+	+	+	+	+	+	-	+	+	-	(+)	+	+	+	+	+	+	+	(+)
LB210.I	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	-	(-)
LB780*	+	+	-	+	+	+	+	-	+	-	-	+	+	-	+	-	-	-	+	+	-	(+)
LB780.6*	-	+	-	+	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	+	-	-
LB780.8*	-	+	-	+	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	+	-	-
LB600	+	+	+	+	+	+	+	(+)	+	-	(-)	+	+	+	+	+	+	+	-	+	+	-
LB630	+	+	(-)	-	+	+	(+)	+	+	-	+	-	+	+	+	(+)	+	+	-	+	+	(-)
LB860	+	+	-	+	+	+	-	-	+	-	+	-	+	+	+	-	+	+	+	+	-	(+)
LB330*	+	+	-	+	+	+	-	+	-	-	-	+	-	+	-	-	-	-	+	-	-	-
LB330.TG3*	-	-	-	-	-	+	-	-	-	-	-	-	-	(+)	-	-	-	-	-	-	-	-
LB510.6	+	+	-	+	+	+	-	-	+	-	-	-	-	+	+	+	+	+	+	+	-	-
LB1040*	+	+	-	-	+	+	(-)	+	+	-	-	+	+	+	-	-	+	+	-	+	-	+
LB1040.TG1*	-	-	-	-	-	+	-	+	-	-	-	-	+	(+)	-	-	+	+	-	+	-	-
Number of discordant clones[d]:		4	8	6	**0**	3	4	5	3	9	4	6	6	2	5	6	3	3	4	5	6	4

TABLE 7-1. Assignment of the *D1Mgh18* SSLP locus to rat chromosome 3.

[a] Independent hybrid clones are those derived from distinct fusion events. Nonindependent clones are indicated by asterisks.

[b] A + or a - indicates the presence or absence of the PCR product, respectively.

[c] A + indicates that the rat chromosome is present in more than 55% of the metaphases. A (+) indicates that the chromosome is present in 25-55% of the metaphases; (-) indicates that the chromosome is present in less than 25% of the metaphases ([d] in these two latter cases, the hybrid in question has not been taken into account to establish the number of discordancies for that particular chromosome). A - indicates absence of the rat chromosome.

7.3 Discussion.

An informative SSLP locus was found, *D1Mgh18*, which putatively mapped in the region containing the *agu* gene. According to a published map (Jacob *et al.*, 1995) this locus was located between *D1Mit1* and *Cype*. Therefore, this microsatellite marker was expected to map at a distance of less than 7.6 cM (95% upper confidence limit for *D1Mit1*).

Allelic differences were demonstrated between the AS/AGU and the F344 rat strains at this locus. Contrary to previously released data, the *D1Mgh18* allele in the latter was smaller than that in the BN and the AS/AGU strains. Moreover, typing of [AS/AGU x (AS/AGU x F344) F1] progeny with the corresponding primer pair CYSS indicated that this locus was not linked to the *agu* gene nor to two other markers in the region of interest, *D1Cep4* and *D1Mco2*. Correlation of these results with the published *D1Mgh18* locus was confirmed by cloning and sequencing the PCR product. Mapping of this SSLP within the cystatin S (*Cys S*) gene was equally demonstrated.

PCR screening of a panel of mouse x rat somatic cell hybrid clones showed *(D1)Mgh18* to map to rat chromosome 3 and not to chromosome 1. In contrast to the perfect correlation observed between the former and the rat-specific PCR product, more than 50% of the clones were discordant with the expectation of this marker being on rat chromosome 1. This is summarised below, in Table 7-2. This new location of *(D1)Mgh18* explains the failure to detect any linkage of this marker to other chromosome 1 loci, experienced by this group and also reported by Gu and coworkers (1996).

		Chromosome 1		Chromosome 3	
		+	-	+	-
(D1)Mgh18	+	1	7	8	0
	-	1	5	0	6

TABLE 7-2. Correlation of the presence or absence of the *(D1)Mgh18*-PCR product with each of the chromosomes 1 and 3 in a panel of 14 mouse x rat somatic cell hybrid clones.

Further evidence of *(D1)Mgh18* indeed mapping to rat chromosome 3 was obtained from linkage analysis studies. Another SSLP marker which mapped to this chromosome, *D3Mit4*, was clearly demonstrated to be linked to the locus under study, at a distance of about 10 cM. Three point cross analysis between these two markers and a third chromosome 3 locus, *Scn2a*, showed that *(D1)Mgh18* mapped distal to the latter. Comparison of the data obtained with the map published by Jacob and coworkers (1995) showed a close similarity between the distances estimated between *D3Mit4* and *Scn2a*. These were 41.5 cM and around 40 cM, respectively. A more precise location of *(D1)Mgh18* could not be defined, as none of the other SSLP loci analysed to that date were informative for the [AS/AGU x (AS/AGU x F344) F1] cross. Nonetheless, according to the estimated distances, it seems likely that the *Cys S* gene will map in the region flanked by the loci *D3Mit2* and *D3Mgh3*, presented in the map by Jacob *et al.* (1995).

Given its location, *(D1)Mgh18* was renamed *D3Gu1*, according to the rat nomenclature committee.

CHAPTER 8

General discussion.

8.0 Introduction.

The AS/AGU strain of rat constitutes a new genetic model of movement disorder. The phenotypic abnormalities exhibited by these animals are believed to result from mutation at a single locus, *agu*. The main objective of this work was the mapping of the *agu* mutation to a genetic interval of 1 cM or less, which allowed the initiation of the physical mapping. In the absence of any evidence for the function of the product of the *agu* gene, a positional cloning approach was required. Microsatellite loci were selected as genetic markers (Litt and Luty, 1989; Weber and May, 1989), and linkage to *agu* was sought in two sets of backcross progenies. Co-segregation of alleles at *agu* and the kallikrein (*Kal*) gene, located on rat chromosome 1, indicated that *agu* mapped to this chromosome. Linkage analysis studies with other available SSLP markers allowed the establishment of a 25-30 cM interval containing the *agu* gene. Attempts to develop new informative markers through the amplification of rat genomic DNA with mouse SSLP primers failed. In contrast, screening of a rat P1 library with dinucleotide repeats resulted in the isolation of various new microsatellite markers. The *agu*-containing region was consequently narrowed down to ca. 11 cM. This was still too large an interval to attempt the physical mapping of the region.

The publication in 1996 of two genetic maps of rat chromosome 1 (Gu *et al.*, 1996; Ding *et al.*, 1996) revealed that the protein kinase C, type I (*Pkc*) gene mapped within the *agu*-containg interval. The microsatellite locus *R158*, identified within *Pkc*, was shown to be informative for both backcrosses in this project. Typing of backcross progeny demonstrated that this SSLP locus maps at less than 0.5 cM from *agu*, thus making it possible to start the physical mapping of this region. The final genetic map is shown in Figure 8-1.

8.1 Precise genetic mapping of the *agu* mutation.

8.1.1 Approaches to the chromosomal localisation of agu.

A genome-wide search was required to identify the location of the *agu* locus on the rat genome. Microsatellites were selected as the genetic markers, on the basis of

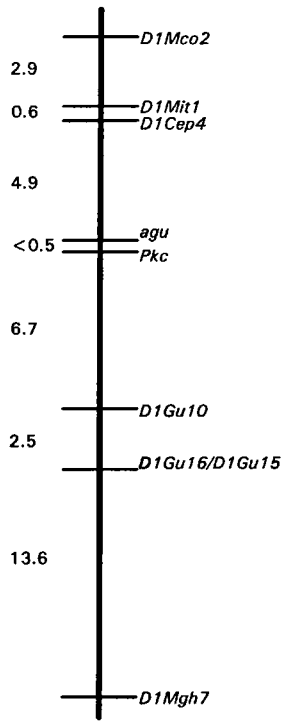


FIGURE 8-1. Composite map of the *agu*-containing region in rat chromosome 1. Distances between loci (in cM) were calculated using the Kosambi's map function, and are indicated to the left of the chromosome. The microsatellite loci *D1Mit1* and *D1Mgh7* were used to type BN backcross offspring. Their position should be regarded as approximate since it does not correspond to the number of recombination events observed between these and the rest of the loci, mapped on the F344 backcross.

their widespread distribution on the genome, the high rate of allelic variation, and the ease with which these loci can be typed by PCR amplification (Luty *et al.*, 1990; Beckmann and Weber, 1992; Serikawa *et al.*, 1992). Nevertheless, high-resolution maps of the rat genome are not available yet, in comparison to the mouse (Dietrich *et al.*, 1994) and man (Murray *et al.*, 1994). At the time this work was started, between 50 and 70% of the rat genome was covered by genetic markers (Serikawa *et al.*, 1992), only a proportion of which were informative in either or both of the backcrosses used. Therefore, linkage of the available rat markers to the *agu* locus might not have been so readily detected, if at all. Typing of backcross progeny with a very large number of SSLPs might have been required in order to establish linkage of at least one marker to *agu*. Automated genotyping, not used in this work, could have then been a crucial factor in increasing the efficiency of the screening process, as combinations of the various fluorescent dyes and of amplification reactions which yield products of different size, allow a higher throughput (Edwards *et al.*, 1991; Smith, 1995). In addition, separate pools could have been prepared with DNAs extracted from normal and from affected offspring, respectively (Michelmore *et al.*, 1991; Williams *et al.*, 1993; Sheffield *et al.*, 1995). Random assortment should be observed between *agu* alleles and alleles at SSLP loci which are not linked to *agu*. Therefore, amplification with these markers should yield two signals of approximately the same intensity in each DNA pool, one corresponding to the AS/AGU allele and the other to the allele from the reference strain in the backcross (i.e. BN or F344). In contrast, amplification of the pool prepared with DNAs from affected progeny with SSLP markers which are linked to *agu*, should yield a single main product, corresponding to the AS/AGU allele. The signal given by the allele from the reference strain should weaken as the markers get closer to the *agu* gene. Confirmation of a possible linkage would then require the typing of the individual animals at the particular microsatellite locus.

A proportion of 25% of the SSLPs which were initially studied did not show allelic differences between the parental strains in the backcrosses. Williams *et al.* (1997) have demonstrated the use of SSCP analyses to reveal polymorphism at microsatellite loci which do not differ in the number of repeats. Therefore, in the absence of linkage of any of the informative markers to *agu*, the uninformative loci could have been further analysed in SSCP assays. Moreover, analysis of a larger number of rat strains prior to the establishment of the backcrosses may have revealed a

greater divergence between AS/AGU and some strains other than BN and, especially, F344, than between these two latter and AS/AGU.

Genome scan analyses could also have employed other types of genetic markers, such as random amplified polymorphic DNA (RAPDs). These polymorphisms result from the amplification of genomic DNA with single 9-10 nucleotide primers of arbitrary sequence (Williams *et al.*, 1990). A series of DNA fragments are obtained with each primer which can be easily analysed on agarose gels. Advantages of these markers are the simplicity of the technique and, above all, the fact that no sequence information is required for the development of amplification primers. This is in contrast to SSLPs, where establishment of new markers requires identifying repeat sequences in cloned DNA fragments and obtaining the unique sequences flanking both sides of the repeat, as already described in Chapter 5. Moreover, the amplification of various DNA fragments in a single reaction renders RAPDs a fast and efficient method to detect differences between genomes. A further increase in the level of polymorphism has been observed following digestion of the templates prior to PCR (Koebner, 1995). This may be due to a reduction in the secondary structure, which in turn allows the primers to anneal more efficiently to the template. Disadvantages of this approach are problems with reproducibility and with the interpretation of some banding patterns. Careful standardization of the amplification conditions has been reported to solve many of the problems of poor reproducibility (Meunier *et al.*, 1993; Micheli *et al.*, 1994). Resolution on polyacrylamide gels eliminates the apparent co-migration of DNA fragments of very similar size. However, it is not possible to distinguish between fragments which are identical in size but differ in sequence. Another drawback is that heterozygotes can not be distinguished from one of the parental homozygotes. This is due to the fact that RAPDs are the result of amplification of a given DNA fragment from one of the genomes but not from the other. Consequently, for RAPD markers to be useful in the mapping of the *agu* mutation, these should result from the amplification of DNA from the reference strain and not from AS/AGU. In this way, presence of the RAPD fragment in the backcross progeny would be indicative of heterozygosity at that particular locus. An alternative would be the cloning of the informative RAPD fragments and their sequencing (Kurata *et al.*, 1994), thus allowing the designing of appropriate primers at both ends of each fragment. Primer pairs which are developed through this route define what is called sequence-characterised amplified regions

(SCARs) (Naqvi and Chattoo, 1996). These polymorphisms are highly reproducible and allow the differentiation between heterozygotes and both parental homozygotes.

Representational difference analysis (RDA) could well have been an alternative to the sequential screening for linkage of *agu* to other genetic loci. This method has already been described in Chapter 1 (Lisitsyn *et al.*, 1993; Lisitsyn, 1995). Another approach that could be considered is RFLP subtraction (Rosenberg *et al.*, 1994), which results in the purification of restriction fragments from a genome ('tracer') that do not have a counterpart of the same size in a competing genome ('driver'). This method differs from RDA in that the size fractionation of the DNA fragments is carried out through gel purification instead of PCR and that amplification of 'tracer'-specific DNA (i.e. the 'tester' DNA in RDA) is only carried out after the last subtractive hybridisation step, when the sample has been greatly enriched for 'tracer' molecules. A greater power and specificity has been reported for this latter technique (Rosenberg *et al.*, 1994). Although technically demanding, these methods are to be seriously considered in projects such as the positional cloning of the *agu* gene, where dense genetic maps of the organism in question are not available. Search for informative markers which map within the *agu*-containing region could be carried out with a pool of affected backcross or F2 animals as the source of 'driver' DNA and F1 animals from the corresponding cross as those providing the 'tester' DNA. In this case, the 'driver' DNA most likely contains alleles from both AS/AGU and the reference parent at all the loci except at those mapping to the chromosomal region of interest (i.e. the *agu* gene and proximal loci). In contrast, the 'tester' DNA also carries an allele from the reference parent at these loci. Therefore, sequences which only correspond to the reference parent are not subtracted and are instead selectively amplified. These products can subsequently be cloned and analysed.

8.1.2 Approaches to the precise mapping of agu.

Prior to the discovery of the close linkage of the *Pkc* gene to *agu*, the *agu*-containing region could not be narrowed down to less than 25-30 cM, as more informative markers were not available. Despite the failure in establishing new informative microsatellite loci based on the amplification of rat genomic DNA with mouse SSLP primers, the number of loci tested in this project was small. Given the high density maps currently available for the mouse (Dietrich *et al.*, 1994; Brilliant *et al.*,

1996), a rapid assay of all the mouse SSLPs mapping within the region of interest in mouse chromosome 7 would still remain a sensible approach to the identification of new markers in the *agu*-containing fragment in the rat. Integration of newly identified rat products into the rat map would first require their confirmation as new rat microsatellites, by sequencing. Linkage analysis studies would be carried out with those markers which were informative in the backcrosses. The conservation of synteny at these loci between the mouse and the rat would result in the refinement of the *agu* region. The close synteny observed to date between the region of interest on rat chromosome 1 and proximal mouse chromosome 7 would also validate the use of rat homologues of mouse genes on chromosome 7 to be screened for cleaved amplified polymorphic sequences (CAPS) (Konieczny and Ausubel, 1993). Strain differences at these marker loci are revealed following amplification of the same gene fragment in each strain and digestion of the PCR products with a panel of restriction enzymes. The DNA fragments are then analysed on standard agarose gels, in search for differences in the digestion patterns. Inclusion of introns in the sequence to be amplified maximizes the possibility of finding differences between the strains.

Genetically directed representational difference analysis (GDRDA) (Lisitsyn *et al.*, 1994), already described in Chapter 1, and microdissection approaches, could be two other alternatives to the screening of the P1 library with available reference sequences (i.e. uninformative rat SSLP loci; rat homologues to genes mapped in the region of interest on mouse chromosome 7). GDRDA could be attempted on 'driver' and 'tester' DNA pools prepared as described above (section *8.1.1*), with the difference that only affected backcross animals which carried a cross-over between the *agu* gene and either one or both proximal flanking markers (e.g. *D1Mit1* and *Kal* (or *D1Mgh7*), in the case of [AS/AGU x (AS/AGU x BN) F1] offspring), would be selected to create the 'driver' DNA pool. This selection would further refine the chromosomal area within which allelic differences were to be identified.

