

Mapping the Genes for Arylamine N-acetyltransferases

Submitted for the degree of Doctor of Philosophy

Trinity Term 2000

Giannoulis Fakis

St. Peter's College

Department of Pharmacology

University of Oxford

Supervisor: Prof. Edith Sim



Declaration of authorship

I confirm that the work presented in this Thesis is entirely my own. In places where results were obtained in collaboration with colleagues from the Department of Pharmacology or other Institutions, this is clearly indicated and their contribution is gratefully acknowledged.

To Rea and my family

Mapping the Genes for Arylamine N-Acetyltransferases.**Giannoulis Fakis, St Peter's College, University of Oxford.****Submitted for the degree of Doctor of Philosophy, Trinity Term, 2000.****ABSTRACT**

Arylamine N-acetyltransferases are xenobiotic metabolising enzymes, involved in metabolic pathways leading to either detoxification or bio-activation of drugs and carcinogens. In humans, polymorphisms of the two functional NAT genes (*NAT1* and *NAT2*) have been implicated to cancer susceptibility. The genomic region of the *NAT* loci, 8p22, is often deleted in tumours and is believed to contain tumour suppressor genes. Two previously developed cosmid probes, carrying either *NAT1* or *NAT2*, were tested as probes for FISH analysis using nuclei from healthy lymphocytes and the RT112 urothelial carcinoma cell line (Chapter 3). FISH analysis of cells obtained from barbotage samples of bladder cancer patients was performed and deletion of the *NAT* genomic region was a common observation (Chapter 6). To allow refined mapping of the *NAT* region on 8p22, a series of PAC clones carrying the *NAT* loci were isolated by PCR screening of a genomic library (Chapter 3). Furthermore, the *NAT1* and *NAT2* isoenzymes were studied in healthy intestinal and mammary tissue, as a first step towards understanding their potential involvement in these two types of cancer (Chapter 6).

The mouse has been extensively used as a model for studying NAT. Three functional *Nat* genes are present in the mouse genome. A fine restriction map was generated for the *Nat1* and *Nat2* genomic region in mice, using *Nat*-positive genomic DNA plasmid clones previously isolated from the fast acetylator Balb/c and 129/Ola strains. The *Nat1* and *Nat2* genes were mapped 9.4kb apart and no polymorphisms were detected between the two strains for the restriction enzymes used for mapping (Chapter 4). The 129/Ola restriction map has since been the basis for the production of a construct for *Nat2* knockout and transgenic mouse strains. PAC clones positive for *Nat* were isolated from a mouse genomic library and used as FISH probes to map the three murine *Nat* genes on chromosome 8, band B3.1-3.3. One PAC clone contained all three *Nat* loci, establishing co-localisation of *Nat3* with the other two *Nat* genes in mice. The minimum distance of *Nat3* from *Nat1* and *Nat2* was estimated to be 22kb, while the three loci are within 130kb. YAC clones carrying the *Nat* genes were also isolated to facilitate physical mapping of the *Nat* cluster in mice (Chapter 5). This will allow integration of cytogenetic, physical and previous genetic data for comparative studies.

Acknowledgements

I am grateful to my supervisor, Professor Edith Sim, for the opportunity to join her group and for her guidance and advice during my work in the lab and while writing this thesis. I am equally grateful to the Bodossaki Foundation and St Peter's College for the award of a scholarship to fund my study in Oxford.

I would like to thank the senior members of the lab for their co-operation, especially Dr Mike Stacey for his help during my first year. I would also like to acknowledge the kind and generous help of Dr Veronica Buckle and Dr Sabrina Tosi. I would like to thank all past and present members of the Sim group for making my time in the lab so pleasant. I feel that I formed a stronger bond with the people that were already in the lab when I first arrived in Oxford and I wish I had spent more time with the rest of you. John, may your proteins crystallise easily. Valerie and Rupika, thank you for having me at your weddings and I hope you are enjoying your married and professional lives. Nikki, thank you for many laughs and for always being there. Anna, thanks for teaching me the difference between buttery and battery, among other things. Adeel, may The Force be with you. Also thanks to Mark, Katalin, Fred, James, Aileen and Georgina. I consider you all good friends and I hope we can remain in touch wherever we are in the world.

During my time in Oxford, I was fortunate to meet a lot of great people and I made many friends in College. Dimitris, Emmanuel and Konstantinos, with their intelligence and wit, have helped me see life from their different perspectives. Thank you for your friendship!

I would not have been able to come to Oxford without the efforts and sacrifices of my parents. I cannot express how thankful I am for their emotional and financial support during all my years of study and I apologise for causing them headaches. I wish best of luck to Michalis for his PhD and would like to thank my grandmother for her love.

Finally, I am extremely grateful to Rea, my life's companion, for her support in and out of the lab and her endless and unconditional love.

Publications arising from this Thesis

Journal Articles

1. D Hickman, J Pope, SD Patil, G Fakis, V Smelt, LA Stanley, M Payton, JD Unadkat and E Sim. Expression of arylamine N-acetyltransferase in human intestine. *Gut* **42**: 402-409 (1998).
2. M Stacey, M Payton, G Fakis, J Pope and E Sim. Characterisation of arylamine N-acetyltransferase in the human urothelial cell line RT112. *Pathogenesis* **1**: 99-105 (1998).
3. M Stacey, N Matas, M Drake, M Payton, G Fakis, J Greenland and E Sim. Arylamine N-acetyltransferase type 2 (*NAT2*), chromosome 8 aneuploidy, and identification of a novel *NAT1* cosmid clone: an investigation in cancer by interphase FISH. *Genes, Chromosomes & Cancer* **25**: 376-383 (1999).
4. P Thygesen, A Risch, M Stacey, G Fakis, L Takle, M Knowles and E Sim. Genes for human arylamine N-acetyltransferases in relation to loss of the short arm of chromosome 8 in bladder cancer. *Pharmacogenetics* **9**: 1-8 (1999).
5. JA Williams, EM Stone, G Fakis, N Johnson, E Sim and DH Phillips. Relationship between N-acetyltransferase status and 2-amino-3-methylimidazo [4,5-f] quinoline (IQ)-DNA adduct formation in human breast tissue. *Cancer Research* (submitted).
6. G Fakis, S Boukouvala, V Buckle, M Payton, C Denning and E Sim. Chromosome mapping of the genes for murine arylamine N-acetyltransferases (NATs), enzymes involved in the metabolism of carcinogens: Identification of a novel upstream non-coding exon for *Nat2*. *Cytogenetics and Cell Genetics* **90**: 134-138 (2000).

Abstracts and Presentations

1. M Stacey, M Payton, VA Smelt, G Fakis, M Drake, J Greenland and E Sim. Loss of NAT2 in bladder cancer. Human Genome Organisation Human Genome Meeting 1997, Toronto, Canada.
2. E Sim, J Pope, G Fakis, VA Smelt and M Payton. Expression of arylamine N-acetyltransferases in human intestine. Molecular Genetics Of Cancer 1997, Oxford, UK
3. G Fakis, M Payton, C Denning and E Sim. Mapping the genes for arylamine N-acetyltransferases. British Toxicology Society Autumn Meeting 1998, York, UK.
4. G Fakis, N Matas, P Thygesen, M Lehman, M Payton, VA Smelt, C Denning, S Kelly, A Risch, and E Sim. Mapping of the human and mouse genomes in the region of the arylamine N-acetyltransferase genes. First International Workshop On The Arylamine N-Acetyltransferases 1998, Cairns, Queensland, Australia.
5. N Johnson, M Payton, G Fakis, A Risch, M Stacey, M Knowles, M Wallace, M Drake and E Sim. Human arylamine N-acetyltransferases in bladder cancer. First International Workshop On The Arylamine N-acetyltransferases 1998, Cairns, Queensland Australia.
6. D Hickman, J Pope, G Fakis, VA Smelt, LA Stanley, M Payton and E Sim. Arylamine N-acetyltransferase in normal human intestine-implications for colorectal cancer. First International Workshop On The Arylamine N-Acetyltransferases 1998, Cairns, Queensland, Australia.
7. S Boukouvala, G Fakis, N Johnson, M Payton, VA Smelt, A Upton, J Adjaye, F Brook, H Priddle and E Sim. Investigation of the role of arylamine N-acetyltransferase, an enzyme of folate metabolism, in development. 9th Mammalian Genetics And Development Workshop Of The Genetical Society 1998, London, UK.
8. G Fakis, M Payton, C Denning and E Sim. Mapping the human and mouse genes for arylamine N-acetyltransferases. Human Genome Organisation Human Genome Meeting 1999, Brisbane, Australia.
9. S Boukouvala, G Fakis, K Pinter, V Buckle, H Priddle and E Sim. Expression and localisation of the genes for arylamine N-acetyltransferases in mice. 10th Mammalian Genetics And Development Workshop Of The Genetical Society 1999, London, UK.

List of Contents

Abstract	i
Acknowledgements	ii
Publications arising from this Thesis	iii
List of Contents	v
List of Figures	ix
List of Tables	xi
List of Abbreviations	xii
 Chapter 1: Introduction	 1
1.1 Xenobiotic metabolising enzymes	1
1.1.1 Phase I reactions	4
1.1.2 Phase II reaction	5
1.2 Pharmacogenetics	7
1.3 Arylamine N-acetyltransferases in humans	10
1.3.1 Human arylamine N-acetyltransferases in drug metabolism	14
1.3.2 Mechanism of action of arylamine N-acetyltransferases	16
1.3.3 The NAT polymorphism in humans	17
1.3.3.1 The NAT2 polymorphism	17
1.3.3.2 The NAT1 polymorphism	22
1.3.4 Tissue-specific expression of human arylamine N-acetyltransferases	27
1.3.5 Human arylamine N-acetyltransferase in development	28
1.3.6 Arylamine N-acetyltransferase and disease in humans	31
1.3.6.1 NAT and diabetes mellitus	31
1.3.6.2 NAT and neurodegenerative disorders	32
1.3.6.3 NAT and AIDS	33
1.3.6.4 NAT and endometriosis	34
1.3.7 Arylamine N-acetyltransferase and Cancer	35
1.3.7.1 NAT in cancers of the larynx, pharynx and mouth	39
1.3.7.2 NAT in lung cancer	40
1.3.7.3 NAT in prostate cancer	42
1.3.7.4 NAT in hepatocellular carcinoma	43
1.4 Animal models for arylamine N-acetyltransferase	44
1.4.1 The mouse model for arylamine N-acetyltransferase	48
1.4.1.1 The arylamine N-acetyltransferase polymorphism in mice	48
1.4.1.2 Tissue-specific expression of murine arylamine N-acetyltransferases	50
1.4.1.3 Murine arylamine N-acetyltransferases in development	50
1.4.1.4 Murine arylamine N-acetyltransferases as model for carcinogenesis	52
1.4.1.5 Murine arylamine N-acetyltransferase 3: A functional isoenzyme?	53
1.5 Genome Projects	54
1.5.1 The Human Genome Project	54
1.5.2 The Mouse Genome Project	58
1.6 Genome Mapping	60
1.6.1 Genetic Mapping	60
1.6.2 Physical Mapping	63
1.6.3 Radiation Hybrid Mapping	65

1.6.4 Cytogenetic Mapping	68
1.6.5 Strategies for mapping disease genes	70
1.7 Aims	71
Chapter 2: Materials and Methods	72
2.1 Reagents	72
2.1.1 Water	72
2.1.2 Suppliers of chemicals	72
2.1.3 Primers	72
2.2 General molecular biology methods	74
2.2.1 Isolation and purification of plasmid and PAC DNA	74
2.2.2 Isolation and purification of cosmid DNA	74
2.2.3 Isolation and purification of genomic DNA	75
2.2.4 Measuring concentration and purity of DNA solutions	76
2.2.5 DNA gel electrophoresis	76
2.2.5.1 Standard agarose gel electrophoresis	76
2.2.5.2 DNA electrophoresis for Southern blotting	77
2.2.5.3 Pulse field gel electrophoresis (PFGE)	77
2.2.6 Purification of DNA from agarose and solutions	78
2.2.7 Polymerase chain reaction (PCR)	78
2.2.8 Restriction enzyme digestion	79
2.2.9 Transformation of competent cells	79
2.2.10 Bacterial culture	80
2.2.11 Southern blotting	80
2.2.12 Southern hybridisation	81
2.2.12.1 Hybridisation using radioactive probes	81
2.2.12.1.1 Radioactive probe labelling	81
2.2.12.1.2 Hybridisation	82
2.2.12.2 Hybridisation using digoxigenin-labelled probes	82
2.2.12.2.1 Digoxigenin labelling of probes	82
2.2.12.2.2 Mock hybridisations	83
2.2.12.2.3 Hybridisation and chemiluminescence detection of the signal	85
2.3 Cytogenetic methods	86
2.3.1 Cell culture	86
2.3.1.1 Culture of the RT112 cell line	86
2.3.1.2 Culture of T-lymphocytes from whole blood	86
2.3.2 Preparation of metaphase spreads for FISH analysis	87
2.3.3 Preparation of interphase spreads for FISH analysis	87
2.3.4 Preparation of APES-coated slides for microscopy	88
2.3.5 Labelling of probes for FISH analysis	88
2.3.6 Fluorescence <i>in situ</i> hybridisation (FISH)	89
2.3.6.1 FISH analysis of human T-lymphocyte, RT112 and barbotage cells	89
2.3.6.2 FISH analysis of mouse spleen cells	91
2.4 General methods for protein analysis	92
2.4.1 Measuring protein concentration	92
2.4.2 Dot blot	92
2.4.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)	93

2.4.4 Western blotting	94
2.5 NAT activity assays	94
Chapter 3: Arylamine N-acetyltransferase Genes as Markers for Chromosomal Region 8p22 in Humans	96
3.1 Introduction	96
3.1.1 Cloning and characterisation of the genes for human NAT	96
3.1.2 Chromosomal localisation and mapping of the human <i>NAT</i> genes	99
3.1.3 The Human Genome Project and Chromosome 8	100
3.2 Strategies for genotyping the human <i>NAT</i> genes	101
3.2.1 <i>NAT1</i> genotyping	101
3.2.2 <i>NAT2</i> genotyping	103
3.3 Aims and study design	105
3.3.1 The <i>NAT1</i> and <i>NAT2</i> cosmid clones	105
3.3.2 Sample of healthy lymphocytes	105
3.3.3 The RT112 cell line	106
3.3.4 The RPCI 1 human genomic DNA PAC library	107
3.3.4.1 Pooling information and screening scheme	107
3.4 Results	110
3.4.1 Genotyping of the <i>NAT</i> cosmids	110
3.4.1.1 Genotyping of the <i>NAT1</i> cosmid	110
3.4.1.2 Genotyping of the <i>NAT2</i> cosmid	112
3.4.2 Use of the <i>NAT</i> cosmids for FISH analysis	113
3.4.2.1 FISH analysis of healthy lymphocytes	113
3.4.2.2 FISH analysis of the RT112 cell line	115
3.4.2.3 <i>NAT2</i> genotyping of the RT112 cell line	117
3.4.3 Screening the RPCI 1 library for <i>NAT1</i> , <i>NAT2</i> and <i>NATP</i>	121
3.4.4 Genotyping the identified <i>NAT1</i> and <i>NAT2</i> positive PAC clones	128
3.5 Discussion	128
Chapter 4: Mapping the Region of the <i>Nat1</i> and <i>Nat2</i> Genes in the Mouse	133
4.1 Introduction	133
4.2 Materials and study design	137
4.3 Method for genotyping the mouse <i>Nat2</i> gene	137
4.4 Results	138
4.4.1 Investigation of the <i>Nat</i> content of the plasmid Balb/c clones 1 and 2 and the 129/Ola clone A	138
4.4.2 Genotyping the mouse <i>Nat2</i> plasmid clones	142
4.4.3 Restriction mapping of the mouse genomic plasmid clones	144
4.4.3.1 Mapping of Balb/c clone 1	144
4.4.3.2 Mapping of the Balb/c clone 2	149
4.4.3.3 Mapping of the 129/Ola clone A	154
4.5 Discussion	159
Chapter 5: Chromosomal Localisation of the <i>Nat</i> Genes in the Mouse	162
5.1 Introduction	162
5.2 Materials and study design	165
5.2.1 The RPCI 21 mouse PAC genomic library	166
5.2.2 The WI/MIT mouse YAC genomic library	166
5.2.3 Chromosome 8 YAC contigs	167

5.2.4 D8Mit markers	168
5.3 Results	170
5.3.1 Screening of the RPCI 21 mouse PAC genomic library for <i>Nat</i> genes	170
5.3.2 Genotyping of the <i>Nat2</i> -positive mouse PAC genomic clones	175
5.3.3 Cytogenetic mapping of the mouse <i>Nat</i> genes	176
5.3.4 Restriction analysis of the 511-O17 PAC clone	180
5.3.5 Screening of <i>Nat</i> -positive mouse PAC genomic clones for chromosome 8 specific markers	184
5.3.6 Screening of chromosome 8 specific YAC contigs for murine <i>Nat</i> genes	185
5.3.7 Screening of the WI/MIT mouse YAC genomic library	187
5.4 Discussion	191
 Chapter 6: Arylamine N-Acetyltransferase - A Study in Healthy Colon, Breast and Cancerous Bladder Material	198
6.1 Introduction	198
6.1.1 NAT in colon cancer	198
6.1.2 NAT in bladder cancer	200
6.1.3 NAT in breast cancer	203
6.2 NAT1 and NAT2 in healthy human intestine	206
6.2.1 Study design and samples	206
6.2.2 Results	206
6.2.2.1 <i>NAT1</i> genotyping	207
6.2.2.2 <i>NAT2</i> genotyping	209
6.2.3 Discussion	214
6.3 NAT1 and NAT2 in human bladder cancer material	218
6.3.1 Study design and samples	218
6.3.2 Results	219
6.3.3 Discussion	225
6.4 NAT1 and NAT2 in healthy breast tissue	230
6.4.1 Materials and study design	230
6.4.2 Results	231
6.4.2.1 Measurement of protein concentration in breast cytosols	231
6.4.2.2 NAT1 enzymatic activity in human breast cytosols	232
6.4.2.3 NAT2 enzymatic activity in human breast cytosols	238
6.4.2.4 Detection of NAT protein in healthy human breast tissue	238
6.4.3 Discussion	241
 Chapter 7: Conclusions	246
 List of References	252

List of Figures

1.1	Schematic overview of the metabolism of xenobiotics	3
1.2	External and internal factors affecting drug disposition in humans	9
1.3	Chemical structure of common NAT2 substrates in humans	12
1.4	Chemical structure of common NAT1 substrates in humans	13
1.5	Chemical structure of 2-aminofluorene, a common substrate for both NAT1 and NAT2 in humans	13
1.6	Chemical structures of folate and its catabolites, pterate and p-aminobenzoyl glutamate (pABGlu)	30
1.7	Schematic representation of the metabolic reactions catalysed by NAT	38
1.8	Alignment of the amino acid sequences of different mammalian NATs	46
1.9	Outline of the radiation hybrid mapping procedure	67
2.1	Example of mock hybridisation, following labelling of a mouse <i>Nat2</i> -specific probe with digoxigenin	84
2.2	Labelling of a human <i>NAT1</i> -specific cosmid probe with biotin 16-dUTP by nick translation	89
3.1	Strategy for genotyping human <i>NAT1</i>	102
3.2	Strategy for genotyping human <i>NAT2</i>	104
3.3	Strategy for screening the RPCI 1 human genomic DNA PAC library	108
3.4	Genotyping of the <i>NAT1</i> cosmid	111
3.5	Genotyping of the <i>NAT2</i> cosmid	112
3.6	FISH analysis of healthy T-lymphocytes using the characterised <i>NAT1</i> cosmid clone as probe	114
3.7	FISH analysis of RT112 cells using the characterised <i>NAT1</i> cosmid clone as probe	115
3.8	<i>NAT2</i> genotyping of RT112 cells	118
3.9	Screening the RPCI 1 human genomic DNA PAC library for <i>NAT</i> genes	123
3.10	PAC clones positive for <i>NAT</i> genes	125
4.1	Investigation of the <i>Nat</i> content of the three mouse genomic plasmid clones used for restriction mapping	139
4.2	Specificity of amplification from the mouse <i>Nat</i> genes	142
4.3	Genotyping of the mouse genomic plasmid clones carrying the <i>Nat2</i> gene	143
4.4	Isolation of the insert from the Balb/c clone 1	145
4.5	Restriction enzyme digestions (a) and Southern blot hybridisation (b) for mapping of the 4.8kb <i>Nat1</i> -positive Balb/c clone 1	148
4.6	Isolation of the insert from the Balb/c clone 2	149
4.7	Restriction enzyme digestions (a) and Southern blot hybridisation (b) for mapping of the 8kb <i>Nat2</i> -positive Balb/c clone 2	153
4.8	Isolation of the insert from the 129/Ola plasmid clone	154
4.9	Restriction enzyme digestions (a) and Southern blot hybridisation (b) for mapping of the 14.3kb 129/Ola clone A	157
4.10	Restriction maps of the three mouse genomic plasmid clones	158

5.1	Integrated linkage and physical map of the region 29-33cM from the centromere of mouse chromosome 8	169
5.2	Screening of the RPCI 21 mouse PAC genomic DNA library by hybridisation	171
5.3	Investigation of the <i>Nat</i> content of the PAC clones	173
5.4	Specificity of amplification of the mouse <i>Nat3</i> gene	174
5.5	Genotyping of the <i>Nat2</i> -positive mouse PAC genomic clones	175
5.6	Chromosomal localisation of the mouse <i>Nat</i> genes	177
5.7	Partial <i>KpnI</i> digestion of the mouse PAC genomic clone 511-O17	182
5.8	Position of the murine <i>Nat</i> genes relative to their closest <i>KpnI</i> restriction sites	183
5.9	Amplification of SSR marker D8Mit69 from the three <i>Nat</i> -positive mouse PAC genomic clones	185
5.10	Screening of mouse chromosome 8-specific YAC contigs for presence of the <i>Nat</i> genes	186
5.11	Screening of the WI/MIT mouse YAC genomic library for <i>Nat</i> genes	188
5.12	Identification of mouse YAC genomic clones carrying the <i>Nat</i> genes	189
6.1	<i>NAT1</i> genotyping of human intestine samples	207
6.2	<i>NAT2</i> genotyping of human intestine samples	210
6.3	Fluorescence <i>in situ</i> hybridisation with <i>NAT1</i> cosmid probe on bladder cancer cells from barbotage fluid: An example of cells with diploid number of <i>NAT1</i> genes	222
6.4	Fluorescence <i>in situ</i> hybridisation with <i>NAT1</i> cosmid probe on bladder cancer cells from barbotage fluid: An example of cells with loss of one copy of the <i>NAT1</i> gene	223
6.5	Fluorescence <i>in situ</i> hybridisation with <i>NAT1</i> cosmid probe on bladder cancer cells from barbotage fluid: An example of cells with a triploid number of <i>NAT1</i> genes	224
6.6	Diagram showing the NAT1 specific activity (nmol/min/mg) measured for the eighteen breast cytosol samples, following incubation with pABA for 10min	234
6.7	Histogram of frequency distribution for the eighteen breast cytosolic samples, based on their calculated NAT1 specific activity	234
6.8	Comparison of the breast cytosolic NAT1 activity for pABA with the <i>NAT1</i> genotype of the eighteen individuals	236
6.9	Comparison of breast cytosolic NAT1 activity with the number of <i>NAT1*10</i> alleles in the genotype of the eighteen individuals	237
6.10	Detection of NAT1 protein in human breast cytosols by dot blot	239
6.11	Denaturing polyacrylamide gel electrophoresis (SDS-PAGE) of human breast cytosolic samples	240
6.12	Detection of NAT1 protein in the pooled human breast cytosol by Western blotting	240
6.13	Detection of NAT1 protein in samples of human breast cytosols by Western blotting	241
6.14	Correlation between NAT1 specific activity with pABA and the number of IQ-DNA adducts	245

List of Tables

1.1	A list of the most important chemical reaction associated with Phase I and II of the metabolism of xenobiotics	2
1.2	Common drugs metabolised by humans	15
1.3	The identified <i>NAT2</i> alleles in humans	21
1.4	The most common <i>NAT1</i> alleles in the Caucasian population	26
1.5	Percentage (%) identity of the amino acid sequences of rabbit, mouse, rat, Syrian hamster and human NAT isoenzymes	47
1.6	Summary of the achievements of the HGP during the period 1993-98 and goals for period 1998-2003	57
1.7	DNA sequence databases available on the internet	59
1.8	Polymorphic markers used for genetic mapping	62
1.9	A comparison of cloning systems used for physical mapping	64
1.10	Resolution of various traditional and high-resolution FISH techniques, in comparison with the resolution of other mapping methods	69
2.1	Primers used for PCR amplifications	73
3.1	Summary of the <i>NAT</i> positive PAC clones identified following screening of the RPCI 1 human genomic DNA PAC library and their <i>NAT</i> content	127
3.2	Summary of the digestions used to confirm the <i>NAT</i> content of the PAC clones identified as positive for at least one of the <i>NAT</i> loci	127
4.1	Percentage identity between the human and mouse genes for NAT, based on their nucleotide sequence	135
4.2	Summary of the <i>Nat</i> content of the mouse genomic plasmid clones used for restriction mapping	140
4.3	Summary of the restriction enzyme digestions used to confirm specificity of amplification from the mouse <i>Nat</i> genes	141
4.4	Restriction patterns obtained for mapping of the Balb/c clone 1	147
4.5	Restriction patterns obtained for mapping of the Balb/c clone 2	152
4.6	Restriction patterns obtained for mapping of the 129/Ola clone A	156
5.1	Markers contained in the four YAC contigs spanning the 29.5-32.8cM region of murine chromosome 8	170
5.2	Summary of the content of the three <i>Nat</i> -positive mouse PAC genomic clones (584-N8, 393-K20 and 511-O17)	174
5.3	Size of restriction fragments hybridising to different <i>Nat</i> -specific probes, following digestion of PAC clone 511-O17 with <i>KpnI</i>	181
5.4	Simple sequence repeat (SSR) markers used for screening the <i>Nat</i> -positive mouse PAC genomic clones	184
5.5	Summary of the content of the <i>Nat</i> -positive clones isolated from the WI/MIT mouse YAC genomic library	190
6.1	The <i>NAT1</i> and <i>NAT2</i> genotypes and the predicted NAT2 acetylator status of the four intestinal samples	213
6.2	NAT1 and NAT2 activities of the four intestine samples, measured with pABA and SMZ respectively	215
6.3	Tumour samples investigated for the loss of <i>NAT1</i> and <i>NAT2</i> genes	220
6.4	Samples with <i>NAT1</i> or <i>NAT2</i> gene deletion classified according to their observed frequency of loss	221
6.5	Protein concentration of the human breast cytosols	231
6.6	NAT1 specific activity for pABA of the eighteen human breast cytosols and the control PS4 sample	233

List of Abbreviations

6-AN	6-Aminonicotinamide
Acetyl-CoA	Acetyl coenzyme A
AHH	Aryl hydrocarbon hydroxylase
AIDS	Acquired Immunodeficiency Syndrome
APES	3-Aminopropyltriethoxysilane
APS	Ammonium persulphate
ASO	Allele-specific oligonucleotide
AS-PCR	Allele specific polymerase chain reaction
BAC	Bacterial artificial chromosome
bp	Base pairs
BSA	Bovine serum albumin
c.v.	Coefficient of variation
cDNA	Complementary DNA
CEPH	Centre d' Etude du Polymorphisme Humain
CGH	Comparative genomic hybridisation
CHO	Chinese hamster ovary
CL/P	Cleft lip with or without palate
CoA	Coenzyme A
CP	Cleft palate
Cy3	Cyanine 3
DAPI	4',6'-diamino-2-phenylindole
DDT	Dichlorodiphenyltrichloroethane
DGGE	Denaturing gradient gel electrophoresis
DHMHD	Dysmorphic human-mouse homology database
DIG	Digoxigenin
DMAB	Dimethylaminobenzaldehyde
DMSO	Dimethyl sulphoxide
DOP-PCR	Degenerate oligonucleotide-primed polymerase chain reaction
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetate
ELISA	Enzyme-linked immunosorbent assay
ES cells	Embryonic stem cells
EST	Expressed sequence tags
FISH	Fluorescence <i>in situ</i> hybridisation
FITC	Fluorescein isothiocyanate
GDB	Genome Database
gDNA	Genomic DNA
GST	Glutathione-S-transferase
HGMP RC	Human Genome Mapping Project Resource Centre
HGP	Human Genome Project
HIV	Human Immunodeficiency Virus
HPLC	High pressure liquid chromatography
HPRT	Hypoxanthine phosphoribosyl transferase
HUGO	Human Genome Organisation
IDDM	Insulin-dependent diabetes mellitus
IMB	Institute of Molecular Biology
INH	Isoniazid
IQ	2-Amino-3-methylimidazol[4,5-f] quinoline

IRS	Interspersed repeat sequence
JFCR	Japanese Foundation for Cancer Research
kb	Kilobase pairs
LINE	Long interspersed element
lod	Logarithm of odds
LOH	Loss of heterozygosity
LPL	Lipoprotein lipase
LPS	Lipopolysaccharide
MAC	Mammalian artificial chromosome
Mb	Megabase pairs
MeIQx	2-Amino-3,8-dimethylimidazo-(4,5-f)-quinoxaline
MEM	Minimum essential medium
MGD	Mouse Genome Database
MGI	Mouse Genome Informatics
MGP	Mouse Genome Project
MP	Multi-purpose
mRNA	Messenger RNA
MSR	Macrophage scavenger receptor
mt	Mutant type
NAT	Arylamine N-acetyltransferase enzyme
NCBI	National Center for Biotechnology Information
NCI	National Cancer Institute
NHGRI	National Human Genome Research Institute
NIDDM	Non-insulin-dependent diabetes mellitus
NIH	National Institute of Health
NTDs	Neural tube defects
OLA	Oligonucleotide ligation assay
pABA	p-Aminobenzoic acid
pABGlu	p-Aminobenzoyl glutamate
PAC	P1 Artificial Chromosome
PAHs	Polycyclic aromatic hydrocarbons
PAPS	3'-phosphoadenosine-5'phosphosylphate
pAS	p-Aminosalicylic acid
PCR	Polymerase Chain Reaction
PFGE	Pulsed field gel electrophoresis
PhIP	2-Amino-1-methyl-6-phenylimidazo[4,5-b]pyridine
RAPD	Randomly amplified polymorphic DNA
RDA	Representational difference analysis
RFLP	Restriction fragment length polymorphism
RT-PCR	Reverse transcriptase polymerase chain reaction
s	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SINE	Short interspersed element
SMZ	Sulphamethazine
SNP	Single nucleotide polymorphism
SSCP	Single strand conformational polymorphism
SSLP	Single sequence length polymorphism
SSR	Simple sequence repeat
ST	Sulphotransferase

STR	Short tandem repeats
STS	Sequence tagged sites
TCA	Trichloroacetic acid
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
TK	Thymidine kinase
TSG	Tumour suppressor gene
UDP	Uridine diphosphate- α -1-glucuronic acid
UDP-GT	UDP-glucuronosyl transferase
UTR	Untranslated region
UV	Ultraviolet
wt	Wild type
YAC	Yeast artificial chromosome

CHAPTER 1: Introduction

Arylamine N-acetyltransferase (NAT) has been identified as the enzymatic activity responsible for the N-acetylation of aromatic amines and hydrazines. These include a wide range of xenobiotic compounds, such as drugs and carcinogens. In this introductory chapter, a review of NATs in humans and mice will be presented. In addition, the methodologies used for characterisation and mapping of genetic loci will be described, a task greatly assisted by the ongoing Human and Mouse Genome Projects.

1.1 Xenobiotic metabolising enzymes

Xenobiotics are foreign compounds that enter the organism, but do not serve as nutrients or essential factors in endogenous metabolism. Such compounds may be drugs, environmental pollutants, constituents of drink and food or industrial chemicals, and may have beneficial pharmacological or dangerous toxic properties (Caldwell et al, 1995). It is of great importance to understand how xenobiotics are metabolised, as they can undergo modifications affecting their chemical properties and action within the body (La Du, 1992).

Metabolism of xenobiotic compounds is mediated by a number of enzymes, referred to as xenobiotic metabolising enzymes, although some of them may also be involved in the metabolism of endogenous substrates. For example, apart from their role in drug metabolism, cyclooxygenase, leukotriene C₄ synthase, thromboxane synthase (CYP5) and prostacyclin synthase (CYP8), are also involved in the metabolism of arachidonic acid (reviewed in Nebert, 1997). Most of the xenobiotic metabolising enzymes are expressed in the liver, although they can also be found in other tissues, such as the lungs, kidneys and intestine (Windmill et al, 1997). Many theories have attempted to explain how enzymes specialised in the metabolism of recently

synthesised chemicals have evolved, but none of them is conclusive (reviewed in La Du, 1992 and Nebert 1997). It seems more likely that they may have evolved as a response of the animals to toxins naturally produced by plants (Gonzalez and Nebert, 1990; Nebert, 1997).

The process of xenobiotic metabolism from internalisation of the exogenous compound to its excretion can be divided into two phases (Williams, 1959). Phase I reactions are often referred to as “functionalisation reactions” and include oxidation, reduction, hydrolysis (table 1.1). The role of these reactions is to introduce functional groups to exogenous compounds, often increasing their polarity. The products of Phase I reactions may then enter Phase II metabolism, but they may also have a direct cytotoxic effect, as they are highly reactive. Phase II reactions, also known as “conjugation reactions”, include glucuronidation, sulphation, acetylation, methylation (table 1.1). Phase II reactions usually lead to detoxification of the chemically reactive compounds produced during Phase I metabolism, but it is also possible that Phase II enzymes catalyse bioactivation of exogenous compounds. In most cases, Phase II reactions further increase solubility of hydrophobic compounds and lead to their elimination from the body (Idle et al, 1992; Gibson and Skett, 1994; Caldwell et al, 1995; Delaforge, 1998). A simplified overview of the metabolic pathways of xenobiotics is illustrated in figure 1.1.

Table 1.1 A list of the most important chemical reactions associated with Phase I and II of the metabolism of xenobiotics (modified from Gibson and Skett, 1994).

Phase I (Functionalisation) Reactions	Phase II (Conjugative) Reactions
Oxidation involving cytochrome P450	Glucuronidation
Oxidation not involving cytochrome P450	Sulphation
Reduction	Acetylation
Hydrolysis	Methylation
Hydration	Glutathione conjugation
Isomerisation	Amino acid conjugation
Dethioacetylation	Fatty acid conjugation
	Condensation

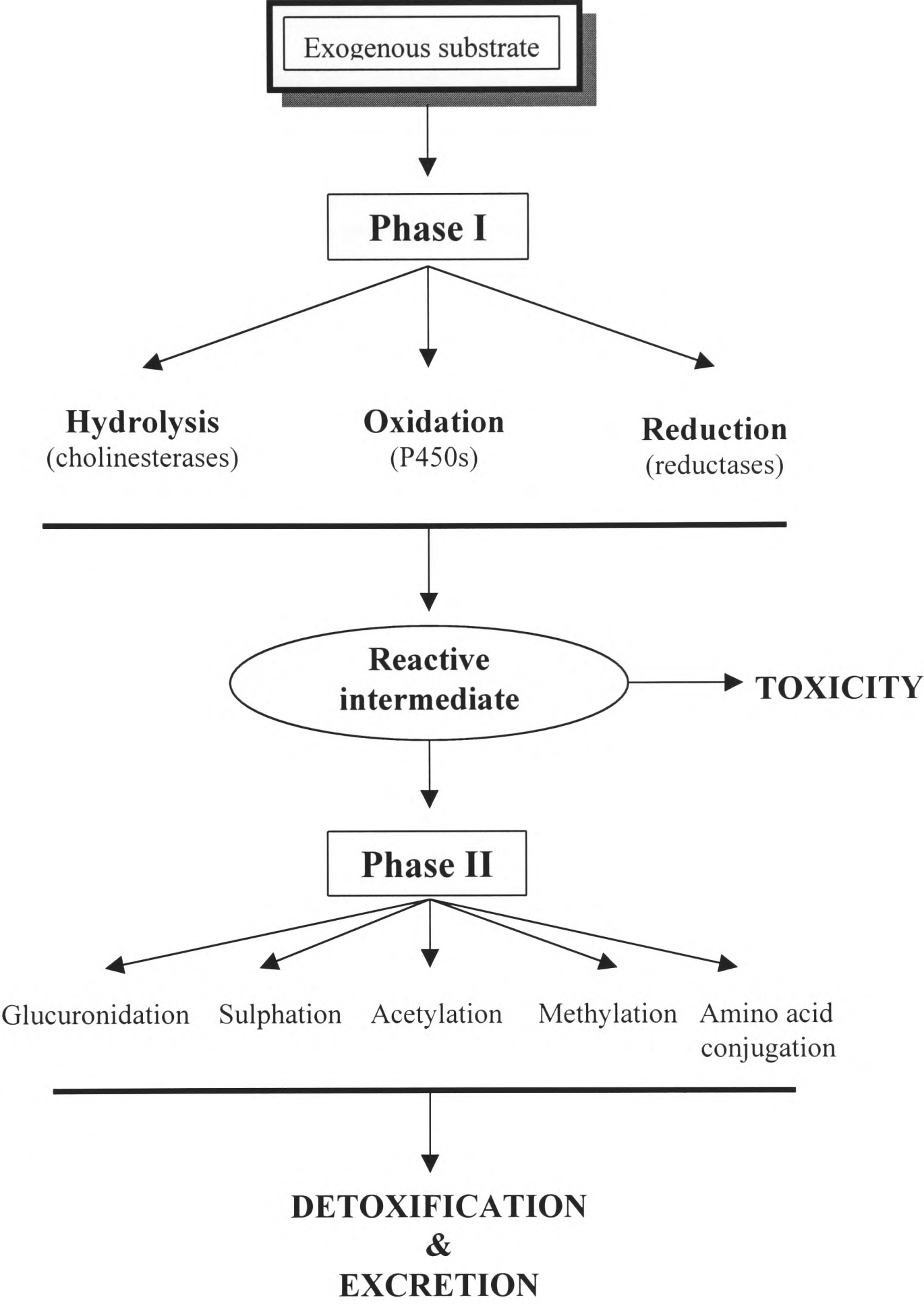
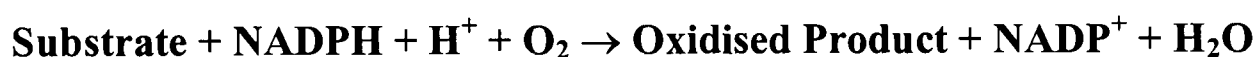


Figure 1.1 Schematic overview of the metabolism of xenobiotics (adapted from Idle et al, 1992; Delaforge, 1998). It is recognised today that Phase I reactions may also directly lead to detoxification of the foreign compound and excretion, while Phase II enzymes can also catalyse bioactivation of xenobiotics into toxic compounds.

1.1.1 Phase I reactions

Only one group of Phase I xenobiotic metabolising enzymes will be discussed in this section, the microsomal mixed-function oxidase (cytochrome P450) system, because of its involvement in the oxidation of a wide range of exogenous substrates (Wong, 1998; Wormhoudt et al, 1999), several of which may also be substrates for NAT (Grant, 1993). Other Phase I enzymes are also involved in oxidation reactions, such as the alcohol and aldehyde dehydrogenases, amine oxidases and aromatases. Some enzymes are responsible for hydration (e.g. cholinesterases and arylesterases) and reduction (reductases) (Idle et al, 1992; Gibson and Skett, 1994).

Cytochrome P450s comprise the most important enzymatic system of functionalisation reactions. They are located on the membranes of the endoplasmic reticulum, mitochondria and microsomes of several cell types and have been identified as the terminal oxidase activity of membrane-based electron transfer systems within these organelles. All cytochrome P450 enzymes contain a heme group, the ferric ion of which undergoes reduction/oxidation during a NADPH-mediated catalytic process, the general scheme of which may be described as follows (Schenkman and Gibson, 1983; Delaforge, 1998; Wong, 1998):



Cytochrome P450s are responsible for Phase I oxidation of many xenobiotics and may be involved in carcinogenesis via activation of procarcinogens (Delaforge, 1998; Wormhoudt et al, 1999). They form a large superfamily of enzymes and efforts have been made for their classification into families and subfamilies, based on sequence similarity. Polymorphisms have been identified in several genes encoding for P450s, including the *CYP1A1*, *CYP1A2*, *CYP2A6*, *CYP2D6* and *CYP3A4*. These polymorphisms may affect the enzymatic activity of the P450s involved, which may in turn be associated with different predisposition of individuals to chemical carcinogenesis (reviewed in Gonzalez and Nebert, 1990; Delaforge, 1998; Hirvonen, 1999; Wormhoudt et al, 1999).

1.1.2 Phase II reactions

The main Phase II xenobiotic metabolising enzymes, which will be described in this section, are the UDP-glucuronosyltransferases, sulphotransferases and glutathione S-transferases. N-acetyltransferases, which form another major group of conjugation enzymes, will be reviewed in section 1.3.

Glucuronidation involves transfer of glucuronic acid from uridine diphosphate- α -1-glucuronic acid (UDP-glucuronic acid) to a variety of endogenous and xenobiotic compounds with amine, hydroxyl sulphydryl, or carboxylic groups. Glucuronidation is the most widespread of the conjugation reactions and is catalysed by a family of enzymes called UDP-glucuronosyltransferases (UDP-GTs). Xenobiotic substrates of these enzymes include phenols, alcohols and carboxylic acids, which undergo O-glucuronidation, aromatic amines, amides and sulphonamides, which undergo N-glucuronidation, and other thiols undergoing S-glucuronidation (Bock and Lilienblum, 1994; Gibson and Skett, 1994; Caldwell et al, 1995; McGurk et al, 1998; Green and Tephly, 1998). An important endogenous role for UDP-GT is the glucuronidation of bilirubin in the catabolism of heme. UDP-GT enzymes are also involved in the metabolism of common drugs, like paracetamol and morphine, as well as in the metabolism of aromatic amine carcinogens of the bladder, such as benzidine, 4-aminobiphenyl and α - or β -naphthylamine. Following N-glucuronidation, these carcinogens are transported to the bladder, where they are converted into the ultimate carcinogen (Bock and Lilienblum, 1994; Zenser et al, 1998; Green and Tephly, 1998).

Conjugation with sulphate is an important phase II reaction leading to the detoxification of a range of xenobiotics or their primary metabolites. The reaction is catalysed by a group of cytosolic enzymes called sulphotransferases (ST) and involves transfer of a sulphate group from 3'-phosphoadenosine-5'-phosphosulphate (PAPS) to the substrate. Sulphation is important in the metabolism of endogenous and exogenous substrates carrying a hydroxyl, amine, or thiol group. Compounds metabolised by sulphation include alcohols, phenols, arylamines, steroids and bile acids (Caldwell et al, 1995; Windmill et al, 1998; Dooley, 1998). STs are also involved in the detoxification of carcinogens. However, compared with other

conjugation reactions, sulphation is more likely to lead to the formation of reactive products. Genotoxic effects mediated by ST activity have been described for many aromatic hydroxylamines and benzylic or allylic alcohols by *in vitro* studies (Glatt, 1997; Surh, 1998; Dooley, 1998; Glatt et al, 2000). Of the ten isoforms of xenobiotic metabolising ST enzymes identified in humans to date, two represent splice variants of the same gene, while the rest are encoded by individual loci. Polymorphisms have been detected in two types of ST and their importance in toxicity and carcinogenesis is currently under investigation (Glatt et al, 2000).

Glutathione is a tripeptide whose sequence is γ Glu-Cys-Gly. In a reaction catalysed by a group of enzymes called glutathione-S-transferases (GST), glutathione is conjugated to electrophilic compounds, such as products of Phase I reactions. GST enzymes have been detected in all aerobic organisms and the importance of glutathione conjugation in protecting macromolecules from electrophilic attack has been demonstrated. In humans, GSTs are expressed in virtually all tissues, including the liver, kidney and colon. In the liver, they constitute a significant proportion (up to 5%) of the total cytosolic protein (Smith et al, 1994; Caldwell et al, 1995; Wormhoudt et al, 1999). In addition to their role in the detoxification of carcinogenic substrates, GSTs can also metabolise compounds used for cancer chemotherapy, leading to decreased effectiveness of these drugs. As a result, glutathione conjugation is suspected to be involved in multidrug resistance of cancer patients (Smith et al, 1994; Wormhoudt et al, 1999). To date, seven gene families encoding for cytosolic GST isoenzymes have been described. Expression of several of these isoenzymes may be induced by xenobiotic and antioxidant compounds (Smith et al, 1994; Strange et al, 2000). Because of their involvement in cell protection, dietary induction of GST enzymes might be useful as a potential chemopreventive strategy for individuals at high risk for cancer (Clapper and Szarka, 1998). Polymorphisms in the GST genes have been of interest, as they may be associated with variable susceptibility to cancer. So far, five classes of GSTs have been described as polymorphic, some of the polymorphisms being due to gene deletions resulting in complete loss of enzymatic activity (null genotype). The role of GSTs in protection against the genotoxic effects of reactive electrophiles suggests that deficiency in GST activity might confer increased risk for cancer. This hypothesis is generally valid for several types of

cancer, although some conflicting results have also been reported (Strange et al, 1998; Wormhoudt et al, 1999; Strange et al, 2000).

1.2 Pharmacogenetics

A common observation in clinical practice is that the therapeutic effect of certain drugs varies among individuals. Often, the standard dose of a drug is not sufficient to treat some patients, while it may be excessive or even life threatening for others (reviewed in La Du, 1992; May, 1994). The involvement of a genetic factor has been suspected in many cases where the same response is reproducible for the same person or members of the same family. For example, debrisoquine, a drug formerly used for the treatment of hypertension, was observed to exhibit striking inter-individual variation in its pharmacological effect and rate of excretion. It was demonstrated that this was due to the ability of some patients to transform debrisoquine to its inactive 4-hydroxy derivative less efficiently than others. Subjects with low conversion rates (poor metabolisers) were shown to be more sensitive to the anti-hypertensive effect of the drug, as well as more susceptible to idiosyncratic reactions. A genetic basis for this phenomenon was established and the enzyme responsible was later identified to be the CYP2D6, a cytochrome P450 (reviewed in La Du, 1992; Nebert, 1997). Another classical example is that of arylamine N-acetyltransferase, a slow acetylating variant of which has been demonstrated to be responsible for increased susceptibility to the side effects of isoniazid (section 1.3).

The field of *Pharmacogenetics* was created in the 1950s on the basis of such observations and its objective is to study the genetic factors responsible for variations in response to drugs. Later, in the 1970s, the term *Ecogenetics* was introduced to describe the genetic variability in response to other xenobiotic compounds, such as environmental chemicals (La Du, 1992, for a historical review). The term *Xenogenetics* has also been used to describe the genetic control over response of an individual to foreign compounds (Sim et al, 1995). In recent years, the Human Genome Project, as well as other genome projects for model animals, have significantly contributed towards the identification of genes encoding for drug-metabolising enzymes and the understanding of the mechanisms underlying their

expression. Nowadays, more and more people use the term *Pharmacogenomics* to refer to the entirety of genes encoding for xenobiotic metabolising enzymes and affecting predisposition to the side effects of drugs and other compounds (Evans and Relling, 1999). Molecular biological techniques have also enabled detection of polymorphisms in several genes encoding for xenobiotic metabolising enzymes and identification of “high-risk” individuals (Meyer, 1992).

In the future, the discipline of Pharmacogenetics is anticipated to affect medical practices, so that adjustments to the dosage of drugs can be made according to the genetic background of the individual, in addition to age, weight, sex, medical history and dietary and other environmental factors (May, 1994; Iyer and Ratain, 1998; Evans and Relling, 1999), as illustrated in figure 1.2. Because of its clinical impact, Pharmacogenetics has focused more on the adverse reactions to drugs than the natural role of the enzymes involved. However, it is reasonable to assume that the enzymes which can metabolise modern synthetic chemicals originally evolved to metabolise toxic chemicals in food or may also have an endogenous function (Nebert, 1997).

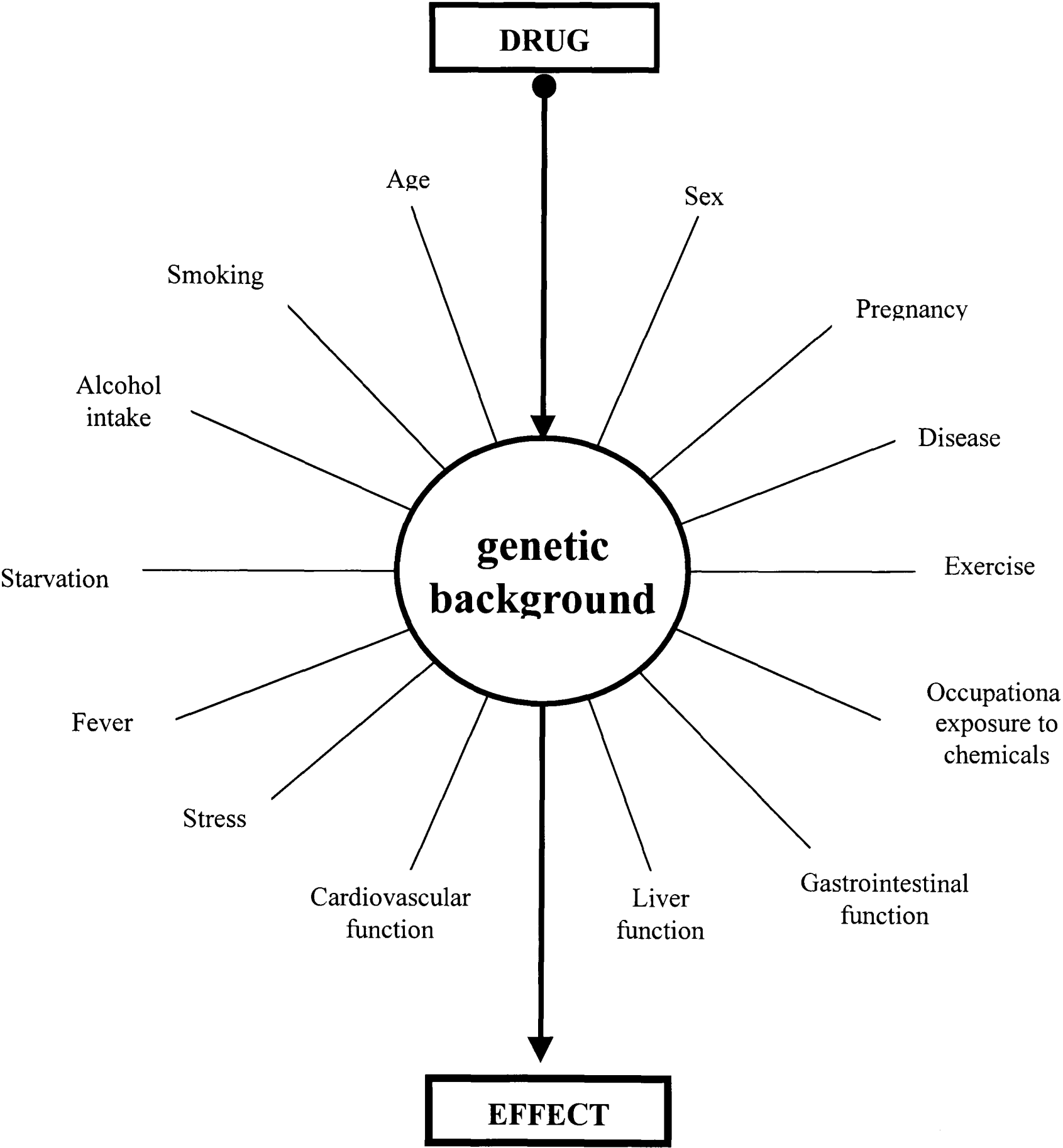


Figure 1.2: External and internal factors affecting drug disposition in humans (modified from Vesell, 1983).

1.3 Arylamine N-acetyltransferases in humans

Arylamine N-acetyltransferases were originally identified as the enzymes responsible for the detoxification of isoniazid, a drug extensively used for its anti-tuberculosis properties for almost fifty years. Following its widespread use, however, it was realised that significant interindividual differences in the metabolism of isoniazid existed. Early data suggested a bimodal distribution of the drug metabolites in different patients, with some individuals showing significantly lower levels of the inactivated drug in their blood and urine, compared with others. These individuals also had higher levels of native drug in their circulation and this was associated with higher predisposition to side effects, such as nerve and hepatic toxicity. It was evident that different individuals varied in their efficiency to metabolise isoniazid and this led Mitchell and Bell (1957) to the conclusion that patients could be classified either as fast or slow inactivators of isoniazid. Thereafter, the genetic basis of this variation was extensively investigated by numerous epidemiological and familial studies (Price Evans, 1992; Sim et al, 1992; Vatsis and Weber, 1994, for historical reviews). The rate of isoniazid inactivation in fast and slow acetylators was later correlated with the consumption of acetyl coenzyme A (acetyl-CoA), indicating that isoniazid metabolism involved an enzyme which catalysed the transfer of an acetyl group from acetyl-CoA to isoniazid (Evans and White, 1964). The term “inactivator” was subsequently replaced by the term “acetylator”.

The enzyme responsible for the acetylation of isoniazid and other drugs of similar chemical structure was later purified 300-fold and its properties were examined (Jenne, 1965). The isoforms isolated from fast and slow acetylators showed very similar physicochemical and kinetic properties, but it was observed that, although isoniazid was polymorphically acetylated, p-aminosalicylic acid was not. This suggested that at least two distinct enzymatic activities, one “polymorphic” and one “monomorphic”, should be responsible for acetylation in human liver (Jenne, 1965). In the years following Jenne’s early study, two distinct isoenzymes were identified as responsible for the N-acetylation of arylamines and hydrazines, including isoniazid and p-aminosalicylic acid, called arylamine N-acetyltransferases (NATs). Grant et al (1989b) purified NAT from human liver 1000-fold and identified two enzymatic

activities in the cytosolic fraction, with very similar properties. The presence of two NAT isoenzymes in humans was later confirmed by cloning the corresponding genes (Blum et al, 1990) and expressing them in transiently transfected CHO (Ohsako and Deguchi, 1990) and COS-1 (Grant et al, 1991) cells. It was demonstrated that NAT1 was responsible for the N-acetylation of “monomorphic” substrates, such as p-aminosalicylic acid and p-aminobenzoic acid (pABA), while NAT2 was responsible for the N-acetylation of “polymorphic” substrates, such as sulphamethazine (SMZ) and procainamide. It was later identified, however, that NAT1 is also polymorphic (Vatsis and Weber 1993), as will be described in section 1.3.3.2. Certain substrates are efficiently metabolised by both NAT isoenzymes, such as the carcinogenic arylamine 2-aminofluorene, which indicated that the two isoenzymes have partially overlapping substrate specificities (Grant et al, 1991). Other substrates metabolised by both NAT1 and NAT2 include benzidine, 4-aminobiphenyl, β -naphthylamine (reviewed in Grant, 1993). The chemical structure of the most common NAT substrates is shown in figures 1.3-1.5 (adapted from Weber and Hein, 1985; Price Evans, 1992; Hanna, 1996).

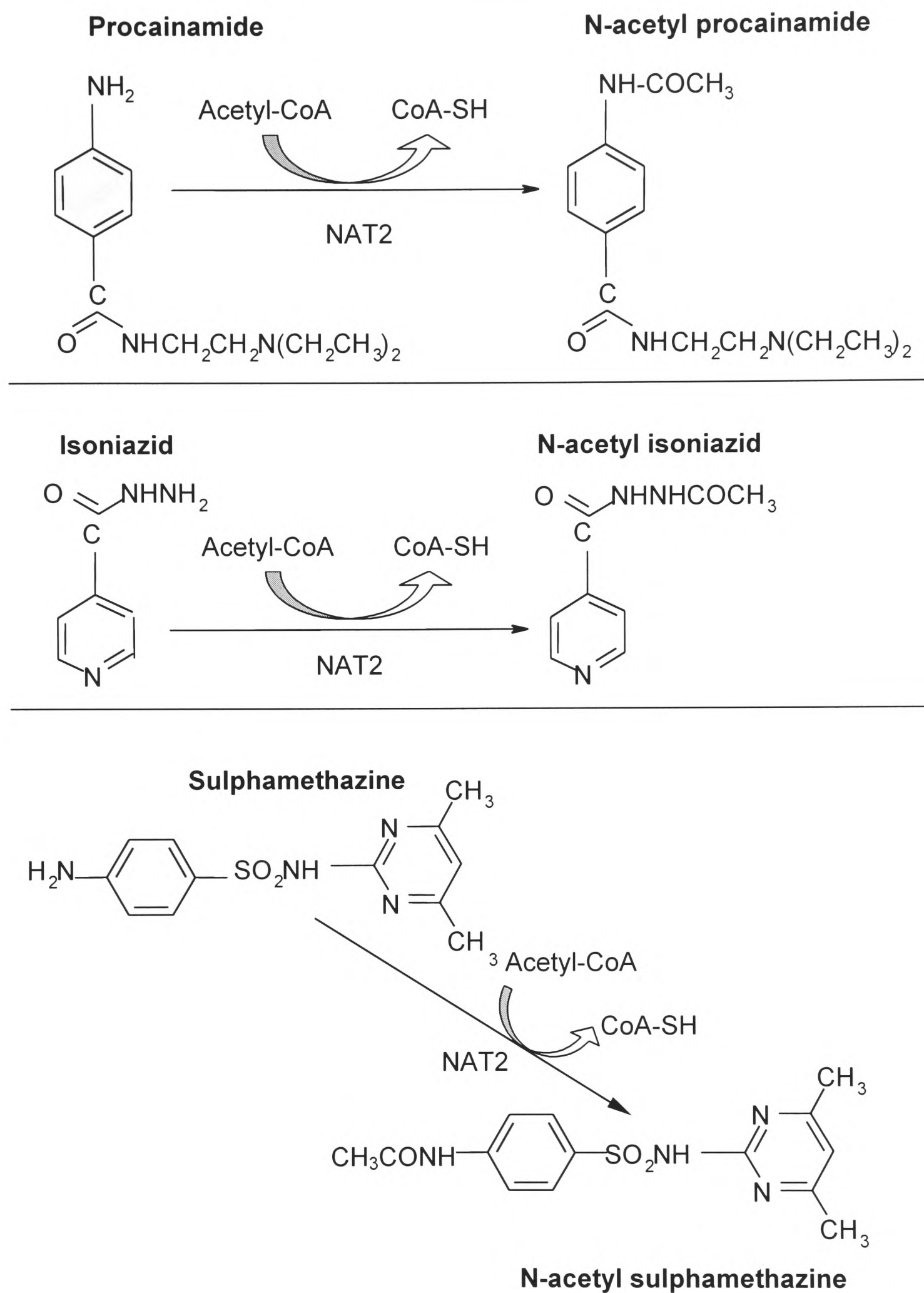


Figure 1.3: Chemical structure of common NAT2 substrates in humans. The N-acetylation reaction, catalysed by NAT2, involves conversion of acetyl-CoA to its non-acetylated form (CoA-SH). The structure of the N-acetylated product of each compound is also illustrated.