Microdissection is widely used as a means to isolate specific chromosomal fragments which can then be further characterised. The dissected DNA can be digested with specific restriction endonucleases and the fragments used to construct libraries, following ligation to adaptor sequences, PCR amplification, and cloning into appropriate vectors (Johnson, 1990). Clones thus derived from the *agu* region could be directly analysed for the presence of repeat sequences. Meltzer and colleagues (1992)

have reported the successful generation of a set of amplification products from specific chromosomal regions, by using a universal primer to initiate the amplification of the microdissected DNA fragment at various sites throughout its length. These products can be labelled and directly used to probe genomic libraries prepared from either whole genomes or specific chromosomes (Flejter *et al.*, 1995). Thus, this approach eliminates the use of any microchemical techniques on the dissected DNA. DNA fragments obtained from the amplification of the *agu* region following microdissection could be employed to screen the rat P1 library. Microdissection of the *agu*-containing region would overcome a limitation inherent to the approach followed in this project, i.e. dependence on the availability of reference sequences to identify genomic clones which derive from the region of interest.

The vast majority of SSLP loci which mapped to the *agu* region had to be analysed on polyacrylamide gels, since the use of 4% MetaPhor gels did not ensure the resolution of 2 bp differences. Search for tri- and tetranucleotide tandem repeats when developing new microsatellite markers could have overcome this requirement, since allelic differences between strains must then correlate to size differences of 3 bp or more.

A large proportion of the microsatellite markers identified to date in the region carrying the *agu* gene are not informative for any of the backcrosses established. Had the *Pkc* marker *R158* not been informative either, the lack of heterogeneity observed between AS/AGU and the F344 and the BN strains at the *agu*-containing region would have become a major obstacle to the precise genetic mapping of this mutation. In this case, search for other rat lines that differ from AS/AGU at SSLP loci in this interval, and establishment of new crosses of the most divergent strains to AS/AGU, would have been critical, and one or several of the methods described above used to isolate new genetic markers close to the *agu* locus. However, this was not necessary, as mapping of the *R158* microsatellite at less than 0.5 cM from *agu* provided a starting point for physical mapping.

8.2 Future work.

8.2.1 Physical mapping of the agu region.

The identification, by Dr.P.Shiels, of a P1 clone which carries the *R158* microsatellite locus [P1(*R158*)], constitutes the first step in the physical mapping of the *agu*-containing region. Probes derived from the ends of the genomic insert are to be used to screen the entire P1 library, in search for overlapping clones. These probes can either be subclones of the end fragments or PCR products amplified following the determination of DNA sequences at the ends of the P1 clone (Kimmerly *et al.*, 1994). Alternatively, riboprobes can be synthesized from the T7 and the Sp6 promoters present at each side of the cloning site on the p*Ad*10*sac*BII vector (Southard-Smith *et al.*, 1994).

Various problems can be encountered during chromosome walking. Presence of repetitive sequences at the ends of the genomic insert precludes these from being used to search for overlapping clones. This may restrict the utility of some of the end sequences, given the abundance of long interspersed elements (LINEs) in the rat genome. In the case that both ends of the genomic insert were constituted by repetitive sequences, a new clone carrying the *R158* sequence could be sought for. Alternatively, further sequencing of the P1(*R158*) insert ends could be carried on until a region of unique sequence was identified. A third possibility would be to use sequence-tagged-sites (STSs) derived from the P1(*R158*) clone to identify overlapping clones. It could also happen that the genomic region which is adjacent in the chromosome to the DNA fragment carried by the P1(*R158*) clone is absent from the library. This would result in the cessation of the walking. In addition, the presence of chimaeric clones in the library could seriously endanger the identification of genomic regions which are adjacent to the *R158*-region in the chromosome. The quality of the genomic library is thus of prime importance.

Successful walking approaches are often based on the use of a combination of different genomic libraries (Rommens *et al.*,1989; Albertsen *et al.*, 1994; Wang *et al.*, 1996). A brief description of these libraries is given in Chapter 1 (section 1.3). Rat genomic libraries have been recently constructed using the YAC (Cai *et al.*, 1997) and the PAC vector systems. Integration of these latter and the P1 library should ensure the coverage of the *agu* region, overcoming problems due to the presence of repetitive sequences or to the lack of representation of a particular genomic region in a library.

A thorough characterisation of the identified overlapping clones is critical. Confirmation of the overlap between clones can be achieved through the analysis of the restriction patterns obtained with each clone, following partial digestion with a panel of restriction endonucleases (Wada *et al.*, 1990; Hamvas *et al.*, 1994). However, detection of overlap between two clones using this method requires that the region in common is considerably large. Another way to detect overlap between clones is to analyze the STSs carried in each clone (Albertsen *et al.*, 1994). The clones must also be compared to the genomic region, in search for any rearrangements that may have taken place when constructing and/or analyzing the library (Feil *et al.*, 1990b). In this context, development of microsatellite loci within the clones serves two purposes, the establishment of STSs which can be used in the analysis of overlapping clones, as stated above, and the identification of possible new informative markers between AS/AGU and the reference strains. In this latter case, typing of the backcross progeny should result in a further refinement of both the genetic and the physical maps. Moreover, these markers would allow the orientation of the walk. Independently from the linkage analyses, confirmation of the chromosomal location of each newly developed STS is essential in any walking approach. As explained in Chapter 1 (section 1.3), mapping of non-polymorphic sequences may be carried out by screening a panel of somatic cell hybrids. A radiation hybrid panel of the rat genome is currently under construction (James and Lindpaintner, 1997). This type of hybrids represent a considerable refinement to the mapping with somatic cell hybrids, as they allow a precise localisation of genetic loci on the chromosomes (James *et al.*, 1994).

8.2.2 Identification of genes in the agu region.

Various strategies can be followed to isolate the genes present in a region of the genome for which a physical map has been constructed (Parrish and Nelson, 1993; Monaco, 1994). Each method presents different advantages and disadvantages. Therefore, a combination of several of these techniques are in general applied in these studies. More traditional approaches are based on the identification of sequences within the genomic clones which are conserved across different species (Monaco, 1986). Disadvantages of this method are the failure to detect genes that have diverged between species and the amount of work involved. A second method relies on the identification of CpG islands, which are often associated with genes (Bird, 1986; Bird, 1987;

Rommens *et al.*, 1989). These are small, hypomethylated CpG-rich regions, normally containing a cluster of recognition sites for rare-cutter restriction endonucleases. Lack of methylation at these sites results in cleavage in both the cloned fragment and the genomic DNA, following digestion with the appropriate restriction enzymes. However, not all genes are associated with CpG islands, and these will be missed. Moreover, even when present, CpG islands may be located at a considerable distance from the transcribed region.

Other methods are based on the hybridisation of the genomic clones to cDNA libraries. Limiting factors in these studies are the possibility that the gene is expressed in a tissue-specific manner and therefore be present in some cDNA libraries but not in others, and the quality of the cDNA library itself. This latter feature is concerned with the representation of all the mRNAs present in the tissue used to construct the library. Pooling various cDNA libraries is often carried out in order to minimise the possibility of a given mRNA not being represented in the cDNA sample. Complex probes such as cosmid, P1, YAC clones, can be used to screen cDNA libraries (Elvin *et al.*, 1990; Albertsen *et al.*, 1994). However, problems arise due to the presence of repetitive sequences. Use of YAC inserts as probes has been shown to be particularly troublesome, yielding poor signals and a high background, and a considerable number of false positives. An alternative approach is known as hybrid selection (or cDNA selection). This has been reported to be a very efficient method for the identification of genes contained in genomic clones (Hattier *et al.*, 1995; Harshman *et al.*, 1995). The clones are immobilised on nylon filters and hybridised, following suppression of repetitive sequences, to cDNA inserts amplified from a cDNA library with vector primers (Parimoo *et al.*, 1991; Lovett *et al.*, 1991). Unbound cDNAs are eliminated and the hybridisation step is repeated 2 or 3 times. The cDNAs which are bound to the array of genomic clones following the last hybridisation step are eluted. These are then amplified with the same vector primers previously used to amplify the cDNA inserts, and cloned. This approach is simpler and less labour intensive than direct screening of cDNA libraries. Manipulations of large genomic clones such as YACs are not required. An improvement in the efficiency of this method has been obtained by carrying out the hybridisation step in solution, following biotinylation of the genomic DNA (Korn *et al.*, 1992). A problem with hybrid selection is that pseudogenes that have not deviated extensively from the active gene also select cDNAs.

Other methods for the identification of expressed sequences are exon amplification (or exon trapping), and screening for splice sites. Exon trapping consists in the cloning of genomic fragments into appropriate vectors carrying either a splice donor site (Duyk *et al.*, 1990) or both a donor and an acceptor site (Buckler *et al.*, 1991; Hamaguchi *et al.*, 1992). Presence of an exon in the cloned genomic DNA results in the splicing of this fragment into mature mRNA following expression in cell lines. Reverse transcription-PCR (RT-PCR) is subsequently carried out on the mRNA sample. The presence of possible exons is recognised by the sizes of the PCR products. The main disadvantages of this approach are the greater complexity of this method, the rate of false positives obtained due to the activation of cryptic splice sites, and the failure to detect genes that do not contain introns. Furthermore, some splicing events are tissue-specific or temporally regulated, and may not occur in the cell lines being used. Screening for splice sites (Melmer and Buchwald, 1992) can be carried out by hybridisation to degenerate oligonucleotides corresponding to splice site consensus sequences. This method is simple and inexpensive and appears to be highly efficient.

An additional procedure towards gene identification is the sequencing of genomic DNA. Coding sequences can then be sought for, based on developed computer algorithms (Fickett, 1996). Sequence comparison through the available databases may reveal sequence homology between species, pointing to the presence of an exon. Analysis of a large region involves the generation of huge amounts of sequence data. Also, a sequence window (ca. 99 bp) is required for detection. Very small exons may thus be missed; however, it is unlikely that all exons of a gene will pass undetected.

Genomic fragments containing putative expressed sequences are normally used to probe Northern blots of mRNA, in order to confirm and analyse expression. These studies precede and/or complement the screening of cDNA libraries.

8.2.3 Identification of the agu mutation.

Searches for the *agu* mutation will be conducted on the genes identified in the physical map, by comparing the sequences in the AS and in the AS/AGU strains. Priority will be given to those genes which identify cDNA molecules in cDNA libraries prepared from total brain or dissected SN, and also to genes which show homology to

sequences with a known function which could be putatively related to the *agu* phenotype.

Initial analyses could focus on identified cDNAs. Alternatively, genomic inserts could be directly sequenced. Isolation of full-length cDNAs may be difficult. A strategy for the production of full-length cDNA clones has been devised by Frohman *et al.* (1988), termed 'rapid amplification of cDNA ends' (RACE). This technique is based on the amplification of the region between a known short sequence in the particular cDNA molecule and its unknown 3' or 5' end.

There are various methods to detect differences in the sequence of DNA fragments. Some of these are RFLP, SSCP and DGGE analyses, which have already been described in Chapter 1 (section 1.3). The sensitivity of the latter two methods decreases with increasing fragment length. Most exons in the mammalian genomes are less than 300 bp in length, which allows their direct analysis. Longer sequences can be analysed by multiplexing several PCR reactions and separating all the resulting products on the same gel (Nigro *et al.*, 1995; Özçelik and Andrulis, 1995; Traystman *et al.*, 1990). Alternatively, Tawata *et al.* (1996) have reported the successful identification of single base changes in PCR products of up to ca. 16 kb in length, by carrying out a long-range PCR and digesting the amplification product prior to SSCP analysis.

Two other methods that detect sequence differences between DNA fragments are thermal gel gradient electrophoresis (TGGE), similar to DGGE except that the DNA is denatured across a temperature gradient (Fischer and Lerman, 1983), and heteroduplex analysis (HA), based on the lower mobility on polyacrylamide of heteroduplex DNA , formed by the annealing of a normal and a mutant DNA strand, compared to that of homoduplex DNA (Delwart *et al.*, 1993).

Presence of mismatches in heteroduplex DNA can also be detected by selective cleavage of the mismatched base pairs following reactivity with appropriate chemicals (chemical cleavage of mismatch, CCM) or enzymes (enzyme mismatch cleavage, EMC). A limited success was initially reported for these approaches. However, new variants of these methods are achieving increasingly higher levels of sensitivity. An example is the mapping of mutations using T4 endonuclease VII (Youil *et al.*, 1995). In the work reported by Youil *et al.* (1995), this endonuclease recognized 17 out of 18

single bp mutations. The mismatch A•A/T•T was not detected; in addition, some nonspecific nuclease activity of the enzyme was observed.

DNA sequencing is the most direct way to detect mutations. Availability of fluorescence-based automated sequencers allows the rapid analysis of long stretches of DNA. This is, however, an expensive approach when attempting to sequence large DNA regions. Consequently, DNA sequencing is often used in combination with some of the methods described above. Notwithstanding, DNA sequencing is always the ultimate step to define a mutation once a difference in the DNA sequence has been detected by any given method.

Should a single nucleotide change be consistently detected when analysing a certain DNA fragment from both AS and AS/AGU animals, various features of this variation in the sequence could add supportive evidence of its putative identity as the *agu* mutation. Were this base change found in the coding regions, this must be confirmed as a non-silent mutation, that is, the nucleotide change results in a new codon, corresponding to a different amino acid from that found in the normal protein, or to a translation termination signal. Alternatively, the mutation could occur in either a splice site or in a regulatory region.

Final confirmation of the putative identity of a gene as *agu* would ultimately come from targeted gene experiments. Successful gene replacement in the rat has not been reported, although further developments in transgenic technology could make this a feasible approach (Campbell *et al.*, 1996b; Clouthier *et al.*, 1996). Alternatively, gene targeting could be carried out in mouse, and the effects on the phenotype assessed.

8.3 Final remarks.

8.3.1 The agu gene and the molecular basis of movement disorder.

The AS/AGU strain of rat is likely to be a good model for certain movement disorders. Neurochemical and histological data, together with the general improvement of the condition in these animals following L-dopa administration or fetal midbrain transplants, further suggest a particular role of this mutant as a genetic model for

Parkinson's disease. This would constitute an important finding, given the scarcity of genetic animal models for this pathology.

A genetic etiology in Parkinson's disease has not been favoured until recently, but has now been accepted for a subset of cases. Clear evidence comes from studies of multicase families. Genetic factors have already been proposed and even characterised (Harris *et al.*, 1995; Polymeropoulos *et al.*, 1996, 1997; Wilhelmsen *et al.*, 1997). Characterisation of the *agu* gene is expected to shed new light on the mechanisms involved in dopaminergic deficiency in the basal ganglia. Comparative mapping analysis suggests that the human homologue of this gene resides on chromosome 19. Correlation of PD cases to this chromosome has not been reported to date.

A series of different factors have been proposed in the pathogenesis of PD, such as deficient antioxidative mechanisms. These abnormalities can lead to disruption of the cell membrane and intracellular signalling pathways, resulting in cell death. Polymeropoulos *et al.* (1997) recently identified a mutation in the α-synuclein gene, in affected individuals in a large Italian kindred. This genes codes for a presynaptic nerve terminal protein thought to be involved in neuronal plasticity. Interestingly, the reduction in dopamine levels recorded in AS/AGU animals, compared to normal AS rats, is much smaller in the gross analyses than in microdialysis studies of the extracellular fluid of the striatum in these animals. This suggests that the *agu* mutation is associated with a defect in the functional release of dopamine in the striatum.

8.3.2 Future development of the rat genetic map.

Positional cloning of the *agu* gene will contribute to the development of the rat genetic map. New genetic markers have already been isolated. Additionally, the large size of the backcrosses being used in this project ensures reliable estimates of the genetic distance between loci. These are thus a suitable substrate for the construction of genetic maps of the rat genome which integrate markers already available from different sources and newly established loci. Similar approaches are already being carried out (Courvoisier *et al.*, 1997; Bihoreau *et al.*, 1997). Panels of backcross animals carrying a crossover between any two adjacent loci could be established, as has been done with the markers *R191* and *D1Gu10*, flanking the *agu* locus. Typing of the appropriate panel with any new internal marker would allow the mapping of this locus

with a high level of resolution, eliminating the need to screen very large numbers of animals.

The establishment of a high-resolution genetic map of the rat genome is essential for the full exploitation of this animal as a genetic model. Although traditionally used in anatomical and physiological studies, the rat has become a rich source of models for polygenic and common multifactorial traits. A wide array of inbred strains of rat is currently available (Greenhouse *et al.*, 1990). Rat models exist for hypertension (Hilbert *et al.*, 1991; Jacob *et al.*, 1991; Gu *et al.*, 1996), diabetes (Gauguier *et al.*, 1996; Galli *et al.*, 1996), immunological responses (Ben-Nun *et al.*, 1982), various infectious diseases, drug and chemical addiction, renal failure (Brown *et al.*, 1996), and cancer (Hsu *et al.*, 1994). Moreover, in some cases the laboratory rat has shown to be a more appropriate model than, for example, the mouse, traditionally favoured by mammalian geneticists. Very clear examples are found in the role of rat models in hypertension studies as well as in organ transplantation experiments.

Sequencing of short cDNA fragments and their mapping on the genome greatly facilitates gene identification. These sequences are known as 'expressed sequence tags' or ESTs (Nelson, 1995). A considerable effort is being directed towards the development of ESTs in the human (Adams *et al.*, 1992) and the mouse genomes (Davies *et al.*, 1994b; Stewart *et al.*, 1997). It is logical to assume that the increasing interest in the rat as a genetic model will also lead to the establishment of rat ESTs. This will, in turn, enhance the potential of the rat genetic map to contribute to the further development of the mouse and the human maps, through comparative mapping analyses (Eppig and Nadeau, 1995). Development of a large array of ESTs located throughout the genome will provide a ready source of candidate genes in regions suspected of carrying a disease-gene. Identification and characterisation of genes in the *agu*-containing region will contribute new ESTs to the rat map.

APPENDIX A.

Rat genetic maps.

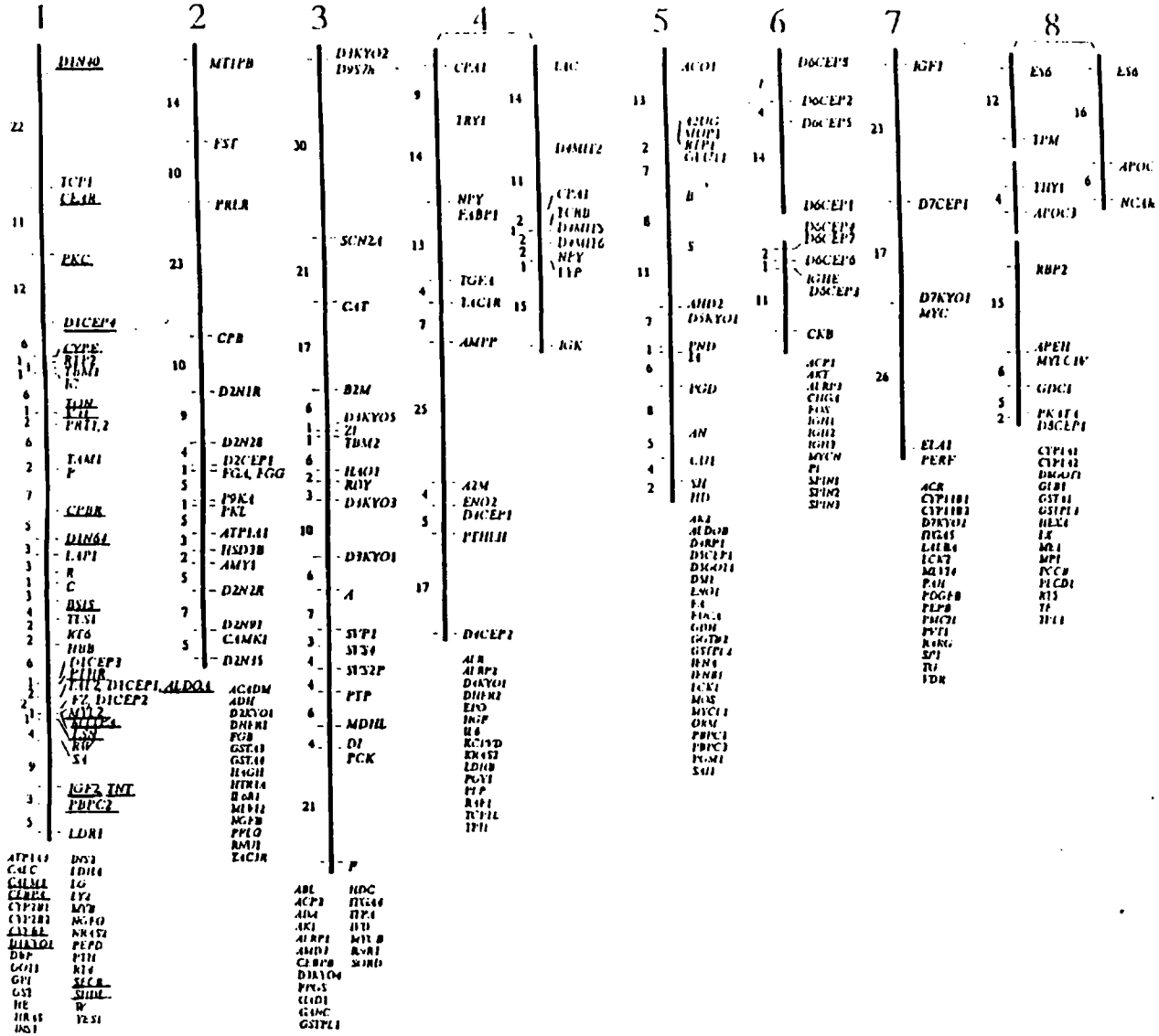


FIGURE A-1. Rat genetic linkage map taken from Yamada *et al.* (1994).
Estimated distances (in cM) between adjacent loci are indicated to the left of the chromosomes. Genes only assigned to specific chromosomes are shown under the respective chromosomes. Chromosome 1 microsatellite-containing loci are underlined.

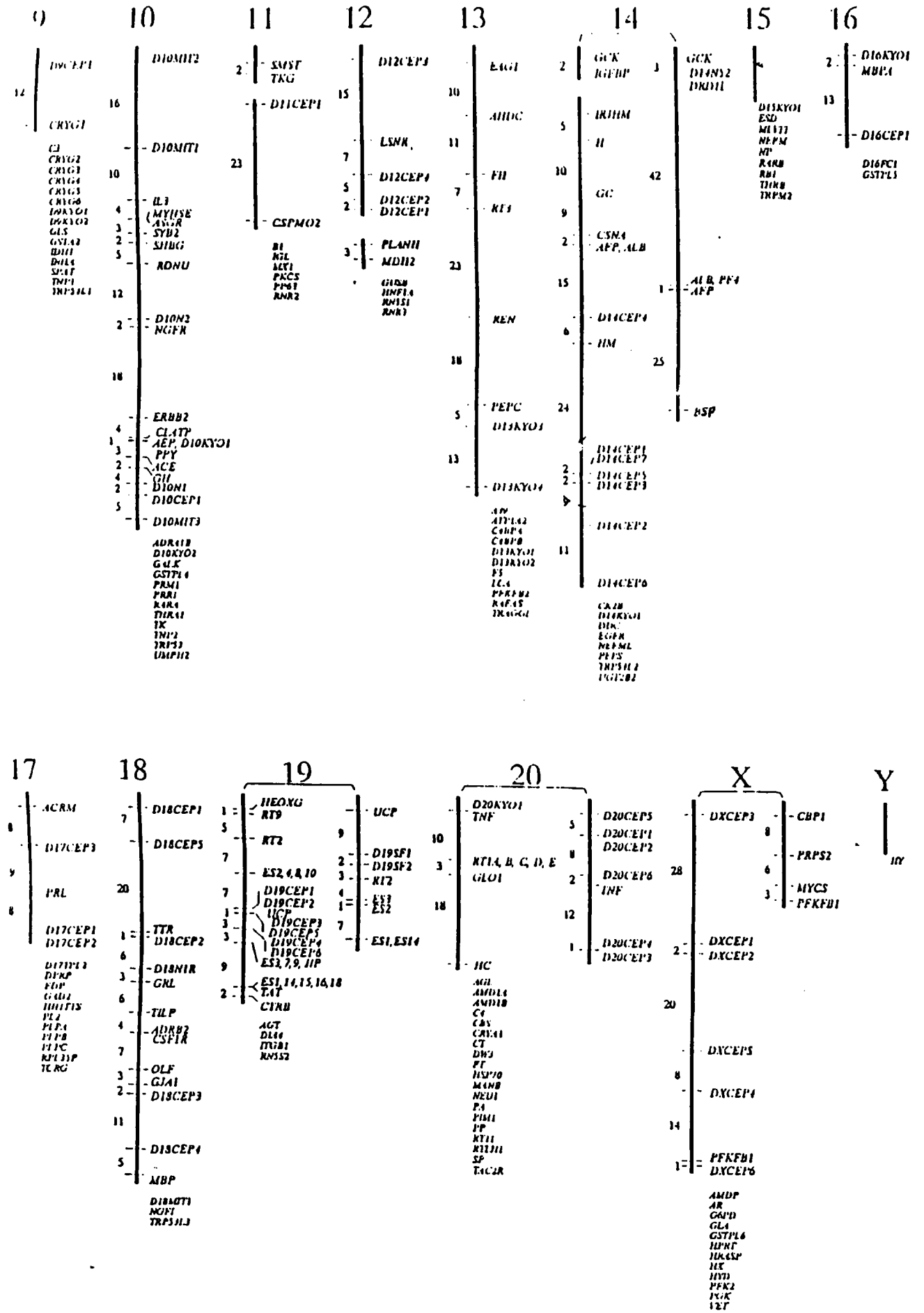


FIGURE A-1. Continued.

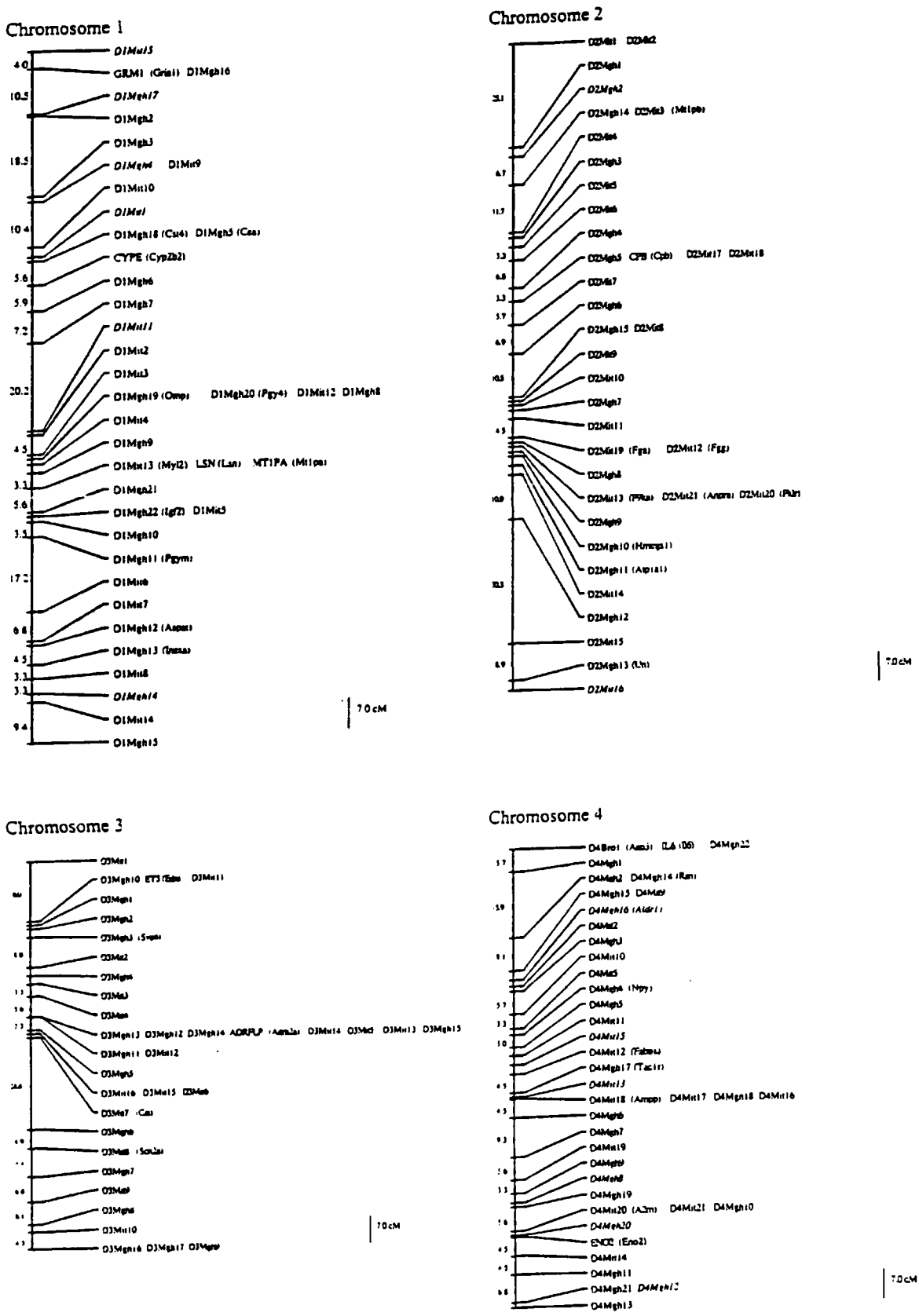
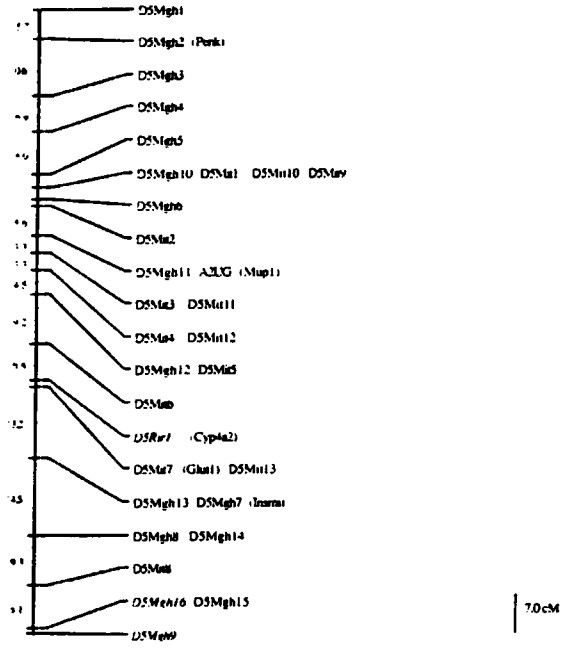
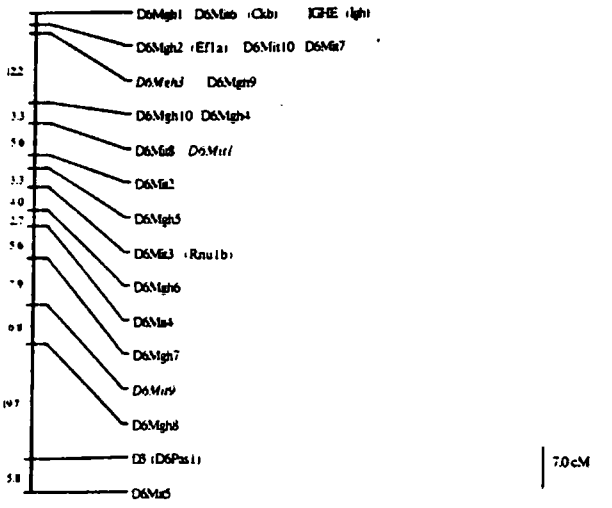


FIGURE A-2. Rat genetic linkage map taken from Jacob *et al.* (1995).
A total number of 430 SSLPs have been integrated into 21 linkage groups, each corresponding to one of 21 rat chromosomes. Gene names are shown in parentheses. Markers placed on the same horizontal line did not show recombination in any of the 46 animals studied. The numbers to the left of the chromosomes are estimated distances (in cM) between adjacent loci. The orientation of the chromosomes is not known except for chromosomes 1 and 2, where the short arms are shown at the top.