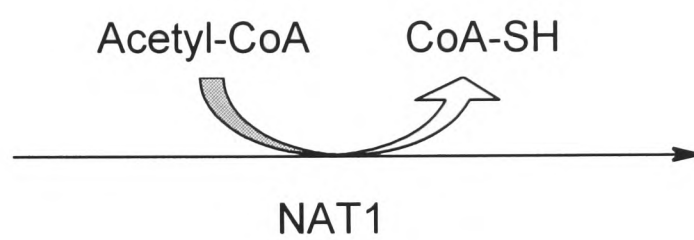
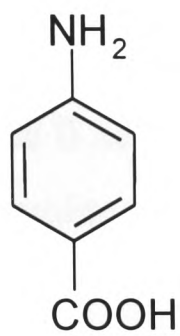
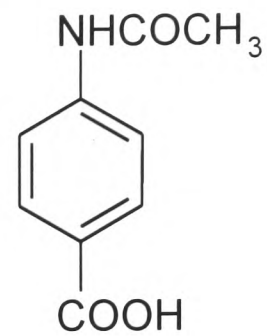
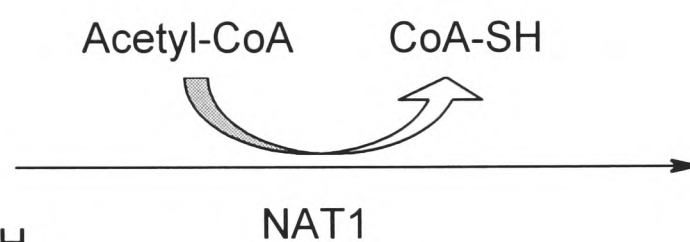
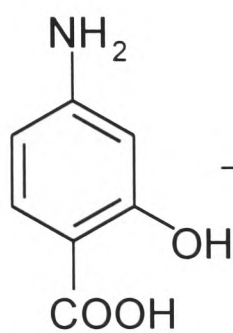
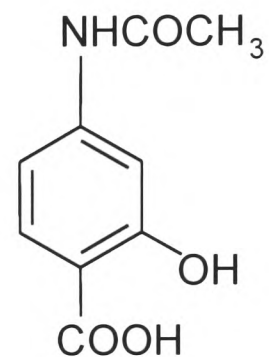
p-aminobenzoic acid**N-acetyl p-aminobenzoic acid****p-aminosalicylic acid****N-acetyl p-aminosalicylic acid**

Figure 1.4: Chemical structure of common NAT1 substrates in humans. The N-acetylation reaction, catalysed by NAT1, involves conversion of acetyl-CoA to its non-acetylated form (CoA-SH). The structure of the N-acetylated product of each compound is also illustrated.

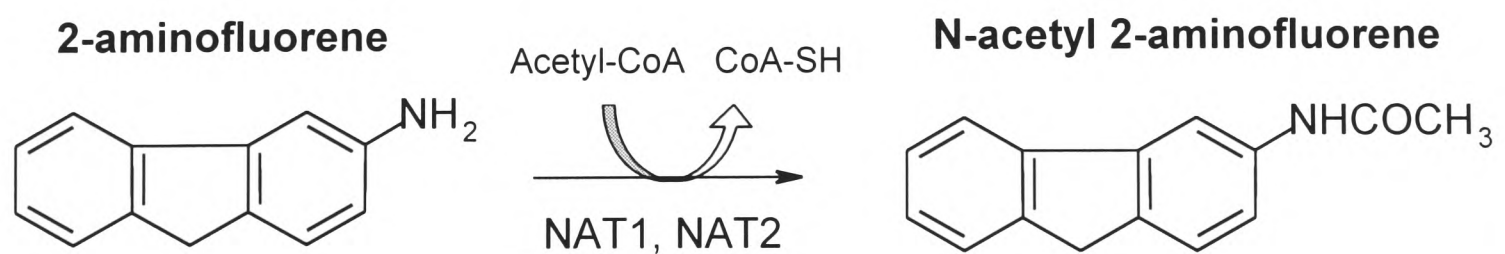


Figure 1.5: Chemical structure of 2-aminofluorene, a common substrate for both NAT1 and NAT2 in humans. The N-acetylation reaction involves conversion of acetyl-CoA to its non-acetylated form (CoA-SH). The structure of the N-acetylated product is also illustrated.

1.3.1 Human arylamine N-acetyltransferases in drug metabolism

Human NATs are cytosolic enzymes, 31kDa in size. They are classified as typical phase II xenobiotic metabolising enzymes, responsible for the metabolism of various arylamine and hydrazine drugs and carcinogens (reviewed in Weber and Hein, 1985; Sim et al, 1992). To date, all substrates identified for human NATs are exogenous, with the exception of p-aminobenzoic glutamate (pABGlu) which is a folate breakdown product and has been proposed as a potential endogenous substrate for human NAT1 (Ward et al, 1995; Minchin, 1995), as will be described in section 1.3.5.

Several drugs are N-acetylated by human NATs and this generally leads to their inactivation and production of metabolites that are excreted in urine. Polymorphic N-acetylation has been associated with differential susceptibility of individuals to the side effects of these drugs, and therefore NATs have long been the focus of pharmacologists and pharmacogeneticists. NATs are involved in the metabolism of primary arylamine (e.g. procainamide and the sulphonamides dapsone and sulphamethoxazole) and hydrazine (e.g. isoniazid and hydralazine) drugs, but they can also N-acetylate drugs which are converted to arylamines or hydrazines inside the body. These include sulphasalazine, nitrazepam and clonazepam, and caffeine (reviewed in Weber and Hein, 1985; Price Evans, 1992; Hanna, 1996). Many of these drugs are subject to polymorphic N-acetylation by human NAT2 and slow acetylators may be more susceptible to possible toxic effects. Examples are summarised in table 1.2.

Table 1.2: Common drugs metabolised by NAT2 in humans. The original use and the potential side effects associated with the acetylator phenotype are described (adapted from Weber and Hein, 1985; Price Evans, 1992).

Drug	Therapeutic Use	Side Effects associated with NAT2 polymorphism	NAT2 phenotype responsible for side effects
Primary Arylamine Drugs			
Procainamide	Cardiac arrhythmia prevention and therapy	Systemic lupus erythematosus-like syndrome	Slow
Dapsone	Leprosy and malaria treatment	Haemolytic anaemia	Slow
Primary Hydrazine Drugs			
Isoniazid	Tuberculosis treatment	Neurotoxicity, hepatotoxicity	Slow
		Insufficient treatment	Fast
Hydralazine	Hypertension treatment	Systemic lupus erythematosus-like syndrome	Slow
Phenelzine	Anti-depressant, monoamine oxidase inhibitor	Slower psychiatric response	Fast
Secondary Arylamine and Hydrazine Drugs			
Sulphasalazine	Treatment of chronic inflammatory disorders (e.g. ulcerative colitis, Crohn's disease, rheumatoid arthritis etc)	Nausea, headache, vomiting, rashes, haemolysis, haematological disorders	Slow
Nitrazepam	Insomnia and anxiety treatment	Drowsiness, hangover feeling, confusion	Relationship with acetylator phenotype unclear

1.3.2 Mechanism of action of arylamine N-acetyltransferases

The N-acetylation reaction catalysed by NAT is acetyl-CoA dependent and proceeds in a “ping-pong Bi-Bi” mechanism (Weber and Cohen, 1967; Riddle and Jencks, 1971; Jencks et al, 1972). The reaction involves two steps: First, acetyl-CoA binds to the sulphhydryl group of cysteine at position 68 (Cys⁶⁸) of the NAT protein, to form a thio-ester intermediate (Andres et al, 1983). Then, the substrate interacts with the active site of the enzyme and the acetyl group is transferred from acetyl-CoA to the substrate. Two very well conserved arginine residues may also be important for enzymatic function, as it has been demonstrated that mutation of Arg⁹ and Arg⁶⁴ of human NAT1 may affect stability of the enzyme (Deloménie et al, 1997). Recent evidence has further indicated that these arginine residues should be responsible for maintaining the overall three-dimensional fold of the protein, rather than for catalysis (Sinclair et al, 2000).

Human NAT1 and NAT2 show 81% identity at the amino acid level. Most of the differences are clustered near the carboxy-terminal end of the molecule, while the amino-terminal part is better conserved among different NATs. This would indicate that the carboxy-terminal part might be responsible for substrate recognition (Vatsis et al, 1995). Cloning and expression of the first 204 amino acids of human NAT1 showed that the truncated protein could efficiently bind acetyl-CoA, but could not bind the substrate and was not capable of N-acetylation (Sinclair and Sim, 1997).

Recently, the three-dimensional structure of one member of the NAT family from *Salmonella typhimurium* was resolved. This revealed that the active site of the enzyme consists of a highly conserved catalytic triad formed by residues Cys⁶⁹, His¹⁰⁷ and Asp¹²² (Sinclair et al, 2000). The highly conserved Cys⁶⁹ residue (Cys⁶⁸ in human NAT isoenzymes) is essential for function of the enzyme, as it binds the acetyl group during catalysis, to form the acetylated intermediate from which the acetyl group is then transferred to the substrate (Andres et al, 1983). Substitution of Cys⁶⁹ with an alanine residue (Ala⁶⁹) by site-directed mutagenesis resulted in a form of *S. typhimurium* NAT that lacked acetyltransferase activity (Watanabe et al, 1992).

Although N-acetylation of arylamines is generally considered to lead to detoxification of these potentially carcinogenic compounds, it has now become clear that NATs may also be responsible for the bioactivation of procarcinogens via O-acetylation of hydroxylamines or intramolecular N,O-acetyltransfer of arylhydroxamic acids (reviewed in Weber and Hein, 1985; Grant, 1993; Hanna, 1996), as will be described in section 1.3.7. These catalytic properties of NAT suggest an important role in chemical carcinogenesis. There is evidence for association of NAT activity with predisposition to various forms of cancer (section 1.3.7). The potential of NATs to bioactivate procarcinogens via O-acetylation has been extensively studied using various models, including O-acetyltransferase deficient strains of *S. typhimurium* expressing human NATs (Oda et al, 1999) and inbred or congenic mouse strains (section 1.4.1.4).

1.3.3 The NAT polymorphism in humans

Polymorphisms of the *NAT1* and *NAT2* genes, which encode for the two NAT isoenzymes in humans, have been extensively studied, mainly because of their possible involvement in human disease. The *NAT1* gene was initially considered to be monomorphic, but it has now been established that it is multi-allelic, similarly to the *NAT2* gene (reviewed in Hein et al, 2000b). For historical reasons, however, the *NAT2* alleles will be described first.

1.3.3.1 The NAT2 polymorphism

Human NAT2 is the enzyme responsible for the classical acetylation polymorphism originally observed with isoniazid (Hughes et al, 1954). Early studies had suggested the presence of at least two alleles for *NAT2*, one causing the slow and the other the fast acetylator phenotype (Evans et al, 1960). However, the full extent of the NAT2 polymorphism at the genetic level only became apparent following the isolation, cloning and sequencing of the *NAT2* gene (Blum et al, 1990; Ohsako and Deguchi, 1990; for reviews Sim and Hickman, 1991; Sim et al, 1992). To date, twenty-seven NAT2 alleles have been identified (summarised in table 1.3) and are listed on the arylamine N-acetyltransferase nomenclature web page

(www.louisville.edu/medschool/pharmacology/NAT.html), which is updated regularly.

The majority of the *NAT2* alleles confer a slow acetylator phenotype and only a few are responsible for the fast phenotype. *NAT2*4* is the most common of the *NAT2* alleles, associated with the fast phenotype and is considered as the wild type (Hickman and Sim, 1991). All other alleles are defined by comparison of their sequence with that of the *NAT2*4* allele, as shown in table 1.3 (Vatsis et al, 1995). There are three more alleles identified as fast, designated *NAT2*12A*, *NAT2*12B* and *NAT2*13* (Lin et al, 1993). *NAT2*13* carries one C to T nucleotide substitution at position 282 of the coding region. This is a silent mutation and does not affect the activity of the produced enzyme (Doll et al, 1995). *NAT2*12A* and *NAT2*12B* have a G, instead of an A, at nucleotide position 803, which causes substitution of Lys by Arg at position 268 of the amino acid chain. The *NAT2*12A* allele carries only this mutation, but the *NAT2*12B* carries an additional C to T silent mutation at position 282, similarly to the *NAT2*13* allele (Cascorbi et al, 1996a; Agundez et al, 1996c).

Mutations have been described for a total of twelve different positions within the *NAT2* coding region (table 1.3). The *NAT2* alleles that are responsible for the slow acetylation phenotype carry up to three mutations in any of these twelve “hot spots”, causing at least one amino acid change in the encoded enzyme (Vatsis et al, 1995; Grant et al, 1997). Often, different alleles have one or more mutation in common and can be grouped on the basis of these similarities. For example, all six *NAT2*5* alleles (*NAT2*5A* to *NAT2*5F*) share the characteristic T to C point mutation at nucleotide position 341, that results in a Thr substituting an Ile at amino acid position 114. Similarly, all *NAT2*6* alleles contain an A, instead of a G, at nucleotide position 590, which causes substitution of Arg by Gln (table 1.3).

Genotyping methods for the detection of the *NAT2* alleles have been established, but require revision on a regular basis to include newly discovered alleles (Blum et al, 1991; Hickman and Sim, 1991; Vatsis et al, 1991; Hickman et al, 1992; Agundez et al, 1994; Doll et al, 1995). Although these genotyping methods have been efficiently used in the past, the large number of different alleles, as well as unidentified alleles may be a common source of error in genotyping studies (Hein et al, 2000a).

There is a good correlation between specific *NAT2* genotypes and the resulting phenotypes, so it is possible to accurately predict the acetylator status of an individual, based on his/her genotype alone. As screening of large populations for all twenty-seven *NAT2* alleles described to date may be difficult or impractical, it has been suggested that detection of mutations at positions 282 and 341 of the *NAT2* coding region could be sufficient to distinguish between different phenotypes (Cascorbi and Roots, 1999). These two mutations are present in the most common of the slow alleles in Caucasian populations, including the *NAT2*5A*, *NAT2*5B* and *NAT2*5C* alleles, which contain the T³⁴¹→C mutation, and the *NAT2*6A*, *NAT2*7B* and *NAT2*14B*, which contain the C²⁸²→T mutation. Other simplified assays for studying the *NAT2* polymorphism have been proposed and they can be very accurate in predicting the phenotype (Smith et al, 1997). A reasonable balance between accuracy of results and investment in time and laboratory resources should be achieved when designing genotyping projects. Previous knowledge of the allelic frequencies in populations of similar ethnic or disease background may be of great help, as well.

Genotyping is usually performed by a combination of PCR and RFLP, but allele specific PCR (AS-PCR) has also been used. AS-PCR methods carry the risk of non-specific amplification, while PCR-RFLP may suffer from incomplete digestion or star activity by the restriction enzyme. Although PCR-RFLP is more specific, neither method is infallible, therefore detection of rare or novel alleles ought to be independently verified by cloning and sequencing (Cascorbi and Roots, 1999). Other genotyping methods have been described, based on oligonucleotide ligation assays (OLA) and multiplex PCR combined with allele-specific oligonucleotide (ASO) hybridisation (Bigler et al, 1997; Labuda et al, 1999). These methods can be automated and are particularly useful and cost-effective when processing large number of samples. Therefore, they may be the methods of choice when conducting large-scale epidemiological studies.

Allele	Nucleotide and Amino Acid Changes											
NAT2*4	T ¹¹¹ Phe ³⁷	C ¹⁹⁰ Arg ⁶⁴	G ¹⁹¹ Arg ⁶⁴	C ²⁸² Tyr ⁹⁴	T ³⁴¹ Ile ¹¹⁴	A ⁴³⁴ Gln ¹⁴⁵	C ⁴⁸¹ Leu ¹⁶¹	G ⁵⁹⁰ Arg ¹⁹⁷	C ⁷⁵⁹ Val ²⁵³	A ⁸⁰³ Lys ²⁶⁸	A ⁸⁴⁵ Lys ²⁸²	G ⁸⁵⁷ Gly ²⁸⁶
NAT2*5A					C ³⁴¹ Thr ¹¹⁴		T ⁴⁸¹ Leu ¹⁶¹					
NAT2*5B					C ³⁴¹ Thr ¹¹⁴		T ⁴⁸¹ Leu ¹⁶¹			G ⁸⁰³ Arg ²⁶⁸		
NAT2*5C					C ³⁴¹ Thr ¹¹⁴					G ⁸⁰³ Arg ²⁶⁸		
NAT2*5D					C ³⁴¹ Thr ¹¹⁴							
NAT2*5E					C ³⁴¹ Thr ¹¹⁴			A ⁵⁹⁰ Gln ¹⁹⁷				
NAT2*5F					C ³⁴¹ Thr ¹¹⁴		T ⁴⁸¹ Leu ¹⁶¹		T ⁷⁵⁹ Val ²⁵³	G ⁸⁰³ Arg ²⁶⁸		
NAT2*6A				T ²⁸² Tyr ⁹⁴				A ⁵⁹⁰ Gln ¹⁹⁷				
NAT2*6B								A ⁵⁹⁰ Gln ¹⁹⁷				
NAT2*6C				T ²⁸² Tyr ⁹⁴				A ⁵⁹⁰ Gln ¹⁹⁷		G ⁸⁰³ Arg ²⁶⁸		
NAT2*6D	C ¹¹¹ Phe ³⁷			T ²⁸² Tyr ⁹⁴				A ⁵⁹⁰ Gln ¹⁹⁷				
NAT2*7A												A ⁸⁵⁷ Glu ²⁸⁶
NAT2*7B				T ²⁸² Tyr ⁹⁴								A ⁸⁵⁷ Glu ²⁸⁶
NAT2*12A										G ⁸⁰³ Arg ²⁶⁸		
NAT2*12B				T ²⁸² Tyr ⁹⁴						G ⁸⁰³ Arg ²⁶⁸		
NAT2*12C							T ⁴⁸¹ Leu ¹⁶¹			G ⁸⁰³ Arg ²⁶⁸		
NAT2*13				T ²⁸² Tyr ⁹⁴								

Allele	Nucleotide and Amino Acid Changes										
<i>NAT2*14A</i>			A ¹⁹¹ Gln ⁶⁴								
<i>NAT2*14B</i>			A ¹⁹¹ Gln ⁶⁴	T ²⁸² Tyr ⁹⁴							
<i>NAT2*14C</i>			A ¹⁹¹ Gln ⁶⁴		C ³⁴¹ Thr ¹¹⁴	T ⁴⁸¹ Leu ¹⁶¹				G ⁸⁰³ Arg ²⁶⁸	
<i>NAT2*14D</i>			A ¹⁹¹ Gln ⁶⁴	T ²⁸² Tyr ⁹⁴			A ⁵⁹⁰ Gln ¹⁹⁷				
<i>NAT2*14E</i>			A ¹⁹¹ Gln ⁶⁴							G ⁸⁰³ Arg ²⁶⁸	
<i>NAT2*14F</i>			A ¹⁹¹ Gln ⁶⁴		C ³⁴¹ Thr ¹¹⁴					G ⁸⁰³ Arg ²⁶⁸	
<i>NAT2*14G</i>			A ¹⁹¹ Gln ⁶⁴	T ²⁸² Tyr ⁹⁴						G ⁸⁰³ Arg ²⁶⁸	
<i>NAT2*17</i>						C ⁴³⁴ Pro ¹⁴⁵					
<i>NAT2*18</i>											C ⁸⁴⁵ Thr ²⁸²
<i>NAT2*19</i>			T ¹⁹⁰ Trp ⁶⁴								

Table 1.3: The identified *NAT2* alleles in humans (data from the arylamine N-acetyltransferase nomenclature web page www.louisville.edu/medschool/pharmacology/NAT.html). Mutations in the nucleotide sequence and the resulting amino acid substitutions are shown in comparison with the wild type *NAT2*4* allele. The nucleotide and amino acid numbering are in reference to the *NAT2* open reading frame with A of the AUG initiation codon, and Met of the polypeptide chain representing position 1. The wild type *NAT2*4* allele, as well as the mutant alleles *NAT2*12A*, *NAT2*12B*, *NAT2*12C* and *NAT2*13* are responsible for the fast acetylator phenotype. The phenotypic effect of the *NAT2*17*, *NAT2*18* and *NAT2*19* alleles remains unknown. The rest of the *NAT2* alleles are responsible for the slow acetylator phenotype.

1.3.3.2 The NAT1 polymorphism

The first report on polymorphism at the human *NAT1* locus (Vatsis and Weber, 1993) came much later than the establishment of the *NAT2* polymorphism and only following cloning and sequencing of the *NAT1* gene (Blum et al, 1990; Ohsako and Deguchi, 1990). To date, twenty-four alleles have been identified, named according to chronological order of discovery, as reviewed in Hein et al (2000b). The most common alleles in the Caucasian population are summarised in table 1.4. A complete list of the polymorphisms at the *NAT1* locus can be found in the NAT nomenclature page on the internet (www.louisville.edu/medschool/pharmacology/NAT.html).

The most common of the *NAT1* alleles, designated *NAT1**4 (Vatsis and Weber, 1993), is considered as the wild type (Vatsis et al, 1995) and all other alleles are defined in comparison to it (table 1.4). The *NAT1**3 was the first allele to be identified when the *NAT1* gene was cloned (Blum et al, 1990). It carries an A at position 1095, instead of the C present in the *NAT1**4 allele. This mutation is located in the 3' untranslated region (3'UTR) of the *NAT1* gene and does not cause any amino acid substitution in the protein.

The *NAT1**10 allele was discovered by direct sequencing of a *NAT1*-specific PCR product, using genomic DNA template from liver and leucocytes from individuals of Caucasian origin (Vatsis and Weber, 1993). It contains the same C to A mutation at nucleotide position 1095 as the *NAT1**3 allele, in addition to a T to A mutation at position 1088, within the 3' UTR. This latter nucleotide substitution is proposed to abolish the polyadenylation signal of the gene, by transforming the AATAAA sequence to a series of five As. The *NAT1**11 allele, discovered in the same study (Vatsis and Weber, 1993), carries a total of seven mutations, including nucleotide substitutions in the 5' and 3' UTR of the gene and a nine nucleotides long deletion close to the polyadenylation signal in the 3' UTR. It also contains three point mutations within the coding region, two of which result in amino acid substitutions (Vatsis and Weber, 1993; Doll et al, 1997).

There are two variants of the allele originally named *NAT1*14* (Hughes et al, 1998a; Payton and Sim, 1998). *NAT1*14A* and *NAT1*14B* share the same G to A mutation at position 560, that causes replacement of Arg by Gln at amino acid position 187. In addition, *NAT1*14A* also carries a T to A and a C to A substitution at positions 1088 and 1095, respectively, as described for the *NAT1*10* allele. The *NAT1*14A* and *NAT1*14B* alleles were originally identified in individuals with considerably reduced *NAT1* activity (Hughes et al, 1998a; Payton and Sim, 1998) and were later detected in patients with colorectal cancer (Hubbard et al, 1998).

The codon for Arg at position 187 of NAT1 is also mutated in the *NAT1*15* allele. In this case, there is a T, instead of a C at nucleotide position 559, introducing a stop codon at amino acid position 187 (Hughes et al, 1998a; Hubbard et al, 1998). Because of this premature stop codon, *NAT1*15* encodes for a truncated and completely inactive protein of 186, rather than 290, amino acids. The truncated protein has been expressed in *E. coli* and Western blot analysis has shown that it has a molecular weight of 23kDa, instead of 33kDa (Hughes et al, 1998a).

The *NAT1*17* allele was originally found in individuals with low NAT1 activity. It is the only allele that carries a point mutation at nucleotide position 190, where C has been replaced by T, causing substitution of an Arg by a Trp at amino acid position 64 (Butcher et al, 1998). The Arg⁶⁴ has been changed by site directed mutagenesis to Ala, Met, Gln and Lys, resulting in reduction of the enzymatic activity of the mutated proteins (Deloménie et al, 1997). This is likely to be related to its role in stabilising the overall structure of the protein (Sinclair et al, 2000).

The significance of *NAT1* polymorphism is unclear. A high (up to six-fold) inter-individual variation in NAT1 activity has been observed, as measured by acetylation of NAT1 specific substrates (Ward et al, 1992; Weber and Vatsis, 1993; Hughes et al, 1998a; Butcher et al, 1998; Bruhn et al, 1999). A bimodal distribution of NAT1 activity has been suggested by Butcher et al (1998), but this was not evident in other studies. A bimodal distribution would be analogous to the NAT2 acetylator polymorphism, but the relationship between genotype and phenotype appears to be more complicated for NAT1.

Although it is difficult to predict the NAT1 phenotype from the genotype based on our current knowledge, some *NAT1* alleles clearly encode for proteins with reduced activity. As already mentioned, the *NAT1*14A* and *NAT1*14B* alleles were originally isolated from individuals with low acetylation activity (Hughes et al, 1998a; Payton and Sim, 1998). In addition, *NAT1*15* produces a truncated inactive protein (Hughes et al, 1998a; Hubbard et al, 1998). Recently, an individual homozygous for this allele was described in a German population (Bruhn et al, 1999). This person was apparently healthy, but lacked detectable NAT1 activity. However, homozygotes for the above slow alleles are very uncommon.

The first report attempting to relate a particular variant of the *NAT1* gene to polymorphism in the acetylator phenotype in humans, referred to the *NAT1*10* allele. Bell et al (1995a) reported that the *NAT1*10* allele was associated with an approximately two fold increase in NAT1 activity, compared with the wild type *NAT1*4* allele. However, subsequent studies have failed to show similar association (Smelt et al, 1998; Payton and Sim 1998; Butcher et al, 1998; Bruhn et al, 1999). *NAT1*11* is another allele where correlation with the acetylator phenotype is unclear. It encodes for a protein with only slightly lower NAT1 activity than the wild type allele (Risch et al, 1996; Hughes et al, 1998a), but this difference has been considered as statistically significant (Bruhn et al, 1999). Although it carries several mutations, two of which result in amino acid substitutions, they appear to have a minor effect on acetylation activity and therefore *NAT1*11* cannot be classified either as a fast or a slow allele.

Correlation between genotype and phenotype for NAT1 may not be clearly established, partly due to limitations in the methodologies used for genotyping. The *NAT1* alleles are usually detected by a combination of PCR-RFLP and AS-PCR (Bell et al, 1995a; Payton and Sim, 1998). These methods can easily detect previously identified mutations, but may misdiagnose unknown alleles. For example, *NAT1*14A* and *NAT1*14B* were originally genotyped as *NAT1*10* and *NAT1*4*, respectively, due to limitations in the genotyping assays used (Payton and Sim, 1998). It is possible that other unknown mutants have been misclassified as known alleles in the past. New NAT1 alleles can be detected by sequencing, especially if their presence is suspected from their phenotypic effect (Vatsis and Weber, 1993; Payton and Sim, 1998; Butcher

et al, 1998). Heteroduplex analysis has also been used for the same purpose (Hubbard et al, 1998).

An interesting observation about *NAT* polymorphism has been the existence of a linkage disequilibrium between the *NAT1**10 and *NAT2**4 alleles (Smelt et al, 1998). This was based on genotyping of placental blood clots and cord blood for *NAT1* and *NAT2*, allowing discrimination between haplotypes of maternal and paternal origin. The haplotype *NAT1**10/*NAT2**4 was found 3.5 times more frequently than expected. Another study investigating the association between *NAT* genotypes and the risk for laryngeal cancer also indicated a linkage disequilibrium between these two alleles (Henning et al, 1999). Although the reason for this linkage disequilibrium is unclear, it has been speculated that this combination of alleles may provide protection against the carcinogenic effects of compounds like 2-aminofluorene, 2-naphthylamine and 4-aminobiphenyl, which are substrates for both NAT1 and NAT2 (Smelt et al, 1998).

Allele	Nucleotide and Amino Acid Changes											
NAT1*4	C ⁻³⁴⁴ 5' UTR	A ⁻⁴⁰ 5' UTR	C ⁹⁷ Arg ³³	C ¹⁹⁰ Arg ⁶⁴	G ⁴⁴⁵ Val ¹⁴⁹	G ⁴⁵⁹ Thr ¹⁵³	C ⁵⁵⁹ Arg ¹⁸⁷	G ⁵⁶⁰ Arg ¹⁸⁷	T ⁶⁴⁰ Ser ²¹⁴	C ¹⁰⁶⁴ to A ¹⁰⁹¹ 3' UTR	T ¹⁰⁸⁸ 3' UTR	C ¹⁰⁹⁵ 3' UTR
NAT1*3												A ¹⁰⁹⁵ 3' UTR
NAT1*10											A ¹⁰⁸⁸ 3' UTR	A ¹⁰⁹⁵ 3' UTR
NAT1*11	T ⁻³⁴⁴ 5' UTR	T ⁻⁴⁰ 5' UTR		A ⁴⁴⁵ Ile ¹⁴⁹	A ⁴⁵⁹ Thr ¹⁵³				G ⁶⁴⁰ Ala ²¹⁴	Δ9 between C ¹⁰⁶⁴ to A ¹⁰⁹¹ 3' UTR		A ¹⁰⁹⁵ 3' UTR
NAT1*14A								A ⁵⁶⁰ Gln ¹⁸⁷			A ¹⁰⁸⁸ 3' UTR	A ¹⁰⁹⁵ 3' UTR
NAT1*14B								A ⁵⁶⁰ Gln ¹⁸⁷				
NAT1*15							T ⁵⁵⁹ Stop ¹⁸⁷					
NAT1*16										AAA insertion after A ¹⁰⁹¹ 3' UTR		A ¹⁰⁹⁵ 3' UTR
NAT1*17				T ¹⁹⁰ Trp ⁶⁴								
NAT1*18A										Δ3 between A ¹⁰⁶⁴ to A ¹⁰⁸⁷ 3' UTR	A ¹⁰⁸⁸ 3' UTR	A ¹⁰⁹⁵ 3' UTR
NAT1*18B										Δ3 between C ¹⁰⁶⁴ to A ¹⁰⁹¹ 3' UTR		

Table 1.4: The most common *NAT1* alleles in Caucasian population (data from the arylamine N-acetyltransferase nomenclature web page www.louisville.edu/medschool/pharmacology/NAT.html). Mutations in the nucleotide sequence and the resulting amino acid substitutions are shown in comparison with the wild type *NAT1**4 allele. The nucleotide and amino acid numbering are in reference to the *NAT1* open reading frame with A of the AUG initiation codon and Met of the polypeptide chain representing position 1. Alleles *NAT1**14A and *NAT1**14B are associated with decreased NAT1 acetylation activity. The *NAT1**15 encodes for a truncated inactive NAT1 protein. The genotype-phenotype relationship for the rest of the alleles is not very well established.

1.3.4 Tissue-specific expression of human arylamine N-acetyltransferases

Human NAT1 and NAT2 show different expression profiles, with NAT1 ubiquitously expressed and NAT2 showing restricted tissue distribution. Both NAT1 and NAT2 were originally described as hepatic enzymes (Jenne, 1965) and efforts to isolate NAT protein (Grant et al, 1989a) and NAT mRNA (Ohsako and Deguchi, 1990) used liver as source of material. NAT1 has been detected in various other tissues, such as the kidney (Pacifici et al, 1986), bladder (Stanley et al, 1996), lung (Pacifici et al, 1986; Macé et al, 1998), mammary gland (Sadrieh et al, 1996; Williams et al, 2000) and intestine (Pacifici et al, 1986; Ilett et al, 1994; Windmill et al, 1997; Hickman et al, 1998; Windmill et al, 2000). In blood, NAT1 has been detected in the erythrocytes (Ward et al, 1992) and leukocytes, including mononuclear leukocytes, monocytes and neutrophils (Cribb et al, 1991). NAT1 expression has also been identified in several cell lines, such as the monocytic U937 cell line (Coroneos et al, 1991; Kelly and Sim, 1991; Coroneos and Sim, 1993), the urothelial U1 and U2 cell lines (Coroneos and Sim, 1991), the bladder EJ138 and RT112 cell lines (Coroneos and Sim, 1991; Stacey et al, 1998), two hepatoma cell lines, HepG2 and Hep3B (Coroneos et al, 1991; Coroneos and Sim, 1993) and tonsil lymphocytes (Coroneos et al, 1991). As it is often difficult to obtain fresh tissue from humans, using cell lines expressing NAT1 as models for N-acetylation can be useful.

Unlike NAT1, extrahepatic expression of human NAT2 has been confirmed only for the intestine (Jenne, 1965). Ilett et al (1994) were able to detect expression of NAT2 in the colorectal mucosa and described differences in acetylation activity between fast and slow acetylators using SMZ as NAT2-specific substrate. Hickman et al (1998) later confirmed presence of NAT2 in the intestine by immunohistochemistry. Both studies agree that NAT2 activity in the intestine is low, compared with NAT1 activity. However, NAT2 could still play some role in the activation of dietary carcinogens and tumorigenesis in the colon (Ilett et al, 1994; Hickman et al, 1998). Using more sensitive *in situ* hybridisation techniques, Windmill et al (1997; 2000) have been able to detect NAT2 mRNA in other parts of the gastrointestinal tract, including the oesophagus and stomach, as well as in bladder, ureter and lung. NAT2 was also detected in the lung by RT-PCR (Macé et al, 1998). However, both techniques may be subject to misinterpretations, as background staining and cross-reactivity of the

probes used for *in situ* hybridisation, or genomic contamination of the RNA template used for RT-PCR may result in false positives. Therefore, detection of NAT2 protein or enzymatic activity would be necessary to confirm these results.

1.3.5 Human arylamine N-acetyltransferase in development

NAT expression during embryonic development in humans has not been extensively studied, mainly due to the limited availability of tissue from human embryos. Pacifici et al (1986) screened a number of tissues from human foetuses at mid-gestation using the NAT1-specific substrate pABA. NAT1 activity was detected in the liver, adrenals, kidney, lung and intestine. However, the presence of NAT2 activity was not investigated in any of these tissues. Later, Pariente-Khayat et al (1991) studied the acetylator phenotype of infants during the first 15 months of age, using caffeine as a NAT2-specific substrate. They reported that all infants appeared as slow acetylators during the first few months of their life and some of them gradually developed the fast acetylator phenotype, while others remained slow acetylators. They concluded that the acetylator status could not be identified with certainty before the first 15 months of life. These results are in agreement with those of Szórády et al (1987) who reported 83% incidence of the slow phenotype in newborn infants, the highest observed among all age groups tested. These studies suggest that NAT2 expression probably begins after birth, although other researchers have detected NAT2 activity between the 9th and 12th week of pregnancy, using procainamide as substrate (Meisel et al, 1986).

Placental expression of human NAT has been studied more thoroughly, as placental metabolism may regulate the nature of environmental chemicals, including drugs and carcinogens, that reach the developing foetus. In term placentas, NAT1 activity is predominant, with NAT2 activity being at least 1000-fold lower. This suggests that biotransformation of arylamines reaching foetal circulation is carried out almost exclusively by NAT1 (Derewlany et al, 1994a; Derewlany et al, 1994b; Derewlany and Koren, 1994; Smelt et al, 1997; Smelt et al, 1998). NAT1 expression has also been detected in placentas at 11-25 weeks of gestation (Pacifici et al, 1986). A recent study demonstrated significant NAT1 activity in first trimester placentas during pregnancy, from as early as 5.5 weeks (the earliest available stage), to the third trimester. NAT2 activity was very low, but detectable (Smelt et al, 2000). The same

study also demonstrated expression of NAT1 in the preimplantation embryo, at the blastocyst stage, by RT-PCR. NAT2 expression was not evident at these early stages of development (Smelt et al, 2000). This suggests that NAT1 may have a role beyond the metabolism of xenobiotics, probably in the metabolism of an endogenous substrate of critical importance during early development. Human NAT1 has previously been demonstrated to N-acetylate the folate catabolite p-aminobenzoyl glutamate (pABGlu), which has been proposed as a potential endogenous substrate for the enzyme (Minchin, 1995; Ward et al, 1995).

Folate catabolism involves cleavage of the molecule at the C9-N10 bond to form a mixture of pteridines and pABGlu (figure 1.6). pABGlu and its N-acetylated derivative N-acetyl pABGlu (figure 1.6) are subsequently excreted in urine, as the major products of folate catabolism (McPartlin et al, 1992; Kownacki-Brown, 1993). There is a dramatic increase in the amount of pABGlu and N-acetyl pABGlu excreted in urine during pregnancy, indicating rapid folate consumption which cannot be attributed solely to weight gain of the pregnant female (McNulty et al, 1993; McPartlin et al, 1993).

N-acetylation of pABGlu by recombinant NAT1 expressed in COS-1 cells, as well as by U937 cells, has been demonstrated (Minchin, 1995). This was further confirmed using pure recombinant NAT1 expressed in *E. coli* (Ward et al, 1995). Furthermore, it has been shown that placental NAT1 activity is sufficiently high to account for the observed increase in the amount of the excreted N-acetyl pABGlu during pregnancy (Smelt et al, 2000). Taken together, these findings suggest that pABGlu could be an endogenous substrate for NAT1 in humans, hence NAT1 could be involved in folate metabolism during development.

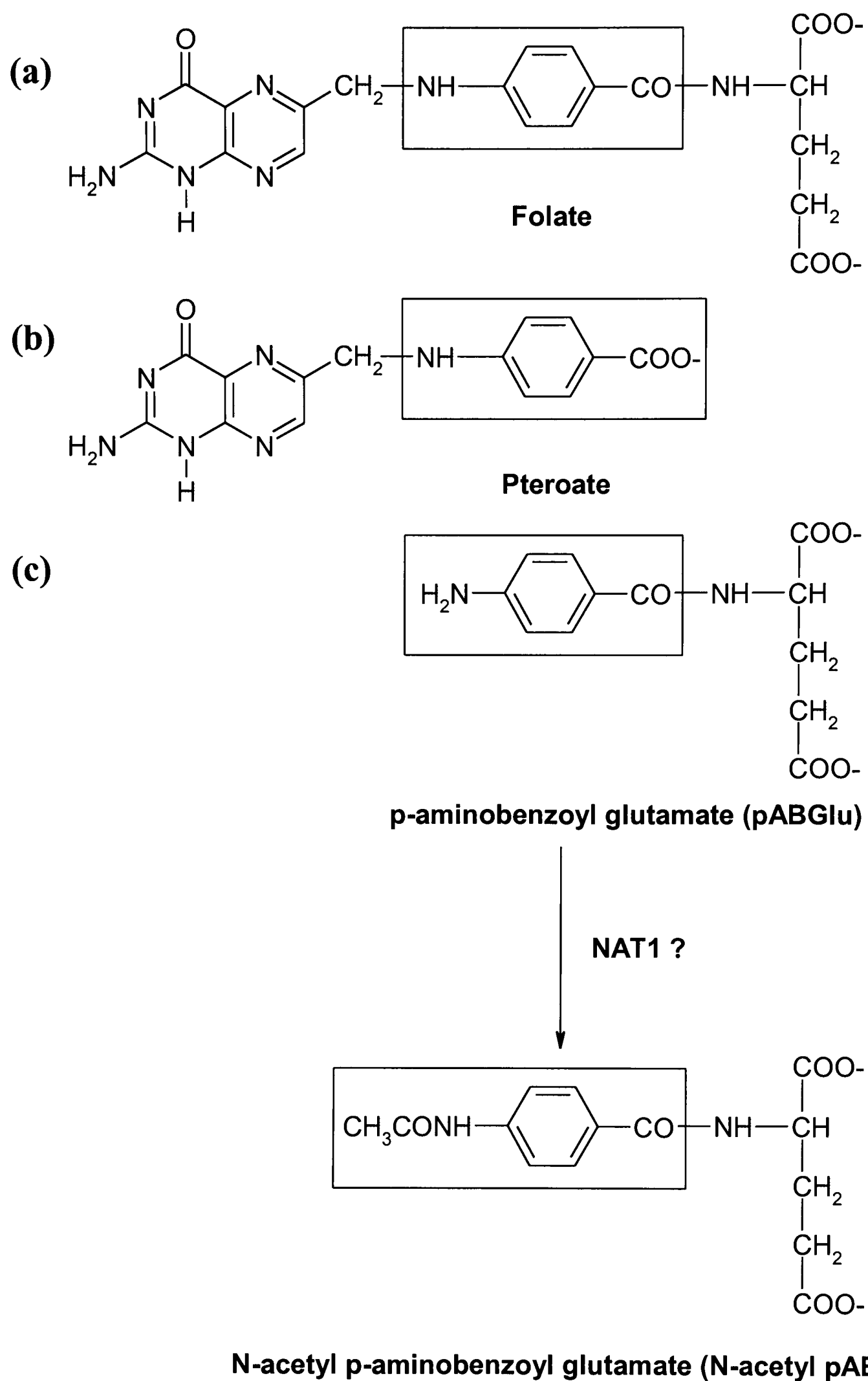


Figure 1.6: Chemical structures of folate (a) and its catabolites, pteroate (b) and p-aminobenzoyl glutamate (pABGlu) (c). N-acetylation of pABGlu, possibly catalysed by NAT1, is also illustrated (c). The boxed structures (a, b and c) are p-aminobenzoic acid (pABA) and its N-acetylated derivative (modified from Ward et al, 1995).

It is well established that folate plays a critical role during early development, protecting the embryo from neural tube defects (NTDs). Periconceptional administration of folate was shown to have a 72% protective effect in women with previous affected pregnancies (MRC Vitamin Study Research Group, 1991), while in women with no previous history, it practically eliminated the risk (Czeizel and Dudás, 1992). However, evidence suggests that folate administration is more likely to overcome some kind of metabolic block in the folate cycle, rather than a deficiency itself (Fleming and Copp, 1998). Therefore, it is likely that disturbance of the balance between folate intake and excretion could be a predisposing factor for NTDs in early development. In that context, human NAT1 could play an important role by regulating the rate of folate clearance during pregnancy. As interindividual variation in placental NAT1 activity has been observed (Smelt et al, 1997; 1998; 2000), it should be of interest to investigate whether polymorphic NAT1 expression could be a risk factor for NTDs.

1.3.6 Arylamine N-acetyltransferase and disease in humans

Complex diseases, such as diabetes, neurodegenerative disorders and many developmental defects are the result of interaction between more than one gene and various environmental factors. This makes it difficult to understand the aetiology of these polygenic disorders, since each individual genetic or environmental susceptibility factor may not be sufficient to initiate disease. The NAT enzymes are involved in the metabolism of foreign compounds, therefore, they could play an important role in the way different individuals respond to the same environmental trigger (Sim et al, 1995). For that reason, the influence of the NAT polymorphism has been investigated in a number of complex disorders, as will be described in the following sections.

1.3.6.1 NAT and diabetes mellitus

There have been several reports about involvement of the NAT2 acetylation polymorphism in diabetes mellitus. In a phenotypic study of a Saudi Arabian population, it was observed that fast acetylators were over-represented among patients suffering from insulin-dependent and non-insulin-dependent diabetes mellitus (IDDM

and NIDDM). This suggested that fast acetylators might be more susceptible to the development of either type of diabetes (El-Yazigi et al, 1992). Another study, also based on phenotyping for the acetylator status using subjects of diverse ethnic origins, has previously agreed with this observation (Ladero et al, 1982). Following cloning of the *NAT2* gene, genotyping methods which could accurately predict the phenotype, became available (section 1.3.3.1). Using these methods, however, researchers have failed to reproduce the results previously obtained by phenotyping alone (Mrozikiewicz et al, 1994; Agundez et al, 1996a). In a study conducted on children with insulin-dependent diabetes mellitus, the fast acetylators were slightly more abundant than in the reference group, but this difference was not statistically significant (Mrozikiewicz et al, 1994). Later work also failed to demonstrate any relationship between the *NAT2* polymorphism and the disease and concluded that the risk of developing diabetes mellitus is independent of the *NAT2* genotype (Agundez et al, 1996a). It would appear that the disease itself might affect the acetylation phenotype, altering the distribution of fast and slow acetylators among the patient group, irrespectively of the *NAT2* genotype.

1.3.6.2 NAT and neurodegenerative disorders

The role of the N-acetylation polymorphism in degenerative diseases of the brain, like Parkinson's and Alzheimer's, has also been investigated. Ladero et al (1989) used sulphamethazine to determine the acetylation phenotype in Parkinson's disease patients and control subjects of similar age and found no relation between the acetylator polymorphism and the risk of developing the disease. However, using genotyping methods, it was later observed that slow acetylators were significantly more common among patients with the familial form of Parkinson's disease than in controls (Bandmann et al, 1997). This indicated that fast acetylators might have an advantage in detoxifying certain neurotoxic compounds involved in the progression of the familial form of the disease. This association was not evident in a study on the sporadic form of Parkinson's disease (Harhangi et al, 1999). Another study found no apparent association between the *NAT2* genotype and Parkinson's disease (Agundez et al, 1998a). However, a higher frequency of slow acetylator genotypes was observed in patients with early-onset Parkinson's disease (onset before fifty years of age), compared with patients with late-onset, as well as with healthy control subjects.

Therefore, the role of NAT2 polymorphism in the aetiology of Parkinson's disease remains controversial.

Research regarding the possible association between N-acetylation and Alzheimer's disease has been less extensive. A study using sulphamethazine to determine the NAT2 phenotype of Alzheimer's patients did not find any relationship between the acetylator status and the risk of developing the disease. Association with the age of onset was not evident either (Ladero et al, 1993). This was further confirmed by a genotyping study on patients with sporadic Alzheimer's, where no difference was observed in the frequency of the fast *NAT2*4* allele between the patient and the control group (Rocha et al, 1999). Therefore, the NAT2 phenotype does not seem to be a susceptibility factor for Alzheimer's disease.

In view of its possible involvement in folate metabolism (section 1.3.5), NAT1 has recently been examined in relation with Alzheimer's disease (Johnson et al, 2000). Low folate levels in the blood, combined with elevated homocysteine levels, have been considered as a risk factor for the disease (Clarke et al, 1998). This suggests a possible link between NAT1 and Alzheimer's. NAT1 has been detected in the Purkinje cells of human brain by immunohistochemistry and preliminary data suggest that the *NAT1*10* allele may be more frequent among Alzheimer's than control subjects. Screening of a larger group of patients is necessary to assess the significance of this result, however (Johnson et al, 2000).

1.3.6.3 NAT and AIDS

In addition to the idiopathic diseases mentioned above, NAT has also been studied in relation to infectious disease. Lee et al (1993) determined the NAT2 acetylation phenotype in a group of patients infected with human immunodeficiency virus (HIV), using caffeine as substrate. A statistically significant over-representation of slow acetylators was observed only among patients with acute illness, compared with HIV-negative controls, asymptomatic HIV-positive individuals and AIDS patients with a stable condition. This suggested that the slow NAT2 phenotype might be a risk factor for the progression of the disease, although this was not evident in a later study by Deloménie et al (1994). A study based on both genotyping and phenotyping of HIV

patients also found no association between the acetylator status and the disease. Furthermore, it suggested that the predominance of slow acetylators observed by Lee et al (1993) was possibly due to interaction of the drugs used for treatment of the disease (Kaufmann et al, 1996). In agreement with the early observations made by Lee et al (1993), a more recent study also found that the slow genotypes were more common among patients, compared with control subjects (O'Neil et al, 1997). However, many discrepancies between the identified genotypes and the observed phenotypes were found, as indicated by the fact that some patients who were phenotyped up to three times on different occasions within a period of six to fourteen months showed changes in their capacity to acetylate caffeine. Their phenotype changed from fast to slow, or from slow to fast, or even from fast to slow to fast again, in one case. Therefore, it was concluded that the *NAT2* genotype does not correlate well with the N-acetylation phenotype in HIV-positive individuals and that prediction of the phenotype based on the *NAT2* genotype should be avoided. The reasons for this discrepancy are unclear, but may be due to the kind of medication used for treatment. Alternatively, the levels of endogenous *NAT2* could change with progression of the disease (O'Neil et al, 1997).

1.3.6.4 NAT and endometriosis

A small number of studies have investigated possible involvement of the *NAT2* polymorphism in endometriosis. This common gynaecological disorder is defined by the presence of viable endometrial tissue outside the uterus and may be responsible for 10-15% of infertility cases among women of reproductive age. Endometriosis is believed to be a disease of complex aetiology, with environmental and inherited genetic factors affecting predisposition (Kennedy, 1999). In a French population, a higher frequency of slow *NAT2* acetylators was observed among endometriosis patients, especially those with milder forms of the disease, compared with the group of controls (Baranova et al, 1999). The same study also indicated a link between the *GSTM1*-null genotype and the disease, although no synergistic effect with the slow *NAT2* genotype was observed. It is believed that genetic polymorphism of both enzymes may be associated with inefficient detoxification of environmental toxins, especially organochlorines and dioxin, affecting susceptibility to endometriosis.

1.3.7 Arylamine N-acetyltransferase and cancer

Cancer is a disease characterised by aberrant and uncontrolled cell proliferation, typically leading to the formation of tumours. It is a complex and multifactorial disease caused by the combined effect of genetic and environmental factors (Vainio and McGregor, 1992; Harrison, 1998).

Two classes of genes, termed oncogenes and tumour suppressor genes (TSGs), have been recognised to play a critical role in the initiation and progression of the disease (Harrison, 1998). When activated, oncogenes promote tumorigenesis. They are normally involved in the regulation of the cell cycle and are silenced when cell growth and proliferation are not required. Activation of an oncogene, leading to uncontrollable cell proliferation, can occur by molecular events such as point mutation, chromosomal rearrangement or viral infection. On the other hand, tumour suppressor genes (also known as anti-oncogenes) prevent the development of tumours, when functioning normally. Tumorigenesis occurs when both alleles of a TSG become inactivated by mutation (Hooper, 1998). This class of genes was originally discovered following family studies on rare types of cancers, where it was evident that members of the affected families inherited a strong susceptibility to a particular type of cancer. These individuals were demonstrated to have inherited one mutated allele of a TSG, while the other allele was subsequently inactivated by somatic mutation, loss of the relevant chromosomal segment (loss of heterozygosity) or epigenetic silencing of transcription (for example, by methylation of CpG islands in the promoter region of the TSG) (Jones and Laird, 1999).

In addition to genetics, the environment can play a major role in the process of carcinogenesis. This has been suggested by the differences observed in the geographical distribution of certain types of cancer, as well as differences in incidence, according to the individual's social background and lifestyle. More importantly, epidemiological evidence has indicated that occupational exposure to certain chemicals may be a significant risk for cancer. Chemical compounds that may have a carcinogenic effect have been identified and they usually act by causing mutations in oncogenes and tumour suppressor genes (Vainio and McGregor, 1992).

Chemical carcinogens are abundant in the environment, particularly in the industrial world. They may be used in industry (like benzidine and 4-aminobiphenyl), or may be present in cigarette smoke (like 4-aminobiphenyl), or food (like 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine, or PhIP, and 2-amino-3-methylimidazo[4,5-f]quinoline, or IQ) (Weber and Hein, 1985; Nagao and Sugimura, 1993). Many carcinogens are substrates of xenobiotic metabolising enzymes which usually catalyse detoxification of these compounds. However, these enzymes may also be responsible for bioactivation of chemicals that would not exhibit a carcinogenic effect in their native form (Hanna 1996). These compounds are often referred to as procarcinogens, while their activated metabolic products are called ultimate carcinogens. The ultimate carcinogens are typically highly reactive electrophilic compounds which bind to cellular DNA, forming adducts that may subsequently introduce mutations in oncogenes or tumour suppressor genes (Hirvonen, 1999).

Despite exposure to environmental carcinogens in our daily lives, only some individuals eventually develop cancer in their lifetime. In other words, there seems to be significant variability in predisposition to chemical carcinogenesis among different individuals. Polymorphisms in genes encoding for xenobiotic metabolising enzymes may be a significant risk factor, as these enzymes are responsible for the metabolism of many carcinogenic chemicals by either detoxifying or bioactivating them (Hirvonen, 1999). In certain cases, particular enzyme phenotypes have been associated with specific forms of cancer. For example, CYP1A1, a member of the P450 family, is involved in the metabolic activation of benzo[a]pyrene and other polycyclic aromatic hydrocarbons (PAHs). An association between the aryl hydrocarbon hydroxylase (AHH) activity of CYP1A1 and lung cancer has been reported, especially among smokers (reviewed in Caporaso et al, 1991). Glutathione S-transferases (GSTs) are detoxification enzymes, protecting the cells from the carcinogenic effect of electrophilic compounds. They are involved in the metabolism of compounds that derive from bioactivation of PAHs found in cigarette smoke. Lack of GSTM1 activity, caused by the *GSTM1* null genotype, has been identified as a moderate risk factor for lung and bladder cancer, particularly among smokers (reviewed in Hirvonen, 1999; Wormhoudt et al, 1999).

Arylamine N-acetyltransferases are good examples of polymorphic xenobiotic metabolising enzymes involved in the metabolism of carcinogens, including 2-aminofluorene, 4-aminobiphenyl, benzidine and PhIP. N-acetylation of arylamine carcinogens, catalysed by NATs, is generally recognised as a detoxification mechanism. However, it has been demonstrated that NAT is also capable of bioactivation of procarcinogens. Apart from N-acetylation by NAT, arylamines may also undergo N-oxidation by P450s, resulting in the formation of hydroxylamines. NAT can subsequently catalyse O-acetylation of these compounds to produce acetoxy esters that decompose quickly to form arylnitrenium ions (figure 1.7). These very electrophilic compounds may then act as ultimate carcinogens, attacking endogenous DNA and causing mutations. NAT may also be involved in carcinogen activation via N-acetylation, followed by N-oxidation by P450s that leads to the formation of arylhydroxamic acids. NAT may subsequently catalyse N,O-acetyltransfer of these compounds, again producing unstable acetoxy esters which decompose to form reactive arylnitrenium ions (figure 1.7) (Grant, 1993; Hanna, 1996). Considering this complex role of NAT in carcinogenesis, it is difficult to predict association between a particular acetylator phenotype and risk for a specific type of cancer. However, it has been demonstrated that some NAT phenotypes may provide protection against certain forms of cancer, while they may be predisposing factors for others.

NAT has been studied in relation with several types of cancer. The involvement of NAT in colon, bladder and breast cancer will be discussed in detail in chapter 6, while other types of cancer are discussed below.

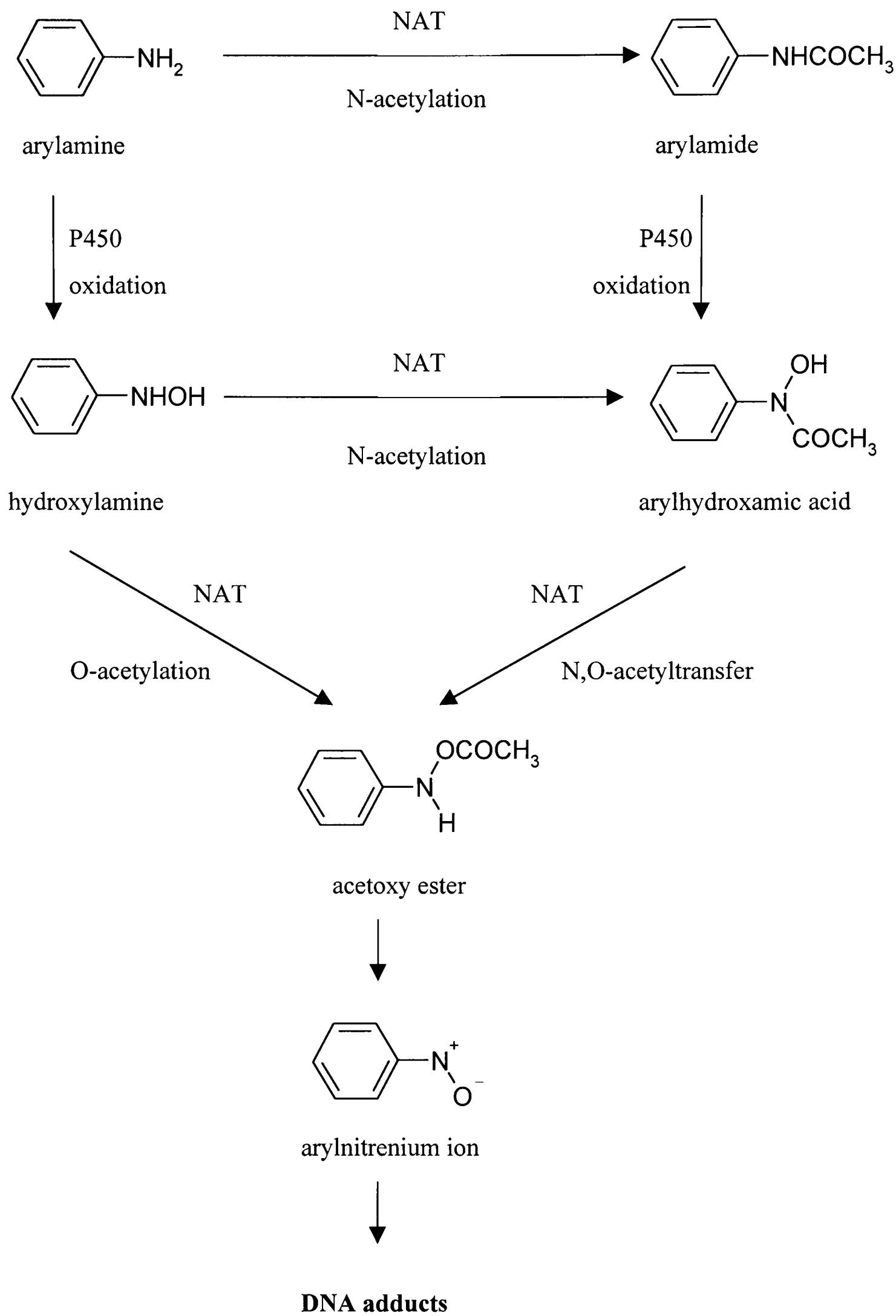


Figure 1.7: Schematic representation of the metabolic reactions catalysed by NAT (modified from Grant, 1993).

1.3.7.1 NAT in cancers of the larynx, pharynx and mouth

Carcinomas of the larynx, pharynx and oral cavity (often referred to as head and neck cancers) have been linked to tobacco smoking and alcohol consumption, as indicated by the fact that their incidence is highest in populations where these habits are common (Gonzalez et al, 1998). To date, relatively few studies have examined the NAT polymorphism as a potential susceptibility factor for these types of cancer.

In a study investigating the role of the NAT2 polymorphisms in head and neck cancers, the genotype of a combined group of Spanish patients affected by pharynx, larynx or oral cancers and of a group of control subjects from the same population was determined. Individuals homozygous for the most common slow *NAT2* alleles were more prevalent among patients, compared with the control group, but the difference between smokers and non-smokers in the patient group was not significant. Slow NAT2 activity appeared as a risk factor for the development of head and neck cancer, presumably as a result of reduced capacity for detoxification of specific chemical carcinogens (Gonzalez et al, 1998). This finding was further confirmed in the Japanese population, for carcinomas of the oesophagus (Morita et al, 1998) and the larynx (Morita et al, 1999).