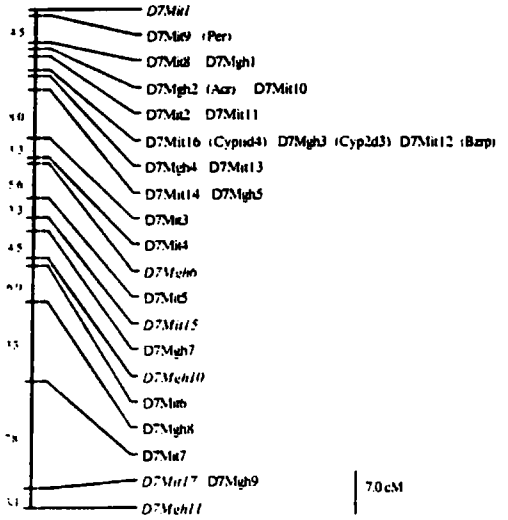
Chromosome 5



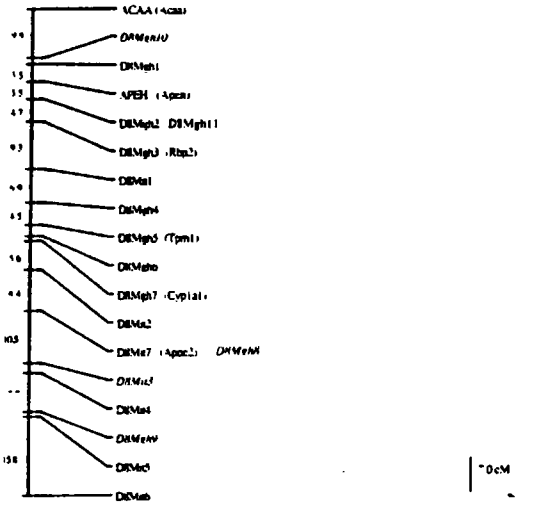
Chromosome 6



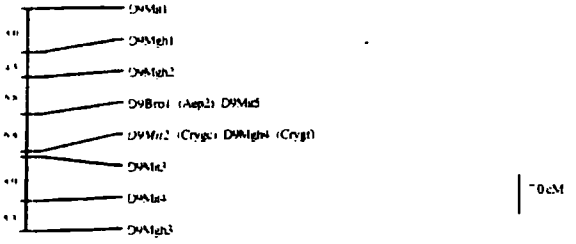
Chromosome 7



Chromosome 8



Chromosome 9



Chromosome 10

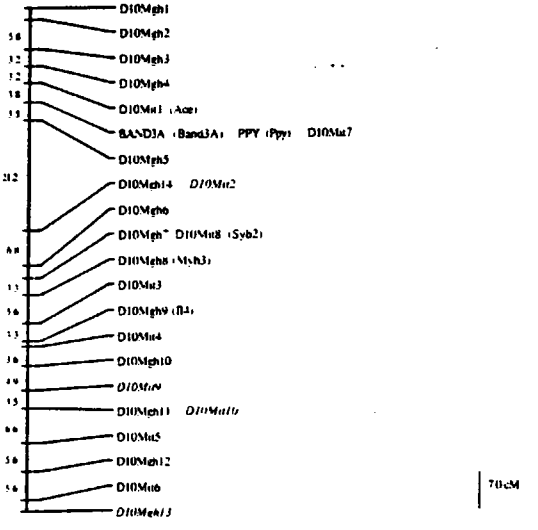
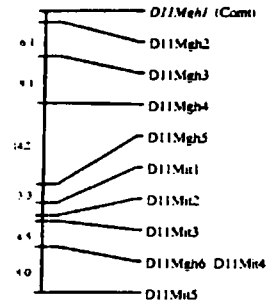
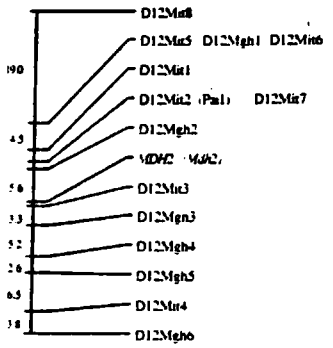


FIGURE A-2. Continued.

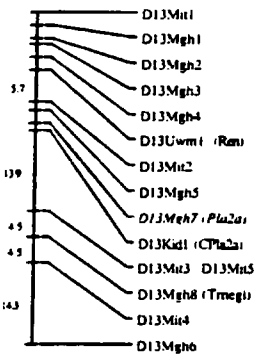
Chromosome 11



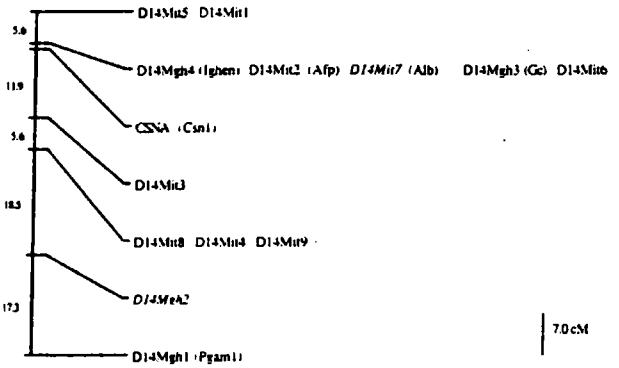
Chromosome 12



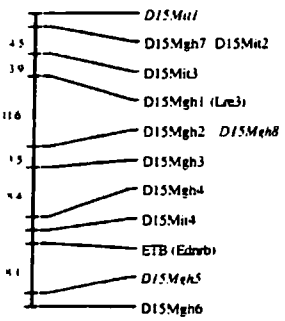
Chromosome 13



Chromosome 14



Chromosome 15



Chromosome 16

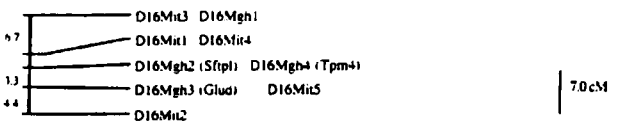
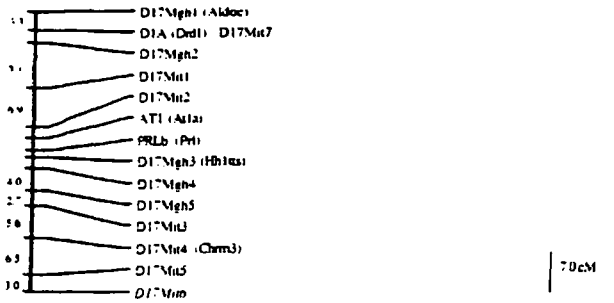
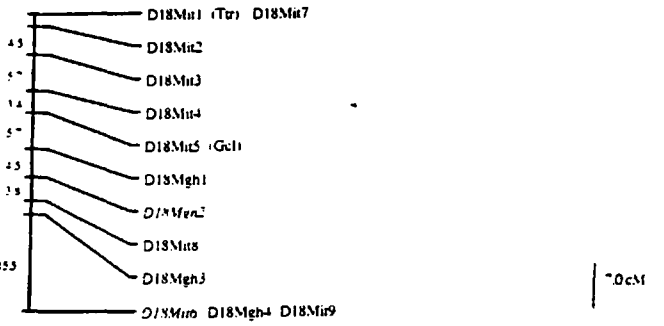


FIGURE A-2. Continued.

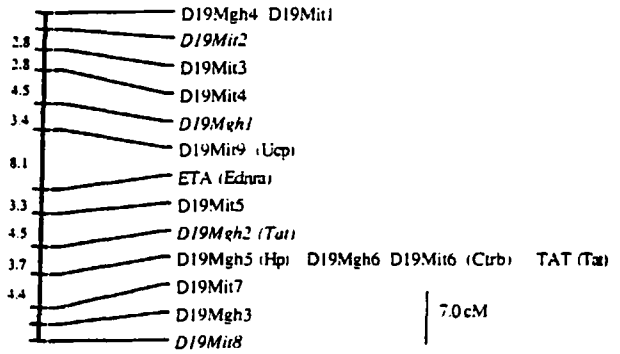
Chromosome 17



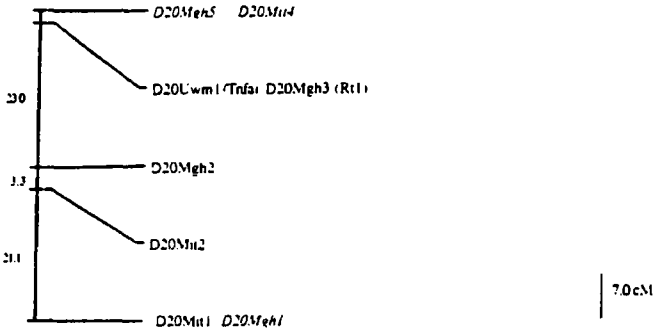
Chromosome 18



Chromosome 19



Chromosome 20



Chromosome X

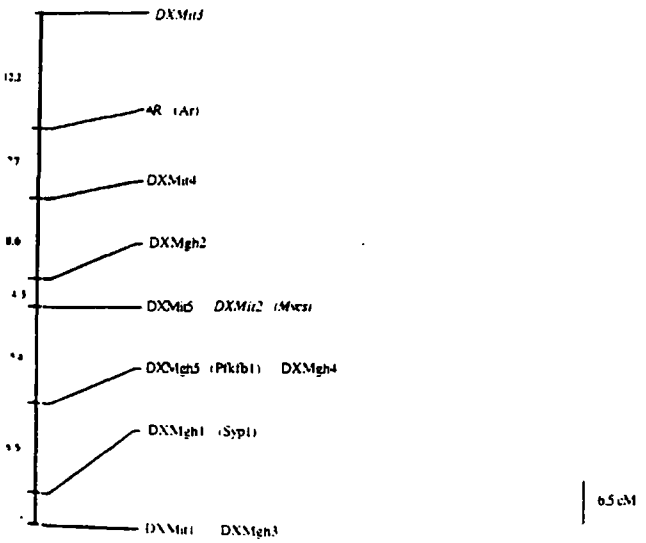


FIGURE A-2. Continued.

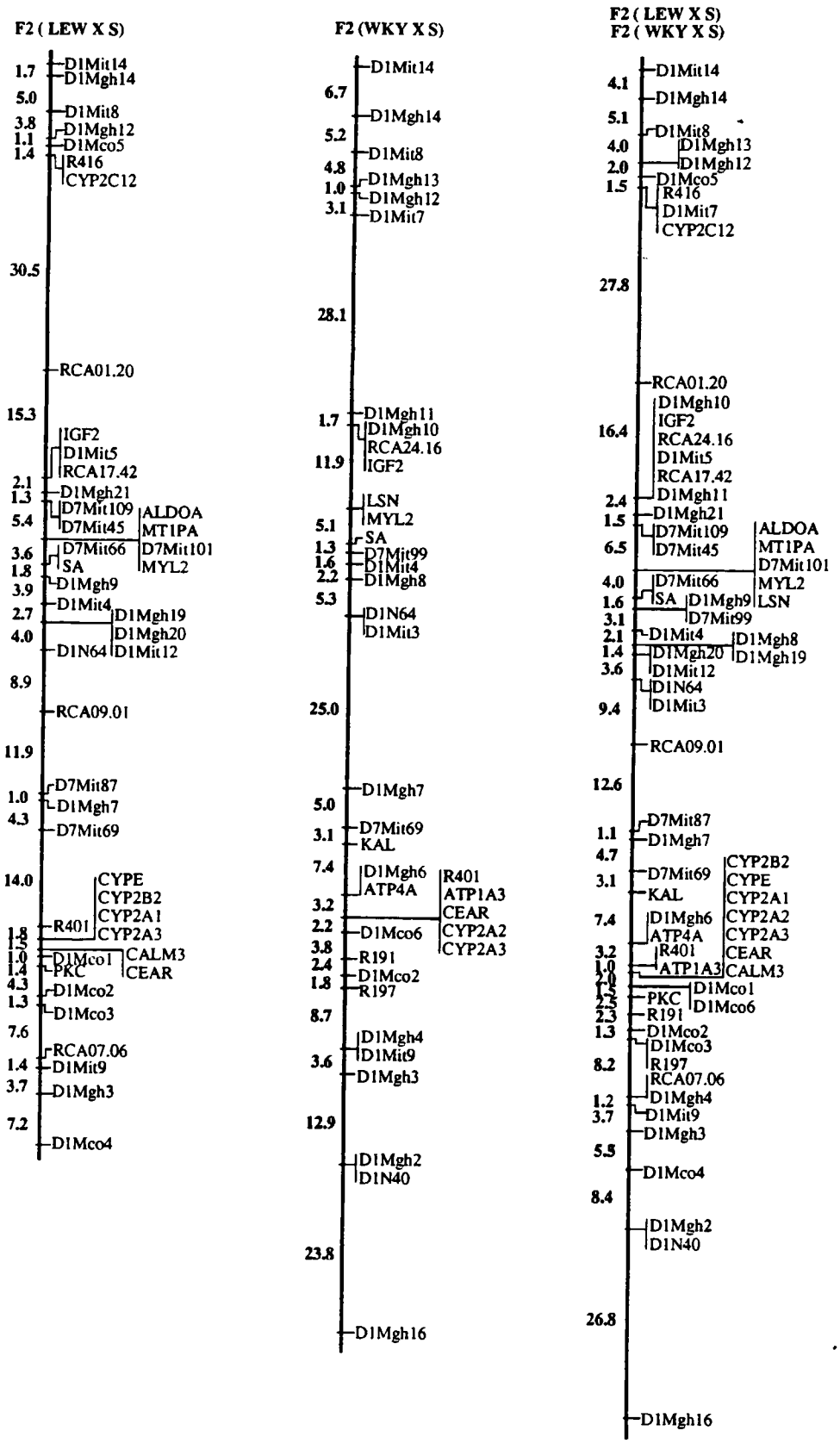


FIGURE A-3. Linkage maps for rat chromosome 1 taken from Gu *et al.* (1996). The map on the left-hand side was constructed with the data from 151 F2 (LEW x S) rats, and that on the centre with data obtained from 159 F2 (WYK x S) rats. The map on the right-hand side is the result of combining both sets of data. Distances (in cM) between adjacent loci are indicated to the left of the chromosome.

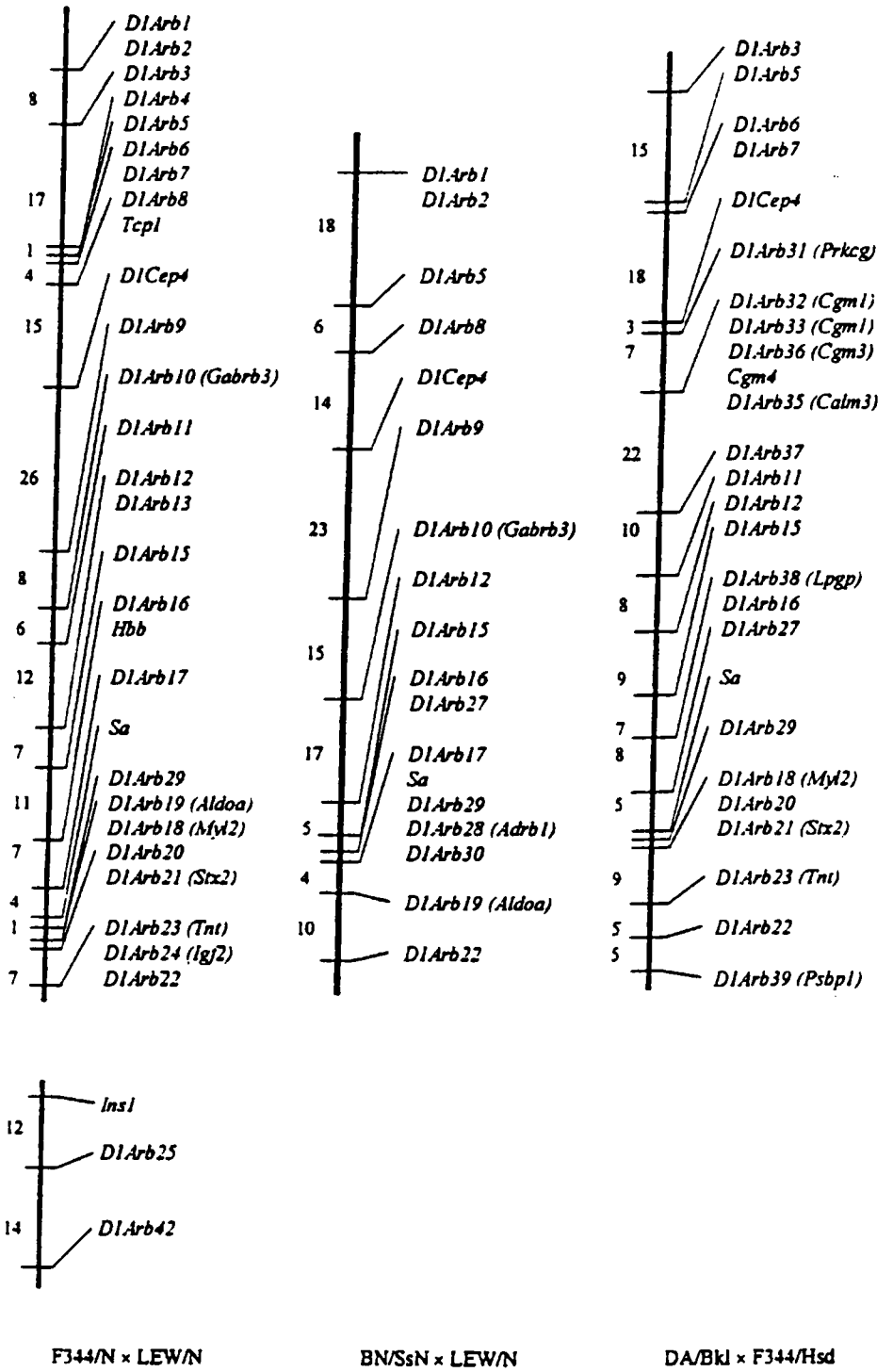


FIGURE A-4. Linkage maps for rat chromosome 1, taken from Ding *et al.* (1996).

From left to right: Linkage maps based on the typing of 40 F2 progeny from the rat crosses F344/N x LEW/N, BN/SsN x LEW/N, and DA/Bkl x F344/Hsd, respectively. Two linkage groups were established with the first cross, which covered 191 cM. Distances between adajacent loci are given in percentage of recombination.

APPENDIX B.

Optimized protocols for typing 75 microsatellite loci in AS, PVG, F344, and BN rats.

REFERENCES

Adams, M., Baverstock, P.R., Watts, C.H.S. and Gutman, G.A. (1984). Enzyme markers in inbred rat strains: Genetics of new markers and strain profiles. *Biochem. Genet.* **22**: 611-629.

Adams, M. and van Zutphen, B. (1990). Biochemical markers. In *Genetic monitoring of inbred strains of rats* (ed. Hedrich, H.J.) (Gustav Fischer Verlag, Stuttgart).

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992). Sequence identification of 2375 human brain genes. *Nature* **355**: 632-634.

Albertsen, H.M. *et al.* (1994). A physical map and candidate genes in the *BRCA1* region on chromosome 17q12-21. *Nature Genet.* **7**: 472-479.

Alexander, G.E. and Crutcher, M.D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci.* **13**: 266-271.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.

Antonarakis, S.E. and Kazazian Jr., H.H. (1988). The molecular basis of hemophilia A in man. *Trends Genet.* **4**: 233-237.

Arenas, O., Wagener, D.K., Kunz, H.W., Gill, T.J. III. (1981). Comparison of the haplotypes of the major histocompatibility complex in the rat. V. A serological, genetic and numerical study of the RT1.A region antigens. *J. Immunogenet.* **8**: 307-314.

Ashworth, L.K., Batzer, M.A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J.A., Gordon, L.A., Lamerdin, J.E., Lennon, G., Mohrenweiser, H., Olsen, A.S., Slezak, T and Carrano, A.V. (1995). An integrated metric physical map of human chromosome 19. *Nature Genet.* **11**: 422-427.

Bailey, M.E.S., Moxley III, R.T. and Johnson, K.J. (1997). Genetic basis of human neuronal disease. In *Molecular biology of the neuron* (eds. Davies, R.W. and Morris, B.J.) (BIOS Sientific Publishers Ltd., Oxford).

Baker, H., Joh, T.H. and Reis, D.J. (1980). Genetic control of number of midbrain dopaminergic neurons in inbred strains of mice: Relationship to size and neuronal density of the striatum. *Proc. Natl. Acad. Sci. USA* **77**: 4369-4373.

Baker, H., Joh, T.H., Ruggiero, D.A. and Reis, D.J. (1983). Variations in number of dopamine neurons and tyrosine hydroxylase activity in hypothalamus of two mouse strains. *J. Neurosci.* **3**: 832-843.

Bandmann, O., Davis, M.B., Marsden, C.D. and Wood, N.W. (1996). The human homologue of the weaver mouse gene in familial and sporadic Parkinson's disease. *Neuroscience* **72**: 877-879.

Baverstock, P.R., Adams, M., Maxson, L.R. and Yosida, T.H. (1983). Genetic differentiation among karyotypic forms of the black rat, *Rattus rattus*. *Genetics* **105**: 969-983.

Beckmann, J.S. and Weber, J.L. (1992). Survey of human and rat microsatellites. *Genomics* **12**: 627-631.

Benfield, P.A., Zivin, R.A., Miller, L.S., Sowder, R., Smythers, G.W., Henderson, L., Oroszlan, S. and Pearson, M.L. (1984). Isolation and sequence analysis of cDNA clones coding for rat skeletal muscle creatine kinase. *J. Biol. Chem.* **259**: 14979-14984.

Benfield, P.A., Henderson, L. and Pearson, M.L. (1985). Expression of a rat brain creatine kinase-beta-galactosidase fusion protein in *Escherichia coli* and derivation of the complete amino acid sequence of rat brain creatine kinase. *Gene* **39**: 263-267.

Ben-Nun, A., Eisenstein, S. and Cohen, I.R. (1982). Experimental autoimmune encephalomyelitis (EAE) in genetically resistant rats: PVG rats resist active induction of EAE but are susceptible to and can generate EAE effector T cell lines. *J. Immunol.* **129**: 918-919.

Bihoreau, M.T., Gauguier, D., Kato, N., Hyne, G., Lindpaintner, K., Rapp, J.P., James, M.R. and Lathrop, G.M. (1997). A linkage map of the rat genome derived from three F_2 crosses. *Genome Res.* **7**: 434-440.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209-213.

Bird, A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**: 342-347.

Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513-1523.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314-331.

Brilliant, M.H., Williams, R.W., Holdener, B.C. and Angel, J.M. (1996). Mouse chromosome 7. *Mamm. Genome* **6**: S135-S150.

Brooks, D.J. (1991). Detection of preclinical Parkinson's disease with PET. *Neurology* **41** (suppl. 2): 24-27.

Brown, P.O. (1994). Genome scanning methods. *Curr. Opin. Genet. Develop.* **4**: 366-373.

Brown, M.D., Voljavec, A.S., Lott, M.T., Torroni, A., Yang, C.C. and Wallace, D.C. (1992). Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. *Genetics* **130**: 163-173.

Brown, D.M., Provoost, A.P., Daly, M.J., Lander, E.S. and Jacob, H.J. (1996). Renal disease susceptibility and hypertension are under independent genetic control in the fawn-hooded rat. *Nature Genetics* **12**: 44-51.

Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. and Housman, D.E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88**: 4005-4009.

Bull, P.C., Thomas, G.R., Rommens, J.M., Forbes, J.R. and Cox, D.W. (1993). The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. *Nature Genet.* **5**: 327-337.

Burke, D.T., Carle, G.F. and Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**: 806-812.

Burn, D.J., Mark, M.H., Playford, E.D., Maraganore, D.M., Zimmerman, Jr., T.R., Duvoisin, R.C., Harding, A.E., Marsden, C.D., and Brooks, D.J. (1992). Parkinson's disease in twins studied with ^{18}F-dopa and positron emission tomography. *Neurology* **42**: 1894-1900.

Burr, B. and Burr, F.A. (1991). Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. *Trends Genet.* **7**: 55-60.

Cai, L., Schalkwyk, C., Schoeberlein-Stehli, A., Zee, R.Y.L., Smith, A., Haaf, T., Georges, M., Lehrach, H. and Lindpaintner, K. (1997). Construction and characterization of a 10-genome equivalent yeast artificial chromosome library for the laboratory rat, *Rattus norvegicus*. *Genomics* **39**: 385-392.

Calabresi, P., Pisani, A., Mercuri, N.B. and Bernardi, G. (1996). The corticostriatal projection: from synaptic plasticity to dysfunctions of the basal ganglia. *Trends Neurosci.* **19**: 19-24.

Campbell, J.M., Payne, A.P., Gilmore, D.P., Byrne, J.E., Russell, D., McGadey, J., Clarke, D.J., Davies, R.W. and Sutcliffe, R.G. (1996a). Neostriatal dopamine depletion and locomotor abnormalities due to the Albino Swiss rat *agu* mutation. *Neurosci. Let.* **213**: 173-176.

Campbell, K.H.S., McWhir, J., Ritchie, W.A. and Wilmut, I. (1996b). Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **380**: 64-66.

Campbell, J.M., Payne, A.P., Gilmore, D.P., Byrne, J.E., Russell, D., Clarke, D.J., Davies, R.W. and Sutcliffe, R.G. Age changes in dopamine levels in the corpus striatum of Albino Swiss (AS) and AS/AGU mutant rats. *Neurosci. Lett.*, submitted.

Campbell, J.M., Tsang, C., Weir, J., Sutcliffe, R.G., Davies, R.W., Russell, D.A., Bennett, N.K., Gilmore, D.P. and Payne, A.P. (1997). The effects of L-dopa replacement on locomotion in the AS/AGU mutant rat. Abstract to be presented at Society for Neuroscience Conference in New Orleans, 1997.

Canzian, F., Ushijima, T., Pascale, R., Sugimura, T., Dragani, T.A. and Nagao, M. (1995). Construction of a phylogenetic tree for inbred strains of rat by arbitrarily primed polymerase chain reaction (AP-PCR). *Mamm. Genome* 6: 231-235.

Carlsson, M. and Carlsson, A. (1990). Interactions between glutamatergic and monoaminergic systems within the basal ganglia - implications for schizophrenia and Parkinson's disease. *Trends Neurosci.* **13**: 272-276.

Carlsson, A. and Fornstedt, B. (1991). Catechol metabolites in the cerebrospinal fluid as possible markers in the early diagnosis of Parkinson's disease. *Neurology* **41** (suppl. 2): 50-51.

Carvey, P.M., McRae, A., Lint, T.F., Ptak, L.R., Lo, E.S., Goetz, C.G. and Klawans, H.L. (1991). The potential use of a dopamine neuron antibody and a striatal-derived neurotrophic factor as diagnostic markers in Parkinson's disease. *Neurology* **41** (suppl. 2): 53-58.

Charlesworth, B., Sniegowski, P. and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215-220.

Chesselet, M.F. and Delfs, J.M. (1996). Basal ganglia and movement disorders: an update. *Trends Neurosci.* **19**: 417-422.

Clarke, D.J. and Payne, A.P. (1994). Neuroanatomical characterization of a new mutant rat with dopamine depletion in substantia nigra. *Eur. J. Neurosci.* 6: 885-888.

Clouthier, D.E., Avarbock, M.R., Maika, S.D., Hammer, R.E. and Brinster, R.L. (1996). Rat spermatogenesis in mouse testis. *Nature* **381**: 418-421.

Committee for a Standardized Karyotype of *Rattus norvegicus* (1973). Standard karyotype of the Norway rat, *Rattus norvegicus*. *Cytogenet. Cell Genet.* **41**: 199-205.

Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J., Miller, J.C., Dietrich, W.F., Weaver, A., Lincoln, S.E., Steen, R.G., Stein, L.D., Nadeau, J.H. and Lander, E.S. (1993). A genetic linkage map of the mouse: Current applications and future prospects. *Science* **262**: 57-66.

Cornall, R.J., Aitman, T.J., Hearne, C.M. and Todd, J.A. (1991). The generation of a library of PCR-analyzed microsatellite variants for genetic mapping of the mouse genome. *Genomics* **10**: 874-881.

Courvoisier, H., Bihoreau, M.T., Gauguier, D., Plomion, C., Mormede, P. and Moisan, M.P. (1997). Integrated genetic mapping of 64 rat microsatellite markers from different sources. *Mamm. Genome* **8**: 282-283.

Cox, J.L. and Shaw, P.A. (1992). Structure, organization and regulation of a rat cysteine proteinase inhibitor-encoding gene. *Gene* **110**: 175-180.

Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J.M. and White, R.L. (1990). Collaborative genetic mapping of the human genome. *Genomics* **6**: 575-577.

Davies, J.L., Kawaguchi, Y., Bennett, S.T., Copeman, J.B., Cordell, H.J., Pritchard, L.E., Reed, P.W., Gough, S.C.L., Jenkins, S.C., Palmer, S.M., Balfour, K.M., Rowe, B.R., Farrall, M., Barnett, A.H., Bain,S.C. and Todd, J.A. (1994a). A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* **371**: 130-136.

Davies, R.W., Roberts, A.B., Morris, A.J., Griffith, G.W., Jerecic, J., Ghandi, S., Kaiser, K. and Savioz, A. (1994b). Enhanced access to rare brain cDNAs by prescreening libraries: 207 new mouse brain ESTs. *Genomics* **24**: 456-463.

DeLong, M.R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends Neurosci.* **13**: 281-285.

Delwart, E.L., Shpaer, E.G., Louwagie, J., McCutchan, F.E., Grez, M., Rubsamen-Waigmann, H. and Mullins, J.I. (1993). Genetic relatioships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 *env* genes. *Science* **262**: 1257-1261.

Dietrich, W., Katz, H., Lincoln, S.E., Shin, H-S., Friedman, J., Dracopoli, N.C. and Lander, E.S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**: 423-447.

Dietrich, W.F., Miller, J.C., Steen, R.G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D.C., Wessel, M., Dredge, R.D., Marquis, A., Stein, L.D., Goodman, N., Page, D.C. and Lander, E.S. (1994). A genetic map of the mouse with 4006 simple sequence length polymorphisms. *Nature Genet.* **7**: 220-245.

Ding, Y.-P., Remmers, E.F., Du, Y., Longman, R.E., Goldmuntz, E.A., Zha, H., Kotake, S., Cannon, G.W., Griffiths, M.M. and Wilder, R.L. (1996). Genetic maps of polymorphic DNA loci on rat chromosome 1. *Genomics* **36**: 320-327.

Donnis-Keller, H. et al. (1987). A genetic linkage map of the human genome. *Cell* **51**:319-337.

Du, Y., Remmers, E.F., Zha, H., Goldmuntz, E.A., Szpirer, J., Szpirer, C. and Wilder, R.L. (1995). A genetic map of microsatellite markers on rat chromosome 7. *Mammalian Genome* **6**: 295-298.

Du, Y., Remmers, E.F., Goldmuntz, E.A., Zha, H., Mathern, P., Ding, Y.-P., Kotake, S., Szpirer, J., Szpirer, C. and Wilder, R.L. (1996). Linkage maps of rat chromosomes 15, 16, 17, 19 and X. *Genomics* **32**: 113-116.

Durán Alonso, M.B., Shiels, P., McCallion, A.S., Bennett, N.K., Payne, A.P., Szpirer, J., Szpirer, C., Brodie, M.J., Davies, R.W. and Sutcliffe, R.G. (1997). The cystatin S gene maps to rat chromosome 3, to which *D1Mgh18* is re-assigned from chromosome 1. *Mamm. Genome* **8**: 946-947.

Duvoisin, R.G., Eldridge, R., Williams, A., Nutt, J. and Calne, D. (1981). Twin study of Parkinson disease. *Neurology* **31**: 77-80.