The above studies suggested that NAT2 polymorphism could be used as genetic biomarker for predicting the risk for development of head and neck cancers. However, some contradictory evidence has also been reported. A study performed on French Caucasians reported that the homozygous slow acetylator genotype was associated with an increased risk for oral and pharyngeal cancer, but not with laryngeal cancer (Jourenkova-Mironova et al, 1999). Furthermore, the *NAT2* genotype was not identified as a risk factor for oral squamous cell carcinoma in a Japanese group of patients (Kato et al, 1998). In contrast, another study has associated the wild type *NAT2*4* allele, responsible for fast acetylation, with increased risk for laryngeal cancer. Although *NAT2*4* homozygotes and heterozygotes carrying any of the slow alleles were all classified as phenotypically fast acetylators, only *NAT2*4* homozygotes were statistically more frequent in that group of laryngeal cancer patients (Henning et al, 1999).

In regard to the role of *NAT1* polymorphism in cancers of the head and neck region, the situation is also unclear. Katoh et al (1998) observed an increased risk associated with the *NAT1*10* allele in oral squamous carcinoma patients, which was independent of the smoking habits of the individuals carrying this allele. A recent study also indicated involvement of *NAT1*10* in head and neck cancer, but in contrast to the study of Katoh et al (1998) it also found association between the genotype and the smoking habits of the subjects (Olshan et al, 2000). Henning et al (1999), however, found that *NAT1* alleles were equally distributed between cases of laryngeal cancer and controls, hence indicating no special role for *NAT1*10*. In agreement with that study, Jourenkova-Mironova et al (1999) failed to demonstrate increased risk for oral, pharyngeal and laryngeal cancers in association with any particular *NAT1* genotype.

The role of NAT1 and NAT2 polymorphism in relation to cancers of the larynx, pharynx and oral cavity remains, therefore, controversial. More evidence is necessary in order to elucidate the potential role of NAT in these types of cancer. The recent establishment of an international collaborative project, aiming to collect all epidemiological data available on the association between polymorphisms of xenobiotic metabolising enzymes and cancer, should allow collation of results from different studies into much larger cohorts (Taioli, 1999).

1.3.7.2 NAT in lung cancer

The factors affecting predisposition to lung cancer seem to be similar to those involved in the development of laryngeal cancer, with smoking being the most important of them. The NAT enzymes have been studied in relation to lung cancer, mainly because of their ability to metabolise polycyclic aromatic hydrocarbons present in cigarette smoke. In the lung, expression of NAT2 has been demonstrated by RT-PCR in the bronchial mucosa only, while NAT1 expression was detected in all parts of the lung, including the bronchus (Windmill et al 1997; Macé et al, 1998; Windmill et al, 2000).

Early epidemiological studies were based on phenotypic analysis alone to investigate the role of NAT2 in lung cancer. Several studies have either failed to identify differences in the distribution of acetylator phenotypes between lung cancer patients

and healthy individuals or found a small over-representation of the fast phenotype in the patient group (reviewed in Hein et al, 2000a). Interest in studying the role of NAT in lung cancer was revived following the introduction of molecular methods for genotyping individuals for the *NAT2* gene (Roots et al, 1992). Using genotyping methods, Cascorbi et al (1996b) found a small but significant increase in the number of fast acetylators in the group of lung cancer patients. This was evident only for the homozygous fast genotype (*NAT2**4/*4), while heterozygotes for *NAT2**4 and any other slow allele (phenotypically also categorised as fast acetylators) were not over-represented in the patient group. It was concluded that the *NAT2**4/*4 genotype conferred a significantly increased risk for lung cancer, particularly among smokers, probably through very high rate of metabolic activation of tobacco carcinogens via O-acetylation. An increased risk of lung cancer among smokers with the fast acetylator genotype was also demonstrated in a study looking at *NAT2* genotypes in combination with the *GSTM1* genotype (Nyberg et al, 1998). Analysis of the interaction between smoking and the *NAT2* genotype suggested that the risk associated with the fast acetylator genotype increases with higher levels of exposure. On the other hand, among patients who had never smoked, the slow acetylators were at increased risk. In addition, non-smokers were at higher risk if they carried a combination of slow *NAT2* and *GSTM1*+ genotypes.

Contrary to the above reports, some studies did not find any association between lung cancer and the fast *NAT2* genotype. An early study, using AS-PCR for genotyping groups of patients and healthy controls, concluded that the acetylator status was not a major factor affecting susceptibility to lung cancer. However, of the fourteen mutant alleles detected, the two slow *NAT2**5B and *NAT2**6B alleles were found at increased frequency among lung cancer patients and could, therefore, be a minor risk factor for the disease (Martinez et al, 1995). No significant association between lung cancer and any particular *NAT2* genotype was observed by Bouchardy et al (1998) either. Finally, there is some evidence suggesting that the slow genotype may have an impact on the risk for some types of lung cancer, depending on age and smoking dosage (Hou et al, 2000).

The role of *NAT1* genotype in lung cancer has been investigated to a lesser extent, although this is the form of NAT predominantly expressed in the lung (Macé et al,

1998). Again, the published reports present different and often controversial results. Ishibe et al (1998) found no association between lung cancer of any histological type and specific *NAT1* genotypes and this finding was independent of the smoking history of the individuals. However, another study reported a possible link between the *NAT1*14* and *NAT1*15* alleles and lung cancer (Bouchardy et al, 1998). It is of interest that these two *NAT1* alleles, which are known to confer slow NAT1 enzymatic activity (section 1.3.3.2), had not been detected by Ishibe et al (1998). In another study, an increased risk for lung cancer was associated with the inheritance of the *NAT1*10* allele (Abdel-Rahman et al, 1998), which has been controversially associated with higher NAT1 activity (section 1.3.3.2).

The NAT enzymes may catalyse both the activation and deactivation of carcinogenic compounds present in tobacco smoke. The apparent inconsistency in the results reported for the role of human NAT1 and NAT2 in lung cancer may be due to the limited knowledge about the fine balance between these two NAT-mediated metabolic pathways *in vivo*. Moreover, efforts to group patients and controls according to many parameters (including age, smoking habits and exposure to carcinogens, site and stage of cancer, etc) may have led to chance associations. A large-scale epidemiological study with a very large number of subjects would be necessary to establish a link between specific NAT genotypes and predisposition to lung cancer.

1.3.7.3 NAT in prostate cancer

Prostate cancer may be either sporadic or familial, indicating some kind of inherited predisposition to certain forms of the disease. A high dietary intake of fat, especially from red meat, has been associated with prostate cancer, suggesting involvement of food-derived mutagenic compounds, which are substrates of xenobiotic metabolising enzymes, including NATs (Wadelius et al, 1999). Interestingly, a genomic region associated with increased susceptibility to prostate cancer has been identified in chromosomal region 8p (Wadelius et al, 1999), where the *NAT* genes have been mapped.

Only a few studies have looked at the possible association between the *NAT1* and *NAT2* genotype and prostate cancer. A Spanish study found a similar allele distribution among patients and control subjects and therefore reported lack of relationship between *NAT2* and predisposition to prostate cancer (Agundez et al, 1998b). The lack of an epidemiological link between *NAT2* polymorphism and prostate cancer was also suggested by Wadelius et al (1999).

A transgenic mouse model over-expressing human *NAT2* only in the prostate has been generated to investigate its potential role in prostate carcinogenesis (Leff et al, 1999). N-acetylation activity for sulphamethazine was increased in the prostate of the transgenic mice by fifteen-fold, compared with controls, and this increase was attributed to over-expression of human *NAT2*. However, prostate cytosols from transgenic and control mice did not differ in their ability to metabolise the carcinogens PhIP and 2-aminofluorene. Furthermore, the levels of PhIP-DNA adduct formation in the prostate did not differ significantly between transgenic and control animals. These results provide good evidence that *NAT2* is unlikely to be a significant predisposing factor for prostate cancer.

NAT1 polymorphism, on the other hand, may have some role in prostate carcinogenesis, as suggested by a recent study demonstrating that the *NAT1**10 allele may confer increased risk for the disease in a Japanese population. However, it should be noted that the incidence of prostate cancer in Japanese is lower than in Caucasians, although the homozygous *NAT1**10/*10 genotypes are more common among Japanese (Fukutome et al, 1999). The mechanisms of prostate cancer development remain unclear, but the above findings suggest that factors other than NAT are more likely to play a role in the initiation and progression of the disease.

1.3.7.4 NAT in hepatocellular carcinoma

There are many environmental, lifestyle, dietary and viral risk factors for hepatocellular carcinoma. Exposure to aflatoxin, cigarette smoking, alcohol consumption, and hepatitis B and C infections represent some of the contributory factors (Chen et al, 1997). Epidemiological work on the role of NAT in the development of this cancer has been limited, although liver is the organ where both

NAT isoenzymes are expressed. One study reported an excess of slow *NAT2* genotypes in the group of cancer patients, compared with healthy controls. The relationship between the slow *NAT2* acetylator genotypes and increased risk for hepatocellular carcinoma was more pronounced in patients negative for hepatitis B and C virus markers in their blood serum, compared with patients positive for these markers (Agundez et al, 1996b).

1.4 Animal models for arylamine N-acetyltransferase

Arylamine N-acetyltransferases are a very well conserved group of proteins. They have been found in prokaryotes, such as *Salmonella* (Watanabe et al, 1992; Sinclair et al, 1998) and *Mycobacterium* (Payton et al, 1999a), as well as in eukaryotes. They are present in all mammals examined to date, with Canids being the only notable exception (Trepanier et al, 1997). Given the role of NAT in the metabolism of various xenobiotics of great pharmacological or toxicological interest, as well as the possible implications of NAT in chemical carcinogenesis, the use of animals as models for NAT and its polymorphisms, has long been desirable (Levy et al, 1992). Most of the common laboratory animals, such as the rabbit, rat, hamster and mouse have been extensively used in NAT-related research during the past four decades and have provided valuable information about the role of NAT *in vivo* (reviewed in Weber and Hein, 1985).

The rabbit has been the classical animal model for NAT, as it was the first mammal after humans demonstrated to show NAT-polymorphism (reviewed in Weber and Hein, 1985). NAT was purified from rabbit liver in the sixties (Weber and Cohen, 1967) and a genetically determined N-acetylation polymorphism was later detected among different inbred strains (Hein et al, 1982b). There are two NAT-isoforms in rabbit, NAT1 and NAT2, the former preferentially metabolising pABA and the latter isoniazid and sulphamethazine. Similarly to humans, rabbit NAT1 was initially described as monomorphic and NAT2 as polymorphic. However, it was later determined that the polymorphism associated with NAT2 activity in rabbits is due to a complete absence of the *nat2* gene in the slow acetylators, instead of the presence of

point mutations within the gene, as is the case with humans (Blum et al, 1989b; Reeves et al, 1991; Sasaki et al, 1991).

Other laboratory animals used as models for NAT include the rat and the Syrian hamster, where fast and slow acetylating strains have been characterised and genes for NATs have been identified and cloned (Abu-Zeid et al, 1991; Hein et al, 1991). Work has also been carried out using the chicken (Rougraff and Paxton, 1987; Ohsako et al, 1988).

All NAT isoenzymes identified to date present great homology in their amino acid sequences, irrespective of whether they belong to the same species or not (figure 1.8 and table 1.5). For example, rabbit NAT1 shows 95% identity to rabbit NAT2, but also 75% identity to its functional homologue human NAT1. Similarly, rabbit NAT2 is 73% identical with its functional homologue human NAT2. However, it is difficult to determine the genetic orthologue for each one of the isozymes, as the amino acid sequence similarity is also very high between rabbit NAT1 and human NAT2 (73%), as well as between rabbit NAT2 and human NAT1 (75%) (table 1.5; Vatsis et al, 1995). Such similarities have often caused difficulty in separating different NAT isoenzymes using standard purification methods (e.g. Hein et al, 1982a; Watson et al, 1990), as well as in generating specific antibodies (e.g. Stanley et al, 1997). On the other hand, it has facilitated identification and cloning of the genes encoding for different NATs, using degenerate primers for PCR amplification (e.g. Martell et al, 1991) or heterologous probes in Southern hybridisations (e.g. Blum et al, 1990).

Figure 1.8: Alignment of the amino acid sequences of different mammalian NATs. Amino acids are represented by the single-letter code. Numbers indicate the amino acid position. Codes on the left represent rabbit NAT1 (rabNAT1) and NAT2 (rabNAT2), mouse NAT1 (mNAT1), NAT2 (mNAT2) and NAT3 (mNAT3), rat NAT1 (ratNAT1) and NAT2 (ratNAT2), Syrian hamster NAT1 (hamNAT1) and NAT2 (hamNAT2), human NAT1 (hNAT1) and NAT2 (hNAT2). Amino acids conserved in all mammalian NATs presented here are shown in purple, while partially conserved residues are shown in blue. The amino acid sequence of the NAT enzymes was deduced from the nucleotide sequence of the corresponding genes, available from the GeneBank.

Figure 1 displays the amino acid sequence alignment of the protein sequences for NAT1, NAT2, and T1 across various species. The sequences are aligned in blocks of 100 residues, with positions 1-100, 101-200, 201-300, 301-400, 401-500, and 501-600 shown. The sequences are color-coded to highlight conserved regions (green) and variable regions (red). The alignment shows high sequence identity between the sequences, particularly in the conserved regions. The sequences are labeled as NAT1, NAT2, and T1, and the positions are indicated by the numbers 1-600.

		Rabbit		Mouse			Rat		Syrian hamster		Human	
		NAT1	NAT2	NAT1	NAT2	NAT3	NAT1	NAT2	NAT1	NAT2	NAT1	NAT2
Rabbit	NAT1	-	95.2	65.2	72.4	62.8	68.6	72.0	69.7	73.4	74.8	72.8
	NAT2		-	66.9	73.1	61.4	70.0	72.4	70.3	74.1	75.5	73.1
Mouse	NAT1			-	81.0	67.2	91.7	81.3	83.8	82.1	74.5	71.7
	NAT2				-	73.1	83.1	94.1	78.3	92.8	81.7	74.5
	NAT3					-	68.6	73.4	67.2	73.1	68.3	65.9
Rat	NAT1						-	83.1	85.8	84.8	81.0	74.8
	NAT2							-	77.2	91.7	81.3	73.7
Syrian hamster	NAT1								-	79.0	71.4	68.6
	NAT2									-	81.7	74.8
Human	NAT1										-	81
	NAT2											-

Table 1.5: Percentage (%) identity of the amino acid sequences of rabbit, mouse, rat, Syrian hamster and human NAT isoenzymes (modified from Vatsis et al, 1995; Doll and Hein, 1995).

1.4.1 The mouse model for arylamine N-acetyltransferase

Using the mouse as an animal model has many advantages. Mice present a developmental pattern similar to that of humans, as other commonly used mammalian models do, but in addition they are easy to maintain, handle and breed. There are many inbred strains with well-established characteristics suitable for testing chemicals and they are easily accessible during all stages of development, including gestation. Mouse genetics has been extensively studied, many genes have been characterised and mapped, and numerous genetic markers and extensive sequence information have become available through the Mouse Genome Project. This wealth of information facilitates analysis of the murine genome and allows comparative studies with humans. In addition, mice can be bred to produce congenic strains, differing from their parental strain by a single genetic locus of interest, and they can also be genetically manipulated or mutated using current transgenic technologies (reviewed in Moore, 1999).

1.4.1.1 The arylamine N-acetyltransferase polymorphism in mice

The mouse has long been considered as an excellent model for studying NAT, especially following the characterisation of fast and slow acetylating strains in the early eighties. In a study by Glowinsky and Weber (1982a), twenty inbred mouse strains were examined for NAT activity using pABA and benzidine as substrates. Seventeen of the strains (including the commonly used C57BL/6J, 129/Sv, Balb/cJ, C3H/HeJ, CBA/J and DBA/J) were identified as fast acetylating, while three of them (A/J, Ahe/J and x/Gf) were identified as slow acetylating strains, showing about ten-fold lower NAT activity than the fast acetylators. Mating between slow A/J and fast C57BL/6J mice produced F1 heterozygotes with intermediate levels of NAT activity, suggesting that a single polymorphic locus was responsible for the acetylation of pABA and benzidine in these mouse strains (Glowinski and Weber, 1982a).

Following these observations, the kinetics of N-acetylation catalysed by murine NAT were extensively studied in both slow and fast acetylating strains (Mattano and Weber, 1987), and it was determined that isoniazid was always N-acetylated monomorphically, in contrast to the polymorphic pattern of N-acetylation with pABA.

This suggested the presence of at least two NAT isoenzymes in mice, one of which was monomorphic and responsible for the acetylation of isoniazid and the other polymorphic and responsible for the metabolism of pABA (Hein et al, 1988). The two isoenzymes were designated NAT1 and NAT2 respectively (Vatsis et al, 1995).

Both murine NAT isoforms are 31-33kD in size and very similar to each other in their amino acid sequence (81% identity) (table 1.5). They seem to represent a small fraction of all cytosolic enzymes, as indicated by the fact that at least five thousand fold purification is required to partially purify them (Watson et al, 1990). In addition to their ability to N-acetylate aromatic amines, they are also capable of catalysing O-acetylation of hydroxylamines and N,O-acetyltransfer of arylhydroxamic acid, to produce bio-activated acetoxy esters (Mattano et al, 1989).

Murine NAT1 and NAT2 isoenzymes behaved very similarly during standard purification procedures, which indicated minor structural differences between them (Mattano et al, 1989; Watson et al, 1990). Following cloning of the genes encoding for the two NAT isoenzymes in mice, it was deduced that NAT2 from the fast C57BL/6J strain was different from the slow A/J strain by only one amino acid. The asparagine at position 99 of NAT2 in the fast strain has been substituted by isoleucine in the slow strain. This substitution changes the hydrophobicity of the protein and also seems to affect its structure and catalytic properties (Martell et al, 1991). Later studies suggested that this amino acid substitution may be associated with an approximately fifteen fold decrease in stability of NAT2 in the slow acetylators and possibly affects its rate of translation (Martell et al, 1992). However, De Leon et al (1995) suggested that the amount of the protein synthesised from both alleles was similar, but slow NAT2 had decreased affinity for its substrates compared with the fast isoform. Neither study detected a difference in the transcription rate from the fast and the slow allele, as indicated by quantitative Northern blot analysis.

1.4.1.2 Tissue-specific expression of murine arylamine N-acetyltransferases

NAT1 and NAT2 isoenzymes are predominantly expressed in the liver, with NAT2 being more abundant as indicated by Northern blot analysis (Martell et al, 1992). NAT1 has also been detected in the intestine (Levy et al, 1993; Ware and Svensson, 1996), while the distribution of NAT2 seems to be ubiquitous. NAT2 activity assayed with pABA has been detected in many tissues, including bladder, brain, heart, kidney, lung, muscle, red and white blood cells, thymus, spleen, skin, colon, pancreas etc (Chung et al, 1993). Wide distribution of murine NAT2 has also been demonstrated using immunohistochemical and immunoblotting methods (Stanley et al, 1997). In most tissues analysed, a marked difference between the amount of NAT2 in fast and slow strains was observed.

An unusual gender-related difference in NAT2 expression has been observed in the kidney, with males showing around three fold higher NAT2-activity measured with pABA than females of the same strain (Glowinski and Weber, 1982b). This difference seems to be associated with testosterone induction of NAT2 in males, as castration of male animals reduced kidney NAT2 to female levels (Smolen et al, 1993). This is consistent with other studies showing induction of NAT by glucocorticoid hormones in other animals, although not specifically in the kidney (Reeves et al, 1988). It has recently been shown that a glucocorticoid response element located upstream of the coding region of the gene encoding for NAT2 may be involved in the regulation of NAT2 expression in the kidney (Estrada-Rodgers et al, 1998a; Estrada et al, 2000).

1.4.1.3 Murine arylamine N-acetyltransferases in development

Evidence supports that murine NAT2 is expressed throughout development, while the developmental pattern of NAT1 seems to be more restricted and remains unclear. It seems that the expression of murine NAT1 begins later than the expression of NAT2 during gestation (Mitchell et al, 1999). Total NAT activity seems to increase with age (Estrada et al, 2000).

Based on criteria including substrate specificity and tissue distribution, murine NAT2 seems to be the functional equivalent of human NAT1 and vice versa (Levy et al,

1992). This is also implied by the amino acid sequence homology between these two isoenzymes, as mouse NAT2 shows 81.7% homology to human NAT1, versus 74.5% homology to human NAT2 (table 1.5). Although such comparisons may not be conclusive, since all mammalian NATs are highly homologous with each other (Vatsis et al, 1995; table 1.5), homology between mouse NAT2 and human NAT1 is further supported by their complete identity at the fourteen carboxy-terminal amino acids (Stanley et al, 1996; 1997; figure 1.8), involved in substrate-binding (Sinclair and Sim, 1997).

A very interesting recent finding that further supports the homology between murine NAT2 and human NAT1, has been the discovery that murine NAT2 can N-acetylate the folate catabolite pABGlu, similarly to human NAT1. N-acetylation activity is different between fast and slow strains using pABGlu as substrate (Payton et al, 1999b). This may imply that murine NAT2 may play a role in folate metabolism, as speculated for human NAT1. If this is the case, mice can be an excellent model for studying the role of NAT in early embryonic development, as murine NAT2 N-acetylation polymorphism is very well characterised, unlike human NAT1 polymorphism where the genotype-phenotype correlation is less well established (Payton et al, 1999b).

The hypothesis that murine NAT2 may be involved in the metabolism of folate, a compound of critical importance for neural tube formation during early embryogenesis (section 1.3.5), is further supported by the fact that only NAT2 (Payton et al, 1999b) and not NAT1 (Boukouvala et al, 1999; Boukouvala and Sim, 2000) is expressed in the preimplantation embryo. As shown by RT-PCR using cDNA from mouse embryonic stem (ES) cells as template, NAT2 transcript is present at the blastocyst stage, prior to the onset of neurulation, while NAT1 is not. Mouse NAT2 expression has also been demonstrated in the developing neural tube during neurulation by immunohistochemical methods (Stanley et al, 1998). However, it remains to be firmly established whether or not murine NAT2 plays a role in development and in the metabolism of an endogenous substrate, such as pABGlu.

1.4.1.4 Murine arylamine N-acetyltransferases as a model for carcinogenesis

Due to their very well characterised N-acetylation polymorphism, fast and slow acetylating mouse strains have been extensively used as models for studying the long suspected relationship between NAT and chemical carcinogenesis (Levy et al, 1992). In parallel with the inbred fast C57BL/6J and slow A/J strains, two lines of congenic strains have also been used in these studies, to eliminate the effect of any genetic differences between the parental strains, other than those related with their N-acetylator status. The first of these congenic lines contains the slow NAT2 allele in a C57BL/6J genetic background, and the other contains the fast NAT2 allele in an A/J genetic background (Mattano et al, 1988).

To avoid time consuming procedures involving chronic administration of carcinogens to mice, followed by evaluation of their potential effect in carcinogenesis, most studies have used measurement of DNA-carcinogen adduct formation as an indication of susceptibility to carcinogenesis due to a particular NAT substrate. Adduct formation is measured by High Pressure Liquid Chromatographic (HPLC) analysis of ³²P-post labelled nucleotides from tissues of treated and untreated animals (Levy et al, 1992). Administration of 2-aminofluorene has been associated with two-fold increase in adduct formation in fast C57BL/6J versus slow A/J strains in the liver.

Interestingly, when the congenic strains were used, there was up to seven-fold increase in hepatic adduct formation in fast acetylators with a slow A/J background, compared to slow acetylators with a fast C57BL/6J background. This indicated that other genetic factors, different between the two parental strains, were also affecting the rate of adduct formation in the liver (Levy and Weber, 1989). The results have been less conclusive for the bladder (Flammang et al, 1992), but it seems likely that slow, rather than fast, acetylators are more prone to adduct formation in bladder after administration of 2-aminofluorene (Levy and Weber, 1992). This is similar to the situation observed in humans, where slow acetylators appear to be at increased risk for bladder cancer, as will be described in section 6.1.2. It appears that in mice susceptibility to adduct formation in the bladder is the reverse of that observed in the liver. A possible reason for this difference may be that the high NAT2 activity of fast acetylators could be responsible for the activation of more carcinogenic compounds in

the liver, but at the same time would allow less carcinogenic compounds to reach the bladder (Levy and Weber, 1992).

It has also been suggested that female mice may be more prone to adduct formation than males and that the levels of DNA adducts are higher in older than younger animals (Levy and Weber 1989; 1992; Flammang et al, 1992). Patterns of adduct formation similar to those observed in the liver have also been described for mononuclear leukocytes, with fast acetylators showing higher levels than slow (Levy et al, 1994). Adduct formation in mouse colon has not been described yet, although it has been known that both NAT1 and NAT2 are expressed in this tissue and may be involved in the metabolism of arylamines such as those present in cooked meat (Levy et al, 1993).

1.4.1.5 Murine arylamine N-acetyltransferase 3: A functional isoenzyme?

Mice differ from other mammals in that they possess three NAT isoenzymes (table 1.5). The gene encoding for the third isoenzyme, designated NAT3, is functional (Kelly and Sim, 1992; 1994), as opposed to the third *NAT* locus in humans which is a pseudogene (Blum et al, 1990). Murine NAT3 has been expressed in *E. coli* (Kelly and Sim, 1994; Fretland et al, 1997) and in mammalian COS-1 cells (Estrada-Rodgers et al, 1998b), but no specific substrate has been identified, to date. NAT3 shows only very low activity with 2-aminofluorene and anisidine and no detectable activity with pABA, pABGlu, sulphamethazine and isoniazid (Kelly and Sim, 1994; Fretland et al, 1997; Estrada-Rodgers et al, 1998b). NAT3-related O-acetylation and N,O-acetyltransfer activities were also undetectable (Fretland et al, 1997).

To date, NAT3 has not been found to be polymorphic. It presents 67% identity in amino acid sequence with mouse NAT1 and 73% identity with mouse NAT2 (table 1.5) and has not been detected *in vivo*, although recent evidence suggests that it may be expressed in the spleen, based on RT-PCR (Boukouvala et al, 1999; Boukouvala and Sim, 2000).

The molecular genetics of murine arylamine N-acetyltransferases have been studied quite extensively. A detailed description of the genes encoding for the three NAT

isoenzymes in mice, as well as information on the molecular genetics of the N-acetylation polymorphism will be provided in chapter 4.

1.5 Genome Projects

1.5.1 The Human Genome Project

The Human Genome Project (HGP) is an international collaborative effort between many research groups that started in 1988 with significant governmental support and funding. HGP is a global initiative aiming to sequence the human genome and ultimately identify and map its estimated 50,000-100,000 genes (Watson, 1990; Watson and Cookdeegan, 1991). The major contributors are 16 leading research institutions from France, Germany, Japan, China, UK and USA, including Génethon (Evry, France, http://www.genethon.fr/genethon_en.html), the Institute for Molecular Biotechnology (Jena, Germany, <http://genome.imb-jena.de>), the Keio University (Tokyo, Japan, <http://www.keio.ac.jp>), the Sanger Centre (Hinxton, UK, <http://www.sanger.ac.uk>), the Baylor College of Medicine (Houston, Texas USA, <http://www.hgsc.bcm.tmc.edu>), and the Washington University Genome Sequencing Centre (St. Louis, USA, <http://genome.wustl.edu/gsc/index.shtml>). The National Centre for Biotechnology Information (NCBI, Bethesda, USA, <http://www.ncbi.nlm.nih.gov/>) and the European Bioinformatics Institute (EBI, Cambridge, UK, <http://www.ebi.ac.uk>) have been providing computational support and analysis of the data generated by the HGP. The contribution from a large number of smaller research organisations has also been significant.

The scope of the HGP is broad and affects a variety of scientific and social fields (Olson, 1993). To achieve its goals, the HGP has had to develop powerful tools for mapping and for integrating information from different experiments. The generation of new cloning systems like the yeast artificial chromosome (YAC) (Burke et al., 1987) and a number of intermediate capacity systems based on *E. coli*, such as bacterial artificial chromosome (BAC) and P1 artificial chromosome (PAC) (reviewed in Monaco and Larin, 1994) greatly facilitated the construction of physical

maps. Also, the identification of sequence-tagged sites (STS) and various polymorphic markers and their introduction as landmarks for physical mapping has expanded the knowledge that could be obtained with previous use of restriction fragment length polymorphism (RFLP) markers (Olson et al, 1989). Substantial improvements have also been made in the methods used for sequencing, to achieve high throughput of samples.

The HGP has been the force behind great improvements in the technology used for genetic research. The necessity for efficient large-scale manipulation, processing and quality control of samples brought together biologists and engineers to develop various automated and robotic systems. It also made evident that the large amount of generated biological data required the application of computer technology and statistics for efficient analysis and management. This necessity established the field of bioinformatics that has rapidly expanded during the past decade. In combination with the development of the World Wide Web, researchers are today able to easily access, display and analyse the wealth of information available (Guyer and Collins, 1995).

The Human Genome Organisation (HUGO, <http://www.gene.ucl.ac.uk/hugo>) was established by leading scientists in the field of genome research in 1989, soon after the dawn of the HGP. This international organisation aims to help public understanding of the project and promote collaboration in the field. Its role is to co-ordinate the HGP effort and this is mainly achieved through regular meetings and workshops where information, new ideas and material are exchanged and issues concerning the HGP are debated.

Since the beginning of the HGP, several ethical issues and concerns have been raised by the public, regarding availability and use of the accumulating genetic information. Programs to study the ethical, legal and social implications of genetic research have been introduced as a response to public demand. The aim of these programs is to educate people on the recent advances and ensure that data produced by the HGP are not misused to discriminate individuals based on their genetic background (Guyer and Collins, 1995). There has also been a lot of debate on whether private companies should be allowed to patent newly discovered genes.

Since the publication of the first human genetic map based on RFLP markers (Donnis-Keller et al, 1987), enormous progress has been achieved by the HGP (Gyapay et al, 1996; Schuler et al, 1996). The goals of the first eight years plan (Collins and Galas, 1993) have been accomplished and a new plan has been constructed to set goals for the period 1998-2003, when the complete reference sequence is anticipated to become available (Collins et al, 1998). The goals and achievements of the HGP are summarised in table 1.6. To date, sequence corresponding to about 24% of the human genome has been published in its “finished” form (99.99% accuracy of the released data), including chromosomes 22 (Dunham et al., 1999) and 21 (Hattori et al, 2000). The recent announcement of the assembly of a “working draft” for the rest of the human genome by the major contributors of the HGP (International Human Genome Sequencing Consortium) has been a major landmark in the history of modern genetics. The “working draft” provides 97% genomic coverage and, although of lower accuracy than the “finished” form, has become an invaluable tool for biomedical research. In a joint announcement, Celera Genomics announced completion of another, privately funded, project for sequencing the human genome (National Human Genome Research Institute at the National Institute of Health [NHGRI/ NIH] press release 26 June 2000, http://www.nhgri.nih.gov/NEWS/sequencing_consortium.html). Collaboration between the two projects is expected to accelerate refinement of the existing data, providing a high quality reference sequence of the human genome before the end of 2003, as initially planned (table 1.6; Collins et al, 1998).

Table 1.6: Summary of the achievements of the HGP during the period 1993-98 and goals for period 1998-2003 (modified from Collins et al, 1998).

Area	Goals 1993-98	Status as of end 1998	Goals 1998-2003
Genetic map	Average 2- to 5-cM resolution	1 cM map published	Completed
Physical map	Map 30,000 STSs	52,000 STSs mapped	Completed
DNA sequence	Complete 80 Mb for all organisms by 1998	180 Mb human plus 111 Mb non-human	Finish 1/3 of human sequence by end of 2001. Working draft of remainder by end of 2001 ¹ . Complete human sequence by end of 2003.
Sequencing technology	Evolutionary improvements and innovative technologies	90 Mb/year capacity at ~\$0.50 per base. Capillary array electrophoresis validated. Microfabrication feasible.	Integrate and automate to achieve 500 Mb/year at \$0.25 per base Support innovation
Human sequence variation	Not a goal	-	100,000 mapped SNPs (single nucleotide polymorphisms). Develop technology.
Gene identification	Develop technology	30,000 ESTs (expressed sequence tags) mapped	Full-length cDNAs
Functional analysis	Not a goal	-	Develop genomic-scale technologies
Model organisms	<i>E. coli</i> : complete sequence Yeast: complete sequence <i>C. elegans</i> : complete sequence <i>Drosophila</i> : begin sequencing Mouse: map 10,000 STSs	Released Sept. 1997 Released Apr. 1996 Released Dec. 1998 12,000 STSs mapped	- - - Sequence by 2002 ² Develop extensive genomic resources. Lay basis for finishing sequence by 2005. Produce working draft before 2005.

¹Released June 2000 (NHGRI/ NIH press release, 26 June 2000).

²Released March 2000 (Adams et al, 2000).

1.5.2 The Mouse Genome Project

En route to its goal of studying the human genome, the HGP has stirred research interests in the genetics of a number of model organisms. The genome mapping and sequencing projects of the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae*, the nematode worm *Caenorhabditis elegans*, the insect *Drosophila melanogaster* and the laboratory mouse *Mus musculus* have been within the stated goals of the HGP (Collins and Galas, 1993). As a model organism within the HGP, the role of the mouse is mainly in the discovery of genes homologous to human disease genes, and in the study of the evolution of mammalian genomes (Meisler, 1996).

The mouse has been a popular experimental organism since geneticists realised that it was possible to produce and maintain pure strains by inbreeding. The first inbred strain was generated in 1909, while most of the strains that are still in use today were introduced in the period 1920-1930. Based on that original work, new inbred strains and a large number of congenic and recombinant inbred strains have been established since then (Nishioka, 1995).

Geneticists have been working on mice for decades and have managed to identify the genes responsible for a number of visible traits, such as eye and coat colour (Nishioka 1995). However, genetic mapping was greatly facilitated after it was realised that polymorphisms in DNA sequence could be used as genetic markers. Polymorphic markers, such as RFLPs and simple sequence length polymorphisms (SSLPs), proved to be highly variable among inbred mouse strains. It has also been possible to further increase polymorphism by interspecific crosses between the laboratory mouse and the wild mouse species *Mus spretus* or *Mus castaneus* (Dietrich et al, 1995).

The aim of the Mouse Genome Project (MGP) is the construction of a combined genetic and physical map of the mouse genome. This is important for the generation of a complete human-mouse comparative map, which will facilitate identification of genes homologous between the two organisms. It is estimated that there are at least 200 chromosomal regions which are syntenic between mice and humans (Carver and

Stubbs, 1997), and it has often been possible to predict the location of a human gene based on mapping information generated from mice. This is particularly important in the case of disease genes, as mouse mutants can be used as models for human disease (Meisler, 1996).

Similarly to the HGP, the MGP has been based on the collaboration of major Institutions, although the contribution from smaller laboratories has also been important. The major contributors are the Jackson Laboratory (Bar Harbor, USA, <http://www.jax.org/>), the MIT Centre for Genome Research, (Cambridge, USA, <http://www-seq.wi.mit.edu/>), the MRC Mouse Genome Centre (Harwell, UK, <http://www.mgc.har.mrc.ac.uk/>) and the European Collaborative Interspecific Mouse Backcross (Cambridge, UK, <http://www.hgmp.mrc.ac.uk/MBx/>). Apart from generating sequencing and mapping data, these institutions are also responsible for running databases, including the Mouse Genome Database (MGD, <http://www.informatics.jax.org/>) and the Dysmorphic Human-Mouse Homology Database (DHMHD, <http://www.hgmp.mrc.ac.uk/DHMHD/dysmorph.html/>) (Blake et al., 1997; Ringwald et al, 1997; Eppig et al, 1998). The bulk of sequencing data generated by all genome projects is submitted to a number of major databases and becomes available to the public via the internet (table 1.7).

Table 1.7: DNA sequence databases available on the internet.

Database name	Database WWW address
GenBank, NCBI, Bethesda, USA	http://www3.ncbi.nlm.nih.gov/
EMBL, EBI, Cambridge, UK	http://www.ebi.ac.uk/Databases/index.html
DDBJ, Center for Info Biology, Japan	http://www.ddbj.nig.ac.jp
GSDB, NCGR, Santa Fe, USA	http://www.ncgr.org
dbEST, NCBI, Bethesda, USA	http://www.ncbi.nlm.nih.gov/dbEST/index.html

1.6 Genome Mapping

1.6.1 Genetic Mapping

Early efforts to map genetic loci within the genome involved investigation of linkage patterns between simple Mendelian traits. Linkage analysis is the method used to measure the genetic distance between two distinct loci, provided that they co-localise on the same chromosome. The genetic distance is calculated based on the frequency of recombination between the two loci during meiosis. Alleles of loci in close proximity with each other tend to co-segregate during meiosis and are passed from parents to offspring together, as recombination between them is less likely to occur (Ott, 1999). Genetic mapping of traits by linkage analysis relies on measurement of the recombination events between the loci examined, within a large number of related individuals. From the rate of recombination, the logarithm of odds (lod) score can then be calculated. This index estimates the likelihood of two loci segregating due to recombination between them during meiosis, rather than by chance. Linkage between two loci is indicated by a lod score of 3.0 or higher, which corresponds to 1000:1 odds that the loci are linked (Chotai, 1984).

On rare occasions, linkage between traits may be phenotypically determined, but the resulting linkage pattern is simple and not adequate for mapping. For years, genetic mapping had been difficult, because of the very small number of available polymorphic markers. In recent years, recombinant DNA technology has allowed researchers to identify many types of genetic variation. This has been a major step forward, as construction of high-resolution linkage maps has become possible (reviewed in Curran, 1997). RFLPs were the first type of polymorphic markers used for genetic mapping. They do not always cause a phenotypic effect and can be used without prior knowledge of the sequence or the function of the region they mark (Botstein et al, 1980). However, only some of the genetic polymorphisms affect restriction endonuclease cutting sites, which made the development of additional types of polymorphic markers a requirement for advanced genetic mapping to be achieved.

Microsatellite markers were introduced to complement RFLPs and today they are the most commonly used polymorphic markers. They are short tandem repeats (STRs) found in genomic DNA and they show polymorphism in the number of their repeated elements. STRs usually consist of specific di-, tri- or tetra-nucleotide repeat sequences, of which the dinucleotide (CA)_n repeats are the most common. Similarly to RFLPs, STRs are not usually associated with a particular phenotype, but detection of their polymorphism is easier and involves use of PCR and estimation of the size of the amplified target sequence (Weber and May, 1989). In addition to these, there is currently a variety of other polymorphic markers used for genetic mapping, most of which can be identified by PCR, as summarised in table 1.8. These markers, alone or in combination, have been very useful in constructing human and mouse genome maps (Zietkiewicz et al, 1992; Hudson et al, 1995; Schuler et al, 1996; Brady et al, 1997; Rhodes et al, 1998).

Generating high-resolution genetic maps is a very difficult task, as it requires ordering of a great number of closely linked markers on a chromosomal region, however, it is essential in order to identify the position of new genes. In humans, much of the work on genetic mapping has been possible based on the CEPH (Centre d'Etude du Polymorphisme Humain) reference families. This specific set of families was selected because of their large sibling size, rather than existing evidence of some familial disease, and DNA samples from these families have been available for research (Dausset et al, 1990). To overcome the necessity for a large number of progeny for accurate genetic mapping, much of the genetic analysis has also been performed on model organisms, especially mice. The mouse is a very good system for studying mammalian genetics, because of a large number of characterised mutations and markers and the possibility of performing crosses reflecting specific meiotic events (Schalkwyk et al, 1995).

Table 1.8: Polymorphic markers used for genetic mapping.
(From: Cooper and Smidtke, 1984; Edwards and Caskey, 1991; Nelson et al, 1991; Chapman and Nadeau, 1992; Hunter et al, 1993; Dietrich et al, 1994; O’Brown, 1994; Dietrich et al, 1995; McCarthy et al, 1995; Collins et al, 1997; Dear, 1997).

Type of marker	Type of polymorphism	Type of screening
Previously identified genes	Mutations/ polymorphisms within the gene	Genotyping by allele-specific PCR, restriction enzyme digestion, DGGE, SSCP, ASO-hybridisation etc.
Restriction fragment length polymorphism (RFLP)	Polymorphism at the recognition site of restriction endonucleases	Digestion with specific restriction endonucleases and analysis of fragments by Southern blotting and hybridisation.
Simple sequence length polymorphism (SSLP)	Differences in the number of microsatellite sequence repeats (CA) _n	PCR
Short tandem repeat (STR) polymorphism	Differences in the number of microsatellite tri-or tetra-nucleotide repeats	PCR
Randomly amplified polymorphic DNAs (RAPDs)	Differences in random sequences within the genome	PCR using random combination of short forward and reverse primers of arbitrary sequence at low stringency
Interspersed repeat sequence (IRS) polymorphism	Differences in LINEs, SINEs, <i>Alu</i> -elements.	PCR
Single strand conformational polymorphism (SSCP)	Differences in single base pairs	SSCP
Genomic mismatch polymorphism	Differences in single base pairs	Heteroduplex analysis
Single nucleotide polymorphism (SNP)	Difference in single base pairs	Techniques for identifying single nucleotide polymorphisms, e.g. ASO-hybridisation.
CpG islands	Differences in methylation of the cytidine of CG-dinucleotides within CpG islands	Digestion with methylation-sensitive restriction endonucleases.

Abbreviations: DGGE (denaturing gradient gel electrophoresis), SSCP (single strand conformational polymorphism), ASO (allele-specific oligonucleotide), LINEs (long interspersed elements), SINEs (short interspersed elements).

1.6.2 Physical Mapping

Physical mapping involves the construction of overlaps between cloned fragments of genomic DNA, by identifying suitable markers that are common between two or more clones. A set of overlapping DNA clones put in order is called a contig and this can provide coverage of a region, which would normally be too long to be cloned as one fragment in a single vector. Contigs of well-characterised clones, covering many megabases (Mb) can be useful for determining the position of genes, as well as for sequencing extended regions of the genome (Olson, 1993). This method can miss parts of the genome that have not been cloned as yet, but the markers need not be polymorphic, as in genetic mapping and linkage analysis.

The accuracy of physical maps largely depends on the type of clones used for their construction. Since the aim is to cover long distances with little effort, it makes sense to choose clones that contain as large inserts as possible. The largest cloning systems currently available are the yeast artificial chromosomes (YACs) which typically carry fragments in the order of 1Mb, although larger fragments are also possible (table 1.9). The YAC vector is linear and typically consists of a centromere and telomeres obtained from the protozoon *Tetrahymena*, plus a yeast autonomously replicating sequence and two genes for selection of the transformed cells (Burke et al, 1987). YAC clones have been commonly used for years to construct large physical maps of the human (Bellane-Chantelot et al, 1992; Chumakov et al, 1995), mouse (Haldi et al, 1996; Nusbaum et al, 1999) and other genomes (Waterston and Sulston, 1995).

The use of the YAC system has several disadvantages, however. YAC clones can be genetically unstable and therefore they tend to lose parts of their inserts. In addition, YAC libraries, particularly of the human genome, can contain up to 40-60% of chimaeric clones, even if they have been constructed in recombination-deficient yeast strains (Anand et al, 1990; Chartier et al, 1992). Instability and chimaerism can cause significant problems in chromosome walking and gene isolation procedures, and it is quite possible that some physical maps may be inaccurate. Working with YACs requires special technology, but even then DNA may be recovered at low yield and it is often difficult to isolate it from the yeast chromosome. In addition, some yeast cells

have been found to carry more than one type of YAC clones making isolation of one desired clone problematic (Monaco and Larin, 1994).

Table 1.9: A comparison of cloning systems used for physical mapping (modified from Monaco and Larin, 1994).

Cloning system	Host organism	Vector structure	Insert size (kb)
Cosmids	<i>E. coli</i>	Circular	35-45
P1	<i>E. coli</i>	Circular	70-100
BAC	<i>E. coli</i>	Circular	up to 300
PAC	<i>E. coli</i>	Circular	100-300
YAC	<i>S. cerevisiae</i>	Linear	100-2000

Despite their shortcomings, YACs have been a very useful tool in genetic research, significantly expanding genomic coverage. To overcome the YAC-related problems, a number of cloning vectors based on *E. coli*, are now extensively used, in addition to YACs (table 1.9).

The P1 cloning system contains many elements of the bacteriophage P1, including a replicon to maintain a single copy of the vector per cell and a kanamycin resistance gene for selection of the transformants. It is possible to clone up to 100kb of genomic DNA in this circular system, which makes it twice as useful as cosmids (Sternberg, 1990; Sternberg, 1992). The P1 vector has been successfully used to construct human (Shepherd et al, 1994) and mouse (Pierce et al, 1992) genomic libraries. Another *E. coli* based system is the bacterial artificial chromosome (BAC). This contains a replication origin from *E. coli* (F-factor) and can incorporate inserts up to around 300kb, but unfortunately it lacks an element for positive selection of recombinants. BACs are circular constructs and they are introduced into cells by electroporation, where they are maintained as a single copy per cell (Shizuya et al, 1992). A set of BAC clones covering the whole human genome has recently been developed, integrating genetic and cytogenetic mapping data (Korenberg et al, 1999).

The P1 artificial chromosome (PAC) cloning system has been designed to combine the best features of the P1 and BAC systems. It contains a P1 replicon and the ability for positive selection of recombinants, like the P1 system, but in addition it can

incorporate fragments 100-300kb in size. PACs are delivered into cells by electroporation, in common with BACs, and have shown no problems related with instability or chimaerism (Ioannou et al, 1994). A number of mammalian genomic libraries have been constructed in pCYPAC2 and pPAC4, the most commonly used PAC vectors, and are available from the originators of the system (<http://bacpac.med.buffalo.edu/>).

Cloning systems of intermediate size, like P1, BACs and PACs, have been widely used to clone intact large genes. They have also been used together with YACs to facilitate and confirm mapping of the genome, as well as to convert long-range YAC contigs into contigs of higher resolution. Furthermore, they are suitable for sequencing large fragments of genomic DNA using shotgun sequencing strategies, a task not possible if using YACs (McPherson, 1997). Sequencing of YACs requires an extra subcloning step into cosmids and may produce unreliable results due to the genetic instability and the chimaerism of the YAC system. In contrast, the smaller P1, BAC or PAC inserts can be directly cloned into M13 vector and sequenced. It has been estimated that around 2500 sequencing reactions, each producing 300-400bp reading, are required for a six-fold coverage of an *E. coli* clone carrying a typical 130kb insert (Mallon and Strivens, 1998).

1.6.3 Radiation Hybrid Mapping

Radiation hybrid mapping is a strategy used to determine the order of markers on a chromosome and to provide estimations for the distances between them. Although it was first described twenty five years ago (Goss and Harris, 1975; 1977a; 1977b), it attracted very little attention until the early nineties (Cox et al, 1990).

Radiation hybrid mapping involves irradiating of donor cells (e.g. human) with a large dose of X-rays, in order to break the chromosomes into several fragments. The irradiated cells are then fused with a rodent host cell line and the chromosomal fragments of the donor are rescued by recombination with the rodent genome. Non-recombined donor cells do not survive the radiation treatment. The host cells contain a selectable marker, typically thymidine kinase deficiency (TK⁻) or hypoxanthine phosphoribosyl transferase deficiency (HPRT⁻). The TK⁻ and HPRT⁻ recipient cells

die when grown in media containing hypoxanthine, aminopterin and thymidine (HAT medium). Therefore, the only cells that can survive fusion and HAT selection are the hybrids of the host cells containing donor DNA, including the donor *TK* or *HPRT* genes (figure 1.9). The surviving hybrids are further cultured and screened for the presence of genetic markers contained in the donor genome. The further apart two markers were originally on the donor chromosome, the more likely it is that a given radiation dose would separate them. Therefore, it is possible to determine the order of these markers by estimating the frequency of breakage (Cox et al, 1990).

Although analysis of the generated data usually requires complicated statistical methods (Lawrence et al, 1991; Barret, 1992), radiation hybrid mapping has the advantage of not relying on polymorphic markers for the construction of high-resolution maps. As many types of markers can be used for mapping, this method has become very useful in integrating genetic, physical and cytogenetic maps (McPherson et al, 1997). Whole genome radiation hybrid maps have also been constructed, indicating the importance of this strategy in the human (Hudson et al, 1995; Gyapay et al, 1996) and mouse (Van Etten et al, 1999) mapping projects.

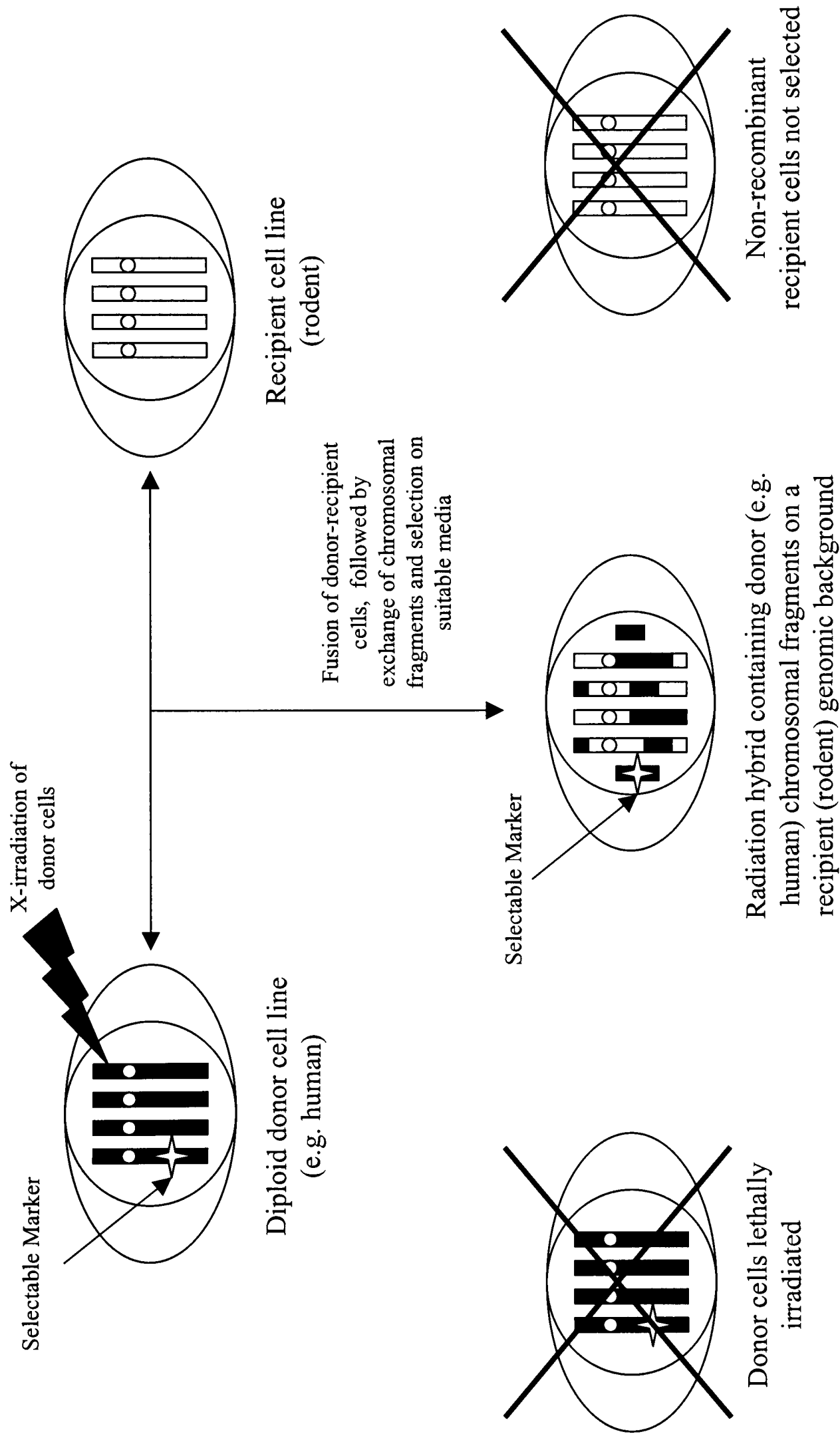


Figure 1.9: Outline of the radiation hybrid mapping procedure. The donor cell line (e.g. human), containing the DNA of interest and a selectable marker, is lethally irradiated and fused with a recipient (rodent) cell line. Chromosomal fragments are exchanged between the donor and the recipient cells, so that DNA fragments of the dead donor cells (including the selectable marker) are rescued on a recipient genomic background. The non-recombinant recipient cells die, following treatment selecting for the marker present in the donor cells, but not in the recipient cells (modified from McCarthy, 1996).

1.6.4 Cytogenetic Mapping

In situ hybridisation using an appropriately labelled DNA probe is a direct method for determining the position of a complementary sequence on nuclear DNA template. It differs from other nucleic acid hybridisation methods, like Southern and Northern, in that the target sequences are at their native cytological position. In the past, this technique used to be performed with radioactively labelled probes, but its use was limited by several factors, including the low resolution of autoradiographic techniques (Mitchell et al, 1992). These problems have now been overcome with the development of non-isotopic methods involving probe labelling with a fluorescent dye and immunocytochemical detection of the signal (Trask et al, 1985; Lawrence et al, 1988).

Fluorescence *in situ* hybridisation (FISH) is the ideal method for assignment of a particular probe sequence (i.e. a cosmid or PAC clone) to a specific chromosomal band. Unlike physical and genetic mapping, it can produce results very quickly and does not require identification of polymorphic markers or rare restriction sites in the area of interest. It also overcomes the need for crosses and pedigrees for performing linkage analysis. However, some sequencing information regarding the target locus is essential, in order to isolate a suitable probe (Trask, 1991).

FISH has been extensively used in combination with physical mapping to confirm clone integrity and exclude chimaerism during the construction of contigs. This is particularly important in view of the problems associated with the YAC system, as mentioned earlier (section 1.6.2). FISH can also be used to assess the region covered by a particular contig, as well as to compile cytogenetic maps covering whole chromosomes (Lichter et al, 1990). It has also been used to characterise somatic cell hybrid mapping panels (Williamson et al, 1995).

Recent technical advance has greatly improved the resolution of FISH and expanded its application to many areas (table 1.10). Sophisticated high-resolution FISH methods generally involve stretching of the target DNA from its condensed chromatin structure to a linear state (Palotie et al, 1996). The simplest way to produce this effect on chromosomal DNA is by allowing a drop of lysed cells to run down the surface of a

tilted slide. Cells can be lysed on the same slide by SDS (sodium dodecyl sulphate) treatment and the viscous lysate is allowed to flow downwards gently (Windle et al, 1995). Mechanical methods for stretching chromosomes by centrifugation have also been described and they can achieve a signal resolution of 170kb (Palotie et al, 1996).

Currently, the maximum resolution can be achieved using a technique called Fibre FISH (table 1.10). This involves hybridisation of probes to free DNA strands released from the chromosomal protein scaffold (Senger et al, 1994). Free DNA fibres can be prepared by alkaline or detergent lysis of nuclei (Fidlerova et al, 1994). Alternatively, high molecular weight cellular DNA may be prepared in an agarose block, which is then melted on the surface of a slide, allowing the DNA to spread in a uniform manner (Heiskanen et al, 1995).

Table 1.10: Resolution of various traditional and high-resolution FISH techniques, in comparison with the resolution of other mapping methods (modified from Palotie et al, 1996).

Technique	Resolution
Restriction mapping	100kb - 4Mb
Radiation hybrid mapping	500kb - 10Mb
FISH on metaphase chromosomes	around 1Mb
FISH on interphase nuclei	50kb - 1Mb
FISH on stretched chromosomes	around 200kb
Fibre FISH	1kb - 500kb

FISH has replaced previous cytogenetic techniques in the study of genomes at the microscopic level. Due to the speed of the method and the easy visual interpretation of results, it has become popular as a very efficient tool for mapping, either on its own or in combination with other mapping strategies. Following great technological advances improving resolution, FISH has become even more powerful and its current potential has become almost comparable to that of other methods used in physical mapping.

1.6.5 Strategies for mapping disease genes

The methodologies described in the previous sections have been particularly useful for identification and mapping of genes associated with specific abnormal phenotypes. The combination of methods applied depends on the amount of information already available about the sequence and function of these genes. When no such information exists, the strategy of choice is positional cloning, also referred to as reverse genetics. This involves identification and cloning of a gene responsible for a particular phenotype, based on its map position only, without previous knowledge of its product (Collins, 1992a). The landmark in the field of positional cloning has been the identification of the gene for cystic fibrosis in humans, a task completed in 1989 (Riordan et al, 1989; Kerem et al, 1989), following almost ten years of effort from a large number of research groups (reviewed in Collins, 1992b). In mice, positional cloning has been greatly facilitated by construction of high-resolution genetic maps, based on analysis of large genetic crosses between defined strains. Large numbers of backcross progeny derived from interspecific backcrosses between laboratory strains (e.g. C57Bl/6J) and wild mouse species (e.g. *Mus castaneus* or *Mus spretus*) have been extensively used to provide the required numbers of meioses for high resolution genetic mapping (Rhodes et al, 1998).

When there is at least some information about the sequence or function of the gene of interest, its cloning and localisation are less complicated, although the individual steps may vary depending on the kind of information available. Many disease genes, such as those responsible for the Lesch-Nyhan syndrome (Shows and Brown, 1975), β -thalassaemia (Sanders-Haigh et al, 1980) and phenylketonuria (Lidksy et al, 1984) have been localised based on previous information about their function. The human genes for NAT have been cloned and mapped by similar approaches, as will be described in detail in chapter 3.

1.7 Aims

A long-term aim of the work on arylamine N-acetyltransferase would be to investigate the mechanisms regulating expression of the NAT isoenzymes in the mouse and human and understand their role in human carcinogenesis. The immediate aims of the work presented here were:

- To provide tools for refined mapping of the human *NAT* loci, extending previous maps of the region.
- To optimise a FISH-based method for detection of chromosomal aberrations involving the *NAT* region in human cancer cells.
- To map the genomic region surrounding the mouse *Nat1* and *Nat2* genes and provide the basis for generation of mouse *Nat2* transgenic and knock-out strains.
- To determine the chromosomal position of the three mouse *Nat* genes.
- To localise the mouse *Nat3* gene in relation to *Nat1* and *Nat2* or to other physical markers.
- To study the expression and activity of the human NAT isozymes in human tissues where they are believed to play a role in carcinogenesis.

CHAPTER 2: Materials and Methods

2.1 Reagents

2.1.1 Water

A Purite reverse osmosis system linked to a deioniser was used to produce deionised water for general use. Where necessary, double deionised water was produced by further purification (StillPLUS by Purite). For certain sensitive applications, including some molecular biology reactions, tissue culture water was purchased from Sigma.

2.1.2 Suppliers of chemicals

All chemicals were purchased from Sigma-Aldrich or Merck-BDH, unless otherwise stated. Reagents for bacterial culture were from Difco, and for cell/tissue culture from GibcoBRL. Enzymes and their buffers were purchased from Promega, New England Biolabs or Boehringer Mannheim, unless otherwise stated.

2.1.3 Primers

Primers were provided by Genosys, Oswell or GibcoBRL. Working stock solutions of primers in water were stored at -20°C , at a concentration of $10\text{pmol}/\mu\text{l}$. A list of the primers used and details of their nucleotide sequence, specificity and annealing temperature is provided in table 2.1.

Primer name	Orientation	Sequence (5'→3')	Specificity	Annealing temperature (°C)
N769 ¹	sense	(769)- ACT CTG AGT GAG GTA GAA ATA -(789)	Human <i>NAT1</i>	50, 54
N1113 ¹	antisense	(1113)- ACA GGC CAT CTT TAG AA -(1096)	Human <i>NAT1</i>	50
N1110a ¹	antisense	(1110)- GGC CAT CTT TAA AAT ACA TTT A -(1089)	Human <i>NAT1</i>	54
N1110b ¹	antisense	(1110)- GGC CAT CTT TAA AAT ACA TTT T -(1089)	Human <i>NAT1</i>	54
NAT-Hu7 ²	sense	(32)- GGC TAT AAG AAC TCT AGG AAC -(52)	Human <i>NAT2</i>	55
NAT-Hu8 ²	antisense	(862)- AAT AGT AAG GGA TCC ATC ACC -(842)	Human <i>NAT2</i>	55
NAT-Hu14 ²	sense	(4)- GAC ATT GAA GCA TAT TTT GAA AG -(26)	Human <i>NAT2</i>	55
NAT-Hu16 ²	antisense	(1002)- GAT GAA AGT ATT TGA TGT TTA GG -(980)	Human <i>NAT2</i>	55
NAT-Hu24 ³	sense	(835)- CTC GTG CCC AAA CAT GGT G -(853)	Human <i>NATP</i>	56
NAT-Hu27 ³	antisense	(1140)- TTC TCA CAC AAA TCA CTA TGG -(1120)	Human <i>NATP</i>	56
mNAT2 278 ⁴	sense	(278)- GTC TTT AAC ACT CCA GCC -(295)	Mouse <i>Nat2</i>	54
mNAT2 409 ⁴	antisense	(409)- ATT CCA GAG GCT CCC AC -(393)	Mouse <i>Nat2</i>	54
mNAT1 516 ⁵	sense	(516)- TTC CAA ATG AAG AAT TTG -(533)	Mouse <i>Nat1</i>	44
mNAT1 867 ⁵	antisense	(867)- CTA AAT AGT AAA AAC TAG T -(849)	Mouse <i>Nat1</i>	44
mNAT2 516 ⁵	sense	(516)- TTC CAA ACC AAG AAT TTA -(533)	Mouse <i>Nat2</i>	46
mNAT2 867 ⁵	antisense	(867)- TAA ATG GTA AAA AAT CGA -(850)	Mouse <i>Nat2</i>	46
mNAT3 516	sense	(516)- TTT CAA ACC AAG AAT TCA -(533)	Mouse <i>Nat3</i>	50
mNAT3 867	antisense	(867)- CTA AAT AGT AAA AAA GAC A -(886)	Mouse <i>Nat3</i>	50
mNAT2 1 ⁵	sense	(1)- ATG GAC ATC GAA GCG TAC TTT G -(22)	Mouse <i>Nat2</i>	56
mNAT2 910 ⁵	antisense	(910)- TTC CAA GTA CAT GGA AGG ACA CC -(888)	Mouse <i>Nat2</i>	56
Mus12 ⁶	sense	(-14)- TGC CTT AGG GAC ATA TGG ACA T -(8)	Mouse <i>Nat1, Nat3</i>	56, 50
Mus13 ⁶	antisense	(894)- AGA AGA ATT CTG CTC CTT ACC C -(873)	Mouse <i>Nat1</i>	56
Mus15 ⁶	antisense	(893)- AGA TCG GAT CCC CTT ATT ACT C -(872)	Mouse <i>Nat3</i>	50

¹Payton and Sim, 1998; ²Hickman and Sim, 1991; ³Matas et al, 1997; ⁴Payton et al, 1999; ⁵Fakis et al, 2000; ⁶Kelly and Sim, 1994.

Table 2.1: Primers used for PCR amplifications. The nucleotide site of primer annealing is indicated in reference to the start of the coding region, using the adenosine of the ATG codon as 1.

2.2 General molecular biology methods

2.2.1 Isolation and purification of plasmid and PAC DNA

Isolation of plasmid DNA was performed using the Wizard Plus Mini-/Midi-Preps DNA Purification System by Promega from either 10ml bacterial culture (section 2.2.10) for a “mini-prep” or 50-100ml culture for a “midi-prep”, according to the manufacturer’s instructions. Isolation of PAC DNA was performed from 100ml culture using the Plasmid Midi Kit by QIAGEN, according to the manufacturer’s instructions for very low copy number and large size constructs.

2.2.2 Isolation and purification of cosmid DNA

Cosmid DNA was extracted from 20ml culture, following centrifugation at 2,000xg for 10min at 4°C. The derived cell pellet was resuspended in 200µl cell resuspension buffer (50mM glucose, 10mM EDTA, 25mM Tris-HCl pH 8.0) and transferred to a 1.5ml tube. After 10min, 400µl of lysis solution (0.2M NaOH, 1%SDS) were added and mixed gently. The mixture was placed on ice for 5min and then ice-cold 3M sodium acetate pH 5.2 (300µl) was added. The tube was inverted gently several times and the mixture was left on ice for 10min, then centrifuged at 13,000rpm for 5min in a microcentrifuge. The supernatant was transferred to a clean tube, spun as above and again the supernatant was transferred to a new tube to which isopropanol (600µl) was added. The DNA was precipitated at -70°C for at least 10min, followed by centrifugation at 13,000rpm for 5min at 4°C, in a microcentrifuge. The supernatant was removed and the pellet was air-dried briefly. The pellet was then resuspended in 200µl of 3M sodium acetate pH 6-7 and 200µl phenol were added. The mixture was vortexed and then centrifuged at 13,000rpm for 5min. The upper (aqueous) phase was then transferred to a clean tube and, following addition of 50µl of 3M sodium acetate pH 5.2, the lower (phenol) phase was spun again at 13,000rpm for 2min. The recovered aqueous phase was pooled with the product of the previous extraction. Isopropanol (200µl) was added and the tube was placed at -70°C for at least 10min. The cosmid DNA was then sedimented by centrifugation at 13,000rpm for 5min and the pellet was rinsed with 70% (v/v) ethanol. The pellet was finally air-dried and resuspended in 50µl TE buffer (10mM Tris-HCl, 1mM EDTA, pH 7.5). The extracted

cosmid DNA was then treated with RNase (10 μ g from a 10mg/ml stock) at 37°C for 15-30min. The concentration and purity of the preparations was determined spectrophotometrically (section 2.2.4) and each preparation was run on an 1% (w/v) agarose gel (section 2.2.5.1) to check the quality of the isolated cosmid DNA.

2.2.3 Isolation and purification of genomic DNA

High molecular weight genomic DNA was isolated from whole blood or homogenised tissue, according to a method by Dilela and Woo (1987). When blood was used, the sample was centrifuged at 1000xg for 10min at 4°C, to separate the plasma (upper layer) from the white blood cells (middle layer) and the red blood cells (lower layer). Genomic DNA was extracted from white blood cells by addition of equal volume of STE buffer (0.1M NaCl, 1mM EDTA, 50mM Tris-HCl, pH 7.5), 100 μ g/ml proteinase K (using stock 100mg/ml in 10mM Tris-HCl pH 7.6, 2mM EDTA, 0.1% [w/v] sodium dodecyl sulphate [SDS]) and 0.5% [w/v] SDS. The mixture was incubated at 50°C overnight and the DNA was extracted by addition of an equal volume of phenol:chloroform:isoamyl alcohol (24:24:1 v/v/v) and gentle mixing for 10min on a rotator at room temperature. The mixture was then transferred to a previously silanised (with 2% dimethyldichlorosilane in 1,1,1-trichloroethane) and autoclaved 30ml Corex tube and incubated on ice for 30min. The mixture was centrifuged at 5000xg for 20min at 4°C and the upper (aqueous) layer was transferred to a clean tube, without disturbing the protein interphase. DNA was precipitated by addition of 2 volumes ice-cold absolute ethanol and 0.1 volume 2M sodium acetate pH 5.5, followed by incubation at room temperature for 5min. The DNA was then spooled at the aqueous-ethanol interphase onto a sterile glass rod and rinsed with 70% (v/v) ethanol. Excess of ethanol was removed by gently squeezing the glass rod against the walls of the tube and then the DNA was released in 1ml TE buffer. The DNA was allowed to dissolve at 4°C overnight and its concentration and purity were then spectrophotometrically determined (section 2.2.4).

For isolation of genomic DNA from cultured cells, a similar approach was undertaken. Cells were recovered from 75cm² culture flasks by trypsinisation (section 2.3.1.1) and pelleted by centrifugation at 400xg for 5min. The supernatant was

removed and the cell pellet was resuspended in 5ml STE buffer. Extraction of genomic DNA was then performed as described above.

2.2.4 Measuring concentration and purity of DNA solutions

Concentration of DNA solutions was spectrophotometrically determined at 260nm, using quartz cuvettes of 1cm path length. The DNA sample was diluted 250-1000 fold in water and the concentration was calculated based on the assumption that absorbance (A_{260}) equal to 1 corresponds to 50 μ g/ml concentration of double stranded DNA (Sambrook et al, 1989). Purity of DNA solutions was spectrophotometrically estimated from the ratio of absorbance at 260nm to the absorbance at 280nm (A_{260}/A_{280}). A ratio of 1.8 or greater is indicative of DNA solutions of high purity (Sambrook et al, 1989).

2.2.5 DNA gel electrophoresis

2.2.5.1 Standard agarose gel electrophoresis

Routinely, DNA fragments ≥ 500 bp were separated by electrophoresis on 0.7-2% (w/v) multi-purpose (MP) agarose gels (Boehringer Mannheim). Small DNA fragments (< 500 bp) were separated on 2-4% (w/v) Metaphor agarose gels (Flowgen). The appropriate amount of agarose was dissolved in 50ml 0.5xTBE buffer (45mM Tris, 45mM borate, 1mM EDTA pH 8.3) by heating in a microwave oven, and ethidium bromide was then added to a final concentration of 0.2 μ g/ml. The gel was allowed to set in a cold room (4°C) and run under a constant electric field of 100V in a Flowgen mini-gel device containing 0.5xTBE buffer. Prior to loading, samples were mixed with gel loading buffer (40% [w/v] sucrose, 0.25% [w/v] bromophenol blue) at a 6:1 (v/v) ratio. At the end of the electrophoresis, the DNA was visualised on the gel and photographed under ultraviolet (UV) light. Molecular weight markers 1kb ladder (Gibco BRL), λ /HindIII+EcoRI ladder (Promega) and BioMarker EXT ladder (CAMBIO) were routinely run alongside the DNA samples to allow estimation of the size of the separated DNA bands. Quantification of the DNA contained in each band was possible by comparison with the bands of the BioMarker EXT ladder in the range of 100-2000bp. When 5 μ l of BioMarker EXT ladder were loaded on the gel, each band contained 50ng of DNA.

2.2.5.2 DNA electrophoresis for Southern blotting

Digested DNA was separated on agarose gels prior to transfer to membranes (section 2.2.11). Agarose gels 0.7-1% (w/v), 19x25cm in size, were prepared in 0.5xTBE or 1xTAE buffer (40mM Tris, 40mM acetate, 2mM EDTA, pH 8.0) and ethidium bromide was added to a final concentration of 0.2µg/ml. Electrophoresis was carried out overnight at 4°C at 30V, with 0.5xTBE or 1xTAE as running buffer. 1kb and λ /HindIII+EcoRI size markers were routinely run alongside the samples. DNA was visualised and photographed under UV light, with a ruler as scale.

2.2.5.3 Pulse field gel electrophoresis (PFGE)

Linear double stranded DNA molecules longer than 20kb in size cannot be resolved by the electrophoretic methods described above. Under constant electric field, the DNA does not migrate through the gel under the sieving effect of the pores and it moves end-on through the gel, in a way that resembles the movement of a snake. PFGE makes use of this type of migration, called “reptation”, to separate very large DNA molecules. Alternating pulsed electric fields are applied to the gel and the DNA molecules are forced to re-orient themselves every time the field changes. The DNA is separated by size, because larger molecules require longer time to re-orient on the gel (Sambrook et al, 1989). PFGE was used for separating digested PAC DNA and was performed with a CHEF-DR II Pulse Field Electrophoresis System (Bio-Rad). PFGE grade agarose (typically 1% (w/v)) was dissolved in 0.5xTBE buffer and poured in the casting tray without the addition of ethidium bromide. The DNA was mixed with gel to a volume of 100µl, transferred into appropriate casts and allowed to solidify and form plugs. These were slotted into the wells of the gel, which was then placed in 0.5xTBE buffer in the electrophoresis chamber. The electrophoresis conditions were as follows: Electric field 6V/cm, switch time 5-15sec and total time 16h.

Electrophoresis was performed at 4°C and the running buffer was re-circulated. At the end of the run, the gel was stained in 0.5µg/ml ethidium bromide, then examined and photographed under UV light with a ruler as scale. Size markers Lambda Ladder PFG Marker (New England Biolabs), Low Range PFG Marker (New England Biolabs) and

λ /*Hind*III+*Eco*RI (Promega) were used to estimate the size of the separated DNA fragments.

2.2.6 Purification of DNA from agarose and solutions

Isolation of specific DNA fragments was carried out from Low Melting Point (LMP) agarose gels (Flowgen) in 0.5xTBE or 1xTAE buffer, following electrophoresis at not more than 70V, as described in section 2.2.5.1. The gel was placed on a glass plate, to minimise damage of the DNA during exposure to UV light, and the DNA band of interest was excised as a gel slice using a sterile scalpel. The DNA was then purified from agarose, using the Wizard DNA Clean-Up System (Promega) or the Gene Clean II Kit (Bio 101). DNA was recovered directly from enzymatic reaction mixtures using the Wizard DNA Clean-up System (Promega) or by phenol-chloroform extraction and ethanol precipitation, as described (Sambrook et al, 1989). The recovered DNA was checked for quantity and quality by agarose gel electrophoresis (section 2.2.5.1).

2.2.7 Polymerase chain reaction (PCR)

PCR amplifications were performed in the presence of 0.2mM of each dNTP (dATP, dCTP, dGTP and dTTP), 0.25 μ M primers (section 2.1.3), 1.5-3.0mM MgCl₂, 1x *Taq*-polymerase buffer and 2.5U *Taq*-polymerase in a reaction volume of 50 or 100 μ l. Genomic DNA (0.5-1 μ g) or recombinant DNA (50-100ng) were used as template and a sample without template was included in each set of reactions, as negative control. The reaction mixture was overlaid with light mineral oil to prevent evaporation during amplification. Reactions were performed using an MJ Research Mini Cyclor or a Programmable Thermal Controller (both by Genetic Research Instruments). A PCR programme typically consisted of 5min hot start at 95°C, followed by 35 cycles of denaturation at 95°C for 30sec, primer annealing at the appropriate temperature (table 2.1) for 30sec and extension at 72°C for 1min per 1kb of target sequence. Final extension took place at 72°C for 5min. PCR products were checked by electrophoresis on agarose gels (section 2.2.5.1).

2.2.8 Restriction enzyme digestion

Restriction enzyme digestions were performed with 10U of restriction enzyme per 1µg of DNA template in 10-20µl reaction volume, in the appropriate buffer (specified by the supplier of the enzyme), supplemented with 0.1mg/ml bovine serum albumin (BSA). The reaction mixture was incubated at the appropriate temperature (specified by the supplier of the enzyme) for at least 2h. A mock reaction, lacking enzyme, was included as control. Where simultaneous digestion with multiple restriction enzymes was possible, the volume was adjusted, so that the total amount of glycerol (contained in the storage buffer of the enzymes) did not exceed the 10% of the total reaction volume. In case of incompatibility of the reaction conditions, multiple digestions were performed sequentially, following purification of the DNA from the reaction mixture (section 2.2.6) after each digestion.

2.2.9 Transformation of competent cells

E.coli JM109 High Efficiency Competent Cells (Promega) were transformed with recombinant plasmid DNA, as follows: Competent cells were thawed slowly on ice and then 50µl aliquots of cell suspension were transferred to chilled 15ml Falcon tubes. Recombinant DNA (1-50ng in a volume of less than 10µl) was mixed with the cells and the tubes were immediately returned on ice for 10min. The cells were then heat-shocked at exactly 42°C for 45-50sec and returned to ice for 2min, without shaking. SOC medium (950µl) was then added and the cells were incubated at 37°C for 1h on a shaker. 100ml of SOC medium contained 2g Bacto-Tryptone, 0.5g Bacto-Yeast Extract, 1ml 1M NaCl, 0.25ml 1M KCl, 1ml 2M Mg²⁺ stock (101.5g MgCl₂·6H₂O and 123.3g MgSO₄·7H₂O per 500ml of solution) and 1ml 2M glucose. Bacto-Tryptone, Bacto-Yeast Extract, 1M NaCl and 1M KCl were mixed together and the volume was adjusted to 100ml with water. The mixture was then autoclaved for 25min and allowed to cool to room temperature. Filter-sterilised (through 0.2µm sterile Acrodisc syringe filters, by Whatman) 2M Mg²⁺ stock and 2M glucose were then added, each to a final concentration of 20mM. All handling was carried out under sterile conditions in a flow-hood. Of the transformed competent cells grown in SOC medium, 200µl were then spread on LB-agar plates (section 2.2.10) and cultured overnight.

2.2.10 Bacterial culture

Bacterial cells were cultured overnight at 37°C, on plates containing solidified LB-agar medium (10g Bacto-Tryptone, 5g Bacto-Yeast Extract, 10g NaCl and 15g Bacto-agar per 1L of medium), supplemented with ampicillin or kanamycin (50µg/ml) for selection of recombinant plasmid, cosmid (ampicillin) or PAC (kanamycin) clones. Alternatively, bacterial cells were grown in suspension in LB medium (10g Bacto-Tryptone, 5g Bacto-Yeast Extract and 10g NaCl per 1L of medium) containing the same concentration of antibiotics. Incubation took place at 37°C on a shaker for 6-16h, depending on the volume of the culture (typically 10-100ml). Liquid cultures were used for isolation of plasmid, cosmid or PAC DNA (sections 2.2.1 and 2.2.2), and for preparation of stock cultures (stored at -70°C) in 15% (v/v) sterile glycerol.

2.2.11 Southern blotting

Following electrophoresis (sections 2.2.5.2 and 2.2.5.3), gels were prepared for transfer of the DNA to a nylon membrane. For hybridisation using radioactive probes (section 2.2.12.1), Hybond N⁺ positively charged nylon membrane (Amersham) was used, while for hybridisation with digoxigenin-labelled probes (section 2.2.12.2) positively charged nylon membranes were purchased from Boehringer Mannheim. PFGE-gels were initially depurinated in 0.25M HCl for 30min, shaking gently, for fragmentation of the DNA to facilitate transfer to the membrane. DNA was denatured by shaking the gel in 0.5M NaOH, 1.5M NaCl for 45-60min, and neutralised in 1M Tris-HCl, 1.5M NaCl, pH 7.5, for 45-60min. Before blotting, the membrane was briefly soaked in water and then in 2xSSC (prepared from a 20x stock containing 3M NaCl and 0.3M sodium citrate, pH 7.0). DNA was transferred to the nylon membranes either overnight by capillary forces (Sambrook et al, 1989), or for 2h using a Posiblot 30-30 Pressure Blotter attached to a Pressure Control Station (Stratagene), according to the manufacturer's instructions. The transfer buffer was 10xSSC, in both cases.

Alternatively, a method for rapid transfer of relatively small size DNA fragments was used, as described in Smith and Summers (1980). If DNA was larger than 5kb, it was depurinated, as above. The gel was soaked twice in 0.5M NaOH, 1.5M NaCl for 30min, and then twice in 1M ammonium acetate, 0.02M NaOH for 30min, shaking

gently. It was then placed on Saran Wrap (the DNA side up) and blotted by capillary forces for 1h, without using transfer buffer.

The DNA was fixed on the membrane by crosslinking with 120mJoules of UV irradiation, using the UV Stratalinker 1800 (Stratagene). Alternatively, the membrane was placed (the DNA side up) on four layers of Whatman 3MM paper soaked in 0.4M NaOH, for 20-30min, to allow fixation of the DNA. The membranes were wrapped in Saran Wrap and stored dry at room temperature, prior to hybridisation, or wet at 4°C or -20°C, following hybridisation.

2.2.12 Southern hybridisation

2.2.12.1 Hybridisation using radioactive probes

2.2.12.1.1 Radioactive probe labelling

DNA probes (typically 300-900bp in size) were labelled as previously described (Feinberg and Vogelstein, 1983). Gel-purified (section 2.2.6) probe (25ng) was labelled with 50µCi [$\alpha^{32}\text{P}$]dCTP (Amersham), using random sequence hexanucleotides to prime DNA synthesis at multiple sites along its length. Extension of these primers was carried out by the large “Klenow” fragment of DNA polymerase I (lacking 5'→3' exonuclease activity). The labelling reaction was performed using the Multiprime DNA Labelling System by Amersham, according to the manufacturer's instructions. The radiolabelled probe was purified from unincorporated radioactive nucleotides by filtration through a Sephadex G50-medium (Pharmacia) column, as described (Sambrook et al, 1989). The efficiency of incorporation of the radioisotope into the probe was determined by measuring the radioactivity associated with 1µl of labelling reaction (diluted 1:10 and mixed with 2ml scintillation fluid) before and after filtration through the Sephadex G50-medium column, using a scintillation counter. Efficiency was expressed as percentage of incorporation of the radioactive isotope and was typically around 50%.

2.2.12.1.2 Hybridisation

The membrane was briefly rinsed in 2xSSC and placed in a hybridisation bottle (Hybaid) with 15ml of hybridisation solution, containing 6xSSC, 5xDenhardt's solution (50x stock of Denhardt's solution contains 1% [w/v] BSA, 1% [w/v] polyvinylpyrrolidone and 1% [w/v] Ficoll), 0.5% (w/v) SDS, 100µg/ml salmon sperm DNA (mechanically sheared by passing through a 19G1 and then a 25G1 gauge needle of a syringe for at least 10 times, and then denatured at 100°C for 5min, prior to addition to the hybridisation solution) and 50% (v/v) formamide. The membrane was pre-hybridised for at least 1h at 42°C, by rotating in a hybridisation oven (Hybaid). The hybridisation solution was then discarded and replaced by 15ml of fresh solution mixed with the radioactive probe, previously denatured at 95°C for 5min. Hybridisation took place overnight at 42°C, by rotating. The membrane was washed for 15min at room temperature in a) 2xSSC, 0.5% (w/v) SDS and b) 2xSSC, 0.1% (w/v) SDS. The filter was then covered with Saran Wrap, placed in a cassette and exposed to autoradiographic film (Kodak X-OMAT) for up to 16h. The film was developed using an X-ograph Compact X2 automated film developer.

It was possible to use the hybridised membrane in subsequent hybridisations, provided that it was maintained wet during all hybridisation and washing steps. To remove the probe, the membrane was immersed in boiling 0.5% (w/v) SDS solution. Once cooled to room temperature, the wet membrane was wrapped in Saran Wrap and stored at 4 or -20°C.

2.2.12.2 Hybridisation using digoxigenin-labelled probes

2.2.12.2.1 Digoxigenin labelling of probes

To label probes with digoxigenin (DIG), DIG-conjugated dUTP (DIG-dUTP, from Boehringer Mannheim) was mixed with dTTP, in a dTTP:DIG-dUTP ratio of 19:1, and the mixture was used in PCR reactions at a final concentration of 0.05mM. The PCR mixture also contained 0.05mM each dATP, dCTP and dGTP and was otherwise identical to that described in section 2.2.7. The cycling conditions were as described (section 2.2.7). The product of the labelling reaction was gel-purified (section 2.2.6), denatured (100°C for 5min) and added to hybridisation solution (50% [v/v]

formamide, 0.1% [w/v] laurosylsarcosine, 0.2% [w/v] SDS, 5xSSC and 2x blocking solution [Boehringer Mannheim]) at a final concentration of 5-25ng/ml. The diluted probe mixture was then stored at -20°C and used in several hybridisation reactions. Prior to each use, the probe mixture was again heated to 100°C for 5min.

2.2.12.2.2 Mock hybridisations

To determine the appropriate working dilution for the probe mixture (originally at a concentration of 5-25ng per ml of hybridisation solution), mock hybridisations were performed, as follows:

A) The probe mixture was diluted 1:2 to 1:50 (figure 2.1a) and 1µl of each dilution was spotted on a strip of membrane, adjacent to 1µl of undiluted probe. The DNA was UV-crosslinked on the membrane (section 2.2.11), which was then developed as will be described in section 2.2.12.2.3 (figure 2.1a, 1-6).

B) Pieces (1x1cm) of membrane were placed in 1ml of hybridisation solution at 42°C for 1h. The hybridisation solution was then discarded and replaced by 1ml of 1:2 to 1:50 dilutions of probe mixture (figure 2.1a). The membranes were hybridised overnight and developed as will be described in section 2.2.12.2.3 (figure 2.1a A-F).

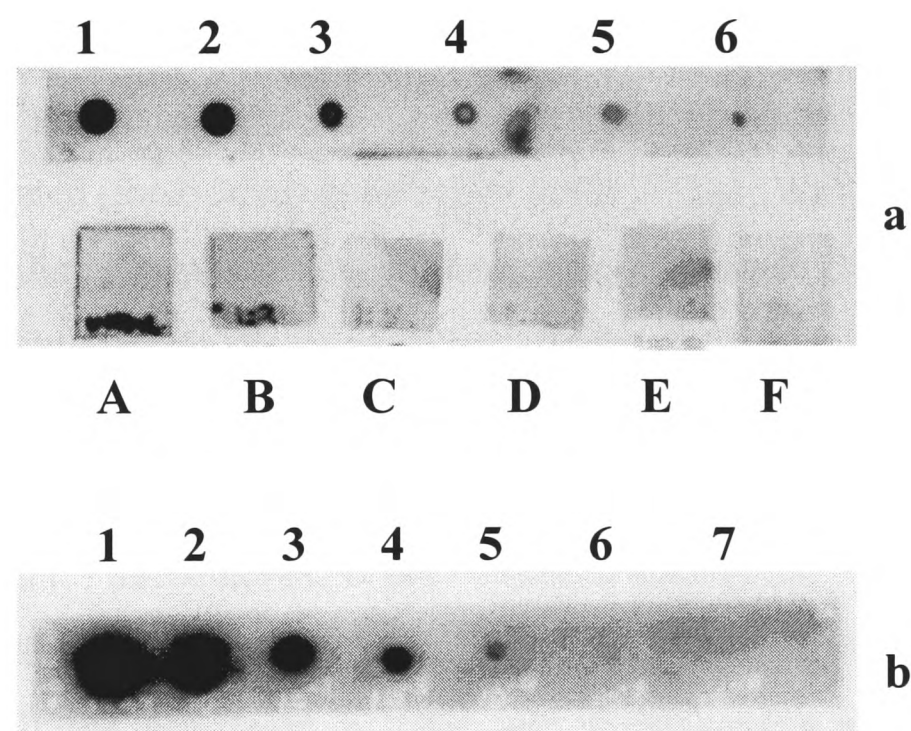
From the above mock hybridisations, the probe concentration providing the strongest specific signal (A) with the lowest background signal (B) was selected as the most appropriate for use in subsequent hybridisations (figure 2.1a).

Finally, to test whether the probe was sufficiently labelled to provide a strong signal on the autoradiogram, the gel-purified (but not diluted in hybridisation solution) DIG-labelled probe was diluted 10^{-1} to 10^{-6} in water (figure 2.1b) and 1µl of each dilution was spotted on a strip of membrane, next to 1µl of undiluted probe, as described above. A “good” probe should provide signal even at the 10^{-6} dilution. However, probes providing signal at dilutions $\geq 10^{-4}$ were considered strong enough to be used in hybridisations, even with templates of high complexity and low representation of the target sequence, such as genomic DNA (figure 2.1b).

Figure 2.1: Example of mock hybridisation, following labelling of a mouse *Nat2*-specific probe with digoxigenin, as described in section 2.2.12.2.1.

2.1a: Mock hybridisation to determine the optimal dilution of the probe providing adequate intensity of specific signal and acceptable level of background signal. Spots 1-6 are 1 μ l of probe (originally diluted in hybridisation solution at a concentration of 25ng/ml) spotted and fixed on the membrane at the following dilutions: “neat” (1), 1:2 (2), 1:5 (3), 1:10 (4), 1:20 (5) and 1:50 (6). A-F are 1x1cm pieces of membrane, without DNA, hybridised overnight with 1ml of probe (originally diluted in hybridisation solution at a concentration of 25ng/ml) at the same dilutions as above. The membranes were developed as described in section 2.2.12.2.3. The probe was not diluted further, as the background signal was not particularly high.

2.1b: Mock hybridisation to determine whether the probe is adequately labelled to provide a “strong” signal in hybridisations. Spots 1-7 are 1 μ l of gel-purified DIG-labelled probe (prior to addition to hybridisation solution) spotted and fixed on the membrane at the following dilutions: “neat” (1), 10^{-1} (2), 10^{-2} (3), 10^{-3} (4), 10^{-4} (5), 10^{-5} (6) and 10^{-6} (7). The membrane was developed, as described in section 2.2.12.2.3. The probe was sufficiently labelled for use in hybridisations, as it provided signal even at the 10^{-5} dilution.



2.2.12.2.3 Hybridisation and chemiluminescence detection of the signal

Pre-hybridisation and hybridisation of membranes took place in 15ml of hybridisation solution and the appropriate dilution of probe mixture (sections 2.2.12.2.1 and 2.2.12.2.2), respectively, as described in section 2.2.12.1.2. The membrane was washed twice in 2xSSC, 0.1% (w/v) SDS, at room temperature for 5min, and then twice in 0.1xSSC, 0.1% (w/v) SDS, at 42°C for 15min. All subsequent steps were performed at room temperature on a shaker. The volumes of the solutions used were adjusted to the size of the membrane, which should remain wet throughout the procedure. The membrane was rinsed briefly (for 1-5min) in washing buffer (0.1M maleic acid pH 7.5, 0.15M NaCl, 0.3% [v/v] Tween 20), and was then incubated in 1x blocking solution (Boehringer Mannheim) for at least 30min. The blocking solution was prepared in maleic acid buffer (0.1M maleic acid pH 7.5, 0.15M NaCl) as a 10x stock solution, according to the manufacturer's instructions. The membrane was then incubated with alkaline phosphatase-conjugated anti-DIG-antibody (Boehringer Mannheim), diluted 1:10,000 in 1x blocking solution. Incubation took place for 30min. The membrane was subsequently washed twice for 15min in washing buffer and then equilibrated (for 2-5min) in detection buffer (0.1M Trizma base, 0.1M NaCl, pH 9.5). The alkaline phosphatase chemiluminescence substrate CSPD[®] (Boehringer Mannheim), diluted 1:100 in detection buffer, was then added to the membrane for 5min. The membrane was then briefly blotted on Whatman 3MM paper, sealed in a plastic bag, incubated at 37°C for 15min, and then exposed to a photographic film (section 2.2.12.1.2). The membrane could be used in subsequent hybridisations, following removal of the probe by incubation in 0.2M NaOH, 0.1% (w/v) SDS at 37°C for 30min. It was then rinsed in 2xSSC and stored wet at 4 or -20°C, wrapped in Saran Wrap.

2.3 Cytogenetic methods

2.3.1 Cell culture

All procedures were carried out under sterile conditions in a flow hood. Cells were grown at 37°C in a 5% CO₂ atmosphere.

2.3.1.1 Culture of the RT112 cell line

The RT112 human urothelial cell line was obtained from the cell bank of the Sir William Dunn School of Pathology, University of Oxford, in the form of 1ml stock stored in liquid nitrogen. The complete culture medium was Minimum Essential Medium (MEM) with Earle's salts, supplemented with 10% (v/v) fetal calf serum, 1% (v/v) non-essential amino acids and 2mM L-glutamine. Freezing medium consisted of 90% (v/v) complete culture medium, plus 10% (v/v) dimethyl sulphoxide (DMSO). The cells were rapidly thawed in a 37°C water bath and then added to 9ml of complete culture medium. The cells were pelleted by centrifugation (250xg, 4min, 20°C) the supernatant was discarded and the cell pellet was resuspended in 4ml of complete culture medium. The cell suspension was then added to a 75cm² culture flask, containing 14ml of complete medium. When confluent, the cells were passed to new 75cm² flasks, as follows: The culture medium was removed from the flask and the cells were rinsed twice with phosphate buffered saline (PBS, containing 137mM NaCl, 2.7mM KCl, 10mM sodium phosphate, pH 7.4). PBS (2ml), containing 0.25% trypsin-EDTA, was then added and the cells were allowed to detach from the flask, by shaking gently for about 5min at room temperature. Trypsinisation was terminated by addition of 4ml complete culture medium and the cell suspension was thoroughly mixed using a sterile 10ml pipette. The subcultivation ratio for 75cm² flasks was typically 1:6. Harvesting of the cells was performed by trypsinisation, as above, but the cells were pelleted by centrifugation (250xg, 4min, 20°C) and resuspended in the appropriate volume of culture medium.

2.3.1.2 Culture of T-lymphocytes from whole blood

T-lymphocytes were cultured from whole blood, as described by Buckle and Rack, 1993. Blood (14ml) was collected in lithium heparin tubes and used within two days.

T-lymphocytes were cultured in RPMI medium, supplemented with 10% fetal calf serum and 2mM L-glutamine. Phytohaemagglutinin (100µl) was also added, to stimulate T-lymphocyte growth. Whole blood (250µl) was added to 5ml of complete culture medium (pre-warmed at 37°C) in 15ml sterile tubes. Cultures were maintained for 48-72h.

2.3.2 Preparation of metaphase spreads for FISH analysis

Colcemid was added to culture medium at a final concentration of 20ng/ml to arrest the cells at the metaphase stage. Incubation took place at 37°C for 15min, and the cells were then centrifuged at 500xg for 5min at room temperature. The cell pellet was resuspended in 5ml of 75mM KCl (hypotonic solution pre-warmed at 37°C) and incubated at 37°C for 10min. This was followed by centrifugation at 500xg for 5min and removal of the supernatant. The cells were then resuspended in 5ml fixative (70% [v/v] methanol, 30% [v/v] glacial acetic acid), by slow addition of the fixative, while vortexing gently to prevent formation of lumps. The cells were centrifuged at 500xg for 5min, the supernatant was removed, the cells were resuspended in 5ml fixative and then spun again. This wash was repeated once more, and the cells were finally resuspended in fixative, at a concentration of approximately 10^4 cells/ml. Using a wide bore pastette, one droplet of this cell suspension was placed on a slide, which was allowed to air-dry and then checked under the microscope for spreading and density of the chromosomes. The slides were used either immediately for FISH analysis or stored in a sealed container with desiccant at -20°C.

2.3.3 Preparation of interphase spreads for FISH analysis.

Each cell suspension was centrifuged at 1,500xg for 10min at 4°C, the supernatant was removed and the cells were resuspended in 20ml water. They were spun again and finally resuspended in 100µl of water per 10^4 cells. 200µl of this cell suspension were then loaded onto a Shandon Cytospin II centrifuge and the cells were spun onto slides for 5min at 500xg. The slides were air-dried and fixed for 5min in a solution containing 70% (v/v) 75mM KCl and 30% (v/v) fixative (section 2.3.2). Fresh fixative was added to excess and the slides were then dried at 65°C. Slides were stored in sealed boxes at -20°C.

2.3.4 Preparation of APES-coated slides for microscopy

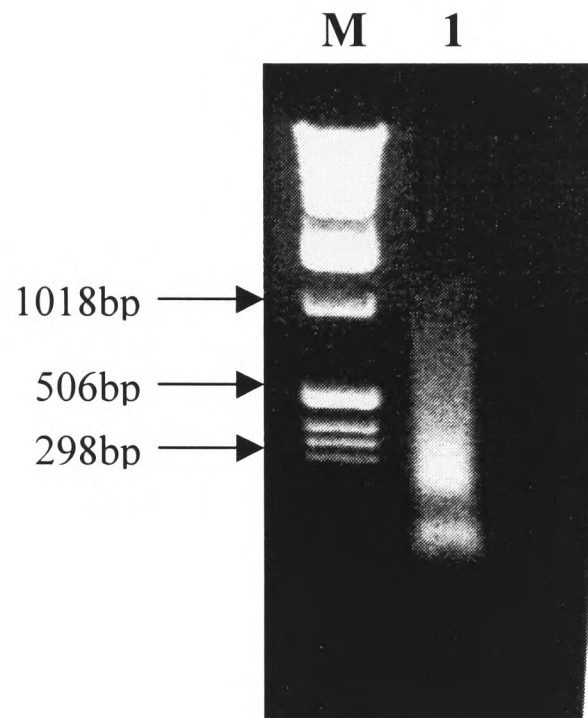
Slides used in microscopy were coated with 3-aminopropyltriethoxysilane (APES), to ensure binding of the cells. The slides were placed in a mixture of 9ml APES and 450ml acetone for 20sec and then washed twice in acetone and twice in water for 20sec. They were air-dried and used to prepare chromosome spreads for FISH analysis.

2.3.5 Labelling of probes for FISH analysis

Nick translation was used for labelling human *NAT1*- and *NAT2*-specific cosmid probes and mouse *Nat*-specific PAC probes with biotin-16-dUTP (Boehringer Mannheim). Probe DNA (1µg) was mixed with 1.2µl 1mM biotin-16-dUTP, 5µl dNTP mix (0.5mM dATP, dGTP, dCTP and 0.1mM dTTP), 5µl 10x nick translation buffer (0.5M Tris-HCl pH 7.5, 50mM MgCl₂, 0.5mg/ml BSA), 5µl 100mM dithiotreitol (DTT), 3µl 2.5ng/µl DNase I and 3µl 3.5U/µl DNA-polymerase I. The reaction volume was adjusted to 50µl with water. The reaction took place at 15°C for 90min, followed by incubation at 75°C for 15-30min to inactivate the enzymes. The reaction product (4µl) was then loaded on a 1.4% (w/v) agarose gel and run to check efficiency of the digestion. Products should form a smear extending from fragments 100 to 500bp in size (Buckle and Rack, 1993), as shown in figure 2.2. When the average size of the produced fragments was too large, nick translation was repeated for 30min longer, using the product of the first reaction as substrate.

The human chromosome 8 centromere-specific probe (Cytocell) was purchased labelled with digoxigenin, while the mouse chromosome 8 paint (CAMBIO) was purchased labelled with Cyanine 3 (Cy3), which can be directly visualised under a fluorescent microscope.

Figure 2.2: Labelling of a human *NAT1*-specific cosmid probe with biotin-16-dUTP, by nick translation. Incubation took place for 2h and the generated fragments ranged from 100-500bp in size, as recommended (Buckle and Rack, 1993). The probe was subsequently used for FISH analysis. The size marker (lane M) is the 1kb ladder. Lane 1 shows 4 μ l of the reaction product, following electrophoresis on an 1.4% agarose gel, as described in section 2.2.5.1.



2.3.6 Fluorescence *in situ* hybridisation (FISH)

2.3.6.1 FISH analysis of human T-lymphocyte, RT112 and barbotage cells

Metaphase spreads prepared from RT112 cells (section 2.3.2) or interphase spreads prepared from T-lymphocytes, RT112 or barbotage cells (section 2.3.3) were used for FISH analysis with biotin-labelled human *NAT1*- or *NAT2*-specific cosmid probes, together with a DIG-labelled human chromosome 8 centromere-specific probe (section 2.3.5). Approximately 100ng of cosmid probe, 10ng of chromosome 8 centromere-specific probe and 2 μ g of human COT-1 DNA (to block non-specific binding) (GibcoBRL) were mixed together and precipitated with 2 volumes of absolute ethanol and 0.1 volume 3M sodium acetate pH 5.2 (Sambrook et al, 1989). The pellet was air-dried and dissolved in 11 μ l of hybridisation mix (50% [v/v] formamide, 2xSSC, 10% [w/v] dextran sulphate). Slides to which cell nuclei were attached (sections 2.3.2 and 2.3.3) were dehydrated by passing them through 70% (v/v), 90% (v/v) and 100% (v/v) ethanol, and dried with an air-jet. They were then

placed in 70% (v/v) formamide, 2xSSC at 70°C for 6min (for uncultured barbotage cells) or 2min (for cultured RT112 and T-lymphocytes) to denature nuclear DNA. The slides were dehydrated through an ethanol series and dried again, as described above, and then placed in a box at 37°C, until the probe was added within 30min. The probe was denatured at 70-80°C for 5min and then added to the slides. A coverslip was placed on the preparation and sealed with rubber solution. Hybridisation took place overnight at 37°C in a humidified chamber.

The following day, the solidified rubber solution was removed and the coverslips were detached from the slides by placing them in 2xSSC. The slides were then washed three times for 5min in 50% formamide, 2xSSC at 42°C and then in 2xSSC at 42°C. They were then placed in 4xSSC, 0.05% (v/v) Tween 20 for 5min at 42°C, and then in the same solution at room temperature. Excess fluid was drained off and 10µl avidin-FITC (fluorescein isothiocyanate) conjugate (2.5µg/ml in 4xSSC, 0.05% [v/v] Tween 20, provided by Vector) was added to each slide, which was covered with parafilm and incubated in the dark for 20min. The parafilm was then removed and the slides were washed three times in 4xSSC, 0.05% (v/v) Tween 20 for 5min, at room temperature, in the dark. Biotinylated anti-avidin antibody (10µl of 5µg/ml solution in 4xSSC, 0.05% [v/v] Tween 20, provided by Vector) was then added to each slide and the incubation and washing steps were repeated as above. Avidin-FITC conjugate (10µl) was finally added to each slide, together with 10µl rhodamine-antidigoxigenin conjugate (200µg/ml, Boehringer Mannheim), and the incubation/washing steps were repeated. This allowed visualisation of both the *NAT*-specific cosmid probe (labelled with two layers of FITC) and the chromosome 8 centromere-specific probe (labelled with rhodamine) on the same preparation.

DAPI (4',6'-diamidino-2-phenylindole, from Boehringer Mannheim) was used to counterstain cellular DNA. A 10µg/ml stock solution of DAPI in 4xSSC, 0.05% (v/v) Tween 20 was prepared and 20µl of this solution were mixed with 1ml of Vectashield antifade mountant (Vector) to create a working concentration of 200ng/ml DAPI. 10µl of this mixture were placed on a coverslip and used to cover the preparation. The

coverslip was sealed with nail polish and the slides were stored in a dark box at 4°C. Slides were viewed under a Zeiss Axioscope fluorescence microscope.

2.3.6.2 FISH analysis of mouse spleen cells

The biotin-labelled mouse *Nat*-specific PAC probes were co-hybridised with the Cy3-labelled mouse chromosome 8 paint to metaphase chromosomal spreads prepared from a culture of normal mouse spleen cells, stimulated with lipopolysaccharide (LPS) for 48h (provided by Dr V. Buckle, Institute of Molecular Medicine, University of Oxford). Slides were treated with RNase to reduce potential background signal, by adding 100µl of 100µg/ml RNase on the chromosomal spreads and incubation of the slides (covered with parafilm) in a humidified chamber at 37°C for 0.5-2h. Slides were dehydrated and cellular DNA was denatured for 5min, as described in section 2.3.6.1.

Biotin-labelled PAC probe (100ng) was vacuum-dried and resuspended in 15µl of Cy3-labelled mouse chromosome 8 paint in hybridisation solution, pre-warmed at 37°C. The probe mixture was denatured at 65°C for 10min and transferred to a 37°C waterbath for 60-90min, to allow binding of the mouse Cot-1 DNA (present in the hybridisation solution) to any repetitive sequences on the probe. Since the Cy3 fluorochrome of the paint is light-sensitive, it was essential to perform as much handling as possible in the dark. The slides carrying the metaphase spreads were pre-heated to 42°C on a warm plate and 15µl of the probe mixture were added to each chromosomal spread. The overnight hybridisation was performed at 42°C and washes were carried out as described in section 2.3.6.1.

Before proceeding with the immunological detection, the slides were incubated in 3% (w/v) BSA, 4xSSC, 0.05% (v/v) Tween 20 for 20min at 37°C, to block non-specific protein binding sites. Avidin-FITC conjugate (100µl of 5µg/ml solution in 3% [w/v] BSA, 4xSSC, 0.05% [v/v] Tween 20) was added onto the slides, which were covered with parafilm and incubated at 37°C for 20min in a humidified chamber. They were then washed three times for 3min in 2xSSC, 0.05% (v/v) Tween 20. Biotinylated anti-avidin antibody (100µl of 5µg/ml solution in 3% [w/v] BSA, 4xSSC, 0.05% [v/v]

Tween 20) was then added to each slide and the incubation and washing steps were repeated. Finally, one more layer of avidin-FITC was applied and the incubation and washing steps were repeated. The slides were finally washed twice for 3min in 2xSSC, 0.05% (v/v) Tween 20 and then once for 3min in PBS. Wet slides were mounted in Vectashield containing 1.5µg/ml DAPI, covered with coverslips and stored (section 2.3.6.1). Images were captured using a CCD camera and Mac Probe version 4.1 software (Perceptive Scientific Instruments). The exact position of each PAC probe was identified by generating black and white images of the DAPI banding and overlaying the fluorescent PAC signal.

2.4 General methods for protein analysis

2.4.1 Measuring protein concentration

The protein concentration was determined using a dye-binding assay (Bradford, 1976). The appropriately diluted protein sample (200µl) was mixed with 50µl of Bio-Rad dye (Bio-Rad Laboratories), the colour was allowed to develop for 10min and the absorbance was read at 580nm in either a microtitre plate reader (Titertek Multiscan PLUS) or in a 1cm cuvette in a spectrophotometer. Gamma globulin (Bio-Rad Laboratories) was used as a protein standard in the range of 0-15.4µg/ml. All samples were prepared in triplicate. A blank without protein was included in each assay. Standard curve construction and handling of the data were carried out using the Delta SOFT II computer programme for the Macintosh.

2.4.2 Dot blot

To determine the appropriate amount of total cytosolic protein and the appropriate dilution of the primary antibody to use in Western blotting experiments, a dot blot was performed. A range (0.5, 1, 2.5, 5, 10, 50, 100, 500 and 1000µg) of total protein from a pooled cytosolic sample was loaded on a dot blot apparatus (Bio-Rad) and transferred to a Hybond C nitrocellulose membrane (Amersham) by applying vacuum. Prior to transfer, the membrane was soaked briefly in water and equilibrated in 1xTBS (prepared from a 10x stock containing 1.5M NaCl, 0.5M Tris-HCl, pH 7.9) for at least

10min. Transfer took place for approximately 1h. The membrane was then rinsed in 1xTBS three times and developed as will be described in section 2.4.4. Anti-sera was diluted 1:1000, 1:1500, 1:2000 and 1:3000, in order to determine the appropriate working dilution in subsequent Western blot analyses.

2.4.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

For optimal separation of proteins with molecular weight around 30kDa, 12% polyacrylamide gels were prepared in mini-gel apparatus (Mighty Small II, Biotech Instruments) and used for SDS-PAGE. The separating gel consisted of 3.7ml acrylamide:bis-acrylamide mix (29:1) (Scotlab), 2.3ml separating gel buffer (0.5M Tris-HCl, pH 8.8, 0.5% [w/v] SDS) and water to a final volume of 9.5ml.

Polymerisation of the gel was initiated by addition of 50 μ l 10% (w/v) ammonium persulphate (APS) and 7.5 μ l N,N,N',N'-tetramethyl-ethylenediamine (TEMED). The stacking gel (poured on top of the separating gel) consisted of 0.38ml acrylamide:bisacrylamide (29:1), 0.63ml stacking gel buffer (0.5M Tris-HCl, pH 6.8, 0.5% [w/v] SDS) and water to a final volume of 2.51ml. Again, 7.5 μ l 10% APS and 5 μ l TEMED were added to initiate polymerisation of the gel. An appropriate comb was used to generate wells on the gel.

Before loading, a volume of sample containing the appropriate amount of total cytosolic protein (determined as optimal by dot blot analysis) was mixed with an equal volume of sample preparation buffer (8M urea, 2% [w/v] SDS, 6mg/ml DTT, 200mM Tris-HCl, pH 7.5) and heated at 90°C for 30min. This was followed by addition of 0.1 volume 83mg/ml iodoacetamide and 0.1 volume loading solution (2.5mg bromophenol blue in 2.5ml glycerol and 5ml stacking gel buffer), and further heating at 90°C for 10min. Electrophoresis took place at constant current of 25-30mA per gel, using 50mM Tris-HCl pH 8.3, 50mM glycine and 1% (w/v) SDS as running buffer. As a molecular weight standard, 5 μ l of pre-stained "Rainbow" ladder (Amersham) was used. The marker was treated similarly to the samples, prior to loading onto the gel. Following electrophoresis, the gels were either blotted (section 2.4.4), or stained for 2min in Coomassie blue (3.75g Coomassie blue dye per litre of 0.5M methanol, 1.87M acetic acid) and destained overnight in 0.5M methanol, 0.75M

acetic acid, to allow visualisation of the protein bands. Stained protein gels were sealed in a plastic bag and stored at room temperature.

2.4.4 Western blotting

Following SDS-PAGE of protein samples, the gel was blotted onto Hybond C nitrocellulose membrane overnight. Transfer was carried out using a Bio-Rad Trans Blot Cell apparatus, according to the manufacturer's instructions, under constant current of 250mA at 4°C. The transfer buffer consisted of 20% (v/v) methanol and 2.8g/l sodium dibasic phosphate. The following day, the membrane was removed from the blotting apparatus and washed twice in TBST (1xTBS containing 0.05% [v/v] Tween 20) for 5min at room temperature. It was then placed in blocking solution (TBST containing 10% [w/v] dried skimmed milk) for 1-2h and then washed three times for 10min in 3% (w/v) dried skimmed milk in TBST. The membrane was subsequently incubated for 2h in primary antiserum, diluted in 3% (w/v) dried skimmed milk in TBST as determined by the dot blot. Excess antiserum was washed three times with 3% (w/v) dried skimmed milk in TBST for 10min. The membrane was subsequently incubated for 1h with a horseradish peroxidase-conjugated goat anti-rabbit IgG secondary antibody (Dako), diluted 1:12,000 in 3% (w/v) dried skimmed milk in TBST. The membrane was finally washed three times, as above, and rinsed briefly in TBST. The antibody binding was visualised by a chemiluminescence method using the ECL reagents (Amersham), according to the manufacturer's instructions. Autoradiography was carried out as described in section 2.2.12.1.2.

2.5 NAT activity assays

NAT activity assays were performed using the method described by Ward et al (1992), but with slight modifications. Cytosols were thawed and immediately placed on ice. Samples were centrifuged at 13,000rpm in a microcentrifuge at 4°C for 15min to remove all cellular debris and then each sample was appropriately diluted in dilution buffer (10mM Tris-HCl, 1mM EDTA, pH 7.5, 1mM DTT). The reaction mixture consisted of 50µl sample and 20µl substrate mix, containing 200mM Tris-

HCl, 4mM EDTA pH 7.5, 4mM DTT and either 1mM of the human NAT1-specific substrate pABA or 0.5mM of the human NAT2-specific substrate sulphamethazine. The tubes containing the reaction mixture were pre-warmed at 37°C for 5min and the reaction was initiated by addition of 30µl acetyl CoA to a final concentration of 0.4mM. The reaction was terminated by addition of 100µl 20% (w/v) trichloroacetic acid (TCA), followed by vortexing. The precipitated protein was pelleted by centrifugation at 13,000rpm for 15min at 4°C. Supernatant (100µl) of each sample was then added to 900µl of 5% (w/v) dimethylaminobenzaldehyde (DMAB) in acetonitrile and vortexed. The DMAB binds to the non-acetylated substrate, producing a yellow complex. Absorbance was measured at 450nm in a 1cm cuvette and was an indication of the remaining amount of non-acetylated substrate in the reaction mix. The blank was 5% DMAB in acetonitrile. All samples were prepared in triplicate and the N-acetylation activity was calculated from the mean of the three absorbance measurements. A control sample containing water, instead of cytosol, was always included. Each sample was analysed for NAT activity following incubation for 0, 10, 30 and 60min, in order to determine the rate of N-acetylation by the breast cytosols.

The concentration of pABA in each sample was calculated using the Beer-Lambert Law:

$$\Delta A_{450} = \epsilon \times \Delta C \times L$$

Where: ΔA_{450} = change in absorbance at 450nm

ϵ = extinction coefficient of pABA at 450nm (2.76 mM⁻¹cm⁻¹, Merck Index)

ΔC = change in concentration (in mM)

L = path length of the cuvette (1cm)

Chapter 3: Arylamine N-Acetyltransferase

Genes as Markers for Chromosomal Region

8p22 in Humans

3.1 Introduction

3.1.1 Cloning and characterisation of the genes for human NAT

Although human NATs have been extensively studied at the phenotypic level for almost fifty years, the molecular genetics underlying variability in the N-acetylation of drugs and xenobiotics became clear only during the past decade. Identification, cloning, genetic manipulation and recombinant expression of the genes encoding for human NATs, as well as characterisation of numerous mutations affecting N-acetylation activity, have allowed understanding of the molecular aetiology of the NAT polymorphism in humans and provided information about the catalytic properties of the NAT isoenzymes (Price Evance, 1992; Vatsis and Weber, 1994 for historical reviews).

Isolation of the genes for human NAT only became possible following molecular genetic studies using laboratory animals as models. Andres et al (1987) purified NAT from the liver of homozygous rapid acetylators rabbits and reported the amino acid sequence of 16 tryptic peptides generated by Edman degradation of the NAT protein. Based on the amino acid composition of these peptides, NAT-specific oligonucleotide probes were synthesised and used to isolate cDNA and genomic clones for NAT from liver of chicken (Ohsako et al, 1988) and rabbit (Blum et al, 1989; Sasaki et al, 1991). Using a rabbit NAT cDNA probe to screen a human genomic library, a 1891bp long

EcoRI fragment, containing the complete open reading frame of a human gene for NAT, was isolated and sequenced (Grant et al, 1989a). Additional NAT-positive genomic clones were later isolated and mapped with restriction enzymes by Blum et al (1990). Based on the restriction pattern they produced, these NAT-positive clones could be classified into three groups, suggesting the presence of three distinct NAT genes in humans. This was further confirmed by Southern blot analysis of genomic DNA. Sequencing analysis revealed that two of these genes, designated *NAT1* and *NAT2*, had an intronless coding region of 870bp. The third gene had a similar structure, but contained a total of eight frameshift and termination mutations, which caused disruption of the open reading frame. Therefore, it was considered as a NAT pseudogene and designated *NATP* (Blum et al, 1990). *NAT1* shows 87% nucleotide sequence identity to *NAT2*, while *NATP* is 79% homologous to *NAT1* and 80% homologous to *NAT2*.

The open reading frames of *NAT1*, *NAT2* and *NATP* were transiently transfected into COS-1 cells, which were then assayed for NAT enzymatic activity using sulphamethazine as substrate. Cells transfected with *NAT1* and *NAT2* constructs provided high levels of NAT-activity, whereas cells transfected with the *NATP* construct only provided background NAT activity. Furthermore, Western blot analysis was able to detect only NAT1 and NAT2 protein in transiently transfected COS-1 cells. These results confirmed that *NAT1* and *NAT2* represented functional NAT genes, while *NATP* was a NAT pseudogene (Blum et al, 1990). The transiently expressed product of the *NAT1*-gene was later demonstrated to N-acetylate pABA and p-aminosalicylic acid, while the product of the *NAT2* gene efficiently N-acetylated sulphamethazine and procainamide. Therefore, it was concluded that the *NAT1* gene encoded for “monomorphic” NAT, while the product of the *NAT2* gene was “polymorphic” NAT (Grant et al, 1991). Following the discovery of *NAT1* polymorphism (Vatsis and Weber, 1993), the terms “monomorphic” and “polymorphic” NAT were abandoned and replaced by NAT1 and NAT2 respectively (Vatsis et al, 1995).

Isolation and analysis of the cDNAs encoding for NAT in humans was achieved by other researchers, at around the same time. Two liver cDNA libraries were screened by Ohsako and Deguchi (1990), using a rabbit NAT cDNA-specific probe (Sasaki et

al, 1991). Three different types of cDNA clones were isolated and mapped by restriction digestion, two of which were identical, apart from a *Bam*HI site, whereas the third was considerably different. Sequencing analysis indicated that the two clones that provided similar restriction patterns only differed at position 857 of the open reading frame, while the nucleotide sequence of the third clone was highly homologous, but distinct. All three cDNAs encoded for a 290 amino acid polypeptide chain. Transient expression of the cDNAs in CHO cells, followed by enzymatic activity assays with pABA and sulphamethazine, demonstrated that one of the clones encoded for “monomorphic” NAT (NAT1), while the two other clones encoded for a fast (NAT2 4) and a slow (NAT2 7A) form of “polymorphic” NAT (NAT2) (Ohsako and Deguchi, 1990). The cDNA clones were subsequently used to isolate and restriction map a series of *NAT*-positive genomic clones (Ebisawa and Deguchi, 1991).

Both *NAT1* and *NAT2* genes in humans have an intronless open reading frame (Grant et al, 1989a; Blum et al, 1990; Ebisawa and Deguchi, 1991). However, the 107 nucleotide 5' untranslated region (UTR) of the *NAT2* cDNA is completely different from the genomic sequence immediately upstream of the *NAT2* coding region. In contrast, the 5' UTR of the *NAT1* cDNA is identical to the genomic sequence 146 nucleotides upstream of the *NAT1* coding region (Grant et al, 1989a; Blum et al, 1990; Ohsako and Deguchi, 1990; Ebisawa and Deguchi, 1991). This indicated that the 5' UTR of human *NAT2* is present in an exon further away from the coding region of the gene. This non-coding exon was localised about 8kb upstream of the coding region of *NAT2* and is connected to it via splicing of the intron (Ebisawa and Deguchi, 1991). The 3' UTR of both *NAT1* and *NAT2* cDNA is identical to the genomic sequence immediately downstream of the stop codon of each one of the genes (Blum et al, 1990; Ohsako and Deguchi, 1990; Ebisawa and Deguchi, 1991).