Duyk, G.M., Kim, S., Myers, R.M. and Cox, D.R. (1990). Exon trapping: A genetic screen to identifiy candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. USA* **87**: 8995-8999.

Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49**: 746-756.

Ellegren, H. (1995). Mutation rates at porcine microsatellite loci. *Mamm. Genome* **6**: 376-377.

Ellenberg, J.H. (1991). Preclinical detection in studies of the etiology, natural history, and treatment of Parkinson's disease. *Neurology* **41** (suppl. 2): 14-20.

Elvin, P., Slynn, G., Black, D., Graham, A., Butler, R., Riley, J., Anand, R. and Markham, A.F. (1990). Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res.* **18**: 3913-3917.

Eppig, J.T. and Nadeau, J.H. (1995). Comparative maps: the mammalian jigsaw puzzle. *Curr. Opin. Genet. Develop.* **5**: 709-716.

Fahn, S. (1997). An overview of novel therapies for Parkinson's disease. *Exp. Neurol.* **144**: 21-23.

Farrer, L.A., Myers, R.H., Connor, L., Cupples, L.A. and Growdon, J.H. (1991). Segregation analysis reveals evidence of a major gene for Alzheimer disease. *Am. J. Hum. Genet.* **48**: 1026-1033.

Faucheux, B.A., Hauw, J.-J., Agid, Y. and Hirsch, E.C. (1997). The density of [^{125}I]-transferrin binding sites on perikarya of melanized neurons of the substantia nigra is decreased in Parkinson's disease. *Brain Res.* 749: 170-174.

Feil, R., Aubourg, P., Heilig, R. and Mandel, J.L. (1990a). A 195 kb cosmid walk encompassing the human Xq28 color vision pigment genes. *Genomics* **6**: 367-373.

Feil, R., Palmieri, G., d'Urso, M., Heilig, R., Oberlé, I. and Mandel, J.L. (1990b). Physical and genetic mapping of polymorphic loci in Xq28 (DXS15, DXS52, and DXS134): Analysis of a cosmid clone and a yeast artificial chromosome. *Am. J. Hum. Genet.* **46**: 720-728.

Fernandez, P.M. and Dujovny, M. (1997). Pallidotomy: Editorial review. *Neurol. Res.* **19**: 25-34.

Festing, M.F.W. and Bender, K. (1984). Genetic relationships between inbred strains of rats. An analysis based on genetic markers at 28 biochemical loci. *Genet. Res.* **44**: 271-281.

Fickett, J.W. (1996). Finding genes by computer: the state of the art. *Trends Genet.* **12**: 316-320.

Fischer, S.G. and Lerman, L.S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory. *Proc. Natl. Acad. Sci. USA* **80**: 1579-1583.

Flejter, W.L., Bennett-Baker, P., Barcroft, C.L., Kiousis, S. and Chamberlain, J.S. (1995). Region-specific cosmids and STRPs identified by chromosome microdissection and FISH. *Genomics* **25**: 413-420.

Fredholm, M. and Winter, A.K. (1995). Variation of short tandem repeats within and between species belonging to the *Canidae* family. *Mamm. Genome* **6**: 11-18.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**: 8998-9002.

Furtado, J.C.S. and Mazurek, M.F. (1991). MPTP-induced neurotoxicity and the quest for a preventative therapy for Parkinson's disease. *Can. J. Neurol. Sci.* **18**: 77-82.

Galli, J. Li, L.S., Glaser, A., Ostenson, C.G., Jiao, H., Fakhrai-Rad, H., Jacob, H.J., Lander, E.S. and Luthman, H. (1996). Genetic analysis of non-insulin dependent diabetes mellitus in the GK rat. *Nature Genet.* **12**: 31-37.

Gauguier, D., Froguel, P., Parent, V., Bernard, C., Bihoreau, M-T., Portha, B., James, M.R., Penicaud, L., Lathrop, M. and Ktorza, A. (1996). Chromosomal mapping of genetic loci associated with non-insulin dependent diabetes in the GK rat. *Nature Genet.* **12**: 38-43.

Gibson, T.J. (1984). "Studies on the Epstein-Barr virus genome". PhD Thesis, Cambridge University, England.

Goldberg, Y.P., Kalchman, M.A., Metzler, M., Nasir, J., Zeisler, J., Graham, R., Koide, H.B., O'Kusky, J., Sharp, A.H., Ross, C.A., Jirik, F. and Hayden, M.R. (1996). Absence of disease phenotype and intergenerational stability of the CAG repeat in transgenic mice expressing the human Huntington disease transcript. *Hum. Mol. Genet.* **5**: 177-185.

Goldmuntz, E.A., Remmers, E.F., Zha, H., Cash, J.M., Mathern, P., Crofford, L.J. and Wilder, R.L. (1993). Genetic map of 12 polymorphic loci on rat chromosome 1. *Genomics* **16**: 761-764.

Goldmuntz, E.A., Remmers, E.F., Du, Y., Zha, H., Mathern, P., Crofford, L.J. and Wilder, R.L. (1995). Genetic map of 16 polymorphic loci forming three linkage groups assigned to rat chromosome 4. *Mamm. Genome* **6**: 459-463.

Grant, S., Jessee, J., Bloom, F. and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. USA* **87**: 4645-4649.

Graybiel, A.M. (1990). Neurotransmitters and neuromodulators in the basal ganglia. *Trends Neurosci.* **13**: 244-254.

Green, E.L. (1981). Linkage, recombination and mapping. In *Genetics and probability in animal breeding experiments* (ed. Green, E.L.) (Macmillan, New York).

Greenhouse, D.D., Festing, M.F.W., Hasan, S. and Cohen, A.L. (1990). Inbred strains of rats. In *Genetic monitoring of inbred strains of rats* (ed. Hedrich, H.J.) (Gustav Fischer Verlag, Stuttgart).

Gross, D.S. and Garrard, W.T. (1986). The ubiquitous potential Z-forming sequence of eukaryotes, $(dT\text{-}dG)_n$. $(dC\text{-}dA)_n$, is not detectable in the genomes of eubacteria, archaebacteria, or mitochondria. *Mol. Cell. Biol.* **6**: 3010-3013.

Gu, L., Dene, H., Deng, A.Y., Hoebee, B., Bihoreau. M-T., James, M. and Rapp, J.P. (1996). Genetic mapping of two blood pressure quantitative trait loci on rat chromosome 1. *J. Clin. Invest.* **97**: 777-788.

Günther, E. (1990a). Immunological markers. In *Genetic monitoring of inbred strains of rats* (ed. Hedrich, H.J.) (Gustav Fischer Verlag, Stuttgart).

Günther, E. (1990b). DNA markers (Restriction Fragment Length Polymorphisms, RFLP). In *Genetic monitoring of inbred strains of rats* (ed. Hedrich, H.J.) (Gustav Fischer Verlag, Stuttgart).

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Philippe, M., Marc, S., Bernardi, g., Lathrop, M. and Weissenbach, J. (1994). The 1993-1994 Généthon human genetic linkage map. *Nature Genet.* **7**: 246-339.

Haldane, J.B.S. (1922). Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12**: 101-109.

Haldi, M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D. and Lander, E.S. (1994). Large human YACs constructed in a *rad*52 strain show a reduced rate of chimaerism. *Genomics* **24**: 478-484.

Hamada, H., Seidman, M., Howard, B.H. and Gorman, C.M. (1984). Enhanced gene expression by the poly(dT-dG) . poly(dC-dA) sequence. *Mol. Cell. Biol.* **4**: 2622-2630.

Hamaguchi, M., Sakamoto, H., Tsuruta, H., Sasaki, H., Muto, T., Sugimura, T. and Terada, M. (1992). Establishment of a highly sensitive and specific exon-trapping system. *Proc. Natl. Acad. Sci. USA* **89**: 9779-9783.

Hamvas, R.M.J., Francis, F., Cox, R.D., Nizetic, D., Goldsworthy, M.E., Brown, S.D.M. and Lehrach, H.R. (1994). Rapid restriction analysis of YAC clones. *Nucleic Acids Res.* **22**: 1318-1319.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**: 557-580.

Harrington, K.A., Augood, S.J., Kingsbury, A.E., Foster, O.J.F. and Emson, P.C. (1996). Dopamine transporter (DAT) and synaptic vesicle amine transporter (VMAT2) gene expression in the substantia nigra of control and Parkinson's disease. *Mol. Brain Res.* **36**: 157-162.

Harris, Z.L., Takahashi, Y., Miyajima, H., Serizawa, M., MacGillivray, R.T.A. and Gitlin, J.D. (1995). Aceruloplasminemia: Molecular characterization of this disorder of iron metabolism. *Proc. Natl. Acad. Sci. USA* **92**: 2539-2543.

Harshman, K. *et al.* (1995). Comparison of the positional cloning methods used to isolate the *BRCA1* gene. *Hum. Mol. Genet.* **4**: 1259-1266.

Hashimoto, L., Habita, C., Beressi, J.P., Delepine, M., Besse, C., Cambon-Thomsen, A., Deschamps, I., Rotter, J.I., Djoulah, S., James, M.R., Froguel, P., Weissenbach, J., Lathrop, G.M. and Julier, C. (1994). Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. *Nature* **371**: 161-164.

Hattier, T., Bell, R., Shaffer, D., Stone, S., Phelps, R.S., Tavtigian, S.V., Skolnick, M.H., Shattuck-Eidens, D. and Kamb, A. (1995). Monitoring the efficacy of hybrid selection during positional cloning: the search for BRCA1. *Mamm. Genome* **6**: 873-879.

Hayashi, J.-I., Yonekawa, H., Gotoh, O., Tagashira, Y., Morikawi, K. and Yosida, T.H. (1979). Evolutionary aspects of variant types of rat mitochondrial DNAs. *Biochimica et Biophysica Acta* **564**: 202-211.

Hedrich, H.J. (1990a). Morphological and physiological traits. In *Genetic monitoring of inbred strains of rats* (ed. Hedrich, H.J.) (Gustav Fischer Verlag, Stuttgart).

Hedrich, H.J. (1990b). Mutant genes and polymorphic loci of the laboratory rat. In *Genetic monitoring of inbred strains of rats* (ed. Hedrich, H.J.) (Gustav Fischer Verlag, Stuttgart).

Hedrich, H.J. (1990c). Variability of strains and sources of strain alterations. In *Genetic monitoring of inbred strains of rats* (ed. Hedrich, H.J.) (Gustav Fischer Verlag, Stuttgart).

Hilbert, P., Lindpainter, K., Beckmann, J.S., Serikawa, T., Soubrier, F., Dubay, C., Cartwright, P., De Gouyon, B., Julier, C., Takahasi, S., Vincent, M., Ganten, D., Georges, M. and Lathrop, G.M. (1991). Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. *Nature* **353**: 521-529.

Hoffman, E.P., Brown, R.H. Jr. and Kunkel, L.M. (1987). Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* **51**: 919- 928.

Holt, I.J., Miller, D.H., Harding, A.E. (1989). Genetic heterogeneity and mitochondrial DNA heteroplasmy in Leber's hereditary optic neuropathy. *J. Med. Genet.* **26**: 739-743.

Horellou, P., Sabaté, O., Buc-Caron, M.-H. and Mallet, J. (1997). Adenovirus-mediated gene transfer to the central nervous system for Parkinson's disease. *Exp. Neurol.* **144**: 131-138.

Hsu, L.-C., Kennan, W.S., Shepel, L.A., Jacob, H.J., Szpirer, C., Szpirer, J., Lander, E.S. and Gould, M.N. (1994). Genetic identification of *Mcs-1*, a rat mammary carcinoma suppressor gene. *Cancer Res.* **54**: 2765-2770.

Hughes, A.E. (1993). Optimization of microsatellite analysis for genetic typing. *Genomics* **15**: 433-434.

Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A. and de Jong, P.J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* 6: 84-89.

Isom, L.L., DeJongh, K.S., Patton, D.E., Reber, B.F.X., Offord, J., Charbonneau, H., Walsh, K., Goldin, A.L. and Catterall, W.A. (1992). Primary structure and functional expression of the beta-1 subunit of the rat brain sodium channel. *Science* **256**: 839-842.

Jacob, H.J., Lindpaintner, K., Lincoln, S.E., Kusumi, K., Bunker, R.K., Mao, Y-P., Ganten, D., Dzau, V.J. and Lander, E.S. (1991). Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* **67**: 213-224.

Jacob, H.J., Brown, D.M., Bunker, R.K., Daly, M.J., Dzau, V.J., Goodman, A., Koike, G., Kren, V., Kurtz,T., Lernmark, A., Levan, G., Mao, Y-P., Pettersson, A., Pravenec, M., Simon, J.S., Szpirer, C., Szpirer, J., Trolliet, M.R., Winer, E.S. and Lander, E.S. (1995). A genetic linkage map of the laboratory rat, Rattus norvegicus. *Nature Genet.* **9**: 63-69.

James, M.R., Richard III, C.W., Schott, J-J., Yousry, C., Clark, K., Bell, J., Terwilliger, J.D., Hazan, J., Dubay, C., Vignal, A., Agrapart, M., Imai, T., Nakamura, Y., Polymeropoulos, M., Weissenbach, J., Cox, D.R. and Lathrop, G.M. (1994). A radiation hybrid map of 506 STS markers spanning human chromosome 11. Nature Genet. **8**: 70-76.

James, M.R. and Lindpaintner, K. (1997). Why map the rat? *Trends Genet.* **13**: 171-173.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**: 67-73.

Jeffreys, A.J., Wilson, V., Kelly, R., Taylor, B.A. and Bulfield, G. (1987). Mouse DNA 'fingerprints': analysis of chromosome localization and germ-line stability of hypervariable loci in recombinant inbred strains. *Nucleic Acids Res.* **15**: 2823-2836.

Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332**: 278-281.

Jenner, P. and Olanow, C.W. (1996). Oxidative stress and the pathogenesis of Parkinson's disease. *Neurology* **47** (suppl. 3): 161-170.

Johnson, D.H. (1990). Molecular cloning of DNA from specific chromosomal regions by microdissection and sequence-independent amplification of DNA. *Genomics* **6**: 243-251.

Johnson, W.G. (1991). Genetic susceptibility to Parkinson's disease. *Neurology* **41** (suppl. 2): 82-87.

Jonsson, J.J. and Weissman, S.M. (1995). From mutation mapping to phenotype cloning. *Proc. Natl. Acad. Sci. USA* **92**: 83-85.

Kashi, Y., King, D. and Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* **13**: 74-78.

Kimmerly, W.J., Kyle, A.L., Lustre, V.M., Martin, C.H. and Palazzolo, M.J. (1994). Direct sequencing of terminal regions of genomic P1 clones. A general strategy for the design of sequence-tagged site markers. *Genet. Anal. Tech. Appl.* **11**: 117-128.

Kish, S.J., Shannak, K. and Hornykiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. *N. Engl. J. Med.* **318**: 876-880.

Koebner, R.M.D. (1995). Predigestion of DNA template improves the level of polymorphism of random amplified polymorphic DNAs in wheat. *Genet. Anal. Biomol. Eng.* **12**: 63-67.

Koller, W.C., Langston, J.W., Hubble, J.P., Irwin, I., Zack, M., Golbe, L., Forno, L., Ellenberg, J., Kurland, L., Ruttenber, A.J., Spencer, P., Tanner, C., Tetrud, J., Wilcox, T., Roman, G., Mayeux, R., Smith, M. and Goetz, C. (1991). Does a long preclinical period occur in Parkinson's disease? *Neurology* **41** (suppl. 2): 8-13.

Koller, W.C. (1997). Neuroprotective therapy for Parkinson's disease. *Exp. Neurol.* **144**: 24-28.

Kondo, Y., Mori, M., Kuramoto, T., Yamada, J., Beckmann, J.S., Simon-Chazottes, D., Montagutelli, X., Guenet, J-L. and Serikawa, T. (1993). DNA segments mapped by reciprocal use of microsatellite primers between mouse and rat. *Mamm. Genome* **4**: 571-576.

Konieczny, A. and Ausubel, F.A. (1993). A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *The Plant Journal* 4: 403-410.