3.1.2 Chromosomal localisation and mapping of the human NAT genes

Chromosomal localisation of the *NAT1* and *NAT2* genes was first reported by Blum et al (1990) by Southern blot analysis of DNA from a panel of human-rodent somatic cell hybrids. The two genes co-localised on human chromosome 8 and additional screening of human-hamster somatic cell hybrids, containing established translocations of human chromosome 8, further refined localisation of *NAT1* and *NAT2* in region 8pter-q11 (Blum et al, 1990). More precise localisation of the *NAT* genes was later accomplished independently by two other groups of researchers, using FISH analysis (Hickman et al, 1994; Franke et al, 1994). The first study (Hickman et al, 1994) used YAC clones as probes for FISH analysis. Two YAC clones, one positive for *NAT1* and the other positive for *NAT2*, were isolated by PCR screening of a YAC library. Both the *NAT1* and *NAT2* probes hybridised to the short arm of chromosome 8, to region 8p21.3-23.1, establishing co-localisation of the two genes on the same chromosomal region. In the other study (Franke et al, 1994), a *NAT2*-positive cosmid clone was used as probe for FISH on metaphase spreads of healthy human lymphocytes and a human-hamster hybrid cell line containing human chromosome 8. This second study independently demonstrated localisation of the *NAT2* gene on the short arm of chromosome 8, in region 8p22, in agreement with Hickman et al, 1994.

Physical mapping of the human NAT genes (Matas et al, 1997) was based on screening of a YAC contig previously constructed for chromosomal region 8p22 by Bookstein et al (1994). Mapping of 36 overlapping YAC clones, isolated from a CEPH YAC library, defined the physical order for a number of polymorphic markers on chromosomal region 8p22. The order of the markers was: telomere-D8S26-D8S511-D8S549-*MSR*-D8S254-D8S233-D8S261-D8S21-*LPL*-D8S258-centromere (Bookstein et al, 1994). The genes for macrophage scavenger receptor (*MSR*) and lipoprotein lipase (*LPL*), previously mapped on 8p22 (Emi et al, 1993a; Mattei et al, 1993), were used to align the YAC contig with cytogenetic maps of the region (Bookstein et al, 1994). The YAC clones comprising this contig were subsequently screened for the presence of the *NAT1* and *NAT2* genes and the pseudogene *NATP* (Matas et al, 1997). All positive clones contained *NATP*, confirming its co-localisation with *NAT1* and *NAT2* on 8p22. Moreover, it was established that the *RsaI*

RFLP marker D8S21 (Steinbrueck et al, 1992) is located within the *NAT2* gene, very close to the known *DdeI* polymorphism at position 803 of the open reading frame, which gives rise to the slow alleles *NAT2*5B* and *NAT2*5C* (Hickman et al, 1992). Restriction mapping of the *NAT*-positive YAC clones also allowed localisation of the microsatellite marker D8S261 on the telomeric side of *NAT1*. The order of the loci on 8p22 was determined to be telomere-D8S261-*NAT1*-*NATP*-*NAT2*/D8S21-centromere. *NAT2* and *NATP* were mapped closer to each other, with *NAT1* lying further apart, within a distance of 170-360kb (Matas et al, 1997).

3.1.3 The Human Genome Project and Chromosome 8

Chromosome 8 represents 4.87% of the human genome and has an estimated size of 155Mb. Washington University and the Whitehead Institute have been coordinating sequencing of chromosome 8 in collaboration with other centres, like the Japanese Foundation for Cancer Research (JFCR) and the Institute of Molecular Biology (IMB) in Jena, Germany. The finished sequence published to date represents only the 5.9% (9.1Mb) of its total size, but additional sequencing information is available in a draft format. However, gaps in the sequence of chromosome 8 still exist (National Centre for Biotechnology Information [NCBI], USA, <http://www.ncbi.nlm.nih.gov/>, June 2000). At least 1200 STSs have been localised by physical mapping (Deloukas et al, 1998), and about 250 microsatellite markers by genetic mapping (Dib et al, 1996) on human chromosome 8. A total of 256 genes have been reported to localise on chromosome 8, of which approximately 30 are in region 8p22 (Genome Data Base [GDB], <http://www.gdb.org/>, June 2000). According to the National Cancer Institute (NCI, USA, <http://nciarray.nci.nih.gov/>, June 2000), at least fifty genes on chromosome 8 have been associated with genetic diseases, including those responsible for the Cohen (*COH1*, 8q22-q23; Horn et al, 2000), Langer-Giedion (*LGCR*, 8q24.11-q24.13; Ludecke et al, 1999) and Werner (*WRN*, 8p12-p11.2; Yu et al, 1996) syndromes. Of the disease genes on chromosome 8, five have been localised close to the 8p22 region. These include the *CLN8* gene associated with progressive epilepsy with mental retardation (8pter-p22), the *MCPH1* gene (8pter-p22) responsible for the autosomal recessive form of microcephaly, and the *KWE* gene (8p23-p22) associated with the Oudtshoorn skin disease (NCI, USA, <http://nciarray.nci.nih.gov/>, June 2000). An extensively studied gene on 8p22 is the

LPL gene encoding for lipoprotein lipase (Mattei et al, 1993), an important enzyme involved in the metabolism of lipids. Deficiency in lipoprotein lipase, due to mutations in the *LPL* gene, results in hyperlipoproteinaemia type I and may be associated with predisposition to obesity and atherosclerosis (reviewed in Beisiegel, 1998). To date, *LPL* is the closest identified gene to the *NAT* locus, located within a maximum distance of 1.9Mb on the centromeric side of the *NAT* genes (Bookstein et al, 1994; Matas et al, 1997). Other important genes in the 8p22 region include those for macrophage scavenger receptor (*MSR*, Emi et al, 1993a), cathepsin B (*CTSB*, Hughes et al, 1998b) and farnesyl diphosphate farnesyltransferase (*FDFT1*, Schechter et al, 1994). These are all polymorphic and have been extensively used as markers for region 8p22. *NATs* could play a similar role as polymorphic markers, since they are genes with a well-defined chromosomal position and a number of well-characterised polymorphisms, adding to the total number of SNPs in the 8p22 region.

3.2 Strategies for genotyping the human *NAT* genes

3.2.1 *NAT1* genotyping

The presence of the most common *NAT1* alleles was investigated using an established genotyping strategy (Payton and Sim, 1998). Alleles *NAT1**4, *NAT1**3, *NAT1**10 and *NAT1**11 (table 1.4) can be identified by a combination of PCR and restriction fragment length polymorphism (PCR-RFLP) or by allele-specific PCR (AS-PCR). A PCR with primers N769 and N1113 (table 2.1) is initially performed to amplify a 344bp segment of the *NAT1* gene, spanning the end of the coding region and part of the 3' UTR (figure 3.1a). Subsequent digestion of the PCR product with *Mbo*II can distinguish between the *NAT1**4, *NAT1**11 and *NAT1**3 or *NAT1**10 alleles (figure 3.1b). However, an AS-PCR amplification is required to distinguish between the *NAT1**3 and *NAT1**10 alleles, as described in figure 3.1a.

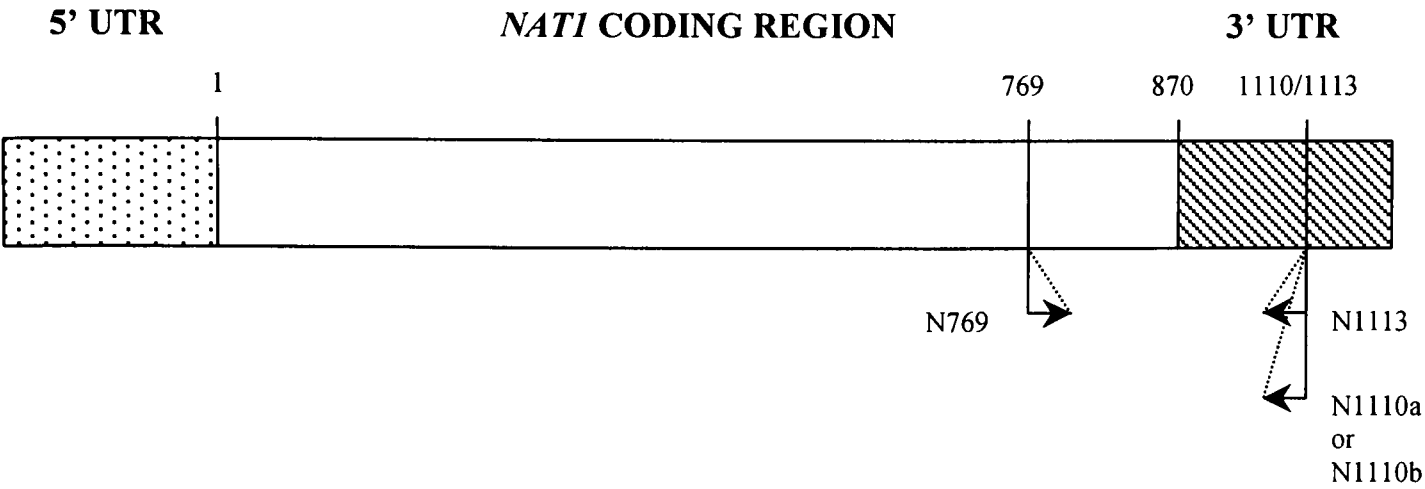


Figure 3.1a: Strategy for genotyping human *NAT1*. PCR amplifications used in *NAT1* genotyping. Primers N769 and N1113 (table 2.1) are used to amplify 344bp of the *NAT1* gene for subsequent RFLP analysis with *Mbo*II (figure 3.1b). Primers N1110a and N1110b (table 2.1) are used in combination with N769 for AS-PCR amplification from the *NAT1**3 and *NAT1**10 alleles respectively. Nucleotide numbering is with respect to the start of the coding region, considering the adenine of the ATG initiation codon as nucleotide position 1. The clear bar represents the *NAT1* coding region, while the dotted and shaded bars represent the 5' and 3' untranslated regions of the gene, respectively.

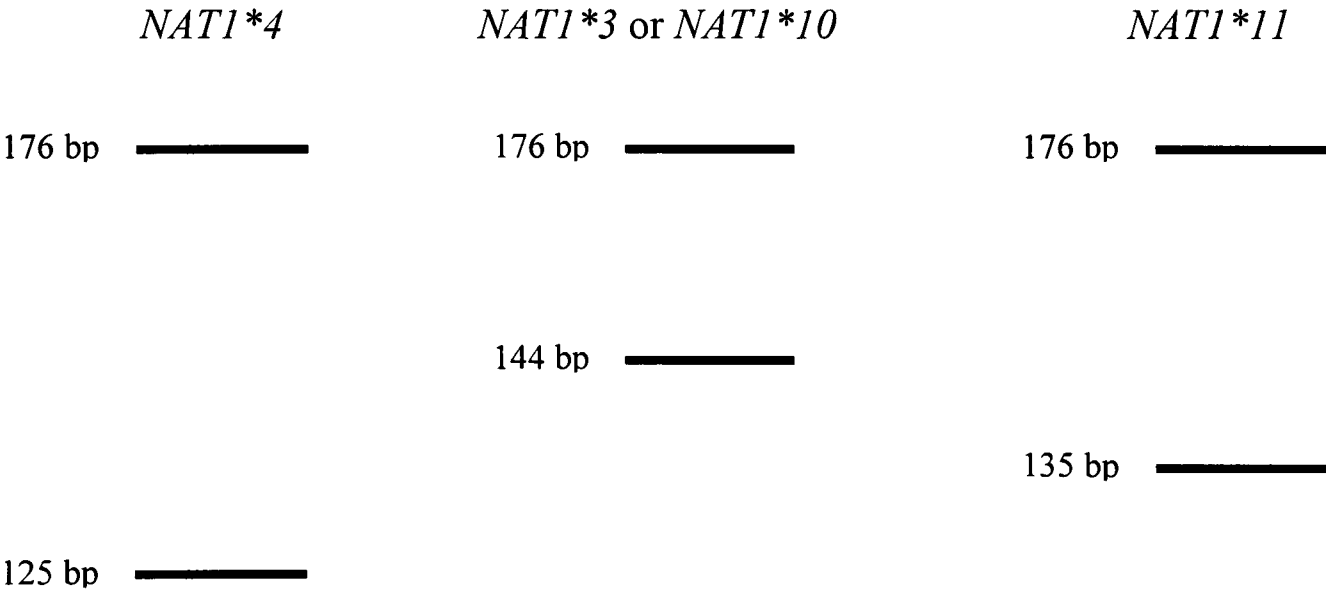


Figure 3.1b: Strategy for genotyping human *NAT1*. Diagram showing the *Mbo*II digestion patterns indicating specific *NAT1* genotypes. Following PCR amplification with primer pairs N769 and N1113 (figure 3.1.a) and digestion of the 344bp PCR product with *Mbo*II, the DNA fragments are separated on a 4% Metaphor gel. The *NAT1**4 allele produces two visible bands 176 and 125bp in size, while two bands 176 and 135bp in size are indicative of the *NAT1**11 allele. The *NAT1**3 and *NAT1**10 alleles both produce two fragments 176 and 144bp in size and can be distinguished from each other by AS-PCR (figure 3.1a).

3.2.2 NAT2 genotyping

Genotyping for the *NAT2* gene was performed according to previously published methods (Hickman and Sim, 1991; Hickman et al, 1992). These methods are capable of detecting the most common alleles in the Caucasian population, namely the *NAT2*4*, *NAT2*5A*, *NAT2*5B*, *NAT2*5C*, *NAT2*6A*, and *NAT2*7B* alleles (table 1.3). Genotyping involves a combination of PCR amplification and RFLP analysis, as demonstrated in figures 3.2a and 3.2b. Two PCR amplifications are performed, one with primer pair NAT-Hu7 and NAT-Hu8 and the other with primers NAT-Hu14 and NAT-Hu16 (table 2.1), producing bands of approximately 840bp and 1000bp respectively (figure 3.2a). The products of the former PCR are subsequently digested with *KpnI* and *TaqI*, while the products of the latter PCR are digested with *BamHI* and *DdeI*. The combination of the produced digestion patterns is then used to determine the *NAT2* genotype, as shown in figure 3.2b. It is possible to check specificity of amplification, with primers NAT-Hu14 and NAT-Hu16, by digestion of the PCR product with *HindIII*. A unique *HindIII* site is present at position 310 of the *NAT2*, but not the *NAT1*, open reading frame.

The strategy presented here cannot distinguish between the *NAT2*4/*5B* and *NAT2*5A/*5C* genotypes. Therefore, it is not entirely accurate in predicting the *NAT2* acetylation phenotype, but it is accurate in the majority of cases. If necessary, the fast *NAT2*4/*5B* and the slow *NAT2*5A/*5C* genotypes can be distinguished by a phenotyping assay of the ambiguous cases.

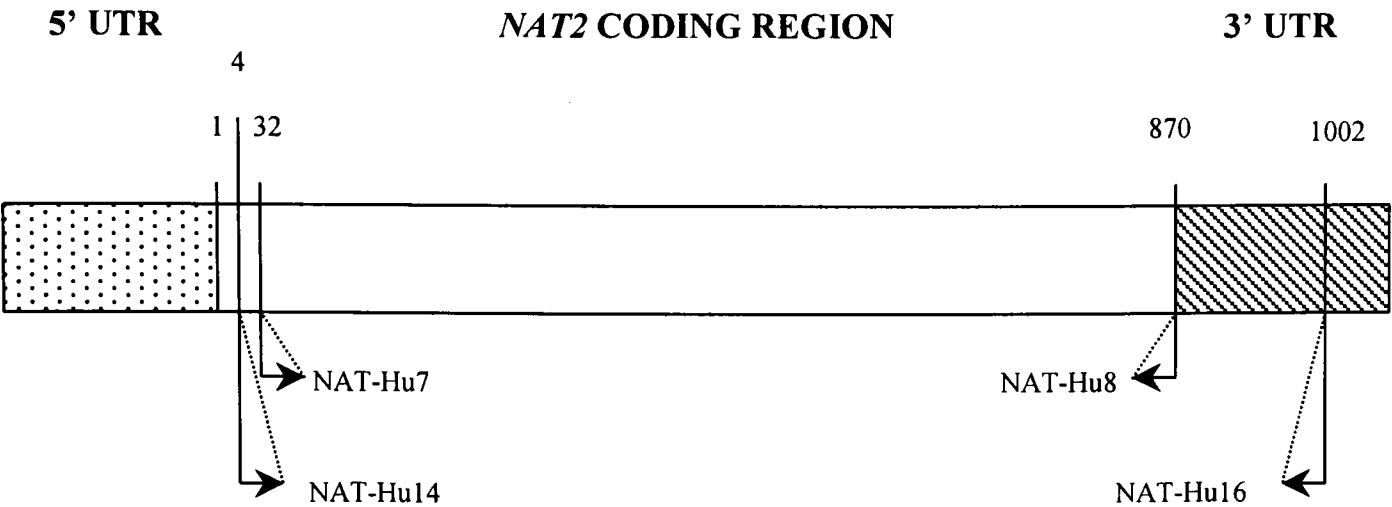


Figure 3.2a: Strategy for genotyping human *NAT2*. PCR amplifications used in *NAT2* genotyping. Primer pairs NAT-Hu7 and NAT-Hu8, as well as NAT-Hu14 and NAT-Hu16 (table 2.1) are used to amplify approximately 840bp and 1000bp of the *NAT2* gene, respectively, for subsequent RFLP analysis (figure 3.2b). Nucleotide numbering is with respect to the start of the coding region, considering the adenine of the ATG initiation codon as nucleotide position 1. The clear bar represents the *NAT2* coding region, while the dotted and shaded bars represent the 5' and 3' untranslated regions of the gene, respectively.

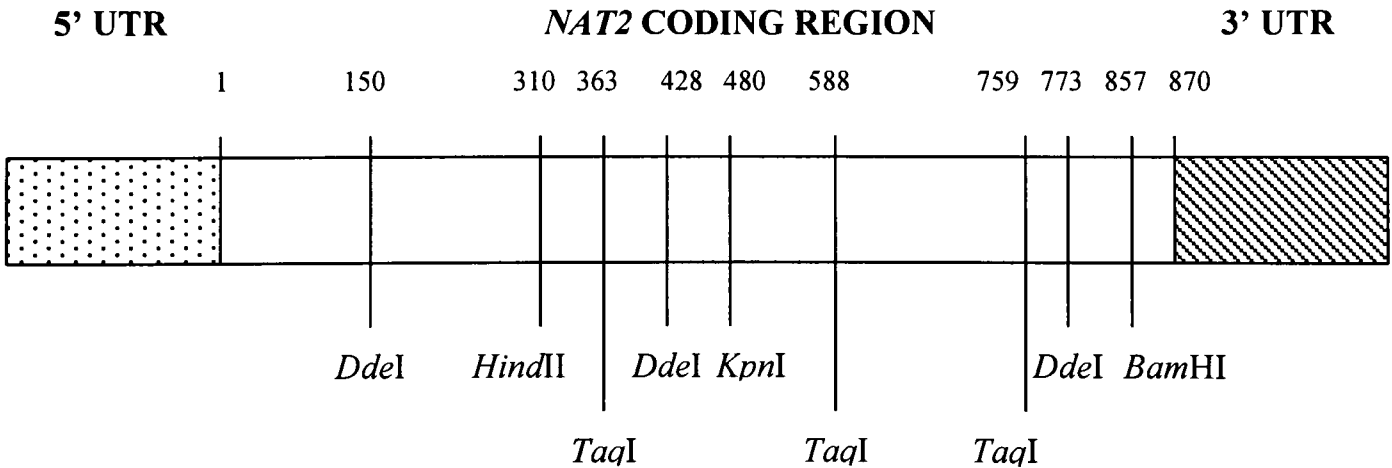


Figure 3.2b: Strategy for genotyping human *NAT2*. The *HindII* site at position 310 is present in all *NAT2* alleles, but not in any of the *NAT1* alleles, therefore it can be used to check specificity of amplification. The *TaqI* site at positions 363 and 759 is present in all *NAT2* alleles, while the one at position 588 is absent in the *NAT2*6A* allele. The *KpnI* site at position 480 is absent in the *NAT2*5A* and *NAT2*5B* alleles. The *BamHI* site at position 857 is absent in the *NAT2*7B* allele. The *DdeI* site at positions 150 and 428 is present in all *NAT2* alleles, while the one at position 773 is absent in alleles *NAT2*5B* and *NAT2*5C*.

3.3 Aims and study design

The aim of the work presented here was to study the genomic region of the human *NAT* genes on the short arm of chromosome 8, using previously isolated *NAT*-positive cosmid clones. This would allow detection of chromosomal aberrations involving the 8p22 region of the *NAT* genes in humans and comparison between healthy and cancerous material by FISH analysis. In addition, new molecular tools, in the form of *NAT*-positive PAC clones, were generated for future analysis and detailed physical mapping of the *NAT* region in humans.

3.3.1 The *NAT1* and *NAT2* cosmid clones

Clone 859-a-7 of the CEPH YAC library has previously been used for physical mapping of the *NAT* genes in humans (Matas et al, 1997). It has been established to be non-chimaeric and to contain all three *NAT* genes. To allow more detailed mapping of the 8p22 region, the YAC clone was subcloned into a cosmid vector and transfected into *E. coli*. Following screening of the produced colonies by hybridisation and PCR, one *NAT1*- and one *NAT2*-positive cosmid clones were isolated (Stacey et al, 1999). In the present study, these two cosmid clones were further characterised and tested for their potential in the study of bladder cancer material, where chromosomal abnormalities of the 8p22 region have been reported (Fujiwara et al, 1993; Knowles et al, 1993; Farrington et al, 1996; Thygesen et al, 1999). Both cosmids were genotyped for *NAT* and used for FISH analysis of healthy cultured lymphocytes and the human urothelial cell line RT112. FISH experiments were performed with the assistance of Dr M Stacey (Department of Pharmacology, University of Oxford, UK), to test the efficiency of the two cosmid clones in detecting chromosomal abnormalities of the 8p22 region.

3.3.2 Sample of healthy lymphocytes

Blood was collected from a healthy volunteer and lymphocytes were cultured as described in section 2.3.1.2. Slides of interphase nuclei were prepared from lymphocytes as described in section 2.3.3.

3.3.3 The RT112 cell line

The RT112 is an immortalised human bladder carcinoma cell line (Marshall et al, 1977). Its morphology, as observed under the light microscope, is indicative of its epithelial origin. The cells lack density-dependent inhibition of growth, but are unable to form colonies on agar or agarose, indicating that they require anchorage. When injected into congenically athymic nude mice, the RT112 cells formed typical papillary tumours of the urinary epithelium. Therefore, the tumours induced by the RT112 cells resembled the original carcinomas, even though many *in vitro* passages had taken place before inoculation into nude mice (Marshall et al, 1977; Hastings and Franks, 1981).

Expression of NAT in the RT112 cell line has previously been investigated. NAT protein purified from RT112 cells has been demonstrated to N-acetylate the NAT1-specific substrate pABA, but not the NAT2-specific substrate sulphamethazine (SMZ) (Coroneos and Sim, 1991). In addition, Western blot analysis, using polyclonal antisera specific for NAT1 or NAT2, has been able to detect only NAT1 protein in cytosols from RT112 cells (Stanley et al, 1996). Immunohistochemistry experiments, using the same antisera, also indicated the presence of NAT1 in these cells, while there was no NAT2-specific staining (Stacey et al, 1998). These results indicated either that the cell line has a slow *NAT2* genotype, so that NAT2 protein and enzymatic activity cannot be detected, or that NAT1 is the only NAT-isoenzyme present in the RT112 cells. In order to clarify this point, the *NAT2* genotype of the cell line was determined, according to the strategy described in section 3.2.2.

3.3.4 The RPCI 1 human genomic DNA PAC library

To generate a more refined map of the region surrounding the human *NAT* genes, it would be useful to develop a system easier to handle than the previously described YACs (Matas et al, 1997). Towards that aim, a human genomic DNA PAC library was screened and a number of PAC clones positive for at least one of the three *NAT* loci were identified.

The RPCI 1 library was constructed by Pieter de Jong and his group at the Roswell Park Cancer Institute, Buffalo, USA (<http://bacpac.med.buffalo.edu>) and was obtained from the HGMP Resource Centre, Cambridge, UK (<http://www.hgmp.mrc.ac.uk>). It was constructed using the pCYPAC2 vector and the source of genomic DNA was a healthy male blood donor. According to the creators of the library, it consists of approximately 115,200 clones and has an estimated three-fold genomic coverage.

3.3.4.1 Pooling information and screening scheme

The strategy for screening the RPCI 1 library is outlined in figures 3.3a and 3.3b. The library consists of twenty-one primary pools, each one representing fifteen 384-well microtitre plates. Following screening of the primary pools and identification of *NAT*-positive plates (figure 3.3a), pools of rows and columns were created for each plate. By screening these pools, the co-ordinates of the *NAT*-positive clones could be determined (figure 3.3b).

All screening steps for the three *NAT* genes were performed by PCR, using 3µl of *E. coli* suspension as template and initial denaturation for 5 min at 95°C, to ensure lysis of the cells. The library was screened using primers N769 and N1113 for *NAT1*, NAT-Hu7 and NAT-Hu8 for *NAT2*, and NAT-Hu24 and NAT-Hu27 for *NATP* (table 2.1).

Figure 3.3a: Strategy for screening the RPCI 1 human genomic DNA PAC library. Identification of positive clones is the result of three screening steps, as illustrated below.

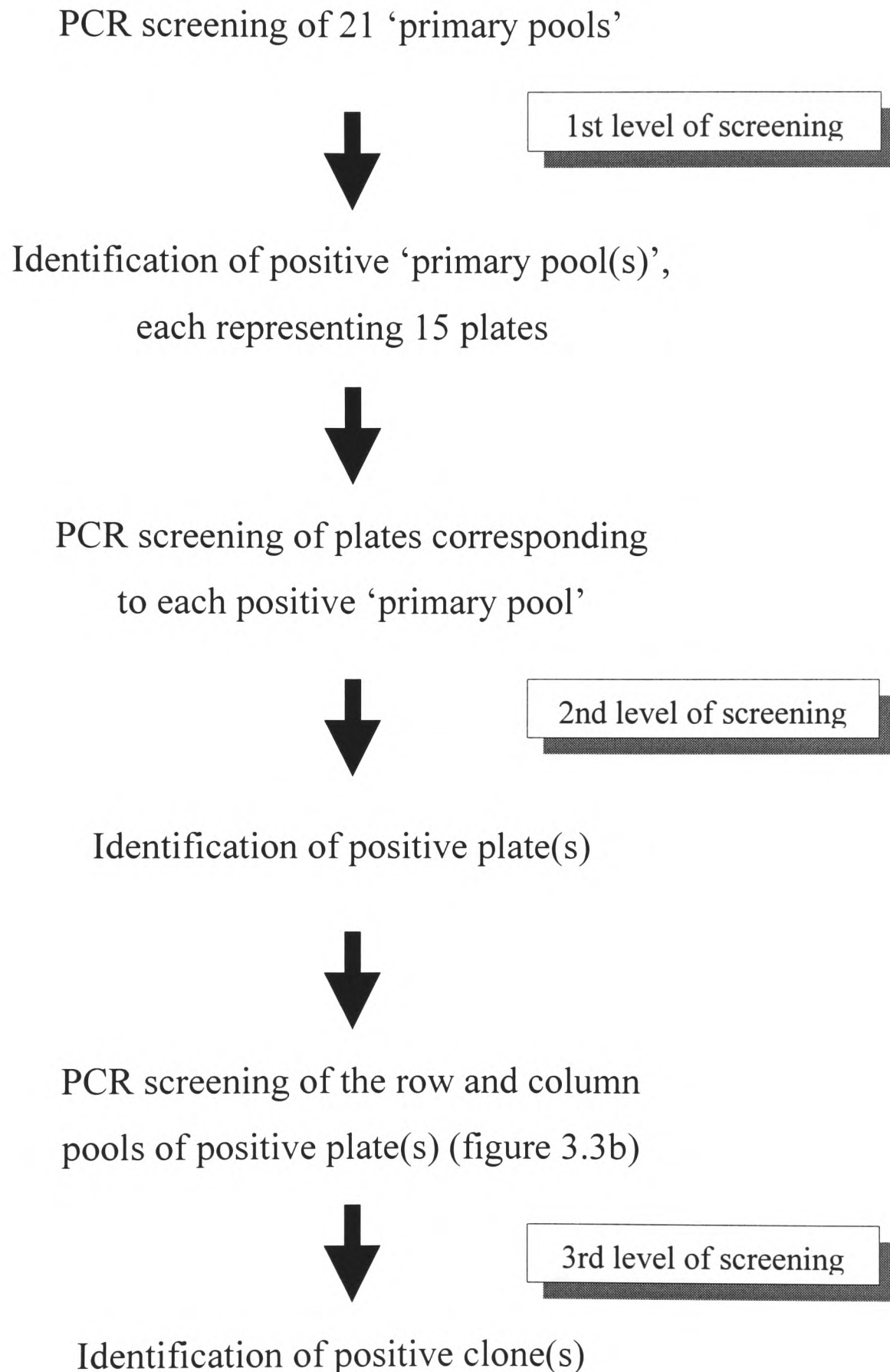
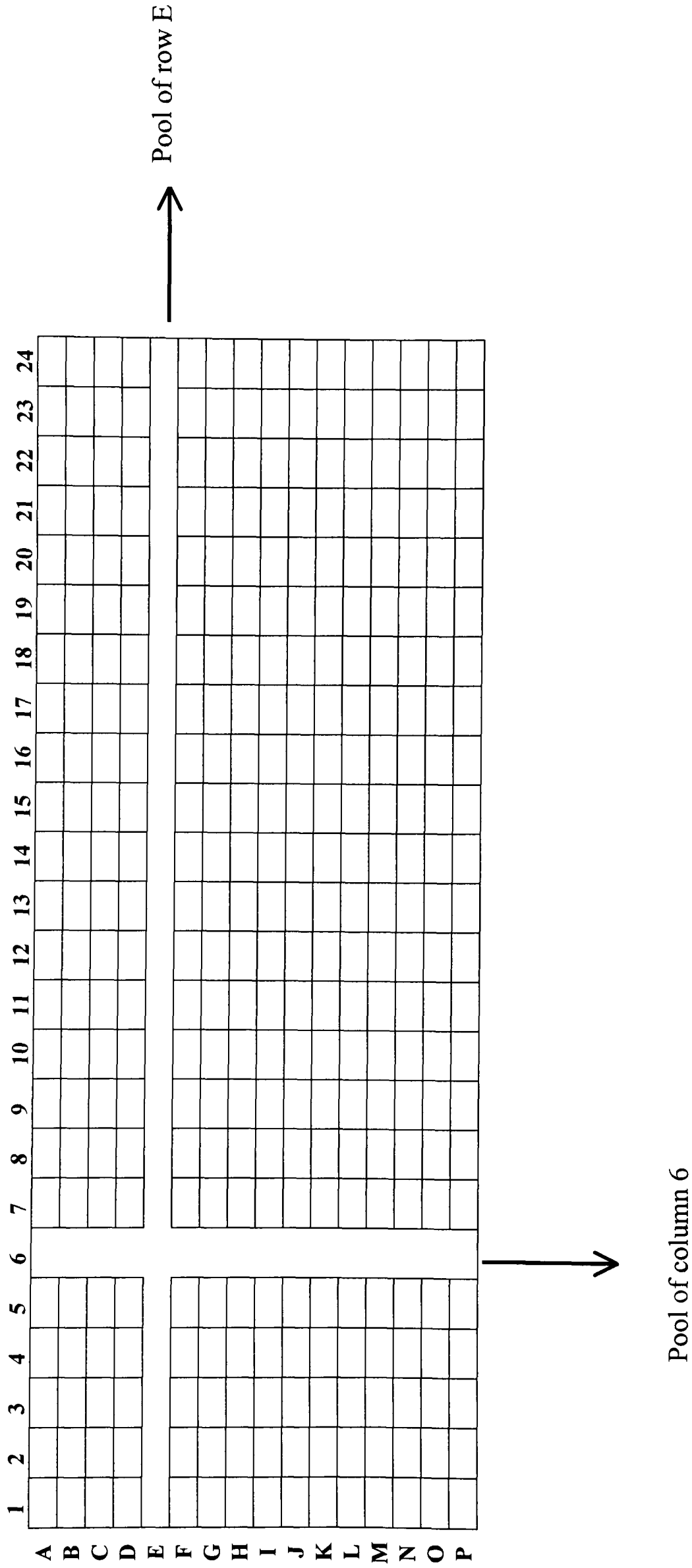


Figure 3.3b: Strategy for screening the RPCI 1 human genomic DNA PAC library. Pooling the rows and columns of each positive plate facilitates identification of the positive clones. Screening the pools of rows and columns by PCR provides the plate co-ordinates of each positive clone.



3.4 Results

3.4.1 Genotyping of the *NAT* cosmids

3.4.1.1 Genotyping of the *NAT1* cosmid

NAT1 genotyping was performed according to the strategy described in section 3.2.1, which can distinguish between the *NAT1**4, *NAT1**3, *NAT1**10 and *NAT1**11 alleles. PCR amplification with the *NAT1*-specific primers N769 and N1113 (table 2.1) produced a 344bp band, confirming the presence of *NAT1* gene in this cosmid clone (figure 3.4a). RFLP analysis of the amplified fragment with *Mbo*II produced four bands (176, 144, 24bp), indicating that the cosmid was carrying either the *NAT1**10 or the *NAT1**3 allele (figure 3.4a). An allele-specific PCR (AS PCR) for *NAT1**10 determined the presence of this allele in this *NAT1*-positive cosmid clone (figure 3.4b).

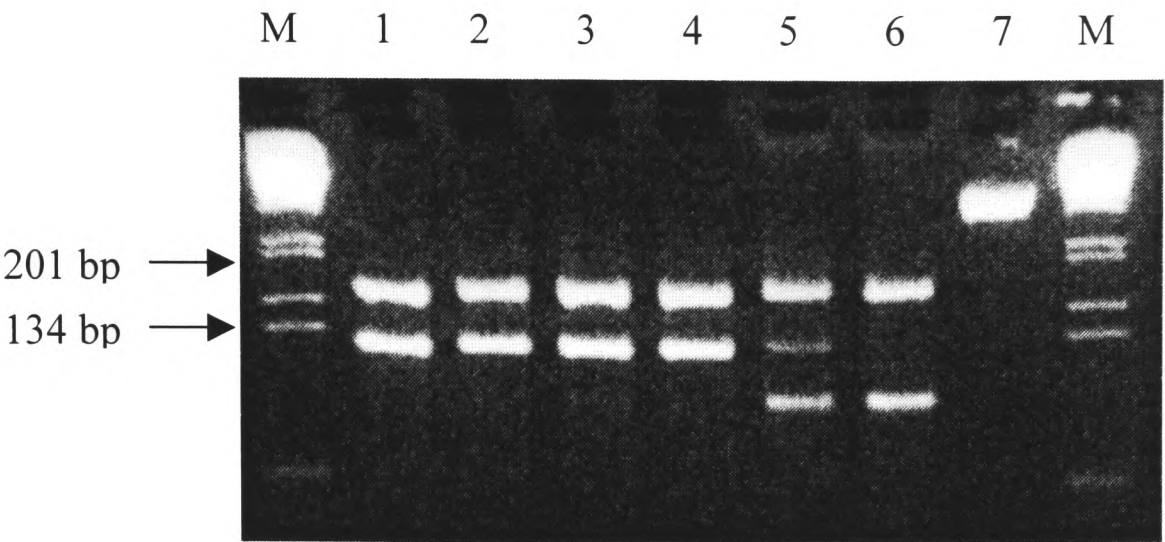


Figure 3.4a: Genotyping of the *NAT1* cosmid. Amplification with primers N769 and N1113 (table 2.1 and figure 3.1a) was followed by *Mbo*II digestion of the 344bp product. Lanes 1 to 4 are *Mbo*II digests of four distinct amplification products from the *NAT1* cosmid. Lanes 5 and 6 are *Mbo*II digests of PCR products from two genomic DNA samples of known genotype (*NAT1**4/*10 and *NAT1**4/*4 respectively), for comparison. Lane 7 was loaded with undigested PCR product (344bp in size, as expected) and lanes M were loaded with 1µg of 1kb ladder. *Mbo*II digestion produced three bands (176, 144, 24bp), of which only the 176 and 144bp ones were visible. This digestion pattern indicated the presence of either the *NAT1**10 or the *NAT1**3 allele (figure 3.1b). Samples were run on a 4% Metaphor gel. Arrows indicate the size of selected marker bands (lanes M).

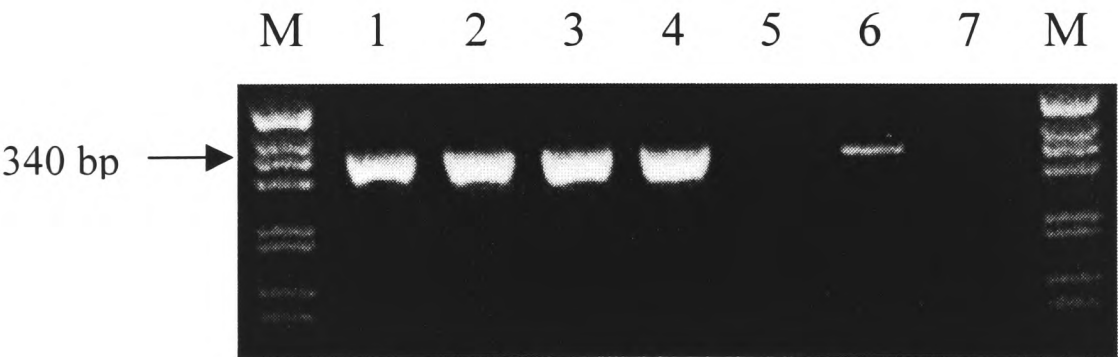


Figure 3.4b: Genotyping of the *NAT1* cosmid. *NAT1**10 allele-specific PCR was performed with primers N769 and N1110b (table 2.1 and figure 3.1a). Lanes 1 to 4 are four distinct amplification products from the *NAT1* cosmid clone. Lanes 5 and 6 are amplification products from two genomic DNA samples of known genotype (*NAT1**4/*4 and *NAT1**4/*10 respectively), to serve as controls. Lane 7 was loaded with PCR negative control (no DNA template in the reaction mix) and lanes M were loaded with 1µg of 1kb ladder. The 340bp band in lanes 1 to 4 indicated the presence of a *NAT1**10 allele in the *NAT1* cosmid clone. Lack of amplification product in lane 5 (indicating absence of a *NAT1**10 allele) and presence of amplification product in lane 6 (indicating presence of at least one *NAT1**10 allele) confirmed specificity of the reaction. Samples were run on a 2% Metaphor gel. The arrow indicates the size of a selected marker band.

3.4.1.2 Genotyping of the NAT2 cosmid

The genotype of the *NAT2* cosmid was determined using a combination of PCR and RFLP analysis, as described in section 3.2.2. Two independent PCR amplifications with primer pairs NAT-Hu7/ NAT-Hu8 and NAT-Hu14/ NAT-Hu16 (table 2.1 and figure 3.2a) were performed, and the product of the first PCR was digested with *KpnI* and *TaqI*, while the product of the second PCR was digested with *BamHI*, and *DdeI* restriction endonucleases. Specificity of amplification was checked by digestion with *HindII*. Digestion with *TaqI* of the 840bp PCR product generated with primers NAT-Hu7 and NAT-Hu8 produced bands 400, 330 and 110bp in size, indicating the presence of a *NAT2**6A allele (figure 3.5). This is one of the alleles responsible for the slow acetylator phenotype. An example of the complete *NAT2* genotyping procedure for all alleles described in section 3.2.2 is presented in figure 3.8a-g for the RT112 cell line.

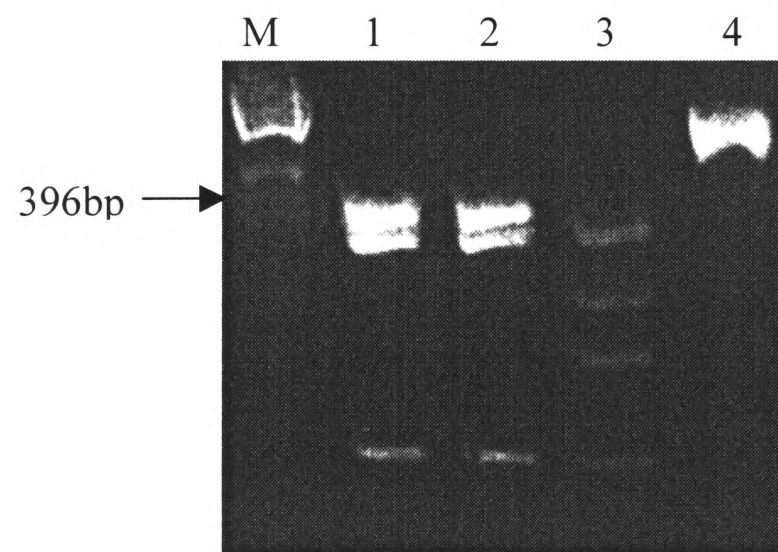


Figure 3.5: Genotyping of the *NAT2* cosmid.

Amplification with primers NAT-Hu7 and NAT-Hu8 was followed by *TaqI* digestion of the 840bp product. Lanes 1 and 2 are *TaqI* digests of two distinct amplification products from the *NAT2* cosmid. Lane 3 is the *TaqI* digest of PCR product from a genomic DNA sample of known genotype (*NAT2**4/*5B), for comparison. Lane 4 was loaded with the undigested PCR product (840bp) and lane M was loaded with 1µg of 1kb ladder. *TaqI* digestion of the cosmid PCR products (lanes 1 and 2) provided bands 400, 330 and 110bp in size, characteristic of a *NAT2**6A allele. Samples were separated on a 3% Metaphor gel. The arrow indicates the size of a selected marker band.

3.4.2 Use of the *NAT* cosmids for FISH analysis

3.4.2.1 FISH analysis of healthy lymphocytes

The presence of the *NAT* genes in T-lymphocytes, obtained from a healthy individual and cultured *in vitro* (section 2.3.1.2), was investigated by FISH analysis. The *NAT*-specific cosmid probes were labelled by nick translation (section 2.3.5) and hybridised to slides of interphase nuclei prepared from T-lymphocytes (section 2.3.3), together with a chromosome 8 centromere-specific probe (section 2.3.6.1).

Approximately one hundred interphase nuclei were examined. All lymphocytes showed two copies of *NAT* (figure 3.6b), as well as of chromosome 8 centromere (figure 3.6a). These experiments were used to test the performance of the *NAT*-specific cosmid probes in FISH analysis and assess the reliability of the technique in detecting the diploid number of copies of *NAT* and the chromosome 8 centromere in cells derived from healthy individuals.

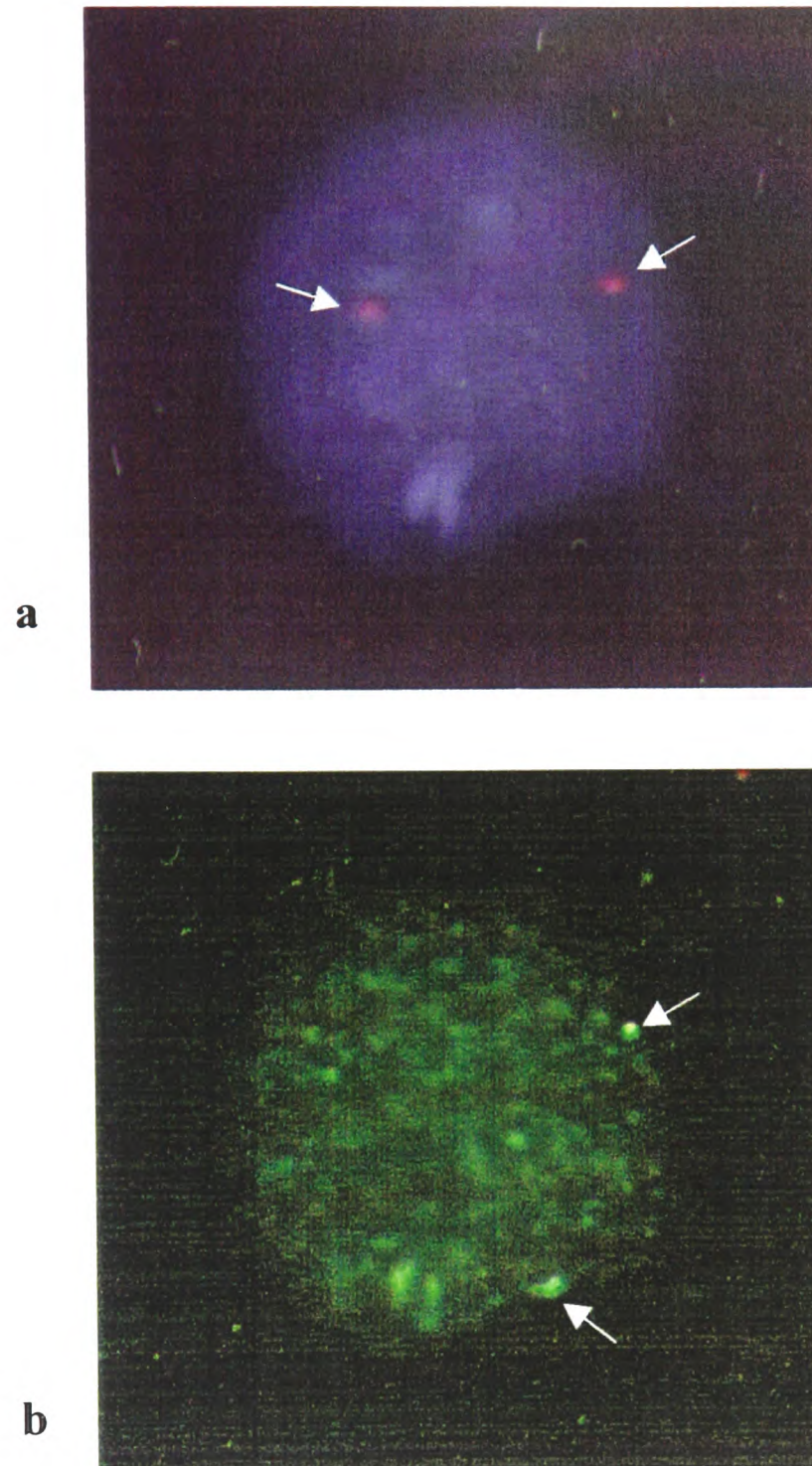


Figure 3.6: FISH analysis of healthy T-lymphocytes using the characterised *NAT1* cosmid clone as probe.

Approximately 100µg of biotin-labelled *NAT1* cosmid probe, 10ng of digoxigenin-labelled chromosome 8 centromere-specific probe and 2µg human COT-1 DNA were hybridised to slides of interphase nuclei at 37°C overnight. Following washes, the chromosome 8 centromere probe was detected through rhodamine-conjugated anti-digoxigenin Fab fragments. *NAT1* signal detection was achieved through FITC-conjugated avidin and amplified with biotinylated goat anti-avidin antibody and a second layer of FITC-conjugated avidin. Nuclei were counterstained with DAPI (section 2.3.6.1). Rhodamine signals were visualised using the red and blue filters (a), while FITC signals were visualised with the green filter (b) under the fluorescence microscope.

In this example of an interphase nucleus, a and b show the same cell in the same orientation. Two red rhodamine signals (arrows) indicate two copies of chromosome 8 centromere (a) and two green FITC signals (arrows) show the presence of two *NAT1* loci (b).

3.4.2.2 FISH analysis of the RT112 cell line

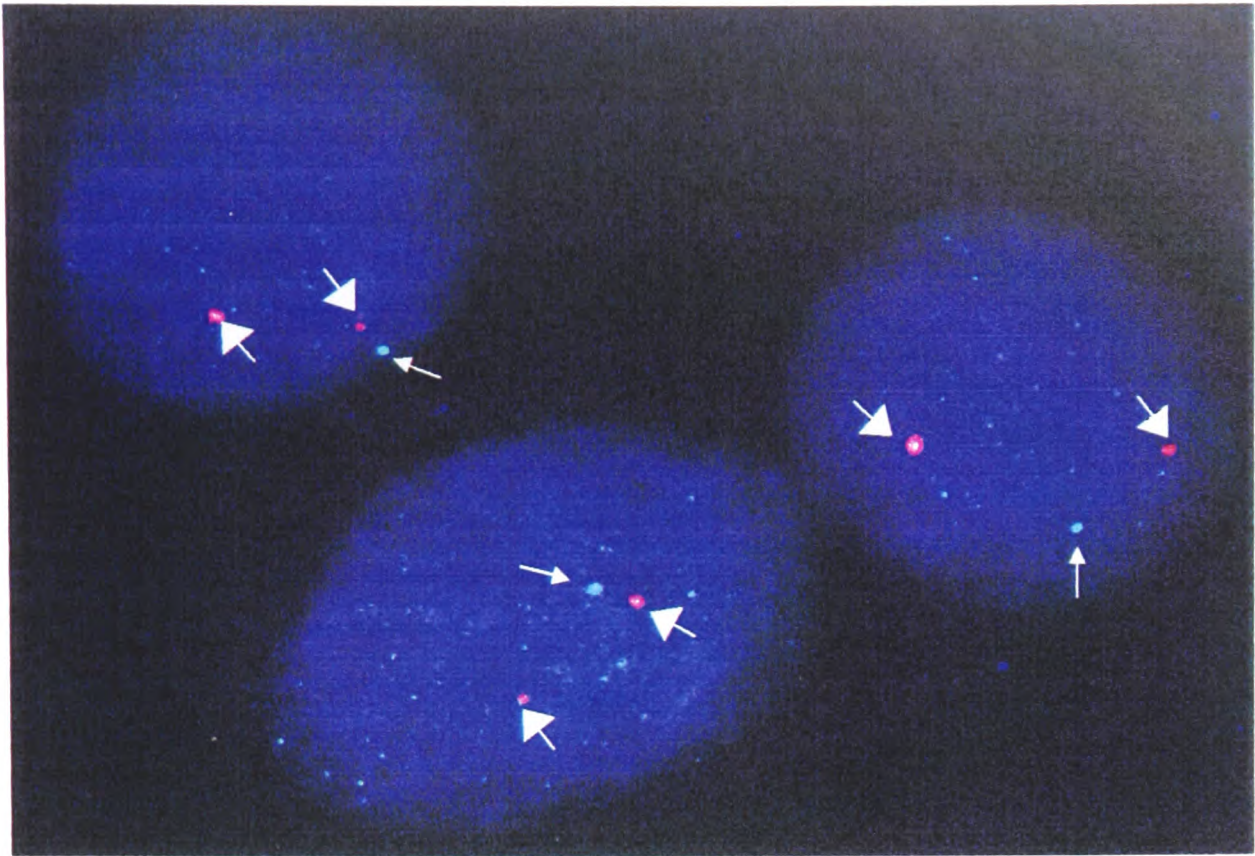
The presence of *NAT* genes in the RT112 cell line was investigated by FISH analysis, following a procedure similar to that described above for cultured T-lymphocytes (section 3.4.2.1). The *NAT*-specific cosmid probes were labelled by nick translation (section 2.3.5) and hybridised to either interphase (section 2.3.3) or metaphase (section 2.3.2) nuclei prepared from RT112 cells, as described in section 2.3.6.1 of Materials and Methods. A centromeric probe was used to generate chromosome 8-specific signals, as described for the T-lymphocytes. Typically, one hundred interphase or metaphase nuclei were examined and they were all identified to lack one *NAT*-specific signal. However, two chromosome 8 centromere-specific signals were observed in all nuclei examined, indicating presence of a normal (diploid) number of chromosome 8 copies in all RT112 cells (figure 3.7a and b). Using metaphase spreads it was demonstrated that the *NAT*-specific probes hybridised to the short arm of one of the chromosome 8 copies (figure 3.7b). Absence of the *NAT* locus from the other copy of chromosome 8 was due to loss of the entire 8p arm and formation of an iso-8q chromosome.

Figure 3.7: FISH analysis of RT112 cells using the characterised *NAT1* cosmid clone as probe.

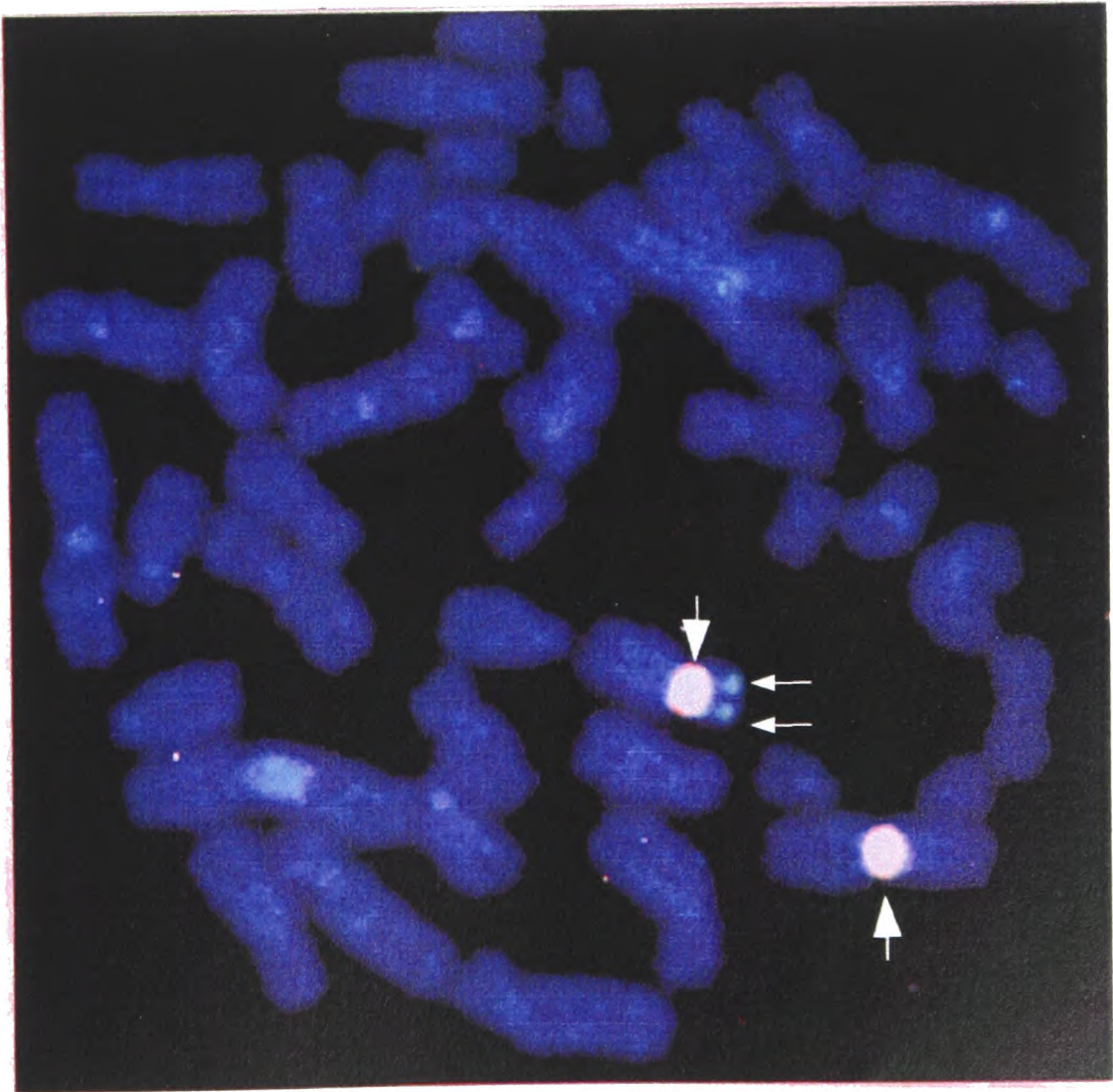
A chromosome 8 centromere-specific probe was used to detect copies of chromosome 8. Hybridisation and signal detection were carried out as described in figure 3.6 and section 2.3.6.1. FITC and rhodamine signals were visualised on DAPI-stained nuclei using the blue, green and red filters under the fluorescence microscope.

Two red centromeric signals were detected (large arrows) both on interphase (a) and metaphase (b) nuclei. However, only one green *NAT1*-specific signal (small arrows) was observed, indicating loss of one *NAT1* locus in RT112 cells. Observation of metaphase chromosomes stained with DAPI indicated that the chromosome 8 lacking the *NAT1* locus was an iso-8q chromosome, formed by loss of the entire short arm and duplication of the remaining long arm.

a



b



3.4.2.3 NAT2 genotyping of the RT112 cell line

Genotyping of the RT112 cell line for *NAT2* was performed as described in section 3.2.2. Genomic DNA was extracted from RT112 cells (section 2.2.3) and used as template for specific amplification of *NAT2* (figure 3.2a). Amplification with the NAT-Hu7/ NAT-Hu8 and NAT-Hu14/ NAT-Hu16 primer pairs produced the expected 840bp and 1000bp bands respectively (figure 3.8a and b). Specificity of amplification from the *NAT2* gene was confirmed by *Hind*III digestion of the 1000bp fragment, which produced bands 700bp and 300bp in size, as expected (figure 3.8c). Digestion of the 840bp PCR product with *Kpn*I provided fragments of 450bp and 390bp (figure 3.8d) and with *Taq*I fragments of 330bp, 230bp, 170bp and 110bp (figure 3.8e). Digestion of the 1000bp PCR product with *Bam*HI provided bands of 850bp and 150bp (figure 3.8f) and with *Dde*I bands of 345bp, 280bp, 220bp and 155bp (figure 3.8g). According to the restriction map of the *NAT2* gene shown in figure 3.2b, these digestion patterns indicated presence of the wild type *NAT2**4 allele, which is responsible for the fast acetylation phenotype.

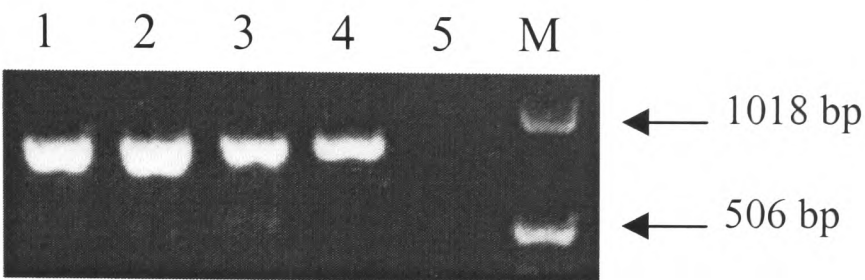


Figure 3.8a: *NAT2* genotyping of RT112 cells. Amplification products from RT112 genomic DNA with primers NAT-Hu7 and NAT-Hu8 (lanes 1 to 3). Lane 4 is PCR product from a genomic DNA sample of known genotype (*NAT2**4/*5*B*), used as positive control. Lane 5 is the PCR negative control (no DNA template in the reaction mix). Lane M was loaded with 1µg of 1kb ladder. A band 840bp in size was produced from all RT112 samples (lanes 1 to 3 and from the positive control (lane 4). Samples were separated on a 1% agarose gel. Arrows indicate marker bands.

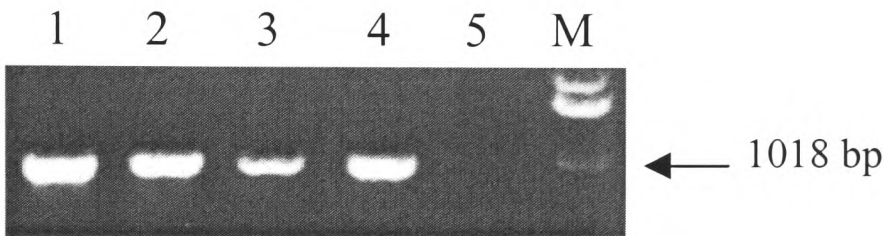


Figure 3.8b: *NAT2* genotyping of RT112 cells. Amplification products from RT112 genomic DNA with primers NAT-Hu14 and NAT-Hu16 (lanes 1 to 3). Lane 4 is PCR product from a genomic DNA sample of known genotype (*NAT2**4/*5*B*), used as positive control. Lane 5 is the PCR negative control (no DNA template in the reaction mix). Lane M was loaded with 1µg of 1kb ladder. A band 1000bp in size was produced from the RT112 samples (lanes 1 to 3) and from the positive control (lane 4). Samples were separated on a 1% agarose gel. The arrow indicates a marker band.

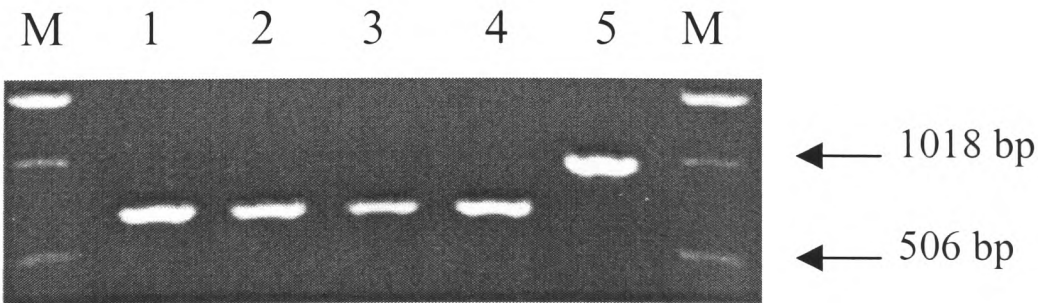


Figure 3.8c: *NAT2* genotyping of RT112 cells. Amplification products with primers NAT-Hu14 and NAT-Hu16 were digested with *Hind*II and the fragments were separated on a 1.5% agarose gel. Lanes 1 to 3 are digests of the RT112 PCR products, while lane 4 is digest of the control sample (*NAT2**4/*5*B*), for comparison. Lane 5 is undigested control and lanes M were loaded with 1µg of 1kb ladder. Digestion of the 1000bp PCR product with *Hind*II provided two fragments of 700bp and 300bp (not visible), which confirmed specific amplification of the *NAT2* gene. Arrows indicate marker bands.

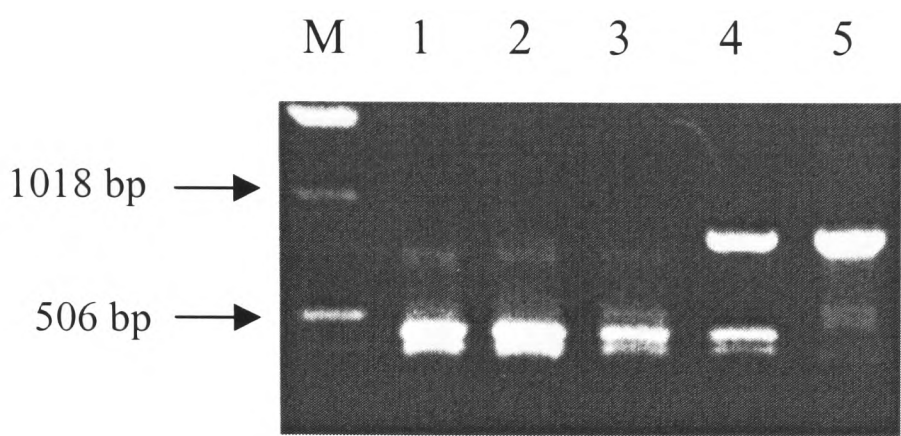


Figure 3.8d: *NAT2* genotyping of RT112 cells. PCR products (10µl) with primers NAT-Hu7 and NAT-Hu8 were digested with *KpnI* and the fragments were separated on a 1.5% agarose gel. Lanes 1 to 3 were loaded with digests of the RT112 PCR products, while lane 4 was loaded with digest of the control sample (*NAT2**4/*5*B*), for comparison. Lane 5 was loaded with undigested control and lane M was loaded with 1µg of 1kb ladder. RT112 samples provided two bands 450 and 390bp in size, while the positive control provided three bands 840, 450 and 390bp in size. The size of the undigested control was 840bp. Arrows indicate marker bands.

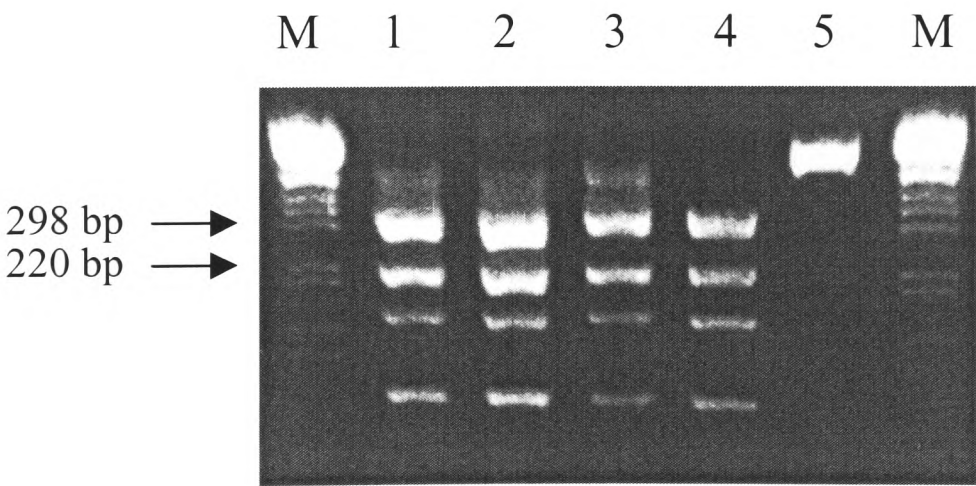


Figure 3.8e: *NAT2* genotyping of RT112 cells. PCR products (10µl) with primers NAT-Hu7 and NAT-Hu8 were digested with *TaqI* and the fragments were separated on a 3% metaphor gel. Lanes 1 to 3 were loaded with digests of the RT112 PCR products, while lane 4 was loaded with digest of the control sample (*NAT2**4/*5*B*), for comparison. Lane 5 was loaded with undigested control and lanes M were loaded with 1µg of 1kb ladder. RT112 samples, as well as the control sample, provided bands 330, 230, 170 and 110bp in size. The size of the undigested control was 840bp. Arrows indicate marker bands.

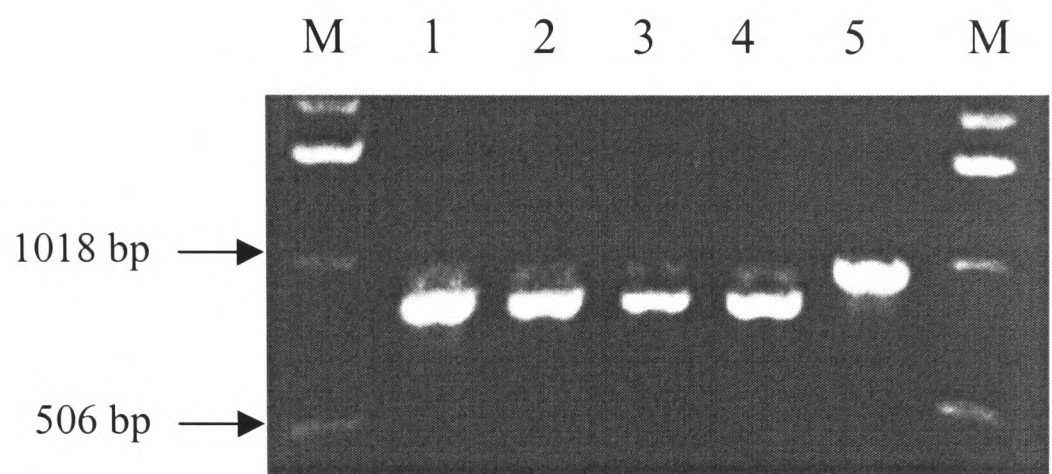


Figure 3.8f: *NAT2* genotyping of RT112 cells. PCR products (10 μ l) with primers NAT-Hu14 and NAT-Hu16 were digested with *Bam*HI and the fragments were separated on a 1.5% agarose gel. Lanes 1 to 3 were loaded with digests of RT112 PCR products, while lane 4 was loaded with digest of the control sample (*NAT2**4/*5*B*), for comparison. Lane 5 was loaded with undigested control and lanes M were loaded with 1 μ g of 1kb ladder. RT112 samples, as well as the control sample, provided bands 850 and 150bp in size, only the largest of which was visible on the gel. The size of the undigested control was 1000bp. Arrows indicate marker bands.

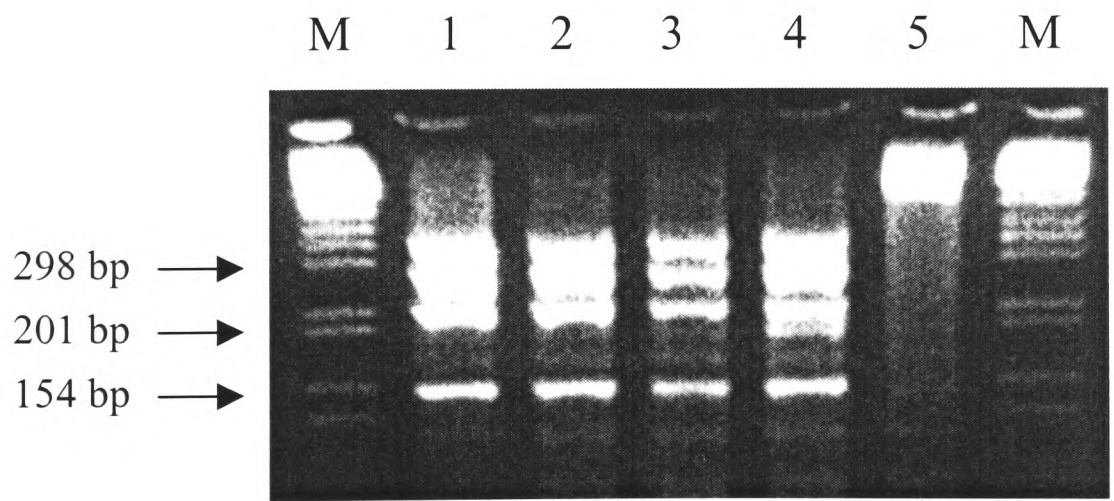


Figure 3.8g: *NAT2* genotyping of RT112 cells. PCR products (10 μ l) with primers NAT-Hu14 and NAT-Hu16 were digested with *Dde*I and the fragments were separated on a 3% metaphor gel. Lanes 1 to 3 were loaded with digests of RT112 PCR products, while lane 4 was loaded with digest of the control sample (*NAT2**4/*5*B*), for comparison. Lane 5 was loaded with undigested control and lanes M were loaded with 1 μ g of 1kb ladder. RT112 samples provided bands 345, 280, 220 and 155bp in size, while the control sample provided bands 345, 280, 220, 200 and 155bp in size. The size of the undigested control was 1000bp. Arrows indicate marker bands.

3.4.3 Screening the RPCI 1 library for *NAT1*, *NAT2* and *NATP*

The library was screened by PCR for each *NAT* locus individually, according to the scheme described in figure 3.3. Amplification was performed using primer pairs N769/ N1113 for *NAT1*, NAT-Hu7/ NAT-Hu8 for *NAT2* and NAT-Hu24/ NAT-Hu27 for *NATP* (table 2.1). Positive clones were identified following screening of three levels of pooled material (figure 3.9a-d) and an additional PCR amplification from the isolated *NAT*-positive clones was used to confirm the results (figure 3.10a-d). The identified clones are summarised in table 3.1.

Screening of the twenty-one primary pools (A-U) for *NAT1* provided four positive pools (C, L, M and U). Secondary level screening of the fifteen plates corresponding to each one of these primary pools indicated that plates 33 and 34 (pool C), 178 (pool L), 185 (pool M) and 308 (pool U) were positive for *NAT1*. Plate 33 was later identified to be a false positive. Following screening of the rows and columns pools of these five plates and subsequent screening of the individual candidate clones, a total of seven clones (34-G18, 34-H20, 178-P8, 185-H1, 308-I7, 308-J8 and 308-K9) were identified as carrying the *NAT1* gene (figure 3.10a; table 3.1).

Screening of the primary pools for *NAT2* (figure 3.9a) identified four of them as positive (pools K, O, S and T). Subsequent screening of the fifteen plates constituting each primary pool identified plates 155 (pool K), 226 (pool O) and 296 (pool T) as positive for *NAT2* (figure 3.9b). It also indicated that the product originally obtained from primary pool S was a false positive, so this pool was excluded from further analysis. Following screening of the rows (figure 3.9c) and columns (figure 3.9d) pools of the three positive plates and subsequent screening of the individual candidate clones, five clones (155-H6, 226-H11, 296-B17, 296-C18 and 296-F16) were identified as positive for *NAT2* (figure 3.10b; table 3.1).

Screening of the twenty-one primary pools of the library for *NATP* indicated five positive pools (H, I, K, O and T). Pools K, O and T were also positive for *NAT2*, therefore the clones that had been identified as containing *NAT2* were also tested for the presence of *NATP*. It was determined that all the clones that were positive for *NAT2* (155-H6, 226-H11, 296-B17, 296-C18 and 296-F16) also carried *NATP* (figure

3.10c; table 3.1). Secondary screening of primary pools H and I indicated that plates 115, 116 and 117 (pool H) and 129 (pool I) were also positive for *NATP*. This was later confirmed for plates 116 and 129, while plates 115 and 117 were shown to be false positives. Following the screening of rows and columns pools of the positive plates (116 and 129) and subsequent screening of the candidate clones, two clones (116-K16 and 129-F16) were identified to contain *NATP* only (figure 3.10d; table 3.1).

A number of clones were randomly selected from the RPCI 1 library to serve as negative controls in all amplifications (figure 3.10d, for example). Following each step of the screening process, specificity of amplification was checked by digestion of the PCR product with *Mbo*II for *NAT1*, *Hind*II for *NAT2* and *Alu*I for *NATP*. All positive clones (table 3.1) provided the expected restriction pattern, described in table 3.2. This confirmed the *NAT*-content of each one of the identified PAC clones.

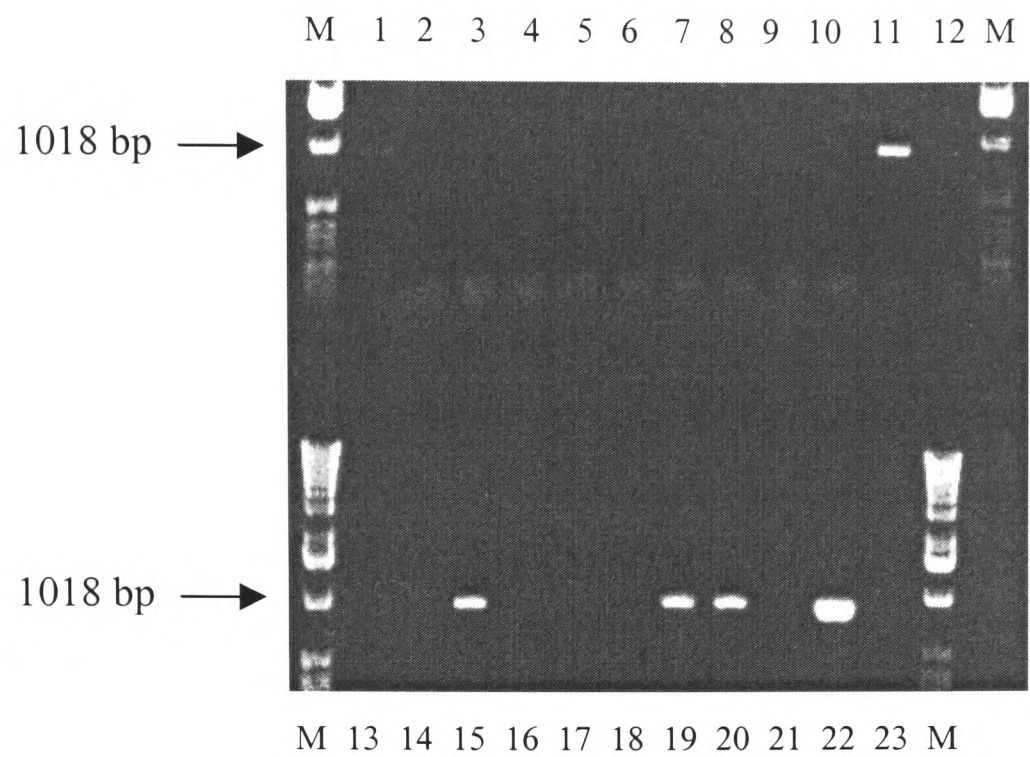


Figure 3.9a: Screening the RPCI 1 human genomic DNA PAC library for *NAT* genes.

The 21 ‘primary pools’ of the library were screened for *NAT2* by PCR, using primers NAT-Hu7 and NAT-Hu8. Amplification products were loaded on a 1% agarose gel. Lanes 1 to 21 were loaded with PCR products from the primary pools A to U. Lanes 22 and 23 were loaded with the PCR positive and negative controls respectively. Lanes M were loaded with 1µg of 1kb ladder. A band 840bp in size was produced from primary pools K, O, S and T (lanes 11, 15, 19 and 20), as well as from the genomic DNA sample used as positive control (lane 22). The arrow indicates a marker band.

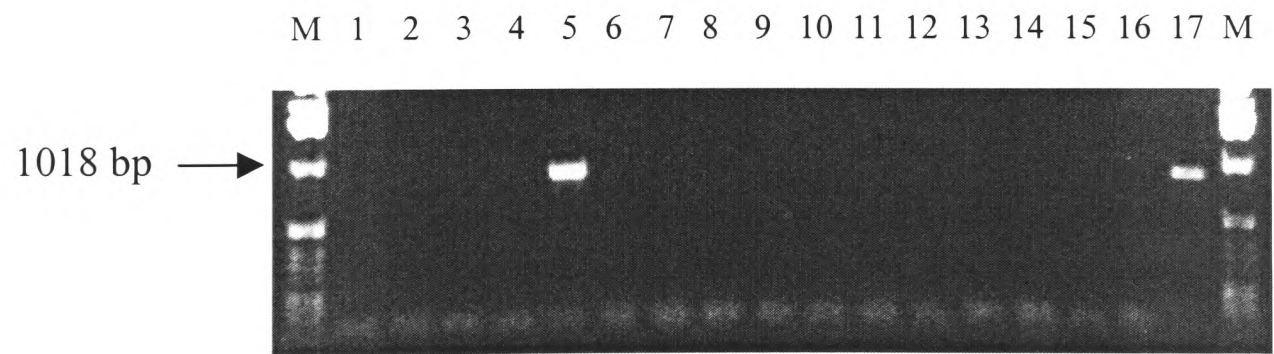


Figure 3.9b: Screening the RPCI 1 human genomic DNA PAC library for *NAT* genes.

Pools of the 15 plates that correspond to primary pool K were screened for *NAT2* by PCR, using primers NAT-Hu7 and NAT-Hu8. Amplification products were loaded on a 1% agarose gel. Lanes 1 to 15 were loaded with PCR products from the pools of plates 151 to 165. Lanes 16 and 17 were loaded with the PCR negative and positive controls respectively. Lanes M were loaded with 1µg of 1kb ladder. A band 840bp in size was produced from the pool of plate 155 (lane 5), as well as from the genomic DNA sample used as positive control (lane 17). The arrow indicates a marker band.

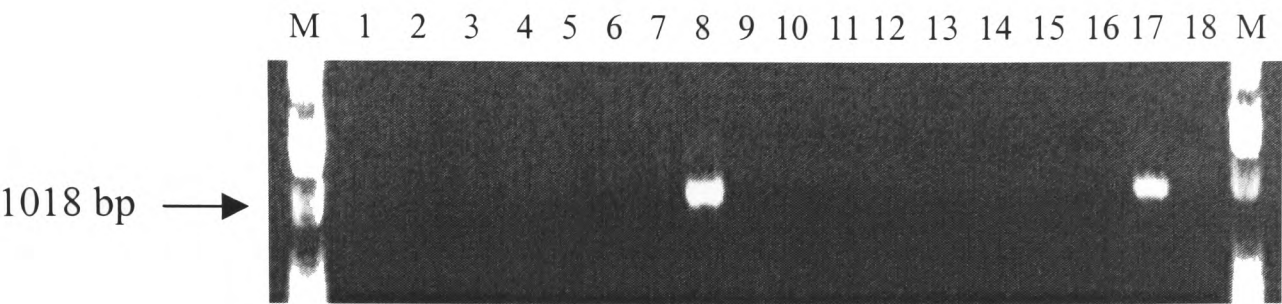


Figure 3.9c: Screening the RPCI 1 human genomic DNA PAC library for the *NAT* genes. ‘Row pools’ of plate 155 were screened for *NAT2* by PCR, using primers NAT-Hu7 and NAT-Hu8. Amplification products were loaded on a 1% agarose gel. Lanes 1 to 16 were loaded with PCR products from the pools of rows A to P. Lanes 17 and 18 were loaded with the PCR positive and negative controls respectively. A band 840bp in size was produced from the pool of row H (lane 8), as well as from the genomic DNA sample used as positive control (lane 17). The arrow indicates a marker band.

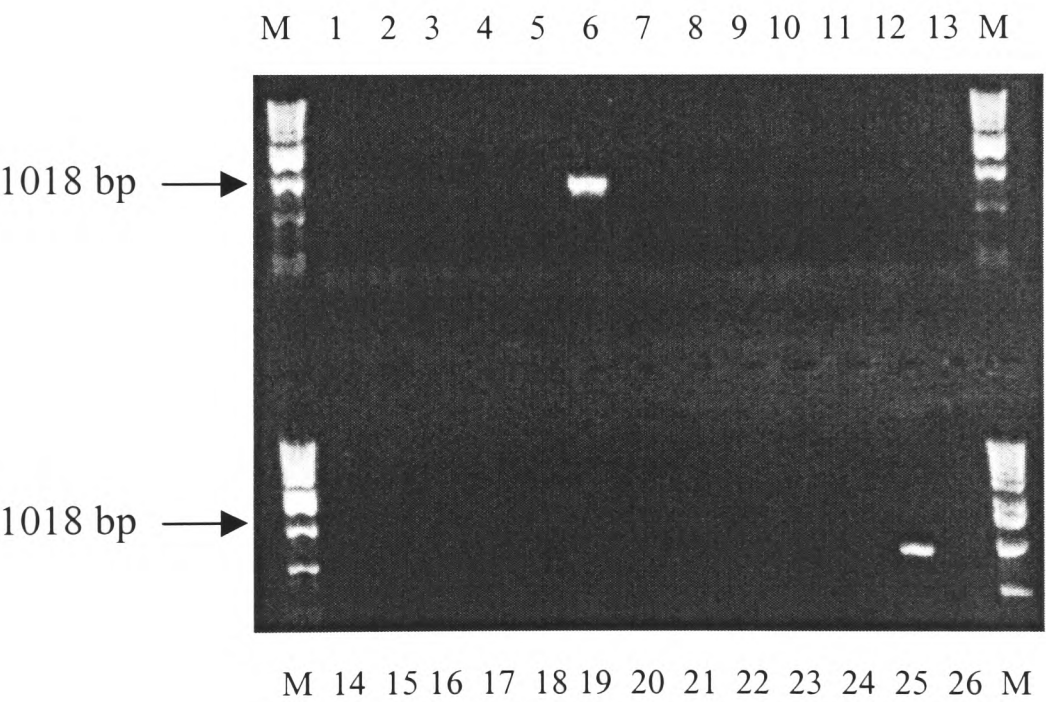


Figure 3.9d: Screening the RPCI 1 human genomic DNA PAC library for the *NAT* genes. ‘Column pools’ of plate 155 were screened for *NAT2* by PCR, using primers NAT-Hu7 and NAT-Hu8. Amplification products were loaded on a 1% agarose gel. Lanes 1 to 24 were loaded with PCR products from the pools of columns 1 to 24. Lanes 25 and 26 were loaded with the PCR positive and negative controls respectively. A band 840bp in size was produced from the pool of column 6 (lane 6), as well as from the genomic DNA sample used as positive control (lane 25). The arrow indicates a marker band.

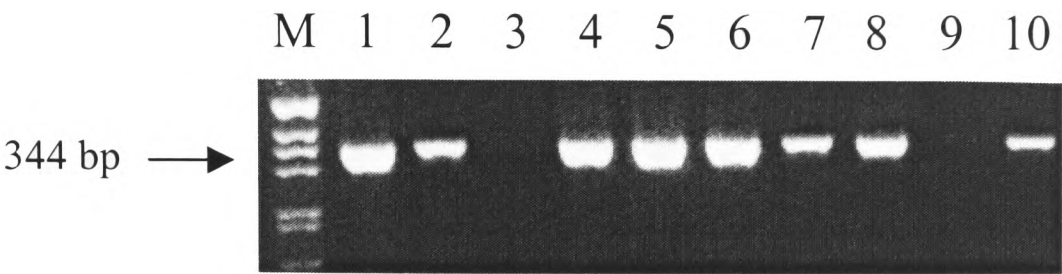


Figure 3.10a: PAC clones positive for *NAT* genes.
The candidate *NAT1*-positive clones identified following three levels of screening the RPCI 1 library were grown in bacterial culture and tested for presence of the *NAT1* gene using primers N769 and N1113. Amplification products were loaded on a 2% agarose gel. Lanes 1 to 8 were loaded with 10µl of PCR product from PAC clones 34-G18, 34-H20, 178-G11, 178-P8, 185-H1, 308-I7, 308-J8 and 308-K9. Lanes 9 and 10 were loaded with the PCR negative and positive controls respectively. Lanes M were loaded with 1µg of 1kb ladder. The 344bp product obtained from seven of the candidate clones confirmed presence of the *NAT1* gene in these PAC clones. Clone 178-G11 (lane 3) did not provide any product and proved to be a false positive. The arrow indicates a marker band.

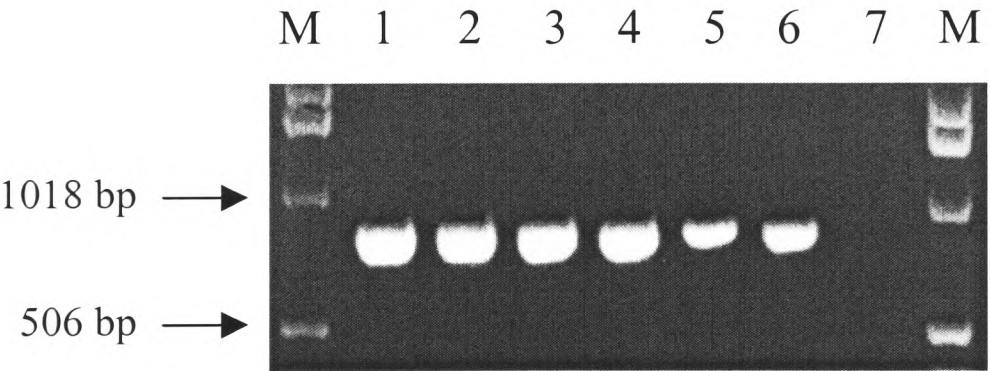


Figure 3.10b: PAC clones positive for *NAT* genes.
The candidate *NAT2*-positive clones identified following three levels of screening the RPCI 1 library were grown in bacterial culture and tested for presence of the *NAT2* gene using primers NAT-Hu7 and NAT-Hu8. Amplification products were loaded on a 1% agarose gel. Lanes 1 to 5 were loaded with 10µl of PCR product from PAC clones 155-H6, 226-H11, 296-B17, 296-C18 and 296-F16. Lanes 6 and 7 were loaded with the PCR positive and negative controls respectively. Lanes M were loaded with 1µg of 1kb ladder. The 840bp product obtained from all five candidate clones confirmed presence of the *NAT2* gene in these PAC clones. The arrows indicate marker bands.

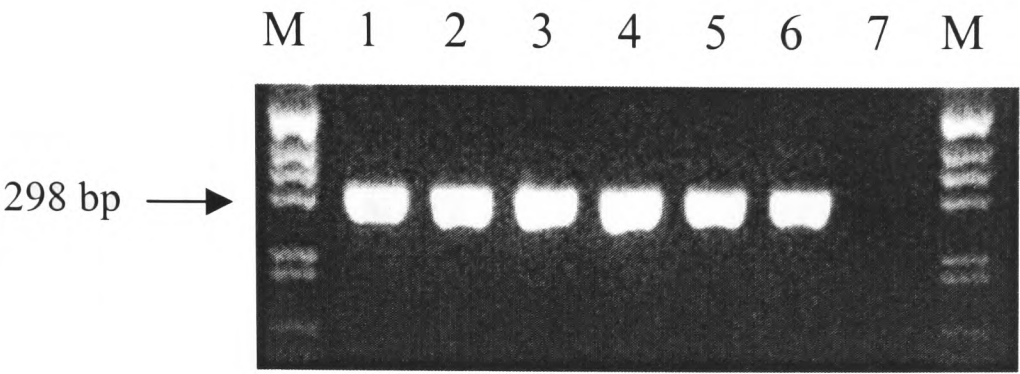


Figure 3.10c: PAC clones positive for *NAT* genes.
The clones identified to carry the *NAT2* gene (figure 3.10b) were also screened for presence of the *NATP* pseudogene using primers NAT-Hu24 and NAT-Hu27. Amplification products were loaded on a 2% metaphor gel. Lanes 1 to 5 were loaded with 10 μ l of PCR product from PAC clones 155-H6, 226-H11, 296-B17, 296-C18 and 296-F16. Lanes 6 and 7 were loaded with the PCR positive and negative controls respectively. Lanes M were loaded with 1 μ g of 1kb ladder. The 300bp product obtained from all five candidate clones confirmed presence of *NATP* in these PAC clones. The arrow indicates a marker band.

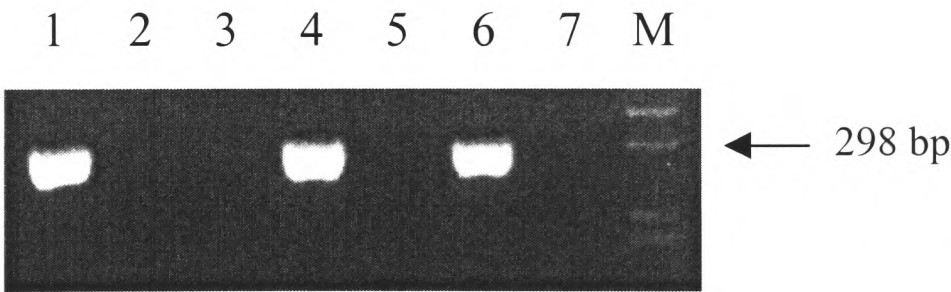


Figure 3.10d: PAC clones positive for *NAT* genes.
The candidate *NATP*-positive clones identified following three levels of screening the RPCI 1 library were grown in bacterial culture and tested for presence of the *NATP* pseudogene, using primers NAT-Hu24 and NAT-Hu27. Amplification products were loaded on a 2% metaphor gel. Lanes 1 to 5 were loaded with 10 μ l of PCR product from PAC clones 116-K16, 117-A5, 117-E2, 129-F16 and 129-N8. Lanes 6 and 7 were loaded with the PCR positive and negative controls respectively. Lanes M were loaded with 1 μ g of 1kb ladder. The 300bp product obtained from candidate clones 116-K16 and 129-F16 (lanes 1 and 6) confirmed presence of *NATP* in these PAC clones. The remaining three clones (lanes 2, 3 and 5) were randomly selected from the library to serve as negative controls for *NATP* amplification. The arrow indicates a marker band.

Table 3.1: Summary of the *NAT* positive PAC clones identified following screening of the RPCI 1 human genomic DNA PAC library and their *NAT* content.

PAC clones	<i>NAT1</i>	<i>NATP</i>	<i>NAT2</i>
34-G18	+	-	-
34-H20	+	-	-
178-P8	+	-	-
185-H1	+	-	-
308-I7	+	-	-
308-J8	+	-	-
308-K9	+	-	-
116-K16	-	+	-
129-F16	-	+	-
115-H6	-	+	+
226-H11	-	+	+
296-B17	-	+	+
296-C18	-	+	+
296-F16	-	+	+

Table 3.2: Summary of the digestions used to confirm the *NAT* content of the PAC clones identified as positive for at least one of the *NAT* loci. Specificity of amplification with each set of primers was checked by restriction digestion of the PCR product. All clones listed in table 3.1 produced the expected digestion pattern.

<i>NAT</i> locus	PCR product (bp)	Restriction endonuclease	Restriction fragments (bp)
<i>NAT1</i>	344	<i>Mbo</i> II	176, 125, 24*, 19*
<i>NAT2</i>	840	<i>Hind</i> II	560, 280
<i>NATP</i>	300	<i>Alu</i> I	223, 77

*Bands not visible on gel.

3.4.4 Genotyping the identified *NAT1* and *NAT2*-positive PAC clones

The genotype of the seven *NAT1*-positive clones (34-G18, 34-H20, 178-P8, 185-H1, 308-I7, 308-J8 and 308-K9) was determined following the approach described in section 3.2.1. All *NAT1* clones were determined to carry the *NAT1**4 allele, indicating that the human male whose DNA the library was constructed from was homozygous for the wild type allele (*NAT1**4/*4). Genotyping of the five *NAT2*-positive clones was performed as described in section 3.2.2. The *NAT2* clone 155-H6 was genotyped as carrying the *NAT2**4 allele, responsible for the fast acetylator phenotype, while the rest of the *NAT2* positive clones (226-H11, 296-B17, 296-C18 and 296-F16) were identified to carry the *NAT2**6A allele, one of the slow alleles. Therefore the DNA donor is a *NAT2**4/*6A heterozygote.

3.5 Discussion

The work presented here was designed to provide additional tools for studying the chromosomal region 8p22, where the *NAT* genes have been localised (Franke et al, 1994; Hickman et al, 1994). Two previously isolated cosmid clones, one positive for *NAT1* and the other positive for *NAT2* (Stacey et al, 1998; 1999), were characterised and assessed for their efficiency as probes in FISH analysis of the 8p22 region using healthy lymphocytes and a urothelial carcinoma cell line (RT112). Furthermore, a series of PAC clones containing one or more of *NAT* loci were identified and characterised, to facilitate investigation of the surrounding genomic region and overcome some of the limitations of cosmid and YAC cloning systems.

YAC clones carrying the *NAT* loci have previously been identified and used to localise the *NAT1* and *NAT2* genes in chromosomal region 8p21.3-23.1 (Hickman et al, 1994). However, these clones were found to be chimeric, carrying regions from chromosomes 10p and 11q. Due to their large size, YAC clones may contain several polymorphic markers and can be difficult to use for fine deletion mapping of chromosomes prepared from tumour cells by FISH analysis. Therefore, the

application of alternative tools for the detection of cytogenetic abnormalities involving the *NAT* region in cancerous material should be particularly useful.

A cosmid clone containing *NAT2* has previously been isolated and used to localise the *NAT2* gene in region 8p22 by FISH analysis (Franke et al, 1994), in agreement with the aforementioned study by Hickman et al (1994). This *NAT2* cosmid probe has also been tested with healthy urothelial cells extracted from urine (Stacey et al, 1996) and used in a preliminary study on loss of 8p22 in bladder cancer (Thygesen et al, 1999). Characterisation of the *NAT1* cosmid clone, described here, was of significance, as it allowed detection of the *NAT1* locus in nuclei from healthy and cancer cells, complementing current (this study) and previous (Franke et al, 1994; Stacey et al, 1996; Thygesen et al, 1999) work using *NAT2* cosmids as FISH probes. Both cosmid probes described in this study performed well in hybridisations with healthy cultured lymphocytes, providing the normal (diploid) number of *NAT*-specific signals. This indicated that they should be reliable in detecting cytogenetic abnormalities of the *NAT* region, as spontaneous loss or amplification of the fluorescent signal was not observed.

The ability of the two *NAT* cosmids to detect chromosomal abnormalities of the 8p22 region was also assessed using the RT112 cell line as a model for loss of the short arm of human chromosome 8 in tumour cells. FISH analysis revealed that the RT112 cell line carries only one copy of the *NAT1* and *NAT2* genes. The presence of an iso-8q chromosome was also confirmed. The results are in agreement with the published karyotype of the RT112 cell line, which describes the presence of an iso-8q chromosome among many other cytogenetic aberrations. The RT112 cells have been identified to contain two intact copies of most chromosomes, except for chromosomes 1, 3, 4, 5, 8, 19 and 21. Seven marker chromosomes of unknown identity (designated M₁ to M₇), originating from complex chromosomal rearrangements, have also been detected (Hastings and Franks, 1981). Using conventional chromosome banding techniques and FISH, the complete karyotype of this cell line has been determined as 47, XX, inv(2)(q32q37), del(3)(p22), i(4)(p10), i(8)(q10), der(11)t(3;11)(3q12;11p15.1), der(17)t(17;21)(17p12;21q21), der(18)t(18;20)(q21;20q11), add(21)(q11)+mar (Wang et al, 1995).

The NAT isoenzymes have been studied quite extensively in the RT112 cell line. Previous work has reported expression of NAT1, but not NAT2, using enzymatic assays (Coroneos and Sim, 1991), Western blotting (Stanley et al, 1996) and immunohistochemistry (Stacey et al, 1998). However, it was unclear whether this was due to very low levels of NAT2 expression, caused by a slow *NAT2* genotype, or to complete lack of NAT2 expression, or even to gene deletion. As reported here, the *NAT2* genotype of the RT112 cells was determined as *NAT2*4*, which is the wild-type allele. The NAT2 product of this allele would have been detectable by Western blot and immunohistochemical analysis, if it was expressed in the RT112 cells, and enzymatic activity would have been measurable, as well. Therefore, it was concluded that NAT2 is not expressed in the urothelial RT112 cell line. The *NAT1* genotype had previously been described as *NAT1*10* (Stacey et al, 1998). Linkage disequilibrium has been reported between specific *NAT1* and *NAT2* alleles, so that the combination *NAT1*10/NAT2*4* has been observed in a higher than expected frequency (Smelt et al, 1998; Henning et al, 1999). As demonstrated here by FISH analysis of interphase and metaphase nuclei, the RT112 cell line is hemizygous for both *NAT* genes and the haplotype of this cell line consists of the same combination of *NAT* alleles. The present work provides further information on the *NAT* status of the RT112 cell line, which can be a useful *in vitro* model for the study of chromosomal loss associated with the *NAT* region in bladder cancer cells.

The genomic region 8p22 that harbours the human *NAT* genes is often lost in bladder cancer, as will be discussed in section 6.3 (Knowles et al, 1993; Wagner et al, 1997; Knowles, 1999). It may also be deleted in other tumours like breast (Yaremko et al, 1996), colon (Gryfe et al, 1997), liver (Emi et al, 1993b), lung (Ohata et al, 1993) and prostate (Macoska et al, 1995). The presence of two tumour suppressor genes has long been suspected in this region, but their exact position has not been determined to date (Fujiwara et al, 1993; 1995; Farrington et al, 1996; Knowles, 1999). Therefore, the involvement of region 8p22 appears to be important in bladder and possibly other types of cancer and the *NAT* genes could serve as useful markers in studying the molecular progression of the disease (section 6.3.3). Based on their efficiency as probes for FISH analysis of RT112 interphase and metaphase cells, the *NAT1* and *NAT2* cosmid clones are reliable tools for the detection of chromosomal loss associated with the *NAT* loci.

Generating an accurate physical map of the region would be of benefit for identifying genes close to the *NAT* locus in humans. A contig covering a 1.2Mb region of 8p21.3-p22 has previously been constructed, using cosmid, BAC and PAC clones, and sequenced by a combination of shotgun and primer walking methods (Isomura et al, 1999). An earlier contig of the region, generated by the same group, consisted of three YAC clones and was used to isolate putative exons (Chinen et al, 1996). The YAC contig by Bookstein et al (1994), previously used for physical mapping of the *NAT* loci (Matas et al, 1997), also covers the same (8p22) region.

Two more contigs that span the 8p22 region have been recently described. A contig of approximately 1.7Mb, consisting of YAC and BAC clones, covers 1.4cM of the 8p22-p23 region and was constructed with the aim of identifying a TSG involved in *in situ* ductal carcinoma of the breast (Wang et al, 1999). This contig focuses on the minimum area of deletion around marker D8S550, which is 13.8cM away from marker D8S261 and 16.7cM away from the *LPL* gene, the markers flanking the *NAT* loci (Matas et al, 1997). In an effort to localise a TSG in a region of deletion in prostate cancer, Arbieva et al (2000) have also reported a contig, 1.0-1.5Mb in size, centred at the *MSR* locus on 8p22. This contig covers the region between markers D8S511 and D8S261. Based on their size and position, both contigs should cover the region on the telomeric side of the *NAT* loci, therefore, they should not contain them.

A small size contig covering the region surrounding the *NAT* loci might be useful for the identification of other genes in the immediate area. Towards that aim, a number of PAC clones positive for the three *NAT* loci were identified. The use of PACs has advantages over other tools used for similar purposes. Unlike YACs, which may exhibit problems associated with re-arrangements of the insert, deletions and chimaerism, PAC clones are much more stable. Their inserts are typically around 100-300kb, a convenient size conveying more genetic information than cosmids, but also allowing a more refined mapping of the region of interest, compared to YACs (Monaco and Larin, 1994). Therefore, PACs and other related *E. coli*-based cloning systems, such as P1s and BACs, should be more flexible tools for higher resolution mapping focusing on specific chromosomal regions of interest.

The PAC clones described here were found to be positive for either *NAT1* or *NAT2*, but the two loci were not found together on the same clone. Clones that contain *NAT2* were also shown to contain the *NATP* locus, although two clones were positive for *NATP* but not *NAT2*. This was a good indication that *NAT2* and *NATP* are closer to each other than to *NAT1*. It was possible to estimate the distance between the three loci, based on the average size of the insert of the PAC clones of the RPCI 1 library. According to the information about the library, available on the web site of the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu/>), the average size of the inserts is 110kb. Approximately 70% of the clones contain inserts in the range of 65-155kb, therefore *NAT2* and *NATP* can be placed within a maximum distance of approximately 155kb. This is in agreement with previous mapping data generated with YACs, which placed *NATP* within 140kb from *NAT2* (Matas et al, 1997). The same study has placed the *NAT1* and *NAT2* genes 170-360kb apart, which can explain why they were not found on the same PAC clone. Screening of the RPCI 1 PAC library and the *NAT*-positive clones in particular for additional known markers in the region would help integrate information from other relevant studies. Although 8p22 is estimated to have a low gene density, more than 35 ESTs have been identified to date in the 10cM region, between markers D8S549 and D8S258, spanning the three *NAT* loci (GeneMap'99, <http://www.ncbi.nlm.nih.gov>). In addition, several CpG islands have been identified in the 800-900kb region spanning the *NAT* loci, indicating the presence of transcriptionally active genes in the area (Matas et al, 1997). Mapping of the PAC clones described here should provide more information about the *NAT* region in humans.

Chapter 4: Mapping the Region of the *Nat1* and *Nat2* Genes in the Mouse

4.1 Introduction

The mouse is an extensively developed animal model for studying human disease, including single-gene disorders, complex multigenic traits, congenital defects, and different types of cancer. The ability to control mouse genetics, either via breeding or production of transgenic strains, has facilitated assessment of the extent to which genetic inheritance and environmental factors contribute towards the development of an abnormal phenotype (Moore, 1999). For over twenty years, the mouse has been widely used to study the biochemistry and molecular genetics of NAT, providing valuable information about its involvement in drug toxicity, chemical carcinogenesis (reviewed in Levy et al, 1992) and more recently development (Stanley et al, 1998).

As in the case of their human homologues, isolation and cloning of the genes for murine NAT was based on the nucleotide sequence previously identified for the cDNA of rabbit NAT (Blum et al, 1989a). A murine *Nat*-specific probe was generated by PCR amplification, using primers specific for rabbit *nat* and genomic DNA from fast acetylating C57Bl/6J mice as template. The probe was subsequently used for screening a mouse genomic library of C57Bl/6J origin. This allowed identification of a genomic clone containing two distinct genes for murine NAT, designated *Nat1* and *Nat2*, which were further subcloned and sequenced (Martell et al, 1991). A similar approach was adopted to isolate the third gene for murine NAT, designated *Nat3* (Kelly and Sim, 1992; 1994). Southern hybridisations of genomic DNA from slow acetylating A/J mice, using a rabbit *nat* probe, indicated the presence of a third *Nat* gene contained in a distinct *Hind*III restriction fragment. Following gel

electrophoresis of genomic DNA digested with *HindIII*, the fragments carrying the *Nat* genes were isolated and the *Nat3* gene was partially amplified using primers designed for rabbit *nat* (Kelly and Sim, 1992). Using this amplification product to screen a genomic library originating from the Balb/c mouse strain provided a clone containing the full length of the *Nat3* coding region, which was subsequently sequenced (Kelly and Sim, 1994).

All three murine *Nat* genes have an intronless coding region of 870bp, each encoding for a 290 amino acid chain (Martell et al, 1991; Kelly and Sim, 1994). They are highly homologous to each other and to their human counterparts, as summarised in table 4.1. Transient transfection of COS-1 cells with expression constructs containing the coding region of either *Nat1* or *Nat2* was used to determine substrate specificity of the protein products of the two genes. Cytosols prepared from cells transfected with *Nat1* preferentially N-acetylated isoniazid, while cytosols from cells transfected with *Nat2* were capable of pABA N-acetylation. Both expression products N-acetylated 2-aminofluorene, but neither of them could acetylate sulphamethazine (Martell et al, 1992). This confirmed that both genes were functional and indicated that *Nat1* encoded for “monomorphic” and *Nat2* for “polymorphic” NAT, the second being responsible for the observed NAT phenotype polymorphism among different inbred mouse strains (Glowinsky and Weber, 1982a). Comparison of the *Nat2* genomic sequence from the fast acetylating strains C57Bl/6J (Martell et al, 1991), Balb/c (Kelly and Sim, 1994), C3H/HeJ (Fretland et al, 1997), CBA/B6 and 129/Ola (Payton et al, 1999b) with the slow acetylating strains A/J (Martell et al, 1991) and A/HeJ (Fretland et al, 1997) revealed that the slow phenotype is caused by a single A to T nucleotide substitution at nucleotide position 296 of the *Nat2* coding region. This mutation changes the amino acid residue at position 99 of the NAT2 polypeptide chain from asparagine in the fast to isoleucine in the slow acetylating strains (Martell et al, 1991). The amino acid substitution is believed to result in a less stable and catalytically impaired form of the enzyme (De Leon et al, 1995; section 1.4.1.1). The fast and slow alleles of murine *Nat2* were designated *Nat2**8 and *Nat2**9 respectively (Vatsis et al, 1995).

Table 4.1 Percentage identity between the human and mouse genes for NAT, based on their nucleotide sequence. The genes for human NAT (*NAT1* and *NAT2*) and the human pseudogene (*NATP*) are compared with the genes for murine NAT (*Nat1*, *Nat2* and *Nat3*). The genes of each species are also compared with each other. For comparison of the amino acid sequences, see table 1.5.

	<i>NAT1</i> ¹	<i>NAT2</i> ²	<i>NATP</i>	<i>Nat1</i>	<i>Nat2</i> ³	<i>Nat3</i>
<i>NAT1</i> ¹	-	87	75	80	83	75
<i>NAT2</i> ²		-	76	79	80	75
<i>NATP</i>			-	68	69	65
<i>Nat1</i>				-	84	74
<i>Nat2</i> ³					-	78
<i>Nat3</i>						-

¹Wild type *NAT1**4 allele (Vatsis and Weber, 1993)
²Wild type *NAT2**4 allele (Grant et al, 1989a)
³*Nat2**8 allele present in fast acetylating strains (Martell et al, 1991)

No polymorphisms have been described for the *Nat1* gene in any of the strains examined to date (Martell et al, 1991; Kelly and Sim, 1994; Fretland et al, 1997). The single *Nat1* allele was designated *Nat1**6 (Vatsis et al, 1995). The *Nat3* gene also appears to be monomorphic, as its nucleotide sequence is identical between Balb/c (Kelly and Sim, 1994), C3H/HeJ, A/HeJ (Fretland et al, 1997), C57Bl/6J, A/J and CD1 (Estrada-Rodgers et al, 1998b) mice. According to NAT nomenclature, the single *Nat3* allele was named *Nat3**1 (Vatsis et al, 1995).

No deleterious mutations have been identified in the coding region of *Nat3* (Kelly and Sim, 1994), suggesting that, unlike the third human *NAT* locus (*NATP*), it is not a pseudogene (Blum et al, 1990). Expression of recombinant *Nat3* in both bacterial (Kelly and Sim, 1994; Fretland et al, 1997) and eukaryotic (Estrada-Rodgers et al, 1998b) systems also indicated that this is a functional gene. However, *Nat3* shows less homology with either *Nat1* or *Nat2*, than *Nat1* and *Nat2* with each other (table 4.1), implying that *Nat3* may have diverged earlier than the two other *Nat* genes and could have a role other than in the metabolism of xenobiotics. This might explain the hardly detectable or absent enzymatic activity of NAT3 with all common NAT-specific substrates examined to date (Kelly and Sim, 1994; Fretland et al, 1997; Estrada-Rodgers et al, 1998b), as well as its very restricted tissue distribution (section 1.4.1.5; Boukouvala and Sim, 2000).

There is limited information on transcription of the mouse *Nat* genes. *Nat1* and *Nat2* mRNA from liver of fast (C57Bl/6J) and slow (A/J) acetylating mouse strains has been analysed by Northern blotting. Both mRNAs were about 1.5kb in size and their levels did not differ between the two strains. However, the *Nat2* mRNA was more abundant in liver than *Nat1* mRNA (Martell et al, 1992; De Leon et al, 1995). Several other studies have employed RT-PCR to detect *Nat1*, *Nat2*, and *Nat3* transcripts in various tissues and at different developmental stages (Mitchell et al, 1999; Estrada et al, 2000; Boukouvala and Sim, 2000), as described in sections 1.4.1.2, 1.4.1.3 and 1.4.1.5.

Two recent studies have investigated the transcription of murine *Nat2*. The first (Estrada-Rodgers et al, 1998a), used RNase protection assays and primer extension analysis to identify the transcription initiation site, which was located 112-151bp upstream of the *Nat2* coding region. The same study used reporter gene assays to characterise a glucocorticoid-dependent hormone response element (section 1.4.1.2), located approximately 580bp upstream of the *Nat2* coding region, and computational analysis to identify possible promoter elements and transcription factor binding sites in the same upstream region. The second study (Fakis et al, 2000) identified an upstream non-coding exon for murine *Nat2*, located about 6kb upstream of the coding region, based on sequence similarity with the upstream non-coding exon previously described for human *NAT2* (Ebisawa and Deguchi, 1991). Amplification from genomic DNA and cDNA, using primers specific for the putative non-coding exon and the *Nat2* coding region, provided products with a size difference of about 6kb. Sequencing of the cDNA amplification product identified two splice sites for the *Nat2* transcript, a donor site located 6,172bp and an acceptor site located only 6bp upstream of the translation initiation site of the gene. This finding suggests that transcription of *Nat2* is initiated upstream of the non-coding exon of the gene. The former study (Estrada-Rodgers et al, 1998a) used kidney and liver from outbred CD1 mice as source of RNA, while the second (Fakis et al, 2000) used RNA from embryonic stem cells of 129/Ola origin. It seems possible that more than one promoter might drive *Nat2* expression in different tissues, but this requires further investigation.

4.2 Materials and study design

Three genomic DNA plasmid clones were used for restriction mapping of the region surrounding the *Nat1* and *Nat2* genes in mice. Two of the clones were of Balb/c origin, while the third originated from the 129/Ola mouse strain. The two Balb/c clones (clones 1 and 2) had previously been isolated by screening of a mouse genomic library with *Nat*-specific probes, and a partial restriction map of the clones had been generated (Kelly and Sim, 1994). The third clone, of 129/Ola origin (clone A), was isolated by Dr C. Denning (Centre for Genome Research, University of Edinburgh, UK) by screening of the E14 TG2a library, constructed with genomic DNA from 129/Ola mouse ES cells (Nehls et al, 1994). Hybridisation was performed with a *Nat2*-specific probe provided by Dr M. Payton (Department of Pharmacology, University of Oxford, UK). The probe was generated by PCR amplification with primers mNAT2 516 and mNAT2 867 (table 2.1) and hybridised to part of the *Nat2* gene, at the 3' end of the coding region.

The plasmid Balb/c clones 1 and 2 and the 129/Ola clone A were analysed for their content of *Nat* genes by PCR. Furthermore, the clones positive for *Nat2* were genotyped, to determine whether they were from a fast or slow acetylator strain. To study the genomic region surrounding the *Nat1* and *Nat2* genes, the three plasmid clones were subsequently mapped with a number of restriction endonucleases, selected to allow comparison with previous work on the Balb/c (Kelly and Sim, 1994) and the C57Bl/6J (Martell et al, 1991) mouse strains. The results of these studies were confirmed and expanded, with the use of several additional restriction enzymes, to produce a higher resolution restriction map of the *Nat1* and *Nat2* region in mice.

4.3 Method for genotyping the mouse *Nat2* gene

The presence of the *Nat2**8 and *Nat2**9 alleles, responsible for fast and slow acetylator phenotype in mice respectively, was investigated by a combination of PCR and RFLP analysis (Payton et al, 1999b). The 131bp amplification product, using primers mNAT2 278 and mNAT2 409 (table 2.1), covers part of the *Nat2* coding region from nucleotide position 278 to 409. Digestion of the PCR product with *Tru9I*

is then used to distinguish between the two *Nat2* alleles. Within the amplified region, the product of the slow *Nat2**9 allele contains a single *Tru9I* restriction site and is digested into two fragments of 115 and 16bp, while the product of the fast *Nat2**8 allele remains undigested. The 16bp difference in the size of the *Tru9I*-digested PCR product amplified from the fast *Nat2**8 and the slow *Nat2**9 allele can be visualised by electrophoresis on a 4% (w/v) Metaphor agarose gel, as described in section 2.2.5.1.

4.4 Results

4.4.1 Investigation of the *Nat* content of the plasmid Balb/c clones 1 and 2 and the 129/Ola clone A

PCR amplification was used to investigate the *Nat* content of the plasmid clones described above. The coding region of the *Nat1*, *Nat2* and *Nat3* genes was amplified using primer pairs Mus12/ Mus13, mNAT2 1/ mNAT2 910 and Mus12/ Mus15 respectively (table 2.1). The two Balb/c plasmid clones (clones 1 and 2) were determined to carry the entire coding region of either *Nat1* (clone 1) or *Nat2* (clone 2), but not both (figure 4.1a and b). Using the same pairs of primers, the 129/Ola clone (clone A) was identified to contain the entire coding region of *Nat2* only (figure 4.1a and b). However, following amplification with primers mNAT1 516 and mNAT1 867 (table 2.1), which amplify the 3' end of *Nat1* from nucleotide position 516 to 867, the 129/Ola clone was demonstrated to contain part of the *Nat1* coding region. This part of *Nat1* extends from at least nucleotide 516 to the end of the coding region (figure 4.1c). The *Nat3* gene was not present in any of the plasmid clones analysed, as indicated by lack of amplification with primers Mus12 and Mus15. The *Nat* content of the three plasmid clones is summarised in table 4.2.

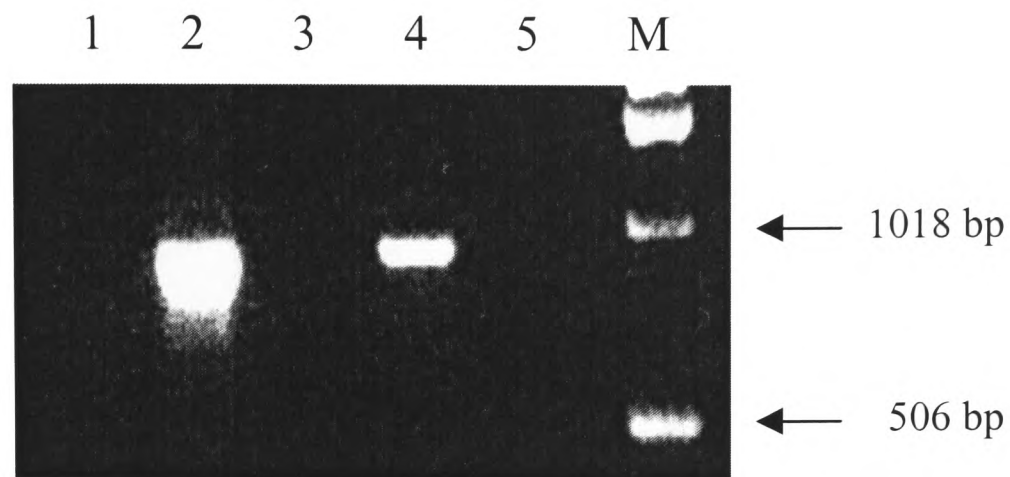


Figure 4.1a: Investigation of the *Nat* content of the three mouse genomic plasmid clones used for restriction mapping.

The coding region of *Nat1* was amplified from plasmid DNA (50ng) using primers Mus12 and Mus13. PCR products were separated on a 1.5% agarose gel. Lanes 1 to 3 were loaded with PCR product from 129/Ola clone A, Balb/c clone 1 and Balb/c clone 2 respectively. Lane 4 is the PCR positive control (0.5 μ g mouse genomic DNA as template), while lane 5 is the reaction negative control (no template). Lane M was loaded with 1 μ g of 1kb ladder. A fragment of 908bp was amplified from Balb/c clone 1, indicating that this is the only clone containing the entire *Nat1* coding region.

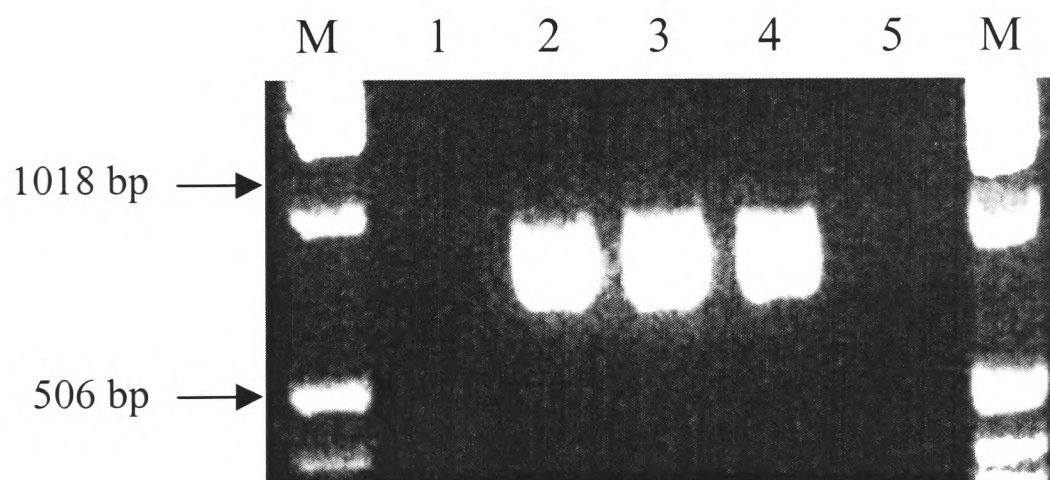


Figure 4.1b: Investigation of the *Nat* content of the three mouse genomic plasmid clones used for restriction mapping.