Korn, B., Sedlacek, Z., Manca, A., Kioschis, P., Konecki, D., Lehrach, H. and Poustka, A. (1992). A strategy for the selection of transcribed sequences in the Xq28 region. *Hum. Mol. Genet.* **1**: 235-242.

Kozak, C., Nichols, E. and Ruddle, F.H. (1975). Gene linkage analysis in the mouse by somatic cell hybridization: assignment of adenine phosphoribosyl-transferase to chromosome 8 and α-galactosidase to the X-chromosome. *Somatic Cell Genet.* **1**: 371-382.

Kumar, R., Agarwal, A. K. and Seth, P.K. (1995). Free radical-generated neurotoxicity of 6-hydroxydopamine. *J. Neurochem.* **64**: 1703-1707.

Kuramoto, T., Mori, M., Hirayama, N., Saburi, S., Yamada, J. and Serikawa, T. (1993). A strategy for rapid construction of genetic and physical maps in the rat. *Acta Histochem. Cytochem.* **26**: 325-332.

Kurata *et al.* (1994). A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet.* **8**: 365-372.

Langston, J.W. and Koller, W.C. (1991). The next frontier in Parkinson's disease: Presymptomatic detection. *Neurology* **41** (suppl. 2): 5-7.

Langston, J.W. (1996). The etiology of Parkinson's disease with emphasis on the MPTP story. *Neurology* **47** (suppl. 3): 153-160.

Lapchak, P.A., Gash, D.M., Jiao, S., Miller, P.J. and Hilt, D. (1997). Glial cell line-derived neurotrophic factor: A novel therapeutic approach to treat motor dysfunction in Parkinson's disease. *Exp. Neurol.* **144**: 29-34.

Lawn, R.M. (1985). The molecular genetics of hemophilia: blood clotting factors VIII and IX. *Cell* **42**: 405-406.

Lazzarini, A.M., Myers, R.H., Zimmerman, Jr., T.R., Mark, M.H., Golbe, L.I., Sage, J.I., Johnson, W.G. and Duvoisin, R.C. (1994). A clinical genetic study of Parkinson's disease: Evidence for dominant transmission. *Neurology* **44**: 499-506.

Levan, G., Klinga, K., Szpirer, C. and Szpirer, J. (1990). Gene map of the rat (Rattus norvegicus). In *Genetic Maps: Locus Maps of complex genomes*, 5th edition, S.J. O'Brien, ed. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 4.80-4.87.

Levan, G., Szpirer, J., Szpirer, C., Klinga, K., Hanson, C. and Islam, M.Q. (1991). The gene map of the Norway rat (Rattus norvegicus) and comparative mapping with mouse and man. *Genomics* **10**: 699-718.

Levan, G., Klinga-Levan, K., Hanson, C., Szpirer, C., Szpirer, J. (1992). Rat gene map 1992. *Rat News Lett.* **27**: 10-34.

Levinson, G. and Gutman, G.A. (1987). High frequency of short frameshifts in poly-CA/GT tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* **15**: 5323-5338.

Levitt, R.C. (1991). Polymorphisms in the transcribed 3' untranslated region of eukaryotic genes. *Genomics* **11**: 484-489.

Levitt, R.C., Kiser, M.B., Dragwa, C., Jedlicka, A.E., Xu, J., Meyers, D. and Hudson, J.R. (1994). Fluorescence-based resource for semiautomated genomic analyses using microsatellite markers. *Genomics* **24**: 361-365.

Li, Y., Togashi, Y., Sato, S., Emoto, T., Kang, J.-H., Takeichi, N., Kobayashi, H., Kojima, Y., Une, Y. and Uchino, J. (1991). Spontaneous hepatic copper accumulation in Long-Evans Cinnamon rats with hereditary hepatitis: a model of Wilson's disease. *J. Clin. Invest.* **87**: 1858-1861.

Lindahl, G., Gersdorf, E., Menzel, H.J., Seed, M., Humphries, S. and Utermann, G. (1990). Variation in the size of human apolipoprotein (a) is due to a hypervariable region in the gene. *Hum. Genet.* **84**: 563-567.

Lindsey, J.R. (1979). Historical foundations in the laboratory rat. In *The laboratory rat* (eds. Baker, H.J., Lindsey, J.R. and Weisbroth, S.H.) 1-36 (Academic Press, New York).

Lisitsyn, N., Lisitsyn, N. and Wigler, M. (1993). Cloning the difference between two complex genomes. *Science* **259**: 946-951.

Lisitsyn, N.A., Segre, J.A., Kusumi, K., Lisitsyn, N.M., Nadeau, J.H., Frankel, W.N., Wigler, M.H. and Lander, E.S. (1994). Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. *Nature Genet.* **6**: 57-63.

Lisitsyn, N.A. (1995). Representational difference analysis: finding the differences between genomes. *Trends Genet.* **11**: 303-307.

Litt, M. and Luty, J.A. (1989). A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**: 397-401.

Liu, L., Xu, W., Harrington, K.A. and Emson, P.C. (1994). The molecular cloning and expression of a human synaptic vesicle amine transporter that suppresses MPP^+ toxicity. *Mol. Brain Res.* **25**: 90-96.

Lord, C.J., Bohlander, S.K., Hopes, E.A., Montague, C.T., Hill, N.J., Prins, J.-B., Renjilian, R.J., Peterson, L.B., Wicker, L.S., Todd, J.A. and Denny, P. (1995). Mapping the diabetes polygene *Idd3* on mouse chromosome 3 by use of novel congenic strains. *Mamm. Genome* **6**: 563-570.

Löscher, W., Richter, A., Nikkhah, G., Rosenthal, C., Ebert, U. and Hedrich, H.J. (1996). Behavioural and neurochemical dysfunction in the circling (*ci*) rat: A novel genetic animal model of a movement disorder. *Neurosci.* **74**: 1135-1142.

Love, J.M., Knight, A.M., McAleer, M.A. and Todd, J.A. (1990). Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Res.* **18**: 4123-4130.

Lovett, M., Kere, J. and Hinton, L.M. (1991). Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88**: 9628-9632.

Luty, J.A., Guo, Z., Willard, H.F., Ledbetter, D.H., Ledbetter, S. and Litt, M. (1990). Five polymorphic microsatellite VNTRs on the human X chromosome. *Am. J. Hum. Genet.* **46**: 776-783.

Maihara, T., Andoh, Y., Yokoi, N., Kuramoto, T. and Serikawa, T. (1995). Fifty-six new microsatellite markers in the rat genetic linkage map. *Transplantation Proceedings* **27**: 1502-1504.

McGeer, P.L., Itagaki, S., Akiyama, H. and McGeer, E.G. (1988). Rate of cell death in parkinsonism indicates active neuropathological process. *Ann. Neurol.* **24**: 574-476.

Melmer, G. and Buchwald, M. (1992). Identification of genes using oligonucleotides corresponding to splice site consensus sequences. *Hum. Mol. Genet.* **1**: 433-438.

Meltzer, P.S., Guan, X-Y., Burgess, A. and Trent, J.M. (1992). Rapid generation of region specific probes by chromosome microdissection and their application. *Nature Genet.* **1**: 24-28.

Meunier, J.R. and Grimont, P.A. (1993). Factors affecting reproducibility of random amplified polymorphic DNA fingerprinting. *Res. Microbiol.* **144**: 373-379.

Micheli, M.R., Bova, R., Pascale, E. and D'Ambrosio, E. (1994). Reproducible DNA fingerprinting with the random amplified polymorphic DNA (RAPD) method. *Nucleic Acids Res.* **22**: 1921-1922.

Michelmore, R.W., Paran, I. and Kesseli, R.V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. USA* **88**: 9828-9832.

Miki *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**: 66-71.

Minna, J.D. and Coon, H.G. (1974). Human x mouse hybrid cells segregating mouse chromosomes and isozymes. *Nature (London)* **252**: 401-404.

Monaco, A.P. (1994). Isolation of genes from cloned DNA. *Curr. Opin. Genet. Devel.* **4**: 360-365.

Monaco, A.P. and Larin, Z. (1994). YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends Biotech.* **12**: 280-286.

Monaco, A.P., Neve, R.L., Colletti-Feener, C., Bertelson, C.J., Kurnit, D.M. and Kunkel, L.M. (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**: 646-650.

Moore, S.S., Sargeant, L.L., King, T.J., Mattick, J.S., Georges, M. and Hetzel, D.J.S. (1991). The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics* **10**: 654-660.

Muramatsu, Y., Yamada, T., Miura, M., Sakai, T., Suzuki, Y., Serikawa, T., Tanzi, R.E. and Matsumoto, K. (1994). Wilson's disease gene is homologous to *hts* causing abnormal copper transport in Long-Evans Cinnamon rats. *Gastroenterology* **107**: 1189-1192.

Murray, J.C. *et al.* (1994). A comprehensive human linkage map with centimorgan density. *Science* **265**: 2049-2054.

Nakao, N., Frodl, E.M., Widner, H., Carlson, E., Eggerding, F.A., Epstein, C.J. and Brundin, P. (1995). Overexpressing Cu/Zn superoxide dismutase enhances survival of transplanted neurons in a rat model of Parkinson's disease. *Nature Med.* **1**: 226-231.

Nam, H.G., Giraudat, J., Dan Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M. and Goodman, H.M. (1989). Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell* 1: 699-705.

Naqvi, N.I. and Chattoo, B.B. (1996). Development of a sequence characterized amplified region (SCAR) based indirect selection method for a dominant blast-resistance gene in rice. *Genome* **39**: 26-30.

Naylor, L.H. and Clark, E.M. (1990). $d(GT)_n \cdot d(CA)_n$ sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. *Nucleic Acids Res.* **18**: 1595-1601.

Nehls, M., Lüno, K., Schorpp, M., Pfeifer, D., Krause, S., Matysiak-Scholze, U., Dierbach, H. and Boehm, T. (1995). YAC/P1 contigs defining the location of 56 microsatellite markers and several genes across a 3.4-cM interval on mouse chromosome 11. *Mamm. Genome* **6**: 321-331.

Nelson, S.F., McCusker, J.H., Sander, M.A., Kee, Y., Modrich, P. and Brown, P.O. (1993). Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nature Genet.* **4**: 11-17.

Nelson, D.L. (1995). Positional cloning reaches maturity. *Curr. Opin. Genet. Develop.* **5**: 298-303.

Nigro, V., Bruni, P., Ciccodicola, A., Politano, L., Nigro, G., Piluso, G., Cappa, V., Covone, A.E., Romeo, G. and D'Urso, M. (1995). SSCP detection of novel mutations in patients with Emery-Dreifuss muscular dystrophy: definition of a small C-terminal region required for emerin function. *Hum. Mol. Genet.* **4**: 2003-2004.

Nojima, H. (1989). Structural organization of multiple rat calmodulin genes. *J. Mol. Biol.* **208**: 269-282.

Nystuen, A., Benke, P.J., Merren, J., Stone, E.M. and Sheffield, V.C. (1996). A cerebellar ataxia locus identified by DNA pooling to search for linkage disequilibrium in an isolated population from the Cayman Islands. *Hum. Mol. Genet.* **5**: 525-531.

Ochman, H. and Buckholtz, L.A. (1995). Physical structure of human chromosome 21: an analysis of YACs spanning 21q. *Mamm. Genome* 6: 84-89.

Olanow, C.W. (1992). An introduction to the free radical hypothesis in Parkinson's disease. *Ann. Neurol.* **32**: S2-S9.

Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**: 874-879.

Owen, A.D., Schapira, A.H., Jenner, P. and Marsden, C.D. (1996). Oxidative stress and Parkinson's disease. *Ann. N. Y. Acad. Sci.* **786**: 217-223.

Özçelik, H. and Andrulis, I.L. (1995). Multiplex PCR-SSCP for simultaneous screening for mutations in several exons of *p53*. *Biotechniques* **18**: 742-744.

Parent, A. (1990). Extrinsinc connections of the basal ganglia. *Trends Neurosci.* **13**: 254-258.

Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D. and Weissman, S.M. (1991). cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc. Natl. Acad. Sci. USA* **88**: 9623-9627.

Parker, Jr., W.D. (1991). Preclinical detection of Parkinson's disease: Biochemical approaches. *Neurology* **41** (suppl. 2): 34-36.

Parrish, J.E. and Nelson, D.L. (1993). Methods for finding genes. A major rate-limiting step in positional cloning. *Genet. Anal. Tech. Appl.* **10**: 29-41.

Patil, N., Cox, D.R., Bhat, D., Faham, M., Myers, R.M. and Peterson, A.S. (1995). A potassium channel mutation in weaver mice implicates membrane excitability in granule cell differentiation. *Nature Genet.* **11**: 126-129.

Paulson, G.W. and Dadmehr, N. (1991). Is there a premorbid personality typical for Parkinson's disease? *Neurology* **41** (suppl. 2): 73-76.

Pawar, H., Bingham, E.L., Lunetta, K.L., Segal, M., Richards, J.E., Boehnke, M. and Sieving, P.A. (1995). Refined genetic mapping of juvenile X-linked retinoschisis. *Hum. Hered.* **45**: 206-210.

Payami, H., Bernard, S., Larsen, K., Kaye, J. and Nutt, J. (1995). Genetic anticipation in Parkinson's disease. *Neurology* **45**: 135-138.

Payne, A.P., Downie, A., Campbell, J., Gilmore, M., Clarke, D.J. (1994). Dopamine deficits in a novel mutant rat with impaired locomotion. *J. Anatomy* **185**: 211.

Perry, T.L., Godin, D.V. and Hansen, S. (1982). Parkinson's disease: a disorder due to nigral glutathione deficiency? *Neurosci. Lett.* **33**: 305-310.

Pierce, J.C. and Sternberg, N.L. (1992). Using bacteriophage P1 system to clone high molecular weight genomic DNA. In *Methods in Enzymology* **216** (ed. Wu, R.) (New York: Academic Press): pp. 549-574.

Pierce, J.C., Sauer, B. and Sternberg, N. (1992). A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. *Proc. Natl. Acad. Sci. USA* **89**: 2056-2060.

Plaha, D.S., Linforth, G.H., Duckett, D.P., Scott-Jupp, R., Tanner, M.S. and Jeffreys, A.J. (1993). Three-base pair resolution of DNA fragments on agarose: Application to Delta F508 analysis in cystic fibrosis. *Biotechniques* **14**: 566-567.

Poewe, W.H. and Wenning, G.K. (1996). The natural history of Parkinson's disease. *Neurology* **47** (suppl. 3): 146-152.

Poirier, J., Kogan, S. and Gauthier, S. (1991). Environment, genetics and idiopathic Parkinson's disease. *Can. J. Neurol. Sci.* **18**: 70-76.

Polymeropoulos, M.H., Higgins, J.J., Golbe, L.I., Johnson, W.G., Ide, S.E., Di Iodrio, G., Sanges, G., Stenroos, E.S., Pho, L.T., Schaffer, A.A., Lazzarini, A.M., Nussbaum, R.L. and Duvoisin, R.C. (1996). Mapping of a gene for PD to chromosome 4q21-23. *Science* **274**: 1197-1199.

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., Stenroos, E.S., Chandrasekharappa, S., Athanassiadou, A., Papapetropoulos, T., Johnson, W.G., Lazzarini, A.M., Duvoisin, R.C., Di Iorio, G., Golbe, L.I. and Nussbaum, R.L. (1997). Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science* **276**: 2045-2047.

Pravenec, M., Gauguier, D., Schott, J.-J., Buard, Kren, V., Bila, V., Szpirer, J., Wang, J.-M., Huang, H., St.Lezin, E., Spence, M.A., Flodman, P., Printz, M., Lathrop, G.M., Vergnaud, G. and Kurtz, T.W. (1996). A genetic linkage map of the rat derived from recombinant inbred strains. *Mamm. Genome* **7**: 117-127.

Rao, V.S., Cupples, A., van Duijn, C.M., Kurz, A., Green, R.C., Chui, H., Duara, R., Auerbach, S.A., Volicer, L., Wells, J., van Broeckhoven, C., Growdon, J.H., Haines, J.L. and Faller, L.A. (1996). Evidence for major gene inheritance of Alzheimer disease in families of patients with and without apolipoprotein E epsilon 4. *Am. J. Hum. Genet.* **59**: 664-675.