The coding region of *Nat2* was amplified from plasmid DNA (50ng) using primers mNAT2 1 and mNAT2 910. PCR products were separated on a 1.5% agarose gel. Lanes 1 to 3 were loaded with PCR product from Balb/c clone 1, Balb/c clone 2 and 129/Ola clone A respectively. Lane 4 is the PCR positive control (0.5 μ g mouse genomic DNA as template), while lane 5 is the reaction negative control (no template). Lanes M were loaded with 1 μ g of 1kb ladder. A fragment of 910bp was amplified from Balb/c clone 2 and 129/Ola clone A, indicating that they both contain the entire *Nat2* coding region.

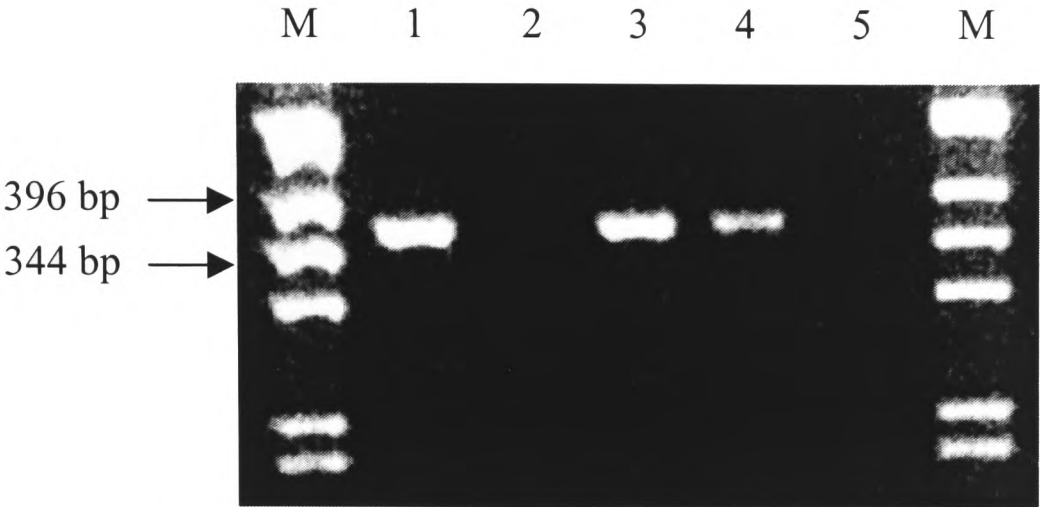


Figure 4.1c: Investigation of the *Nat* content of the three mouse genomic plasmid clones used for restriction mapping. The 3' end of the *Nat1* coding region was amplified from plasmid DNA (50ng) using primers mNAT1 516 and mNAT1 867. PCR products were separated on a 2% metaphor gel. Lanes 1 to 3 were loaded with PCR products from Balb/c clone 1, Balb/c clone 2 and 129/Ola clone A respectively. Lane 4 is the PCR positive control (0.5µg mouse genomic DNA as template), while lane 5 is the reaction negative control (no template). Lanes M were loaded with 1µg of 1kb ladder. A fragment 364bp in size was amplified from Balb/c clone 1 and 129/Ola clone A, indicating that they both contain the 3' end of the *Nat1* coding region.

Table 4.2: Summary of the *Nat* content of the mouse genomic plasmid clones used for restriction mapping. The two Balb/c plasmid clones were positive for either *Nat1* (clone 1) or *Nat2* (clone 2). The 129/Ola plasmid clone (clone A) contains the entire *Nat2* coding region and the 3' end of *Nat1* coding region.

	Balb/c clone 1	Balb/c clone 2	129/Ola clone A
<i>Nat1</i> 3' end	+	-	+
<i>Nat1</i> coding region	+	-	-
<i>Nat2</i> 3' end	-	+	+
<i>Nat2</i> coding region	-	+	+
<i>Nat3</i> coding region	-	-	-

Since all three murine *Nat* genes are very similar to each other in sequence (table 4.1), it was important to confirm specificity of the PCR amplifications performed by restriction enzyme digestions. The amplification product for *Nat1* is specifically cleaved by *AvaI* at nucleotide position 587, while the amplification product for *Nat2* is cleaved by *PstI* at position 673 and the product for *Nat3* is cleaved by *HaeIII* at position 237 (table 4.3). As none of the plasmid clones described here was positive for *Nat3*, specificity was only checked by digestion with *AvaI* or *PstI* (figure 4.2a and b).

Table 4.3: Summary of the restriction enzyme digestions used to confirm specificity of amplification from the mouse *Nat* genes.

	Primer pairs used for amplification	Undigested PCR product (bp)	Restriction fragments (bp) following digestions with:		
			<i>AvaI</i>	<i>PstI</i>	<i>HaeIII</i>
<i>Nat1</i> coding region	Mus12 & Mus13	908	601 307	908	908
<i>Nat2</i> coding region	mNAT2 1 & mNAT2 910	910	910	672 238	910
<i>Nat3</i> coding region	Mus12 & Mus15	908	908	908	657 251

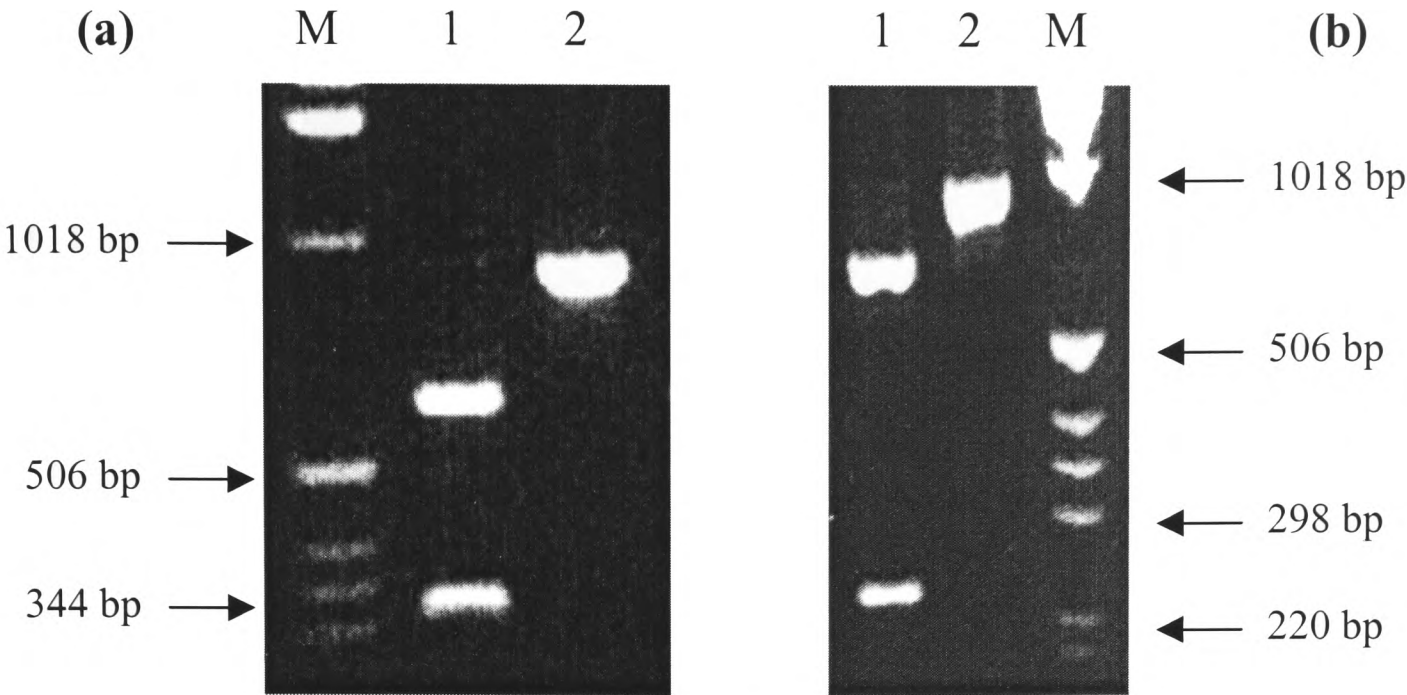


Figure 4.2: Specificity of amplification from the mouse *Nat* genes. The coding region of *Nat1* and *Nat2* was amplified from Balb/c clones 1 and 2 respectively, using the primer pairs shown in table 4.3. The product amplified from clone 1 was then digested with *AvaI* (a) and the product amplified from clone 2 with *PstI* (b). Digestion fragments were separated on a 1.5% agarose gel. Lane 1 was loaded with *Nat1* (a) or *Nat2* (b) amplification product digested with *AvaI* or *PstI*, respectively. Lane 2 was loaded with the same undigested amplification products for comparison. Lane M was loaded with 1µg of 1kb ladder. In both cases, PCR amplification was specific.

4.4.2 Genotyping the mouse *Nat2* plasmid clones

The Balb/c clone 2 and 129/Ola clone A, that were shown to be positive for *Nat2* (table 4.2), were further genotyped as described in section 4.3. DNA from each one of these two clones was used as template for amplification with the *Nat2*-specific primers mNAT2 278 and mNAT2 409, followed by digestion of the generated 131bp product (figure 4.3a) with *Tru9I*. The PCR products from both clones remained undigested, indicating the *Nat2**8 allele, associated with the fast acetylator phenotype in mouse strains (figure 4.3b).

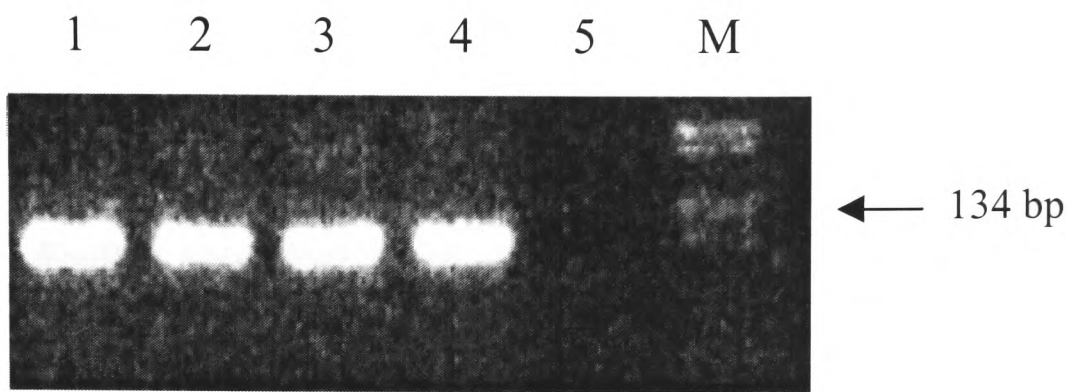


Figure 4.3a: Genotyping of the mouse genomic plasmid clones carrying the *Nat2* gene.

Plasmid DNA (50ng) from the *Nat2*-positive Balb/c clone 2 and the 129/Ola clone A was amplified with primers mNAT2 278 and mNAT2 409. Genomic DNA (0.5µg) from a fast (C57Bl/6J) and a slow (A/J) acetylator mouse strain was also amplified with the same set of primers, for comparison. The PCR products (5µl) were separated on a 4% metaphor gel. Lanes 1 and 2 were loaded with PCR product from Balb/c clone 2 and 129/Ola clone A respectively, while lanes 3 and 4 were loaded with PCR product from A/J and C57Bl/6J genomic DNA, respectively. Lane 5 was loaded with the PCR negative control (no template) and lane M was loaded with 1µg of 1kb ladder. A fragment of 131bp was amplified from the *Nat2* coding region, as expected.

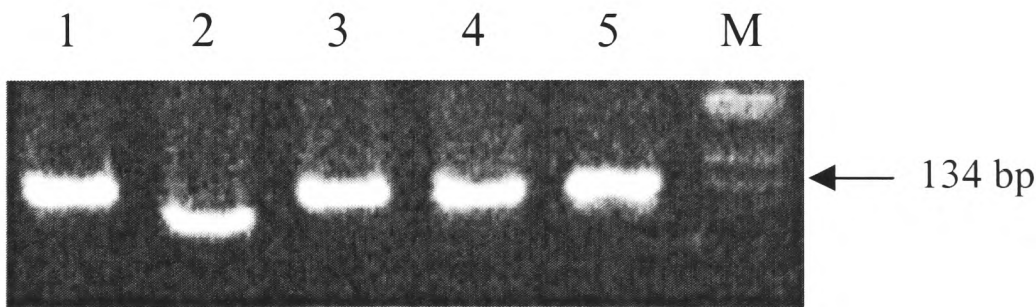


Figure 4.3b: Genotyping of the mouse genomic plasmid clones carrying the *Nat2* gene.

The PCR products (5µl) shown in figure 4.3a were digested with *Tru9I* at 65°C for 3 hours and the digests were separated on a 4% metaphor gel. Lane 1 was loaded with undigested PCR product, as control. Lanes 2 and 3 were loaded with the digested PCR products from A/J and C57Bl/6J genomic DNA, respectively. Lanes 4 and 5 were loaded with the digested PCR products from Balb/c clone 2 and 129/Ola clone A plasmid DNA, respectively. Lane M was loaded with 1µg of 1kb ladder. Only the PCR product from the A/J mouse (lane 2) was cleaved by *Tru9I*, as indicated by a slightly smaller band of 115bp, confirming a homozygous slow *Nat2**9 genotype. The C57Bl/6J, Balb/c clone 2 and 129/Ola clone A amplification products (lanes 3-5) were not cleaved by *Tru9I*, indicating that all three strains were homozygous for the fast *Nat2**8 allele.

4.4.3 Restriction mapping of the mouse genomic plasmid clones

Mapping of the three plasmid clones was carried out using a total of twelve restriction endonucleases. The insert was initially isolated from the vector and then digested with *Bgl*II, *Dra*I, *Hind*III, *Eco*RV, *Nde*I, *Pst*I, *Sma*I, *Acc*I, *Sal*I, *Eco*RI, *Bam*HI or *Not*I. Digestions using combinations of more than one of the above enzymes were also performed (section 2.2.8) to facilitate ordering of the restriction sites and confirm the results obtained from single digestions. The generated restriction fragments were separated on 0.7% (w/v) agarose gels and blotted onto nylon membranes for hybridisation (section 2.2.12.1). The membranes were hybridised to *Nat1*- or *Nat2*-specific probes, designed to recognise the 3' end of each one of the two genes from nucleotide position 516 to 867. This is the region showing the lowest similarity between the *Nat* genes in mice, therefore, these probes are the most appropriate to ensure specific hybridisation to the target *Nat*-gene only.

4.4.3.1 Mapping of Balb/c clone 1

For restriction mapping of the Balb/c clone 1, the insert was excised from the pUC18 vector by digestion with *Sac*I and recovered from a 0.7% (w/v) low melting point agarose gel, as described in section 2.2.6 (figure 4.4). The insert was approximately 4.8kb in size and contained the entire coding region of *Nat1*, as demonstrated in section 4.4.1 (table 4.2).

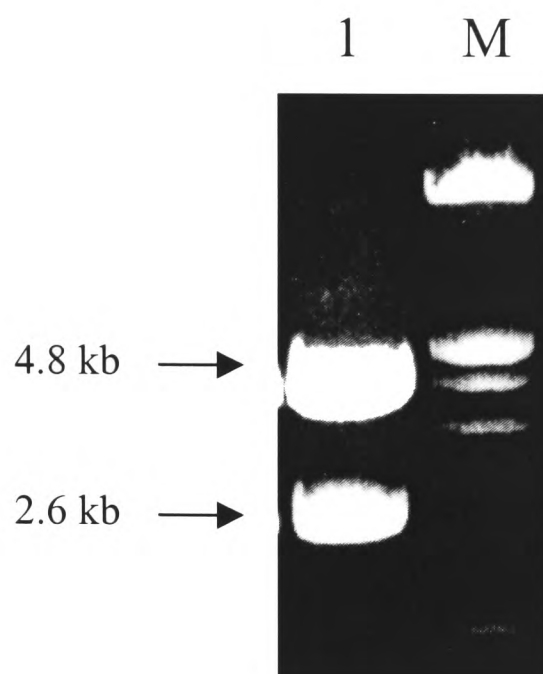


Figure 4.4: Isolation of the insert from Balb/c clone 1.

Approximately 5 µg of plasmid DNA were digested with 50U *SacI*. The digest was separated on a 0.7% low melting point agarose gel (lane 1), next to 1 µg of λ DNA/*EcoRI*+*HindIII* molecular weight markers (lane M). The band of the insert was 4.8kb in size and the band of the pUC18 vector was 2.6kb in size. The insert band was excised from the gel and the DNA was purified as described in Materials and Methods (section 2.2.6).

Digestion of Balb/c clone 1 with *Bgl*II produced two bands of approximately 3.2 and 1.6kb, both of which hybridised to the radioactively labelled *Nat1*-specific probe (table 4.4; figure 4.5a and b). Based on the previously published sequence for the mouse *Nat1* gene (Martell et al, 1991), *Bgl*II should cut within the *Nat1* coding region at nucleotide position 750. Therefore, the *Nat1*-specific probe should hybridise to both restriction fragments generated with *Bgl*II digestion of the Balb/c clone 1, as observed. Digestion with *Nde*I produced two bands, 3.0 and 1.8kb in size, only the latter one hybridising to the probe (table 4.4; figure 4.5a and b). A double digestion with *Bgl*II and *Nde*I produced three bands, 3.0, 1.6 and 0.18kb in size and Southern blot analysis showed that only the smaller two fragments hybridised to the *Nat1*-specific probe (table 4.4; figure 4.5a and b). This indicated that the *Nde*I site should be located approximately 200bp downstream of the *Bgl*II site, which lies within the *Nat1* coding region at approximately 1.6kb from the 5' end and 3.2kb from the 3' end

of the clone. The same combination of digestions also indicated the position of the *Nat1* coding region on the clone (figure 4.10a).

Digestion of the Balb/c clone 1 with *SmaI* produced a single band of approximately 2.4kb, which also hybridised to the *Nat1*-specific probe (table 4.4; figure 4.5a and b). Since the size of this band was half of the estimated 4.8kb size of the insert, it is concluded that there was a unique *SmaI* site located in the middle of the clone. Therefore, digestion appears to have produced two overlapping fragments of similar size, one of which produced hybridisation signal. Digestion of the insert was not observed with *SalI*. Only one band of 4.8kb was visible on the gel and the autoradiogram, indicating that there were no *SalI* restriction sites on Balb/c clone 1 (table 4.4; figure 4.5a and b). Digestion with *DraI* produced three bands, 2.7, 1.6 and 0.5kb in size, with the two largest fragments hybridising to the *Nat1*-specific probe. According to the sequence of the *Nat1* gene (Martell et al, 1991), there is one *DraI* site at nucleotide position 763 of the coding region, very close to the *BglII* site described above. The other *DraI* site should, therefore, be located approximately 0.5kb from the 3' end of the clone (figure 4.10a). This was also confirmed by a double digestion with *DraI* and *SmaI* (table 4.4; figure 4.5a and b).

Digestion of the Balb/c clone 1 insert with *AccI* produced two bands of 0.6kb and 4.0kb, only the largest of which hybridised to the *Nat1*-specific probe. A double digest with *AccI* and *BglII* produced fragments of 0.6, 1.0 and 3.2kb (table 4.4), indicating that the *AccI* site should be located 1.0kb upstream of the *BglII* site. This places the unique *AccI* site at 0.6kb from the 5' end of the clone. Digestion with *PstI* cleaved the insert only once, producing fragments of 3.5 and 1.3kb. Southern blot analysis showed that only the longer fragment hybridised to the *Nat1*-specific probe. Following double digestion with *PstI* and *NdeI*, producing bands 1.8, 1.6 and 1.3kb in size (table 4.4), the position of the *PstI* site was determined at approximately 1.3kb from the 3' end of the clone (figure 4.10a).

When the Balb/c clone 1 insert was digested with *HindIII*, two fragments, 4.6 and 0.2kb in size were generated. The largest fragment produced a signal, following hybridisation with the radioactively labelled *Nat1*-specific probe (table 4.4; figure 4.5a

and b). A combined digestion with *Hind*III and *Pst*I produced three fragments of 3.3kb, 1.3 and 0.2kb, indicating that the *Hind*III site is located about 0.2kb from the 5' end of the clone (figure 4.10a).

Additional double digestions, e.g. *Sal*I+*Acc*I, *Sma*I+*Acc*I, *Pst*I+*Dra*I, *Hind*III+*Sal*I and *Bgl*II+*Pst*I (table 4.4), were used to confirm the above results. The generated complete restriction map of the Balb/c clone 1 is presented in figure 4.10a.

Table 4.4: Restriction patterns obtained for mapping of the Balb/c clone 1. Single or combined enzyme digestions used for mapping and the produced digestion patterns are presented. Where available, fragments hybridising to the *Nat1*-specific probe are indicated with an asterisk (*). The *Nat1* probe used in hybridisations was designed against the 3' end of the coding region from nucleotide position 516 to 867.

Enzyme	Size of restriction fragments
<i>Acc</i> I	0.6kb, 4.0kb*
<i>Bgl</i> II	1.6kb*, 3.2kb*
<i>Dra</i> I	1.6kb*, 2.7kb*, 0.5kb
<i>Hind</i> III	0.2kb, 4.6kb*
<i>Nde</i> I	1.8kb*, 3.0kb
<i>Pst</i> I	3.5kb*, 1.3kb
<i>Sal</i> I	4.8kb*
<i>Sma</i> I	2.4kb*, 2.4kb
<i>Bgl</i> II+ <i>Nde</i> I	3.0kb, 1.6kb *, 0.18kb *
<i>Pst</i> I+ <i>Hind</i> III	3.3kb *, 1.3kb, 0.2kb
<i>Dra</i> I+ <i>Sma</i> I	1.9kb, 1.6kb*, 0.6kb*, 0.5kb
<i>Sal</i> I+ <i>Acc</i> I	0.6kb, 4.0kb*
<i>Acc</i> I+ <i>Bgl</i> II	3.2kb, 1.0kb, 0.6kb
<i>Acc</i> I+ <i>Sma</i> I	2.4 kb, 1.8 kb, 0.6 kb
<i>Pst</i> I+ <i>Dra</i> I	1.9kb, 1.6kb, 1.3kb
<i>Hind</i> III+ <i>Sal</i> I	4.6kb, 0.2kb
<i>Bgl</i> II+ <i>Pst</i> I	1.9kb, 1.6kb, 1.3kb
<i>Nde</i> I+ <i>Pst</i> I	1.8kb, 1.6kb, 1.3kb

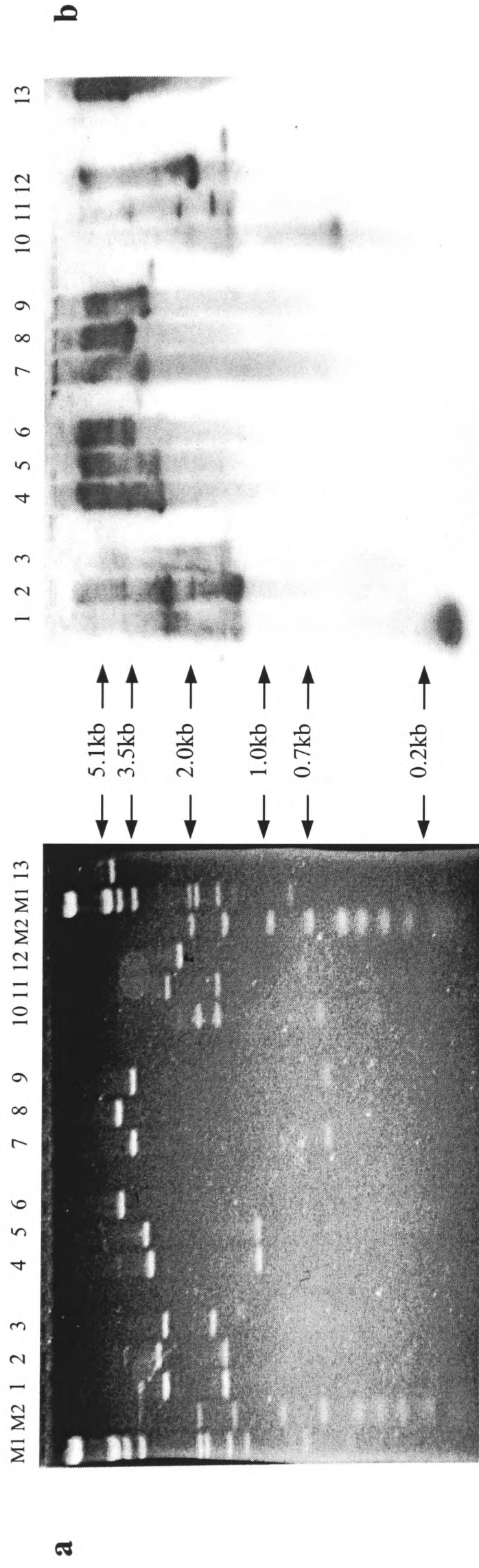


Figure 4.5: Restriction enzyme digestions (a) and Southern blot hybridisation (b) for mapping of the 4.8kb *NatI*-positive Balb/c clone 1. Approximately 500ng of insert DNA were digested with 10U of restriction enzyme or with a combination of two enzymes, and the products were separated on a 0.7% agarose gel (a). Following electrophoresis, the gel was blotted and hybridised with a probe specific for the 3' end of the *NatI* coding region, as described in section 2.2.12.1 (b). In both a and b, lanes were loaded with digests in the following order: *BglII*+*NdeI* (lane 1), *BglII* (lane 2), *NdeI* (lane 3), *PstI*+*HindIII* (lane 4), *PstI* (lane 5), *HindIII* (lane 6), *SalI*+*AccI* (lane 7), *SalI* (lane 8), *AccI* (lane 9), *DraI*+*SmaI* (lane 10), *DraI* (lane 11) and *SmaI* (lane 12). Lane 13 was loaded with 500ng of undigested insert. Lanes M1 and M2 were loaded with 1µg of λ DNA/*EcoRI*+*HindIII* and 1µg Biomarker EXT molecular weight markers respectively. Note that faint bands of less than 300bp in size may not be visible on the gel picture (a).

4.4.3.2 Mapping of the Balb/c clone 2

The same approach was used for restriction mapping of the Balb/c clone 2, following excision of the insert from the pUC18 vector by *SacI* digestion and recovery of the DNA from a 0.7% (w/v) low melting point agarose gel (figure 4.6). The size of the insert was estimated to approximately 8kb and contained the entire coding region of *Nat1*, as demonstrated in section 4.4.1 (table 4.2).

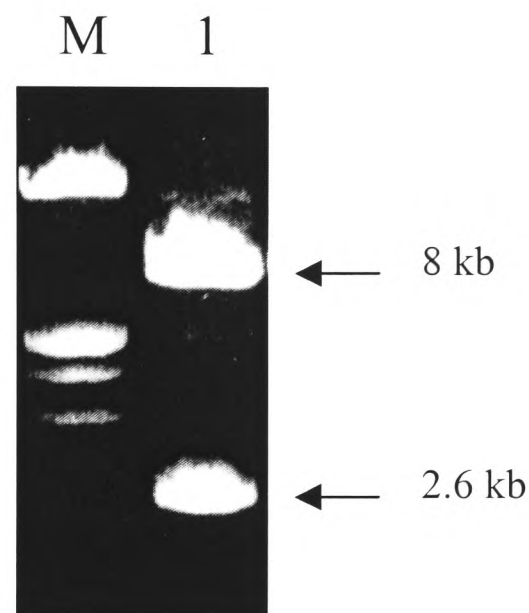


Figure 4.6: Isolation of the insert from the Balb/c clone 2.

Approximately 5µg of plasmid DNA were digested with 50U *SacI*. The digest was separated on a 0.7% low melting point agarose gel (lane 1), next to 1µg of λDNA/*EcoRI*+*HindIII* molecular weight markers (lane M). The insert band was 8kb in size and the band of the pUC18 vector was 2.6kb in size. The band of the insert was excised from the gel and the DNA purified as described in Materials and Methods (section 2.2.6).

Digestion of the Balb/c clone 2 insert with *EcoRV* produced two bands 5.1 and 2.9kb in size, of which only the largest one hybridised to the *Nat2*-specific probe (table 4.5; figure 4.7a and b). Digestion with *HindIII* cleaved the insert into two fragments, 1 and 7kb in size, and Southern blot analysis showed that only the 7kb fragment contained the *Nat2* coding region (table 4.5; figure 4.7a and b). To map the *EcoRV* and *HindIII* restriction sites in relation with each other, a double digestion was performed. This produced bands of 5.1, 1.9 and 1kb, of which only the 5.1kb band was positive for the *Nat2* coding region. This placed the unique *EcoRV* site at approximately 5.1kb from

the 5' end and the *Hind*III site at approximately 1kb from the 3' end of the Balb/c clone 2 (figure 4.10b).

Digestion of the insert with the *Bgl*III produced bands 3.5, 2.2, 1.7 and 0.6kb in size, the largest of which hybridised to the *Nat2*-specific probe (table 4.5; figure 4.7a and b). From sequencing analysis of the *Nat2* gene (Martell et al, 1991), it is known that there is a *Bgl*III restriction site at nucleotide position 534 of the coding region. The probe used is specific for the *Nat2* coding region from nucleotide position 516 to 867, suggesting that only the *Bgl*III fragment containing the 3' end of the gene would provide a signal on the autoradiogram. Therefore, the 3.5kb *Bgl*III fragment that hybridises to the *Nat2*-specific probe (table 4.5; figure 4.7a and b) contains the 3' end of the gene. A combined digestion with *Bgl*III and *Eco*RV produced fragments of 2.3kb, 2.2kb, 1.7kb, 1.2kb and 0.6kb and Southern hybridisation showed the 1.2kb fragment to contain the 3' end of the *Nat2* coding region (table 4.5). These digestion patterns indicated that the *Eco*RV site should be located about 1.2kb downstream of the *Bgl*III site of the *Nat2* coding region. Another *Bgl*III site was placed about 2.2kb from the 5' end of the clone, while a third *Bgl*III site was located 0.6kb from the 3' end of the clone (figure 4.10b). This combination of digestions also placed the *Bgl*III site, contained within the *Nat2* coding region, at approximately 3.9 and 4.1kb from the 5' and the 3' end of the clone, respectively, allowing positioning of the *Nat2* coding region near the middle of Balb/c clone 2 (figure 4.10b).

Digestion of the insert with *Pst*I resulted in three bands of 4, 2.6 and 1.4kb. Two of these bands (4 and 1.4kb) hybridised to the *Nat2* probe, indicating presence of a *Pst*I site within the *Nat2* coding region, in the part covered by the *Nat2*-specific probe (table 4.5). This site is located at nucleotide position 673, as previously established by sequencing analysis of the *Nat2* gene (Martell et al, 1991). To determine the exact localisation of the *Pst*I sites, a combined *Pst*I and *Hind*III digestion was performed. This produced fragments 3, 2.6, 1.4, and 1kb in size, with bands 3 and 1.4kb hybridising to the *Nat2*-specific probe (table 4.5). It was deduced that the *Pst*I site within the *Nat2* coding region was located 3kb upstream of the unique *Hind*III site. The other *Pst*I site was placed 1.4kb upstream of that located within the *Nat2* coding region. In agreement with the results obtained from the *Bgl*III digestions as described

above, the *Pst*I digestions also indicated the position of the *Nat2* coding region near the middle of the Balb/c clone 2. The beginning of *Nat2* coding region was, therefore, placed at approximately 3.4 and 3.3kb from the 5' and the 3' end of the clone, respectively (figure 4.10b).

Digestion of the Balb/c clone 2 with *Acc*I produced fragments 1.8, 5.6 and 0.5kb in size, with the largest one hybridising to the *Nat2*-specific probe (table 4.5; figure 4.7a and b). To determine the position of the *Acc*I sites, a double digestion with *Acc*I and *Eco*RV was performed. The produced bands were 5.1, 1.8 and 0.5kb in size, the smallest one representing two overlapping fragments. Of these products, only the largest one hybridised to the *Nat2*-specific probe. It was, therefore, determined that one *Acc*I site is situated approximately 0.5kb downstream of the single *Eco*RV site. A second *Acc*I site was placed 0.5kb downstream of this site.

Digestion with *Dra*I produced at least five fragments, only one of which was positive for *Nat2* (table 4.5; figure 4.7a and b). Double digestions with *Dra*I and *Eco*RV or *Dra*I and *Pst*I, also provided complex digestion patterns, not allowing ordering of the *Dra*I sites on Balb/c clone 2. Therefore, this enzyme was excluded from further analysis. Finally, incubation with either *Sal*I or *Nde*I did not cleave the insert, indicating absence of sites recognised by these two enzymes from Balb/c clone 2 (table 4.5; figure 4.7a and b). This was further confirmed by additional combined digestions with *Bgl*III+*Nde*I, *Bgl*III+*Nde*I+*Sal*I, *Hind*III+*Sal*I or *Acc*I+*Sal*I, which provided digestion patterns similar to those obtained with *Bgl*III, *Hind*III or *Acc*I alone (table 4.5; figure 4.7a and b). The restriction map of the Balb/c clone 2, deduced from the above data, is presented in figure 4.10b.

Table 4.5: Restriction patterns obtained for mapping of the Balb/c clone 2. Single or combined enzyme digestions used for mapping and the produced digestion patterns are presented. Where available, fragments hybridising to the *Nat2*-specific probe are indicated with an asterisk (*). The *Nat2* probe used in hybridisations was designed against the 3' end of the coding region from nucleotide position 516 to 867.

Enzyme	Size of restriction fragments
<i>AccI</i>	5.6kb*, 1.8kb, 0.5kb
<i>BglII</i>	2.2kb, 1.7kb, 3.5kb*, 0.6kb
<i>EcoRV</i>	5.1kb*, 2.9kb
<i>HindIII</i>	7.0kb*, 1.0kb
<i>NdeI</i>	8.0kb*
<i>PstI</i>	2.6kb, 1.4kb*, 4.0kb*
<i>SalI</i>	8.0kb*
<i>DraI</i>	2.5kb*, 1.8kb, 1.2kb, 1.1kb, 0.8kb
<i>EcoRV</i> + <i>HindIII</i>	5.1kb*, 1.9kb, 1.0kb
<i>BglII</i> + <i>EcoRV</i>	2.2kb, 2.3kb, 1.7kb, 1.2kb*, 0.6kb
<i>PstI</i> + <i>HindIII</i>	3.0kb*, 2.6kb, 1.4kb*, 1.0kb
<i>AccI</i> + <i>EcoRV</i>	5.1kb*, 1.8kb, 0.5kb
<i>BglII</i> + <i>NdeI</i>	2.2kb, 1.7kb, 3.5kb*, 0.6kb
<i>BglII</i> + <i>SalI</i> + <i>NdeI</i>	2.2kb, 1.7kb, 3.5kb*, 0.6kb
<i>HindIII</i> + <i>SalI</i>	7kb*, 1.0kb
<i>SalI</i> + <i>AccI</i>	5.6kb*, 1.8kb, 0.5kb

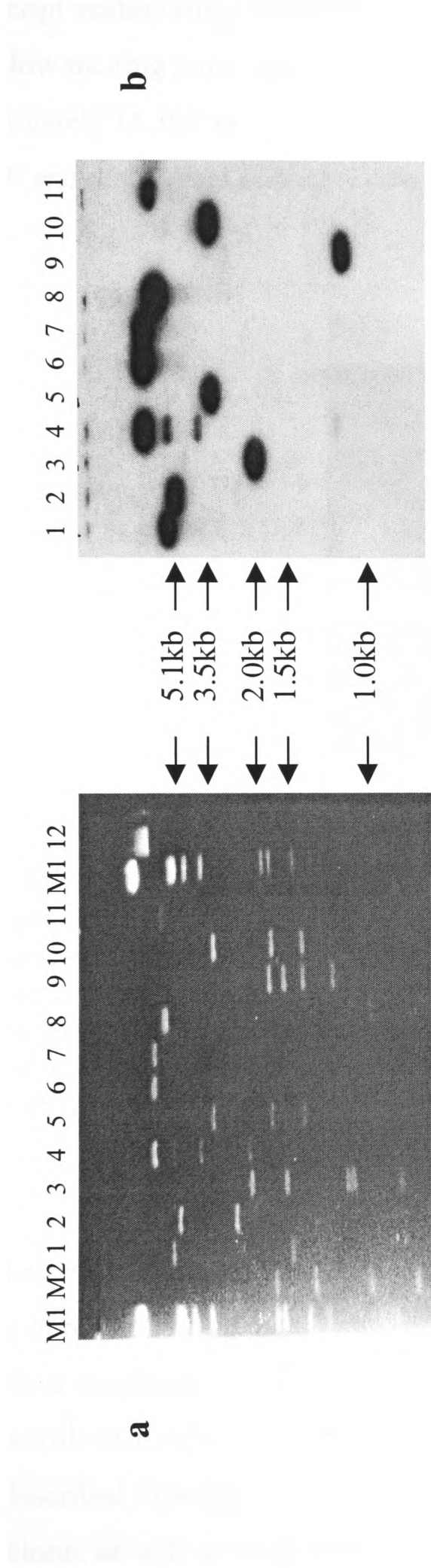


Figure 4.7: Restriction enzyme digestions (a) and Southern blot hybridisation (b) for mapping of the 8kb *Nat2*-positive Balb/c clone 2. Approximately 500ng of insert DNA were digested with 10U of restriction enzyme or with a combination of two enzymes, and the products were separated on a 0.7% agarose gel (a). Following electrophoresis, the gel was blotted and hybridised with a probe specific for the 3' end of the *Nat2* coding region, as described in section 2.2.12.1 (b). Lanes were loaded with digests in the following order: *AccI* (lane 1), *EcoRV* (lane 2), *DraI* (lane 3), *PstI* (lane 4), *BglII* (lane 5), *SalI* (lane 6), *NdeI* (lane 7), *HindIII* (lane 8), *BglII*+*NdeI* (lane 9), *BglII*+*SalI*+*NdeI* (lane 10) and *HindIII*+*SalI* (lane 11). Lane 12 was loaded with 500ng of undigested insert. Lanes M1 and M2 were loaded with 1 µg of λ DNA/*EcoRI*+*HindIII* and 1 µg Biomarker EXT molecular weight markers respectively. The *PstI* and *BglII*+*NdeI* digestions shown here were unsuccessful and were repeated elsewhere (table 4.5 for restriction pattern). Some faint bands of less than 300bp in size may not be visible on the gel picture (a).

4.4.3.3 Mapping of the 129/Ola clone A

For restriction mapping of the 129/Ola clone A, the insert was excised from the pBluescript vector with a combined digestion with *Bam*HI and *Not*I and recovered from a low melting point agarose gel (figure 4.8). The size of the insert was approximately 14.3kb and contained the entire coding region of the *Nat2* gene, as well as the 3' end of the *Nat1* coding region, as demonstrated in section 4.4.1 (table 4.2).

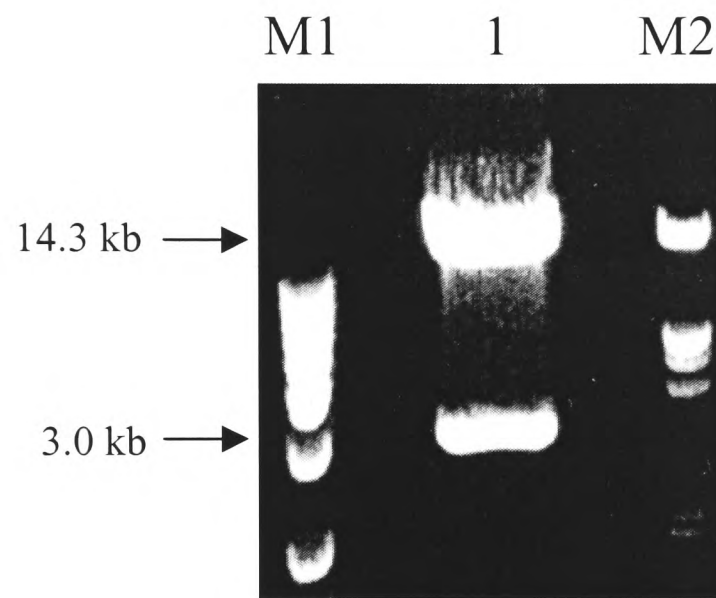


Figure 4.8: Isolation of the insert from the 129/Ola plasmid clone.

Approximately 5µg of plasmid DNA were digested with 25U *Bam*HI and 25U *Not*I. The digest was separated on a 0.7% (w/v) low melting point agarose gel (lane 1), next to 1µg of 1kb ladder (lane M1) and 1µg of λ DNA/*Eco*RI+*Hind*III molecular weight markers (lane M2). The insert band was approximately 14.3kb and the band of the vector was around 3kb. The band of the insert was subsequently excised from the gel and the DNA purified as described in Materials and Methods (section 2.2.6).

As in the case of the two Balb/c clones, the 129/Ola clone A was mapped with a number of restriction endonucleases (table 4.6). Digestions were performed either with one or two enzymes (section 2.2.8), to allow positioning of the restriction sites in relation with each other. Southern blotting and hybridisation using the *Nat2*-specific probe described in section 4.4.3.2 determined the position of the *Nat2* coding region on the clone, as well as its position in relation with the 3' end of the *Nat1* coding region. The preceding mapping of the two Balb/c clones (sections 4.4.3.1 and 4.4.3.2) facilitated the construction of a restriction map for the 129/Ola clone, as it allowed

alignment and comparison between the Balb/c and the 129/Ola maps. The position of restriction sites on clone A was determined according to the digestion and hybridisation patterns described in table 4.6 and shown in figure 4.9a and b. Mapping of the *Bgl*III and *Nde*I sites is described below, as an example.

Digestion with *Nde*I produced three fragments, 0.5, 4 and 9.7kb in size, with only the 9.7kb fragment hybridising to the *Nat2*-specific probe (table 4.6; figure 4.9a and b). Digestion with *Bgl*III produced six fragments, 0.3, 6.5, 2, 1.6, 3.5 and 0.6kb in size (table 4.6; figure 4.9a and b). The probe specific for the 3' end of the *Nat2* coding region hybridised only to the 3.5kb fragment. As it has been established that there is a *Bgl*III restriction site in the *Nat2* coding region, at nucleotide position 534, this 3.4kb fragment should contain the 3' end of the gene. When the same blot was hybridised to a probe specific for the whole coding region of *Nat2*, it became evident that the 5' end of the gene was contained in the 1.6kb *Bgl*III restriction fragment. Another restriction site is present at nucleotide position 750 of the *Nat1* coding region, the 3' end of which is located at the 5' end of clone A (section 4.4.1; table 4.2). This *Bgl*III site could provide either the 0.3 or the 0.6kb restriction fragment observed on the gel. Following comparison with the map of the Balb/c clone 2 (figure 4.10b), it became evident that the 0.6kb fragment should be positioned at the 3' end of clone A. Therefore, the 0.3kb fragment was positioned at the 5' end of the clone. This indicated that clone A should contain part of the *Nat1* coding region starting from at least nucleotide position 450 to the stop codon of the gene. Furthermore, the *Nde*I site located downstream of the *Nat1* coding region on the Balb/c clone 1 (figure 4.10a) could account for the 0.5kb fragment generated by digestion of clone A with *Nde*I (table 4.6) and therefore was also mapped on this clone.

A double digestion of the clone A insert with *Bgl*III and *Nde*I produced fragments of 0.3, 4.0, 2.2, 2.0, 1.6, 3.5 and 0.6kb (table 4.6), with the 1.6kb fragment containing the 5' part and the 3.5kb fragment containing the 3' part of the *Nat2* coding region. As mentioned above, the 0.3 and 0.6kb fragments represent the 5' and the 3' ends of clone A respectively. The 9.7kb *Nde*I fragment was cleaved by *Bgl*III into fragments of 2.2, 2, 1.6, 3.5 and 0.6kb in size, indicating that two more *Bgl*III sites must be located between the second *Nde*I site and the *Nat2* coding region. These would

account for the 2.2, 2 and 1.6kb fragments. The 4.0kb fragment produced by *NdeI* digestion was also generated by combined *BglII* and *NdeI* digestion, therefore, it should not contain any *BglII* sites. The *BglII* and *NdeI* sites were ordered on the map of clone A, as shown in figure 4.10c. Based on the total distance between the *BglII* sites located within the coding regions of *Nat1* and *Nat2*, the distance between the two genes was estimated to approximately 9.4kb. A series of digestions with several other restriction enzymes (table 4.6) completed the restriction map for the 129/Ola clone A (figure 4.10c) and allowed comparison with the two Balb/c clones.

Table 4.6: Restriction patterns obtained for mapping of the 129/Ola clone A. Single or combined enzyme digestions used for mapping and the produced digestion patterns are presented. Where available, fragments hybridising to a probe specific for the 3' end of the *Nat2* coding region are indicated with an asterisk (*) and those hybridising to a probe specific for the whole *Nat2* coding region are indicated with a cross (+). The *Nat2* probe designed against the 3' end of the coding region hybridises to the region from nucleotide position 516 to 867.

Enzyme	Length of restriction fragments
<i>AccI</i>	4.0kb, 1.7kb, 0.6kb (double), 5.7kb*, 1.8kb
<i>BglII</i>	0.3kb, 6.5kb, 2.0kb, 1.6kb ⁺ , 3.5kb* ⁺ , 0.6kb
<i>EcoRV</i>	3.2kb, 0.8kb, 7.5kb*, 2.7kb
<i>HindIII</i>	4.3kb, 9.0kb*, 1.0kb, 0.2kb
<i>NdeI</i>	0.5kb, 4.0kb, 9.7kb*
<i>PstI</i>	2.0kb, 4.4kb, 2.7kb, 1.4kb*, 3.8kb*
<i>SalI</i>	4kb, 10.3kb*
<i>EcoRV</i> + <i>SalI</i>	3.2kb, 0.8kb, 7.5kb*, 2.7kb
<i>NdeI</i> + <i>BglII</i>	0.3kb, 4.0kb, 2.2kb, 2.0kb, 1.6kb ⁺ , 3.5kb* ⁺ , 0.6kb
<i>EcoRV</i> + <i>NdeI</i>	3.2kb, 0.5kb, 2.7kb, 0.8kb 7.0kb*
<i>PstI</i> + <i>HindIII</i>	2.0kb (double), 2.4kb, 2.7kb 1.4kb*, 2.8kb*, 1.0kb
<i>EcoRV</i> + <i>HindIII</i>	3.2kb, 0.8kb, 7.5kb*, 1.4kb, 1.0kb
<i>EcoRI</i>	14.3kb*
<i>BamHI</i>	14.3kb*
<i>NotI</i>	14.3kb*

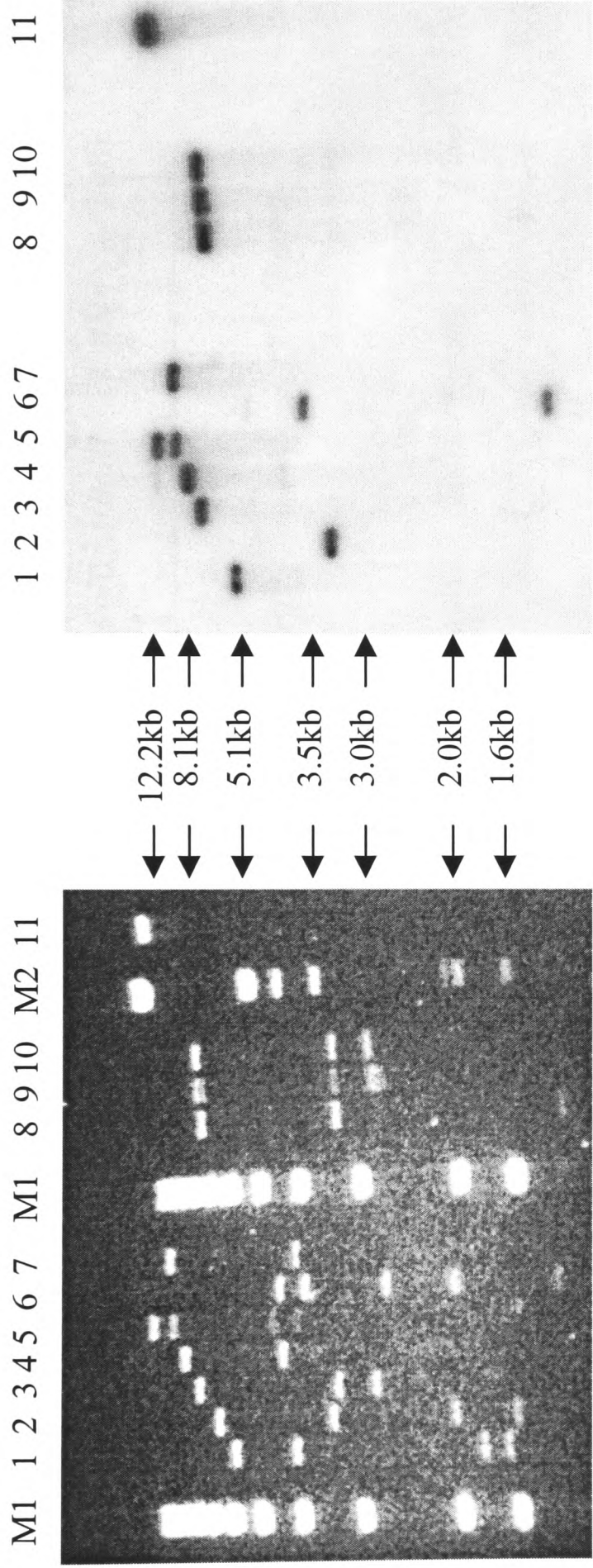


Figure 4.9: Restriction enzyme digestions (a) and Southern blot hybridisation (b) for mapping of the 14.3kb 129/Ola clone A.

Approximately 500ng of insert DNA were digested with 10U of restriction enzyme or with a combination of two enzymes, and the products were separated on a 0.7% agarose gel (a). Following electrophoresis, the gel was blotted and hybridised with a probe specific for the 3' end of the *Nat2* coding region, as described in section 2.2.12.1 (b).

Lanes were loaded with digests in the following order: *AccI* (lane 1), *BglIII* (lane 2), *EcoRV* (lane 3), *HindIII* (lane 4), *NdeI* (lane 5), *PstI* (lane 6), *SaII* (lane 7), *EcoRV+HindIII* (lane 8), *EcoRV+NdeI* (lane 9) and *EcoRV+SaII* (lane 10). Lane 11 was loaded with 500ng of undigested insert. Lanes M1 and M2 were loaded with 1µg of 1kb ladder and λ DNA/*EcoRI+HindIII* molecular weight markers respectively. In lane 5, *NdeI* digestion was not complete, with undigested 14.3kb product and a partially digested fragment of 4.5kb (4.0kb+0.5kb) visible on the gel. Very faint bands of less than 500bp in size may not be visible on the gel picture (a).

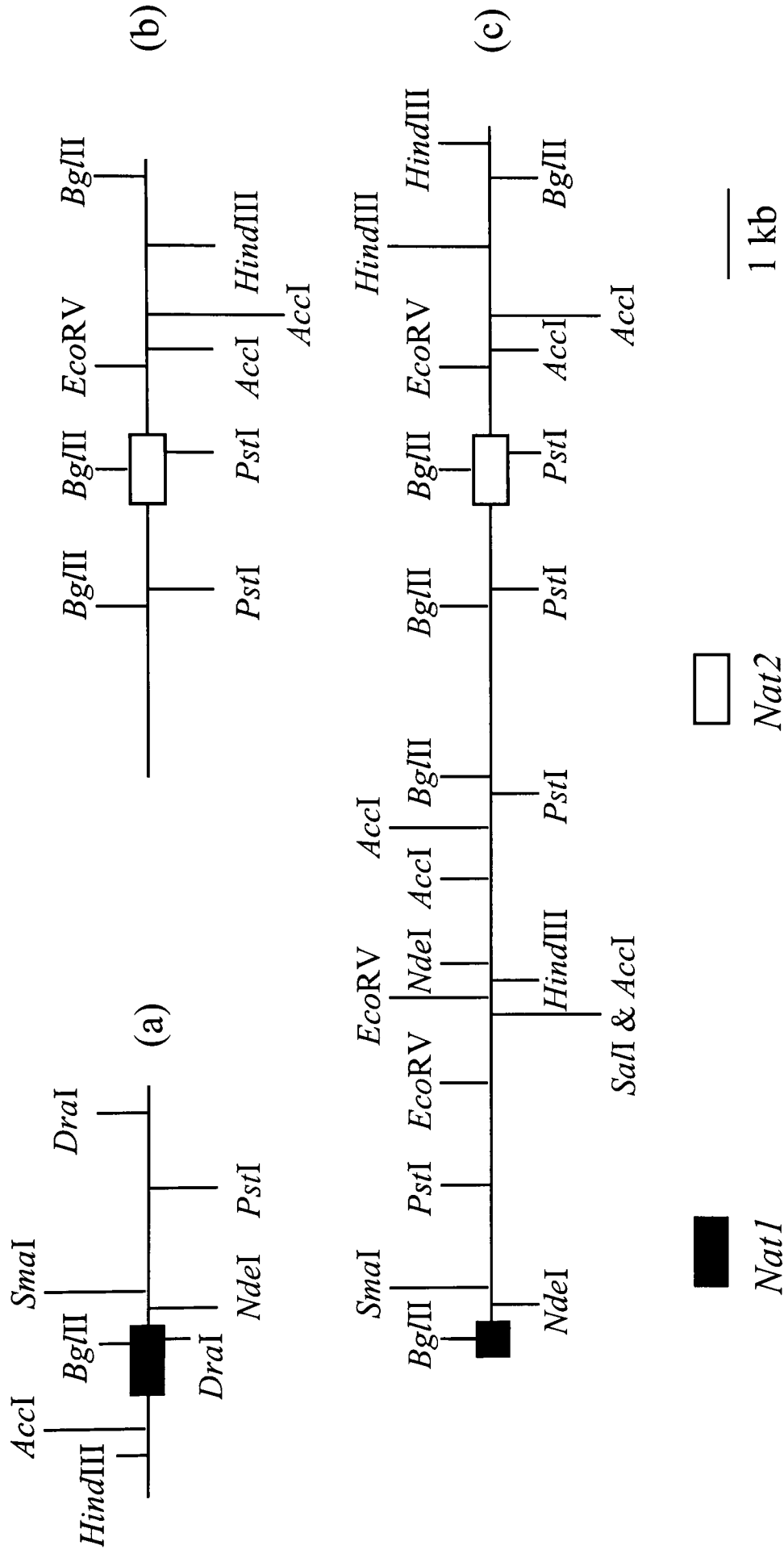


Figure 4.10: Restriction maps of the three mouse genomic plasmid clones.

Analysis of the 4.8kb *Nat1*-positive Balb/c clone 1 (a), the 8kb *Nat2*-positive Balb/c clone 2 (b) and the 14.3kb 129/Ola clone A, carrying the entire coding region of *Nat2* and the 3' end of *Nat1* (c) was performed using a series of restriction enzymes, as described in sections 4.4.3.1 to 4.4.3.3. The order and distances between the restriction sites was determined for each clone and the generated restriction maps were aligned with each other for comparison. The coding region of the *Nat1* and *Nat2* genes is also indicated in the 5'→3' orientation. No polymorphisms between the two strains (Balb/c and 129/Ola) were detected for the enzymes used in mapping.

4.5 Discussion

The work presented here aimed to provide information about the region of the *Nat1* and *Nat2* genes in different mouse strains, in the form of a refined restriction map. Three *Nat*-positive mouse genomic plasmid clones of different strain origin were mapped with a number of restriction enzymes and analysed for possible RFLPs. This work opens the way towards genetic manipulation of the mouse *Nat* genes and production of knockout and transgenic strains.

Restriction mapping is a suitable method for the characterisation of cloned genomic fragments, as it provides information about the physical distance between different loci and detects polymorphisms altering the recognition site of restriction enzymes. It also allows isolation, sub-cloning and modification of DNA sequences of interest. Based on the number and size of the fragments obtained from single and combined enzyme digestions, it has been possible to construct accurate restriction maps for the three plasmid clones (figure 4.10). The three clones together cover a genomic region of about 16kb, from the 5' end of the Balb/c clone 1 to the 3' end of the 129/Ola clone A. Although the Balb/c clones 1 and 2 do not overlap, it has been possible to bridge the gap between them by alignment with clone A, which is the longest of the three plasmid clones. The gap between the two Balb/c clones was estimated to nearly 3.4kb and was demonstrated to be rich in restriction sites.

The *Nat1* and *Nat2* genes were both localised in this short genomic region, with the beginning of *Nat1* estimated to be 1kb and the beginning of *Nat2* coding region nearly 11.2kb from the 5' end of the analysed 16kb region. The distance between the *Nat1* and *Nat2* coding regions was estimated to be around 9.4kb. A previous study has also detected the *Nat1* and *Nat2* genes on the same 16kb clone, derived from the C57Bl/6J mouse strain (Martell et al, 1991). Using the position of the *Nat* genes and some of the restriction sites as anchor points, it has been possible to align the maps presented here for the Balb/c and 129/Ola strains with the map previously constructed for the C57Bl/6J strain (Martell et al, 1991). Comparison of the three strains did not reveal any polymorphisms in this region for the restriction enzymes used, so the only established RFLP in the region remains the *Tru9I* polymorphism within the *Nat2*

gene, associated with phenotypic variation in murine NAT2 acetylation activity. Several inbred mouse strains have been characterised as fast or slow acetylators (Glowinski and Weber, 1982a; Martell et al, 1991; Kelly and Sim, 1994; Fretland et al, 1997; Payton et al, 1999b). Therefore, the *Nat2* polymorphism could be a useful SNP associated with a well-characterised phenotype and could serve as a marker for the chromosomal region harbouring the *Nat* locus in mice.

The work presented here has been the basis for a number of ongoing projects investigating the mechanisms that regulate the expression of the *Nat* genes, as well as the possible role of NAT in development and carcinogenesis. The restriction map described here for the 129/Ola clone A has been the tool for generating a construct containing the inactivated *Nat2* gene, currently being used for the production of knockout strains (Boukouvala et al, 1999). The 9kb *Hind*III fragment containing the *Nat2* coding region (figure 4.10) was isolated and subcloned into pBluescript vector. It was then used to interrupt the *Nat2* gene by insertion of an ablation cassette at the *Bgl*II site of the coding region, as well as to provide flanking arms of homology for targeted replacement of the endogenous *Nat2* locus, via homologous recombination. Restriction sites flanking the 9kb *Hind*III fragment on clone A were also used to generate probes for screening the transfected ES cells for successful incorporation of the knockout construct.