Rajput, A.H. (1993). Environmental causation of Parkinson's disease. *Arch. Neurol.* **50**: 651-652.

Reis, D.J., Fink, J.S. and Baker, H. (1983). Genetic control of the number of dopamine neurons in the brain: Relationship to behaviour and responses to psychoactive drugs. In *Genetics of Neurological and Psychiatric Disorders* (ed. Kety, S.S., Rowland, L.P., Sidman, R.L. and Matthysse, S.W.) (Raven Press, New York).

Remmers, E.F., Goldmuntz, E.A., Cash, J.M., Crofford, L.J., Misiewicz-Poltorak, B., Zha, H. and Wilder, R.L. (1992). Genetic map of nine polymorphic loci comprising a single linkage group on rat chromosome 10: Evidence for linkage conservation with human chromosome 17 and mouse chromosome 11. *Genomics* **14**: 618-623.

Remmers, E.F., Goldmuntz, E.A., Zha, H., Crofford, L.J., Cash, J.M., Mathern, P., Du, Y. and Wilder, R.L (1993a). Linkage map of seven polymorphic markers on rat chromosome 18. *Mamm. Genome* **4**: 265-270.

Remmers, E.F., Goldmuntz, E.A., Zha, H., Mathern, P., Du, Y., Crofford, L.J. and Wilder, R.L. (1993b). Linkage map of nine loci defined by polymorphic DNA markers assigned to rat chromosome 13. *Genomics* **18**: 277-282.

Robinson, R. (1972). *Gene mapping in laboratory mammals*, Part B. (New York: Plenum publishing).

Roffler-Tarlov, S. and Graybiel, A.M. (1984). Weaver mutation has differential effects on the dopamine-containing innervation of the limbic and nonlimbic striatum. *Nature* **307**: 62-66.

Rommens, J.M., Iannuzzi, M.C., Kerem, B-S., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L-C. and Collins, F.S. (1989). Identification of the cystic fibrosis gene: Chromosome walking and jumping. *Science* **245**: 1059-1065.

Rosenberg, M., Przybylska, M. and Straus, D. (1994). "RFLP subtraction": A method for making libraries of polymorphic markers. *Proc. Natl. Acad. Sci. USA* **91**: 6113-6117.

Royle, N.J., Clarkson, R.E., Wong, Z. and Jeffreys, A.J. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3**: 352-360.

Russell, G.C. (1994). Improved single-strand conformation polymorphism analysis by asymmetric polymerase chain reaction with end-labeled primers. *Genet. Anal. Tech. Appl.* **11**: 24-27.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullins, K.B. and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning*, 2nd edition. Cold Spring Harbor Laboratory Press.

Saunders, A.M. and Seldin, M.F. (1990). The syntenic relationship of proximal mouse chromosome 7 and the myotonic dystrophy gene region on human chromosome 19q. *Genomics* 6: 324-332.

Schloetterer, C. and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**: 211-215.

Scott, D.L., Campbell, J.M., Clarke, D.J., Gilmore, D.P. and Payne, A.P. (1994). Noradrenergic deficits in the locus coeruleus of AS/AGU rats. European Neuroscience Association, Vienna, Sept. 1994.

Sengstock, G.J., Olanow, C.W., Dunn, A.J., Barone Jr., S. and Arendash, G.W. (1994). Progressive changes in striatal dopaminergic markers, nigral volume, and rotational behaviour following iron infusion into the rat substantia nigra. *Exp. Neurol.* **130**: 82-94.

Serikawa, T., Kuramoto, T., Hilbert, P., Mori, M., Yamada, J., Dubay, C.J., Lindpainter, K., Ganten, D., Guénet, J-L., Lathrop, G.M. and Beckmann, J.S. (1992). Rat gene mapping using PCR-analyzed microsatellites. *Genetics* **131**: 701-721.

Sever, P. (1995). The heterogeneity of hypertension - why doesn't every patient respond to every antihypertensive drug. *J. Hum. Hyperten.* **9** (suppl. 2): 33-36.

Sheffield, V.C., Cox, D.R., Lerman, L.S. and Myers, R.M. (1989). Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. USA* **86**: 232-236.

Shehee, W.R., Oliver, P. and Smithies, O. (1993). Lethal thalassemia after insertional disruption of the mouse major adult β-globin gene. *Proc. Natl. Acad. Sci. USA* **90**: 3177-3181.

Sheng, Y., Mancino, V. and Birren, B. (1995). Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Res.* **23**: 1990-1996.

Shiels, P., Durán Alonso, M.B., Davidson, A.O., Heeley, R.P., Dominiczak, A.F., Payne, A.P., Davies, R.W. and Sutcliffe, R.G. (1995). Optimized protocols for typing 75 microsatellite loci in AS, PVG, F344, and BN rats. *Mamm. Genome* **6**: 214-215.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**: 8794-8797.

Silverman, S.K., Kofuji, P., Dougherty, D.A., Davidson, N. and Lester, H.A. (1996). A regenerative link in the ionic fluxes through the *weaver* potassium channel underlies the pathophysiology of the mutation. *Proc. Natl. Acad. Sci. USA* **93**: 15429-15434.

Slightom, J.L., Blechl, A.E. and Smithies, O. (1980). Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**: 627-638.

Smith, P.S. and Bolam, J.P. (1990). The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci.* **13**: 259-265.

Smith, R.N. (1995). Accurate size comparison of short tandem repeat alleles amplified by PCR. *Biotechniques* **18**: 122-128.

Smoller, D.A., Petrov, D. and Hartl, D.L. (1991). Characterization of bacteriophage P1 library containing inserts of *Drosophila* DNA of 75-100 kilobase pairs. *Chromosoma* **100**: 487-494.

Sofic, E., Riederer, P., Heinsen, H., Beckmann, H., Reynolds, G.P., Hebenstreit, G. and Youdim, M.B. (1988). Increased iron (III) and total iron content in post mortem substantia nigra of parkinsonian brain. *J. Neural. Transm.* **74**: 199-205.

Southard-Smith, M. and MacDonald, R.J. (1993). Isolating reiterated genes from a rat genomic library constructed with the bacteriophage P1 system. *DuPont Biotech Update* **8**: 36-39.

Southard-Smith, M., Pierce, J.C. and MacDonald, R.J. (1994). Physical mapping of the rat tissue kallikrein family in two gene clusters by analysis of P1 bacteriophage clones. *Genomics* **22**: 404-417.

Stallings, R.L., Ford, A.F., Nelson, D., Torney, D.C., Hildebrand, C.E. and Moyzis, R.K. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics* **10**: 807-815.

Sternberg, N.L. (1990a). Alternative to YACs. *Genet. Anal. Tech. Appl.* **7**: 126-132.

Sternberg, N. (1990b). Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Natl. Acad. Sci. USA* **87**: 103-107.

Sternberg, N.L. (1992). Cloning high molecular weight DNA fragments by the bacteriophage P1 system. *Trends Genet.* **8**: 11-16.

Stewart, A.A., Campbell, J.M., McGadey, J., Clarke, D.J., Gilmore, D.P. and Payne, A.P. (1994). Serotonergic deficits in the raphe nucleus of AS/AGU rats. *J. Anatomy* **185**: 702.

Stewart, G.J., Savioz, A. and Davies, R.W. (1997). Sequence analysis of 497 mouse brain ESTs expressed in the substantia nigra. *Genomics* **39**: 147-153.

Stubbs, L., Carver, E.A., Shannon, M.E., Kim, J., Geisler, J., Generoso, E.E., Stanford, B.G., Dunn, W.C., Mohrenweiser, H., Zimmermann, W., Watt, S.M. and Ashworth, L.K. (1996). Detailed comparative maps of human chromosome 19q and related regions of the mouse genome. *Genomics* **35**: 499-508.

Suggs, S.V., Hirose, T., Miyake, T., Kawashima, E.H., Johnson, M.J., Itakura, K. and Wallace, R.B. (1981). In *Developmental biology using purified genes* (ed. Brown, D.D.) (Academic Press, N.Y.): pp. 683-693.

Szpirer, J., Szpirer, C. and Wanson, J.-C. (1980). Control of serum protein production in hepatocyte hybridomas: Immortalization and expression of normal hepatocyte genes. *Proc. Natl. Acad. Sci. USA* **77**: 6616-6620.

Szpirer, J., Levan, G., Thörn, M. and Szpirer, C. (1984). Gene mapping in the rat by mouse-rat somatic cell hybridization: synteny of the albumin and α-fetoprotein genes and assignment to chromosome 14. *Cytogenet. Cell Genet.* **38**: 142-149.

Szpirer, C., Riviere, M., Cortese, R., Nakamura, T., Islam, M.Q., Levan, G. and Szpirer, J. (1992). Chromosomal localization in man and rat of the genes encoding the liver-enriched transcription factors C/EBP, DBP, and HNF1/LFB-1 (CEBP, DBP, and Transcription Factor 1, TCF1, respectively) and of the Hepatocyte Groeth Factor/Scatter Factor Gene (HGF). *Genomics* **13**: 293-300.

Szpirer, C., Szpirer, J., Klinga-Levan, K., Stahl, F. and Levan, G. (1996). The rat: An experimental animal in search of a genetic map. *Folia Biologica (Praha)* **42**: 175-226.

Tanner, C.M. (1991). Abnormal liver enzyme-mediated metabolism in Parkinson's disease: A second look. *Neurology* **41** (suppl. 2): 89-91.

Tanzi *et al.* (1993). The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. *Nature Genet.* **5**: 344-350.

Tassabehji, M., Read, A.P., Newton, V.E., Harris, R., Balling, R., Gruss, P., Strachan, T. (1992). Waardenburg's syndrome patients have mutations in the human homologue of the *Pax-3* paired box gene. *Nature* **355**: 635-636.

Tatton, W.G. and Chalmers-Redman, R.M.E. (1996). Modulation of gene expression rather than monoamine oxidase inhibition: (-)-Deprenyl-related compounds in controlling neurodegeneration. *Neurology* **47** (suppl. 3): 171-183.

Tawata, M., Iwase, E., Aida, K. and Onaya, T. (1996). A mass screening device of genome by polymerase chain reaction-restriction fragment-single strand conformation polymorphism analysis. *Genet. Anal. Biomol. Eng.* **12**: 125-127.

Thomas, J.R., Bolla, R.I., Rumbyrt, J.S. and Schlessinger, D. (1985). DNase I-resistant nontranscribed spacer segments of mouse ribosomal DNA contain poly(dG-dT) . poly(dA-dC). *Proc. Natl. Acad. Sci. USA* **82**: 7595-7598.

Tong, Y., Wei, J., Zhang, S., Strong, J.A., Dlouhy, S.R., Hodes, M.E., Ghetti, B. and Yu, L. (1996). The weaver mutation changes the ion selectivity of the affected inwardly rectifying potassium channel GIRK2. *FEBS Lett.* **390**: 63-68.

Traystman, M. D., Higuchi, M., Kasper, C.K., Antonarakis, S.E. and Kazazian, Jr., H.H. (1990). Use of denaturing gradient gel electrophoresis to detect point mutations in the Factor VIII gene. *Genomics* **6**: 293-301.

Treco, D., Thomas, B. and Arnheim, N. (1985). Recombination hot spot in the human β-globin gene cluster: meiotic recombination of human DNA fragments in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **5**: 2029-2037.

Wada, M., Little, R.D., Abidi, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Urso, M. and Schlessinger, D. (1990). Human Xq24-Xq28: Approaches to mapping with yeast artificial chromosomes. *Am. J. Hum. Genet.* **46**: 95-106.

Wang, Q., Curran, M.E., Splawaski, I., Burn, T.C., Millholland, J.M., VanRaay, T.J., Shen, J., Timothy, K.W., Vincent, G.M., de Jager, T., Schwartz, P.J., Towbin, J.A., Moss, A.J., Atkinson, D.L., Landes, G.M., Connors, T.D. and Keating, M.T. (1996). Positional cloning of a novel potassium channel gene: *KVLQT1* mutations cause cardiac arrhytmias. *Nature Genet.* **12**: 17-23.

Weber, J.L. and May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388-396.

Weber, J.L. (1990). Informativeness of human $(dC\text{-}dA)_n$. $(dG\text{-}dT)_n$ polymorphisms. *Genomics* **7**: 524-530.

Weber, J.L. and Wong, C. (1993). Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**: 1123-1128.

Weeks, D.E. and Lathrop, G.M. (1995). Polygenic disease: methods for mapping complex disease traits. *Trends Genet.* **11**: 513-519.

Weiss, M.C. and Ephrussi, B. (1966). Studies of interspecific (rat x mouse) somatic hybrids. *Genetics* **54**: 1095-1109 and 1111-1122.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. and Lathrop, M. (1992). *Nature* **359**: 794-801.

Wharton, K.A., Yedvobnick, B., Finnerty, V.G. and Artavanis-Tsakonas, S. (1985). *opa*: A novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D. melanogaster*. *Cell* **40**: 55-62.

Wilhemsen, K., Mirel, D., Marder, K., Bernstein, M., Naini, A., Leal, S.M., Cote, L.J., Tang, M.-X., Freyer, G., Graziano, J. and Mayeux, R. (1997). Is there a genetic susceptibility locus for Parkinson's disease on chromosome 22q13? *Ann. Neurol.* **41**: 813-817.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**: 6531-6535.

Williams, A., Steventon, G., Sturman, S. and Waring, R. (1991). Xenobiotic enzyme profiles and Parkinson's disease. *Neurology* **41** (suppl. 2): 29-32.

Williams, J.G.K., Reiter, R.S., Young, R.M. and Scolnik, P.A. (1993). Genetic mapping of mutations using phenotypic pools and mapped RAPD markers. *Nucleic Acids Res.* **21**: 2697-2702.

Williams, J.L., Olsaker, I. and Teres, V.M. (1997). Using SSCP to facilitate mapping microsatellite loci. *Mamm. Genome* **8**: 79-80.

Wu, C.-L. and Melton, D.W. (1993). Production of a model for Lesch-Nyhan syndrome in hypoxanthine phosphoribosyl-tranferase-deficient mice. *Nature Genet.* **3**: 235-240.

Yamada, J., Kuramoto, T. and Serikawa, T. (1994). A rat genetic linkage map and comparative maps for mouse or human homologous rat genes. *Mamm. Genome* **5**: 63-83.

Yokoi, N., Kitada, K., Kuramoto, T., Maihara, T., Andoh, Y., Takami, S., Higaki, J., Miki, T., Ogihara, T., Lathrop, G.M. and Serikawa, T. (1996). Mapping of 20 polymorphic DNA markers in the rat by somatic hybrid and linkage analysis. *Mamm. Genome* **7**: 71-73.

Yoshida, M.C. (1978). Rat gene mapping by rat-mouse somatic cell hybridization and a comparative Q-banding analysis between rat and mouse chromosomes. *Cytogenet. Cell Genet.* **22**: 606-609.

Youdim, M.B.H. and Riederer, P. (1997). Understanding Parkinson's disease. *Scientific American* **276**: 38-45.

Youil, R., Kemper, B.W. and Cotton, R.G.H. (1995). Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc. Natl. Acad. Sci. USA* **92**: 87-91.

Zha, H., Wilder, R.L., Goldmuntz, E.A., Cash, J.M., Crofford, L.J., Mathern, P. and Remmers, E.F. (1993). Linkage map of 10 polymorphic markers on rat chromosome 2. *Cytogenet. Cell Genet.* **63**: 117-122.

Zhou, E., Favor, J., Silvers, W. and Staubolian, D. (1995). Exclusion of three candidate genes, *Grpr*, *Cxn33*, and *Pdha1*, for the X-linked cataract gene on the distal region of the mouse chromosome X. *Mamm. Genome* **6**: 357-359.

Zigmond, M.J., Abercrombie, E.D., Berger, T.W., Grace, A.A. and Stricker, E.M. (1990). Compensations after lesions of central dopaminergic neurons: some clinical and basic implications. *Trends Neurosci.* **13**: 290-296.

Zuddas, A., Fascetti, F., Corsini, G.U. and Piccardi, M.P. (1994). In Brown Norway rats, MPP^+ is accumulated in the nigrostriatal dopaminergic terminals but is not neurotoxic: A model of natural resistance to MPTP toxicity. *Exp. Neurol.* **127**: 54-61.