A parallel project, attempting over-expression of the human *NAT1* gene in *Nat2* knockout mice, is underway (Dr K. Pinter, Department of Pharmacology, University of Oxford) and has also been based on the mapping data presented here for the 129/Ola clone A. The coding region of *NAT1* was cloned adjacent to a strong transcription promoter and then inserted in the previously knocked out *Nat2* coding region. The knockout and over-expression projects are currently in progress and aim to investigate the potential role of mouse NAT2 and its functional equivalent human NAT1 (Stanley et al, 1997; Payton et al, 1999b) in early embryonic development. These transgenic projects are designed to investigate the possible involvement of NAT in the endogenous metabolism of folate (section 1.4.1.3).

The map of the 14.3kb long clone A has also allowed sequencing of the region around the *Nat1* and *Nat2* genes in mice. A number of restriction fragments were sub-cloned

and sequenced from their ends. These pieces of known sequence were then used as starting points for sequencing the rest of clone A, using a primer walking method (Fakis et al, 2000; EMBL accession number AJ250123). Sequencing of an extended region around *Nat2* opened the way towards detailed analysis of its genomic structure. Similarly to other eukaryotic genes for NAT (Vatsis et al, 1995), the coding region of mouse *Nat2* is intronless (Martell et al, 1991; Kelly and Sim, 1994). This gene structure is the rule for prokaryotes and the yeast, but it is very uncommon in higher eukaryotes (reviewed in Csordas, 1989). Computational analysis of the sequencing data generated for clone A revealed the presence of a non-coding exon approximately 6kb upstream of the *Nat2* coding region (Fakis et al, 2000). Human *NAT2* is also known to have a non-coding exon (Ohsako and Deguchi, 1990; Ebisawa and Deguchi, 1991). Therefore, it is possible that mouse *Nat2* is genetically more related to the human *NAT2* gene, even though its encoded polypeptide product is functionally equivalent to that of human *NAT1*. Recent work, based on RNase protection assays, has indicated the presence of a putative promoter located within 300bp upstream of the *Nat2* non-coding exon (Boukouvala and Sim, 2000). This might be associated with alternative splicing at the 5' end of the *Nat2* transcript and regulation of expression by more than one promoter.

Gene knockout and over-expression studies, similar to those described for the *Nat2* gene, would also be useful for the *Nat1* and *Nat3* genes. The Balb/c clone 1 that carries the coding region of *Nat1* will be useful for generation of sequencing information to facilitate genetic manipulation of the gene and identification of elements regulating its transcription. However, isolation of longer genomic clones carrying *Nat1* might be necessary for that purpose. Similar studies would also be of particular significance for the *Nat3* gene, as they would provide useful information about the possible role of its product, for which very little is known. A Balb/c *Nat3*-positive genomic clone has previously been described and a partial restriction map has been constructed (Kelly and Sim, 1994), but this clone was not available for analysis in the present study. Therefore, isolation of additional *Nat*-positive clones would be essential before attempting genetic modification of the other two members of the *Nat* cluster in mice. For that purpose, mouse PAC genomic clones were isolated and characterised for their *Nat* content, as will be described in the following chapter.

Chapter 5: Chromosomal Localisation of the *Nat* Genes in the Mouse

5.1 Introduction

Sequencing and mapping of the genome of model organisms has become a major aim of the human genome project (HGP) (Collins and Galas, 1993). According to the report on the aims of the second period of the HGP (1998-2003), more than 12,000 STSs had been mapped on the mouse genome by 1998 (Collins et al, 1998). Since then, extensive genomic resources have become available for the mouse. These include high resolution microsatellite (Rhodes et al, 1998) and radiation hybrid (Van Etten et al, 1999) maps, as well as an extensive YAC-based physical map providing the position of around 10,000 loci with an average spacing of approximately 300kb between them (Nusbaum et al, 1999). Mapping data are freely available via the Mouse Genome Database (MGD; <http://www.informatics.jax.org>), which is updated regularly. An encyclopaedia of mouse genes was recently released (Marra et al, 1999), comprising more than 350,000 ESTs. These correspond to 93% of the total number of ESTs available for the mouse at the dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST>). ESTs can be very useful for the identification of novel genes and the development of microarray systems, as well as for comparative sequence analysis and mapping. An action plan, recently published by the National Institute of Health (NIH) in USA (Battey et al, 1999), has ensured adequate support of the Mouse Genome Project (MGP), with the aim of producing a 'working draft' sequence of the mouse genome by 2003, as originally scheduled (Collins et al, 1998).

Large-scale mapping projects, like those described above, provide tools for accelerated discovery and localisation of genes of interest. High resolution genetic maps (e.g. Rhodes et al, 1998) have been extremely useful to researchers using linkage analysis to localise specific genes, as well as for building physical maps of genomic regions. Physical mapping is greatly facilitated by the availability of markers at high density within the genome (Wicking and Williamson, 1991), while established contigs (e.g. Nusbaum et al, 1999) covering a region of interest can be directly screened for determining the position of specific genes.

Genome sequencing projects have been of great importance for the discovery of novel genes, as the generated sequence information can be used to predict genes or elements associated with genes (e.g. transcription control elements and CpG islands) by computational analysis. This can greatly facilitate screening of a cloned region for candidate genes, a procedure exclusively relying in the past upon laborious techniques such as 'zoo blots', identification of CpG islands, cDNA selection and exon trapping (reviewed in Wicking and Williamson, 1991; Collins, 1992; Lovett, 1994). Analysis of the candidate genes for mutations possibly associated with the affected phenotype has currently been carried out by various methods, including SSCP, DGGE, heteroduplex analysis etc. As more and more sequencing information becomes available from different individuals (humans) or strains (model animals), identification of mutations will become possible by sequencing analysis (Kwok et al, 1996).

Cloning and sequencing of the murine *Nat* genes (Martell et al, 1991; Kelly and Sim, 1994) was carried out as described in section 4.1. However, little is known about their physical position in the mouse genome. An early study (Mattano et al, 1988) demonstrated linkage between *Nat2* and the *esterase 1* locus (*Es1*), previously mapped at 43cM from the centromere of mouse chromosome 8 (MGI:95420). The recombination frequency between the two loci was estimated to be $12 \pm 2.8\%$ (mean $\pm$ standard deviation), placing them about 12cM apart. These results indicated that the *Nat2* locus should be located either at 31.0 or 55.0cM from the centromere of mouse chromosome 8. In the former case, the *Nat2* locus would be located on the centromeric side of *Es1* and close to the *hook* (*Hk*) gene, previously mapped at 26cM

(MGI:96102). In the latter case, it would be located on the telomeric side of *Es1* and close to the *esterase 5* (*Es5*) gene, mapped at 52.2cM (MGI:95444). However, linkage of *Nat2* with the *endogenous ecotropic MuLV2* (*Emv2*) locus, also located on chromosome 8 at 67cM from the centromere (MGI:95335), was not as clear as in the case of *Es1*, indicating either independent segregation or very loose linkage between the *Nat2* and *Emv2* loci (Mattano et al, 1988). Therefore, the *Nat2* gene was placed at position 31cM, although this assignment was approximate.

The mouse genome is the most extensively studied of all non-human vertebrate genomes, therefore, it is most appropriate for conducting comparative mapping studies (Carver and Stubbs, 1997). Although rodent chromosomes have undergone a surprisingly high number of chromosomal rearrangements per unit of evolutionary time, compared with other mammalian species, more than 200 conserved linkage groups have been identified between the mouse and human (Carver and Stubbs, 1997; O'Brien et al, 1997; Waddington et al, 2000). Moreover, according to the Oxford Grid (www.informatics.jax.org/searches/oxfordgrid.cgi), about 4500 genes have been identified as homologous between humans and mice to date, of which more than 3000 have been mapped in both species. A significant proportion of these homologous genes have been associated with genetic disease in humans (Andersson et al, 1996).

Apart from its evolutionary interest, comparative mapping (meaning the chromosomal localisation of homologous genes across different species) has significantly contributed to the progress of HGP, as it can be a powerful approach for identification and mapping of novel disease genes, especially those with major effects on quantitative traits. Mutation of the mouse homologues of human genes may also allow a deeper insight into the function of these genes *in vivo* (Andersson et al, 1996). Comparative sequencing between mice and humans has been encouraged as part of the HGP, with the aim of identifying homologies between the two species in the form of individual genes, regulatory elements or evolutionary conserved genomic regions (Mallon and Strivens, 1998; NIH News Release, 1999).

According to the HUGO Committee on Comparative Mapping, homology should be established based on criteria such as high nucleotide sequence similarity of the genes

and amino acid sequence similarity of their products, conserved chromosomal position and coincident tissue or developmental expression. Similar function, antigenic properties, subcellular location, substrate specificity and subunit structure of the protein products of the genes are additional criteria which should be taken into account where possible (Andersson et al, 1996; MGD Mammalian Homology Criteria, www.informatics.jax.org). Most of these criteria are applicable to the murine and human genes for NAT, so that they can be considered as homologous. However, assignment of a mouse *Nat* orthologue for each one of the human *NAT* genes is difficult and the exact evolutionary relationship between the human and murine genes for NAT remains unclear. Generation of cytogenetic, genetic and physical mapping data for murine *Nat* genes, combined with the mapping information already available for their human homologues (Hickman et al, 1994; Matas et al, 1997), would be particularly useful for studying the evolution of this conserved group of genes in mammals.

5.2 Materials and study design

Although linkage has been reported between *Nat2* and the gene for esterase 1 (*Es1*), previously mapped on chromosome 8 (Mattano et al, 1988), the exact localisation of the *Nat* genes has remained unclear. The aim of this work has been to determine the cytogenetic position of the murine *Nat* genes and generate tools to allow physical mapping of the surrounding region, expanding the work described in the previous chapter (Chapter 4) for the three *Nat*-positive plasmid clones. Localisation of the *Nat3* gene was of particular interest, as no previous information was available about its chromosomal position and its position relative to the other two *Nat* genes in mice.

A mouse PAC genomic library was screened by hybridisation with *Nat*-specific probes (section 5.3.1) and the identified positive clones were subsequently used as probes for cytogenetic mapping of the *Nat* loci by FISH analysis (section 5.3.3). The same *Nat*-positive PAC clones were also used to carry out some preliminary physical mapping of the region (sections 5.3.4 and 5.3.5). Screening of a number of YAC contigs spanning the chromosomal region, where the *Nat* genes have been suspected

to localise (Mattano et al, 1988), failed to demonstrate co-localisation with previously established markers (section 5.3.6). To overcome this problem, a number of *Nat*-positive YAC genomic clones were isolated by screening of a library (section 5.3.7), for future analysis of this genomic region of interest.

5.2.1 The RPCI 21 mouse PAC genomic library

The RPCI 21 mouse PAC genomic library was created by K. Osoegawa and P.J. de Jong at the Roswell Park Cancer Institute, Buffalo, New York, USA (<http://bacpac.med.buffalo.edu>) and was obtained from the HGMP Resource Centre, Cambridge, UK (<http://www.hgmp.mrc.ac.uk>). The library was constructed using the pPAC4 cloning vector and genomic DNA prepared from the spleen of a female 129/SvEvTACfBr mouse. According to its creators, the library has a 5-fold genomic coverage, a total number of approximately 130,000 clones and an average insert size of 147kb. The library was available in the form of seven high-density gridded filters, with each individual clone spotted twice on the filter. Positive clones were identified by a single hybridisation step and PCR amplification was then used to confirm the *Nat* content of each one of these clones (section 5.3.1).

5.2.2 The WI/MIT mouse YAC genomic library

The WI/MIT mouse YAC genomic library was created at the Whitehead Institute/MIT Center for Genome Research, Massachusetts, USA, using genomic DNA from female C57Bl/6J mice (Haldi et al, 1996). It consists of approximately 40,000 YAC clones, with an average insert size of 820kb, providing approximately 10-fold coverage of the entire mouse genome. The library was obtained from the HGMP Resource Centre, Cambridge, UK (<http://www.hgmp.mrc.ac.uk>) in the form of YAC DNA pools in agarose plugs, for screening by PCR. The agarose plugs were saturated in 5mM EDTA and were washed twice for one hour and once overnight, in TE buffer at 4°C, to remove components that might inhibit PCR. According to the pooling system, positive clones were identified following two rounds of screening. Primary screening involved specific amplification from fifty primary pools (designated MY1 to MY50), each consisting of eight plate pools. Secondary screening involved specific

amplification from the positive plate pools, and their corresponding row and column pools, so that the plate containing the positive clone and the co-ordinates of that clone on the plate could be determined. The *Nat* content of each one of the identified positive clones was further confirmed by PCR, directly from the culture of each yeast clone.

5.2.3 Chromosome 8 YAC contigs

A subset of approximately 24,000 clones from the WI/MIT mouse YAC genomic library (section 5.2.2) is currently being used by the Whitehead Institute/MIT Centre for Genome Research for physical mapping of the mouse genome (Nusbaum et al, 1999). As a result of this ongoing project, a number of YAC contigs have been constructed across many regions of murine chromosome 8, including part of the 31cM region, where the *Nat* genes have been suspected to be localised (Mattano et al, 1988). The clones of each contig have been anchored to the mouse WI/MIT genetic (linkage) map, based on a series of polymorphic microsatellite markers, identifiable by PCR amplification (<http://www.informatics.jax.org>). The clones comprising these contigs were obtained from the UK HGMP Resource Centre (<http://www.hgmp.mrc.ac.uk>).

Four YAC contigs spanning the region 29.5-32.8cM on mouse chromosome 8 are currently available (figure 5.1 and table 5.1). WC8.14 is the largest of those contigs, covering the region 29.5-31.7cM, and consisting of thirty-eight clones. Contig WC8.16 covers a small region at 31.7cM from the centromere and consists of only six clones. The remaining two contigs cover the 32.8cM region, with WC8.17 consisting of fourteen clones and WC8.18 of twenty-two. All four contigs are singly-linked, which means that the markers have been linked to the contigs with at least one YAC. A gap on the physical map still exists in the region between 31.7-32.8cM from the centromere of murine chromosome 8. This 1.1cM interval may correspond to a physical distance of about 2Mb (Reeves, 1999). Sixteen microsatellite polymorphic markers have been used to align the genetic and physical maps of the region, as illustrated in figure 5.1. Additional STSs have been placed on the physical map and used to order the YAC clones of each contig (table 5.1).

5.2.4 D8Mit markers

The D_Mit markers are simple sequence repeat (SSR) polymorphic microsatellite markers, developed for each murine chromosome by the Whitehead Institute/MIT Center for Genome Research. They can be detected by PCR and exhibit sequence length polymorphism between different mouse strains. This type of marker has been extensively used for the construction of linkage and physical maps of the mouse genome (Rhodes et al, 1998; Nusbaum et al, 1999; <http://www.informatics.jax.org>). Amplification of ten D8Mit markers, localised at the 31cM region of mouse chromosome 8, was carried out using pairs of marker-specific primers obtained from Research Genetics. The sequence of the primers is available from MGD, <http://www.informatics.jax.org>.

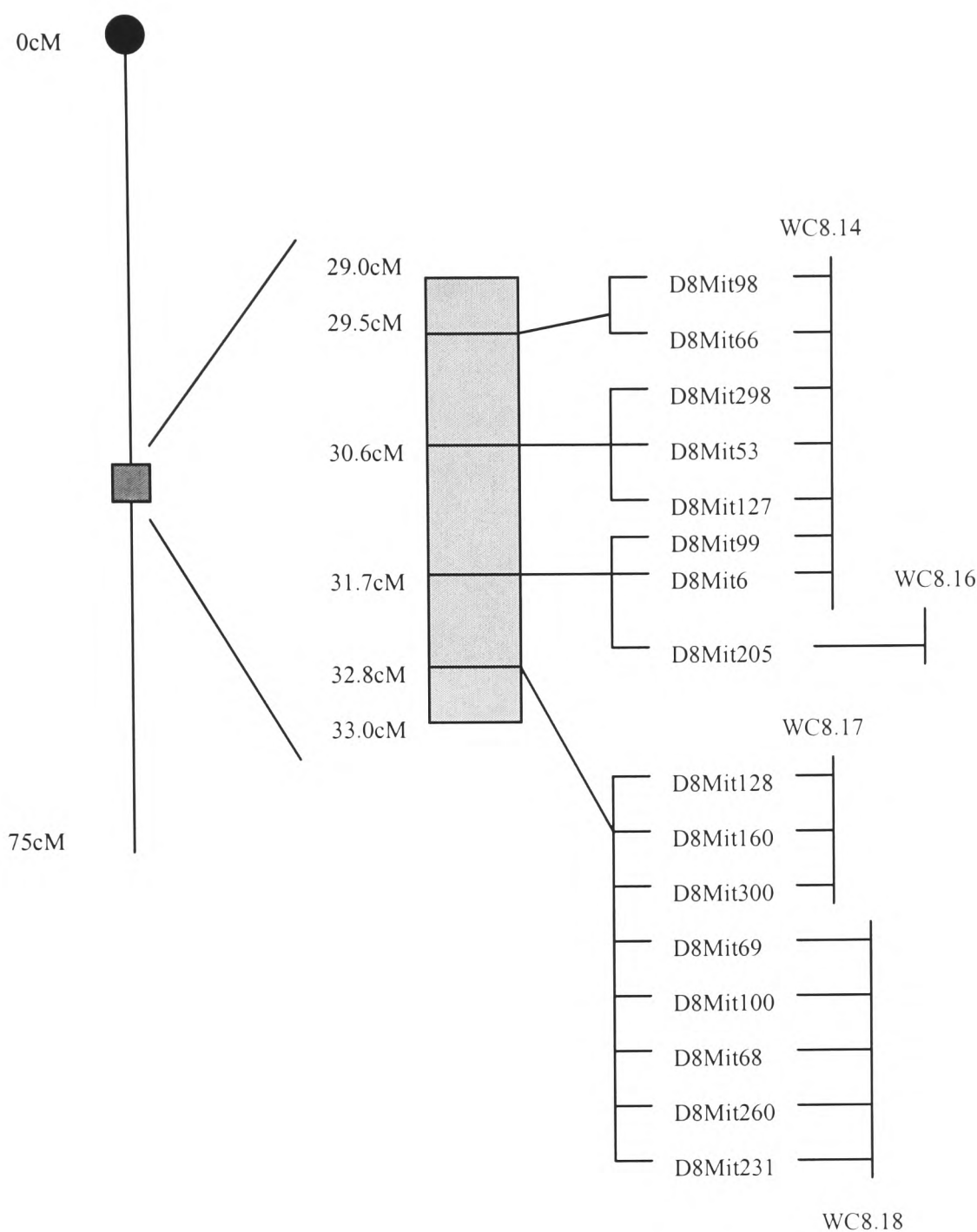


Figure 5.1: Integrated linkage and physical map of the region 29-33cM from the centromere of mouse chromosome 8. A schematic representation of the entire chromosome 8 is shown on the left and the 29-33cM region of interest is indicated with a shaded box. The genetic map of this region is shown in the middle. The four contigs (WC8.14, WC8.16, WC8.17 and WC8.18), spanning the 29-33cM region, are displayed on the right. The polymorphic microsatellite markers (SSRs), contained in each contig, are also indicated (adapted from MGD, <http://www.informatics.jax.org>). There is continuous coverage of the 29.5-32.8cM region by the four YAC contigs, except for a 1.1cM gap (from 31.7cM to 32.8cM) between contigs WC8.16 and WC8.17.

Table 5.1: Markers contained in the four YAC contigs spanning the 29.5-32.8cM region of murine chromosome 8.

Most of the markers are SSR polymorphic microsatellite markers (D8Mit) that have been developed by the Whitehead Institute/MIT Center for Genome Research, but STSs and other random sequences have also been mapped to these contigs. Markers are listed according to their order on the contigs from centromere to telomere of murine chromosome 8. Details about the markers listed below are available from the MGD (<http://www.informatics.jax.org>).

WC8.14	WC8.16	WC8.17	WC8.18
D8Mit98 J03168 MHAa97c12.seq 47.MMHAP9FRH11.seq MDB0589 31.MMHAP81FLC4.seq Results_ER_8_28_95_ 2_23.seq D8Mit66 D8Mit298 D8Mit53 D8Mit127 D8Mit99 D8Mit6 12.MMHAP61FRD9.seq	D8Mit205 0.2MMHAP35FLA2.seq	D8Mit128 D8Mit160 28.MMHAP96FLB4.seq MHAa8f10.seq D8Mit300 11.MHAa49a11.seq	45.MMHAP84FLG6.seq 32.MHAa79T.seq D8Mit69 D8Mit100 D8Mit68 04.MMHAP24FLB1.seq D8Mit260 D8Mit231 12.MMHAP28FRD9.seq

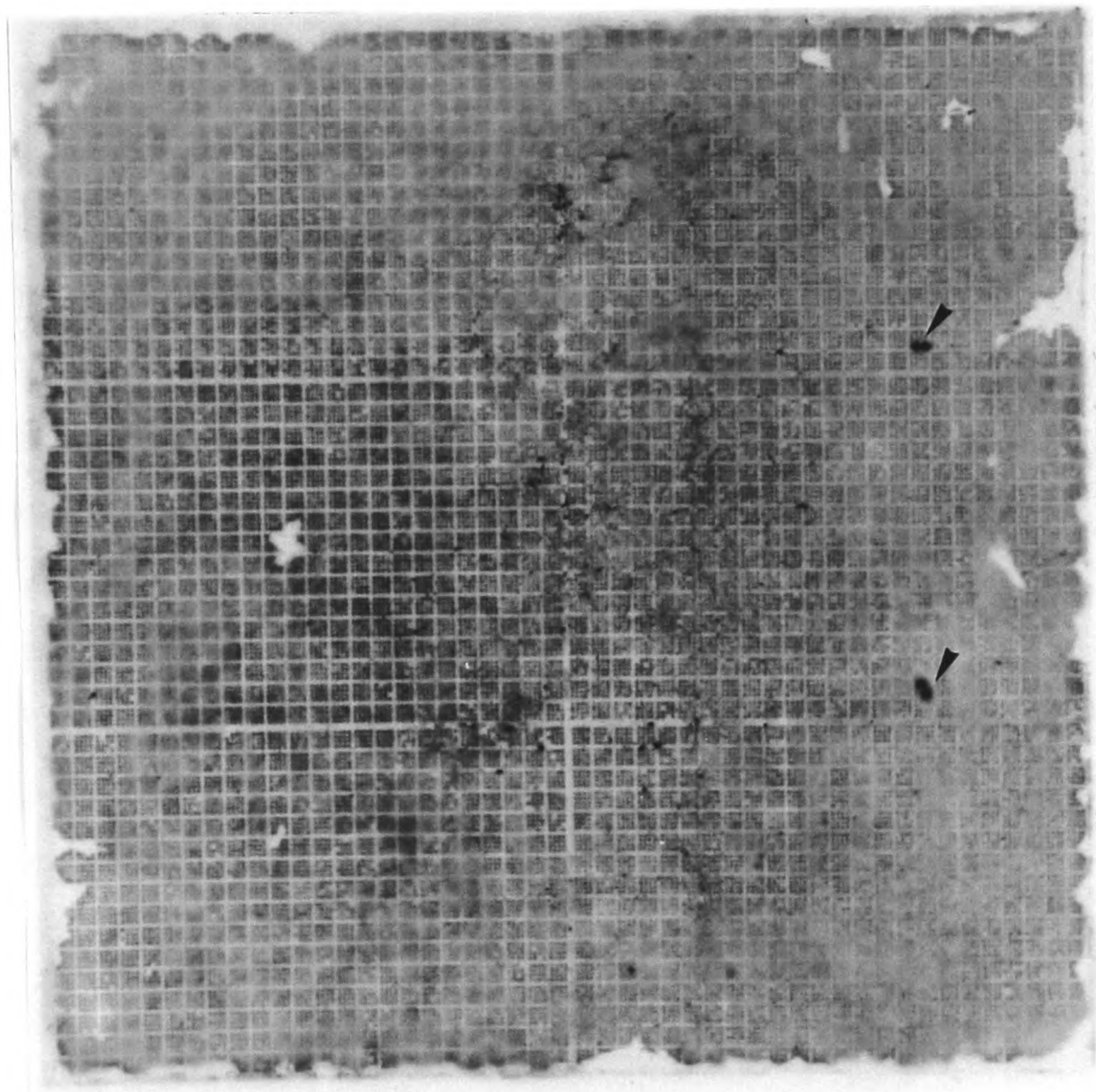
5.3 Results

5.3.1 Screening of the RPCI 21 mouse PAC genomic library for *Nat* genes

The RPCI 21 mouse PAC genomic library (section 5.2.1) was screened by hybridisation with a radioactively labelled probe (section 2.2.12.1). Probes specific for the entire coding region of each one of the three murine *Nat* genes were generated by PCR with primers Mus12/Mus13 for *Nat1*, mNAT2 1/mNAT2 910 for *Nat2* and Mus12/Mus15 for *Nat3* (table 2.1), and combined together in a mixture. This allowed identification of PAC clones positive for any of the three *Nat* genes by a single hybridisation step. Following screening of the seven filters of the library, a total of six clones (362-G17, 393-K20, 489-O17, 584-N8, 587-K16 and 511-O17) were identified as positive for *Nat* (figure 5.2).

Figure 5.2: Screening of the RPCI 21 mouse PAC genomic DNA library by hybridisation.

Each of the seven filters of the library was hybridised to a mixture of radioactively labelled probes, specific for the coding region of the three *Nat* genes. This facilitated screening, as all clones positive for at least one of the *Nat* genes were identified following a single hybridisation step. Each filter of this high-density gridded library carried approximately 18,000 PAC clones, each one of them spotted in duplicate. The two positive signals detected on this filter (arrows) were determined to correspond to clones 489-O17 and 511-O17.



The *Nat* content of the clones, identified as *Nat*-positive by hybridisation, was subsequently established by PCR amplification specific for each one of the three *Nat* genes. Primers Mus12/Mus 13, mNAT2 1/mNAT2 910 and Mus12/Mus15 were used to specifically amplify *Nat1*, *Nat2* and *Nat3*, respectively (table 2.1). Of the six positive clones obtained following primary screening, only three (584-N8, 393-K20 and 511-O17) were confirmed to contain at least one of the *Nat* genes. *Nat1* and *Nat2* were found together on clones 584-N8 and 511-O17 (figure 5.3a and b), as expected by the fact that they are located only 9.4kb away from each other (section 4.4.3.3). Clone 511-O17 also contained *Nat3*, establishing localisation of the gene in the same chromosomal region as *Nat1* and *Nat2* (figure 5.3a, b and c). Only *Nat3* was found on clone 393-K20 (figure 5.3c), suggesting that it might be located further away than *Nat1* and *Nat2* are from each other. Based on the average insert size of the PAC clones (section 5.2.1), *Nat3* should be located within an average maximum distance of 150kb from the other two *Nat* genes. The *Nat* content of the three positive PAC clones is summarised in table 5.2. Specificity of all amplifications was confirmed by digestion with *Ava*I for *Nat1*, *Pst*I for *Nat2* and *Hae*III for *Nat3*, as described in section 4.4.1. Figure 5.4 shows digestion with *Hae*III of the *Nat3*-specific amplification product from clone 511-O17, providing the expected banding pattern (table 4.3). The generated restriction patterns, following digestion of the *Nat1*-specific amplification products with *Ava*I and the *Nat2*-specific amplification products with *Pst*I, were identical to those shown in figure 4.2 for the Balb/c plasmid clones 1 and 2 (section 4.4.1) and are not presented here.

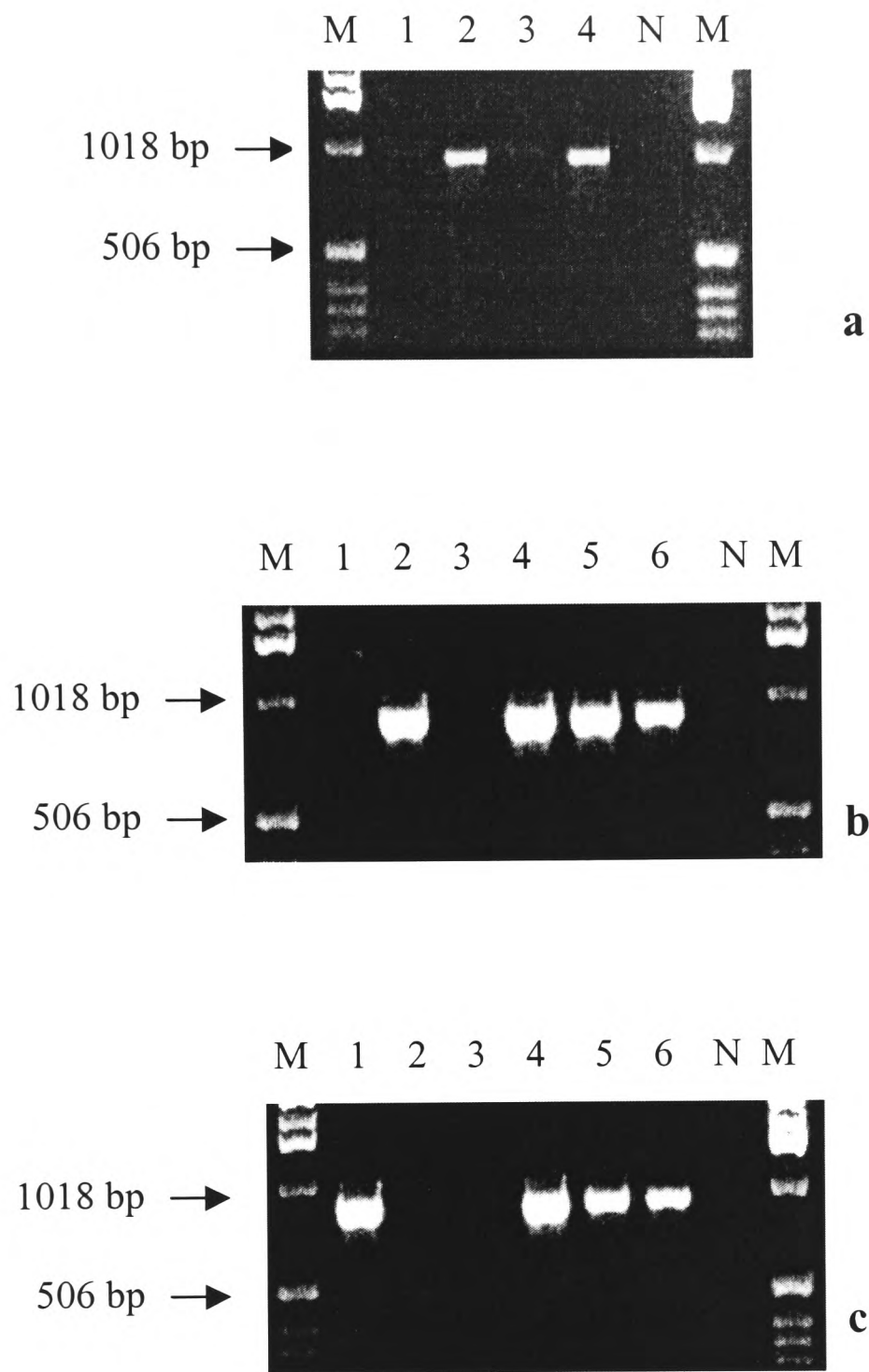


Figure 5.3: Investigation of the *Nat* content of the PAC clones. Specific amplification of the *Nat1* (a), *Nat2* (b) and *Nat3* (c) coding region with primers Mus12/ Mus13, mNAT2 1/ mNAT2 910 and Mus12/ Mus15, respectively (table 2.1). Lanes 1 to 4 were loaded with PCR products from PAC clones 393-K20, 584-N8, 363-P23 and 511-O17. Lanes 5 and 6 were loaded with PCR product from mouse genomic DNA, serving as positive control. Lane N was loaded with the reaction negative control (no template). Lanes M were loaded with 1 μ g of 1kb ladder. Amplification products (10 μ l) were separated on 1.5% agarose gels. Clone 393-K20 (lane 1) was confirmed to contain *Nat3*, clone 584-N8 (lane 2) *Nat1* and *Nat2*, and clone 511-O17 (lane 4) *Nat1*, *Nat2* and *Nat3*. Clone 363-P23 (lane 3) was randomly selected as negative control. *Nat1*-, *Nat2*- and *Nat3*-specific amplification products were 908bp (a), 910bp (b) and 908bp (c) in size, respectively.

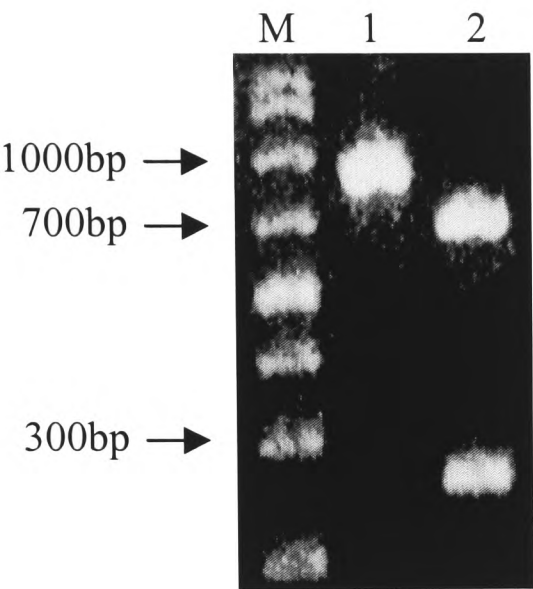


Figure 5.4 Specificity of amplification of the mouse *Nat3* gene. The coding region of *Nat3* was amplified from PAC clone 511-O17 using primers Mus12 and Mus15. The 908bp amplification product was then digested with *Hae*III and the produced fragments were separated on a 1.5% agarose gel. Lanes 1 and 2 were loaded with undigested and *Hae*III-digested *Nat3*-specific amplification product respectively. Lane M was loaded with BioMarker EXT ladder. The expected 657 and 251bp bands were produced, confirming specificity of amplification from the *Nat3* gene.

Table 5.2: Summary of the content of the three *Nat*-positive mouse PAC genomic clones (584-N8, 393-K20 and 511-O17).

<i>Nat</i> content	584-N8 PAC clone	393-K20 PAC clone	511-O17 PAC clone
<i>Nat1</i> coding region	+	—	+
<i>Nat2</i> coding region	+	—	+
<i>Nat3</i> coding region	—	+	+

5.3.2 Genotyping of the *Nat2*-positive mouse PAC genomic clones

Clones 584-N8 and 511-O17, identified as positive for *Nat2* (figure 5.3b and table 5.2), were genotyped as described in section 4.3. A product 131bp in size was amplified from the two clones, using primers mNAT2 278 and mNAT2 409 (table 2.1). Following digestion with *Tru9I*, both PCR products remained undigested, indicating presence of the *Nat2**8 allele (figure 5.5). This demonstrated that the 129/SvEvTACfBr strain is a fast acetylator.

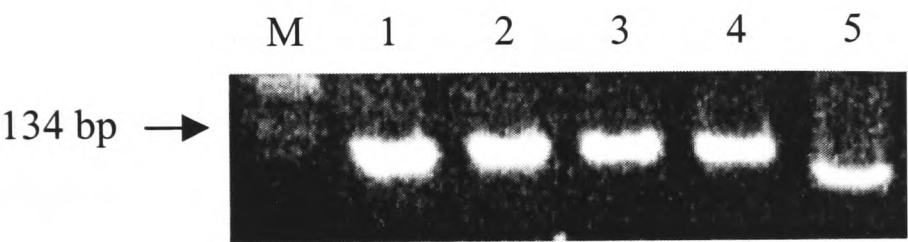
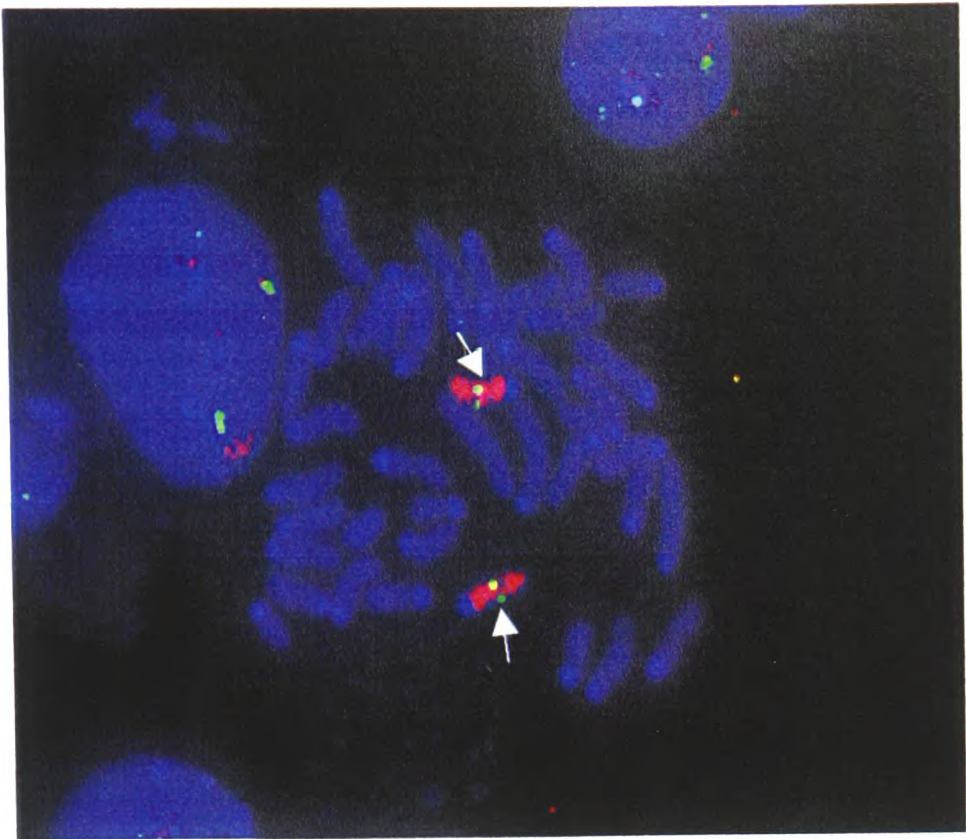


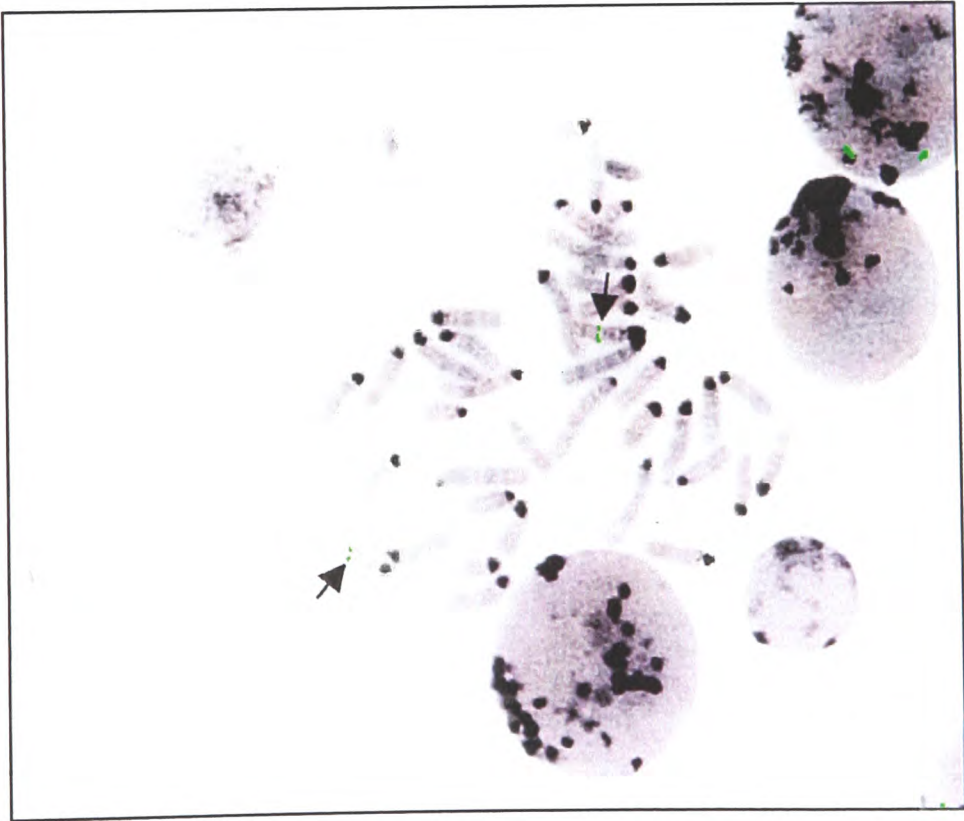
Figure 5.5: Genotyping of the *Nat2*-positive mouse PAC genomic clones. PCR products with primers mNAT2 278 and mNAT2 409 (table 2.1) were digested with *Tru9I* at 65°C for 3 hours and the digests were separated on a 4% metaphor gel. Lane 1 was loaded with undigested control (131bp in size). Lanes 2 and 3 were loaded with digests of the amplification products from clones 584-N8 and 511-O17, respectively. Lanes 4 and 5 were loaded with digests of products amplified from genomic DNA of C57Bl/6J and A/J mouse strain origin. Apart from the product of the A/J, which is known to be a slow acetylating strain, all other products remained undigested indicating presence of the fast *Nat2**8 allele.

5.3.3 Cytogenetic mapping of the mouse *Nat* genes

The three *Nat*-positive PAC genomic clones were used to determine the chromosomal localisation of the *Nat* genes in mice. FISH analysis was carried out in collaboration with Dr Veronica Buckle (Institute of Molecular Medicine, Oxford, UK), using PAC clones 584-N8, 393-K20 and 511-O17 as probes. DNA from each PAC clone was labelled with biotin by nick translation (section 2.3.5) and hybridised to metaphase spreads prepared from mouse spleen cells, as described in section 2.3.6.2. Each PAC clone was used in combination with a mouse chromosome 8-specific paint (section 2.3.6.2), to confirm chromosomal localisation of the *Nat* genes (figure 5.6a, c and e). It was established that all three *Nat* genes are localised on murine chromosome 8. Analysis of the banding pattern of the DAPI-stained mouse metaphase chromosomes, further determined localisation of the three *Nat* genes to bands 8 B3.1-3.3 (figure 5.6b, d and f). FISH analysis also demonstrated that the three PAC clones used as probes are not chimaeric and can, therefore, be considered as reliable tools for subsequent physical mapping of the region surrounding the three *Nat* loci.

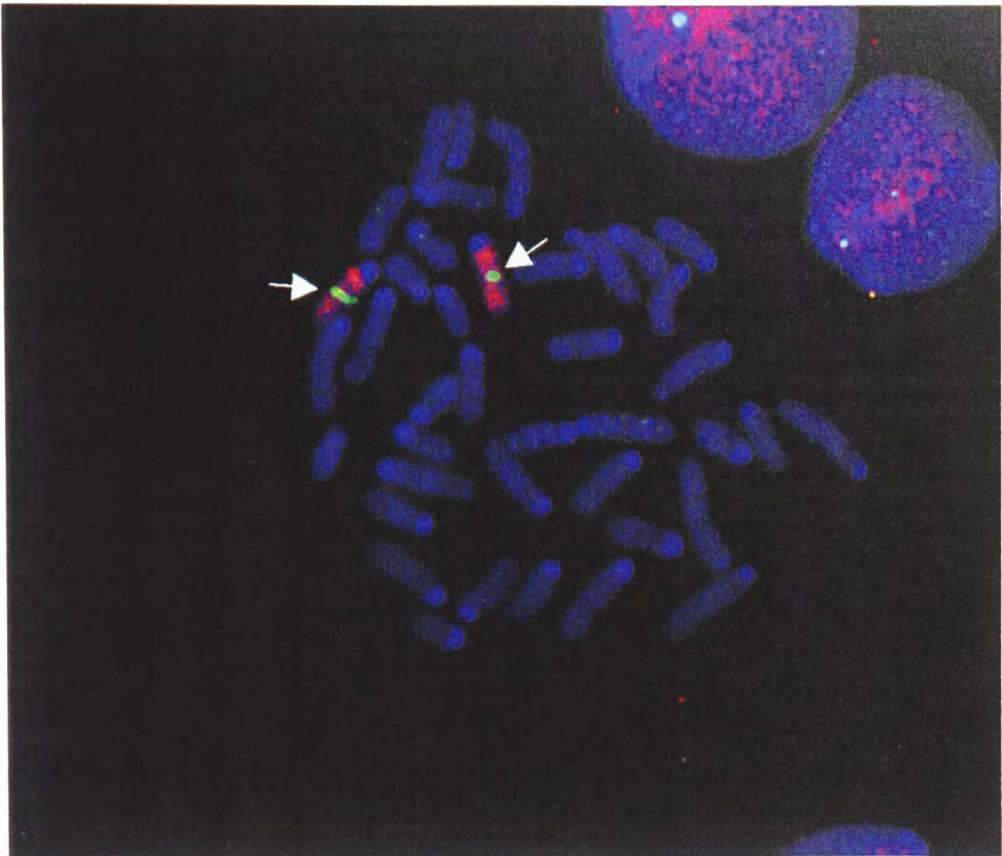


a

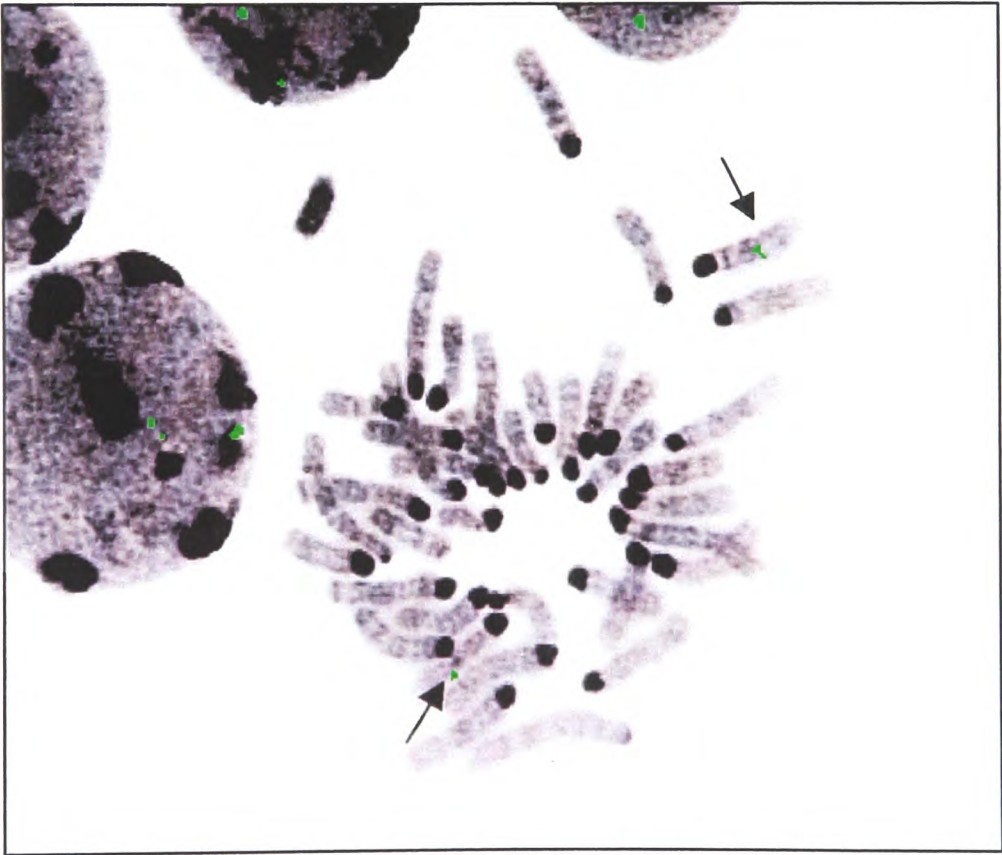


b

c



d



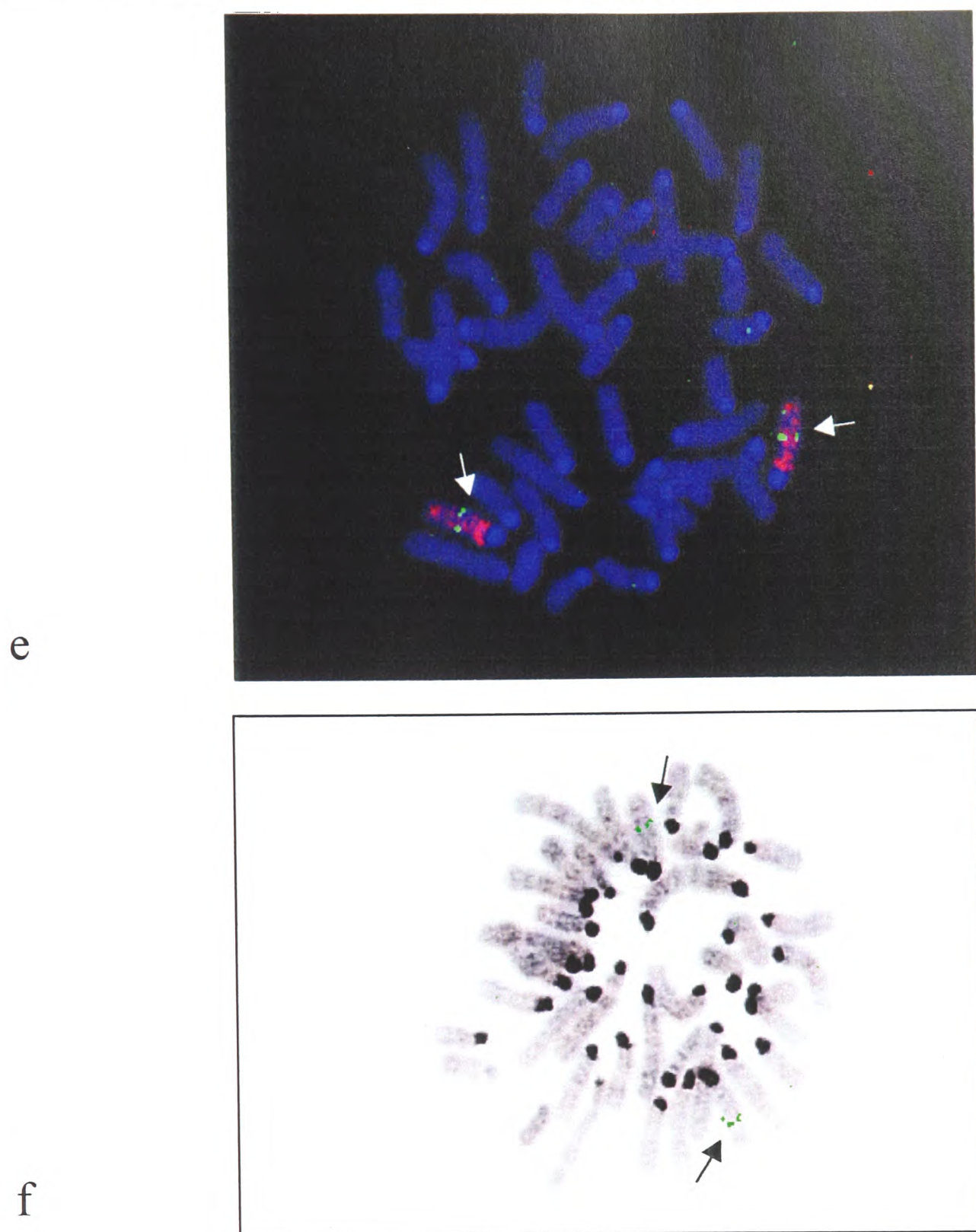


Figure 5.6: Chromosomal localisation of the mouse *Nat* genes. The *Nat*-positive mouse PAC genomic clones 584-N8 (a and b), 393-K20 (c and d) and 511-O17 (e and f) were co-hybridised with a mouse chromosome 8-specific paint to metaphase chromosomes prepared from a culture of healthy spleen cells. The PAC probes were labelled with biotin-16-dUTP by nick translation and the signal was detected with layers of avidin-FITC and biotinylated anti-avidin antibody. The chromosome 8-specific paint was directly labelled with Cy3 fluorochrome. Slides were counter-stained with DAPI in antifade mounting solution. The green FITC *Nat*-specific signal (white arrows) co-localised with the red Cy3 chromosome 8-specific signal, establishing localisation of all three *Nat* genes on mouse chromosome 8 (a, c and e). The PAC probes did not hybridise to any other chromosomes, indicating they were not chimaeric. On the DAPI-banded metaphase chromosomal spreads, the green FITC *Nat*-specific signals (black arrows) were localised to bands 8 B3.1-3.3 (b, d and f).

5.3.4 Restriction analysis of the 511-O17 PAC clone

To expand the information obtained from analysis of the *Nat1*- and *Nat2*-positive mouse plasmid genomic clones (Chapter 4) and to determine the position of the *Nat3* gene relative to *Nat1* and *Nat2*, the 511-O17 mouse PAC genomic clone, positive for all three *Nat* genes, was analysed by restriction digestion. The mapping (Chapter 4) and sequencing (EMBL accession number AJ250123) information previously generated for the region between *Nat1* and *Nat2* was used to identify restriction enzymes separating the two genes. Of the six infrequent cutters identified to cleave between the *Nat1* and *Nat2* coding region (*Cla*I, *Kpn*I, *Nae*I, *Sac*I, *Sal*I and *Sma*I), *Kpn*I was the most suitable because it also cleaved within the coding region of *Nat3* and produced relatively large *Nat*-positive restriction fragments following Southern blot analysis (table 5.3).

Digestions of clone 511-O17 were followed by Southern blot and hybridisation (section 2.2.12.2) with probes specific for each one of the three *Nat* genes. The probes used in all hybridisations were specific for the 3' end of *Nat1*, *Nat2* or *Nat3*, from nucleotide position 516 to 867 of their coding region. Probes were generated by PCR, using primer pairs mNAT1 516/mNAT1 867, mNAT2 516/mNAT2 867 and mNAT3 516/mNAT3 867 respectively (table 2.1). A probe specific for the entire coding region of *Nat3*, generated by PCR with primers Mus12 and Mus15 (table 2.1), was also used.

Complete digestion of the 511-O17 PAC clone with *Kpn*I allowed detection of the three *Nat* genes on distinct restriction fragments. It did not reveal bands carrying *Nat3* and either *Nat1* or *Nat2*, therefore, it did not indicate on which side of the cluster *Nat3* is localised. For that purpose, a series of partial digestions of clone 511-O17 with *Kpn*I was performed, with the aim of generating incompletely digested restriction fragments containing *Nat3* and one of the other *Nat* genes. This would determine the order of the three genes on the PAC clone, as well as the distance of *Nat3* from its closest *Nat* gene. *Kpn*I is known to cleave at nucleotide position 476 of the *Nat3* coding region (Kelly and Sim, 1994), as well as approximately 0.5kb upstream of the *Nat2* coding region (EMBL accession number AJ250123), separating *Nat1* from *Nat2*.

Nat1 was detected on a 22kb *KpnI* restriction fragment, while *Nat2* was found on a fragment of 13-20kb estimated size (table 5.3). As there is a *KpnI* site within the *Nat3* coding region, parts of *Nat3* were found on two restriction fragments. The probe specific for the entire *Nat3* coding region hybridised to two fragments of 13kb and 9kb, while the probe specific for the 3' end of *Nat3* coding region hybridised only to the 9kb fragment (table 5.3; figure 5.7). It was, therefore, possible to establish the 5' to 3' orientation of the *Nat3* gene on these two fragments. It was not possible to detect *Nat3* on the same restriction fragment with any of the other *Nat* genes. It appears that more than one *KpnI* site may separate *Nat3* from its closest *Nat* gene and not all of these sites were lost during partial digestion. Depending on the order and orientation of the *Nat* genes in the mouse genome, the minimum distance was estimated to be approximately 22kb between *Nat3* and *Nat1*, or 26kb between *Nat3* and *Nat2* (figure 5.8). Based on the size of the undigested 511-O17 PAC clone (figure 5.7), *Nat3* should be located within a maximum distance of 130kb from the other two genes.

Table 5.3: Size of restriction fragments hybridising to different *Nat*-specific probes, following digestion of PAC clone 511-O17 with *KpnI*.

Probe specificity	Size of <i>KpnI</i> restriction fragment
<i>Nat1</i> 3' end	22kb
<i>Nat2</i> 3' end	13-20kb
<i>Nat3</i> 3' end	9kb (22kb in partial digestion)
<i>Nat3</i> coding region	13 and 9kb

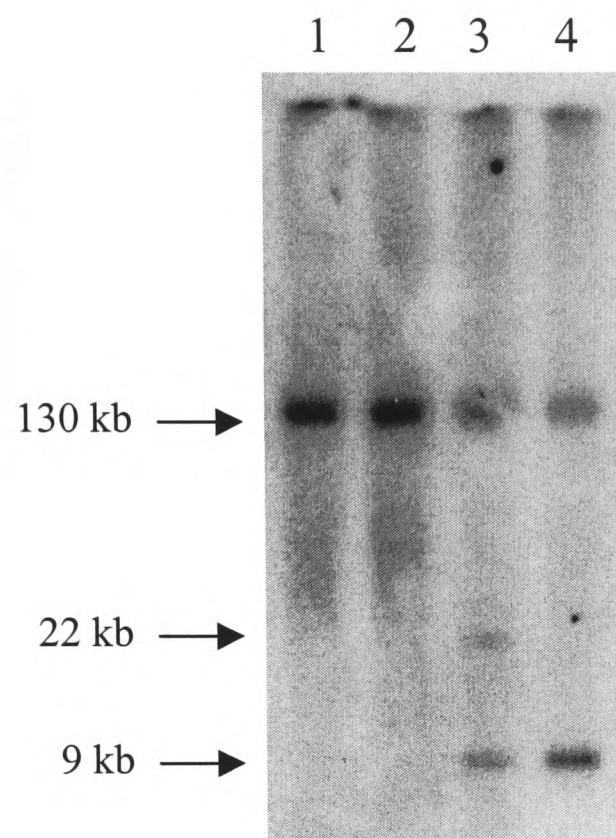


Figure 5.7: Partial *Kpn*I digestion of the mouse PAC genomic clone 511-O17. DNA (approximately 1 μ g) was digested with 0.01U (lane 1), 0.1U (lane 2), 1U (lane 3) and 10U (lane 4) of *Kpn*I restriction enzyme for 2.5 hours at 37°C. The digestion products were separated on a 1% agarose gel at 4°C, by PFGE (section 2.2.5.3). Settings were 6V/cm, with a switch time increasing from 5 to 15 seconds over the 16 hours of the run. The gel was blotted and hybridised to the *Nat3*-specific probe, covering the 3' end of the coding region from nucleotide position 516 to 867. Three bands were visible on the autoradiogram, the largest of which (~130kb) represents the undigested PAC clone. The approximately 9kb band contains part of the *Nat3* gene, from nucleotide position 476 to the end of the coding region. The approximately 22kb band is a partial digestion product, containing the entire coding region of *Nat3*, and is due to loss of the *Kpn*I site located within the *Nat3* gene.

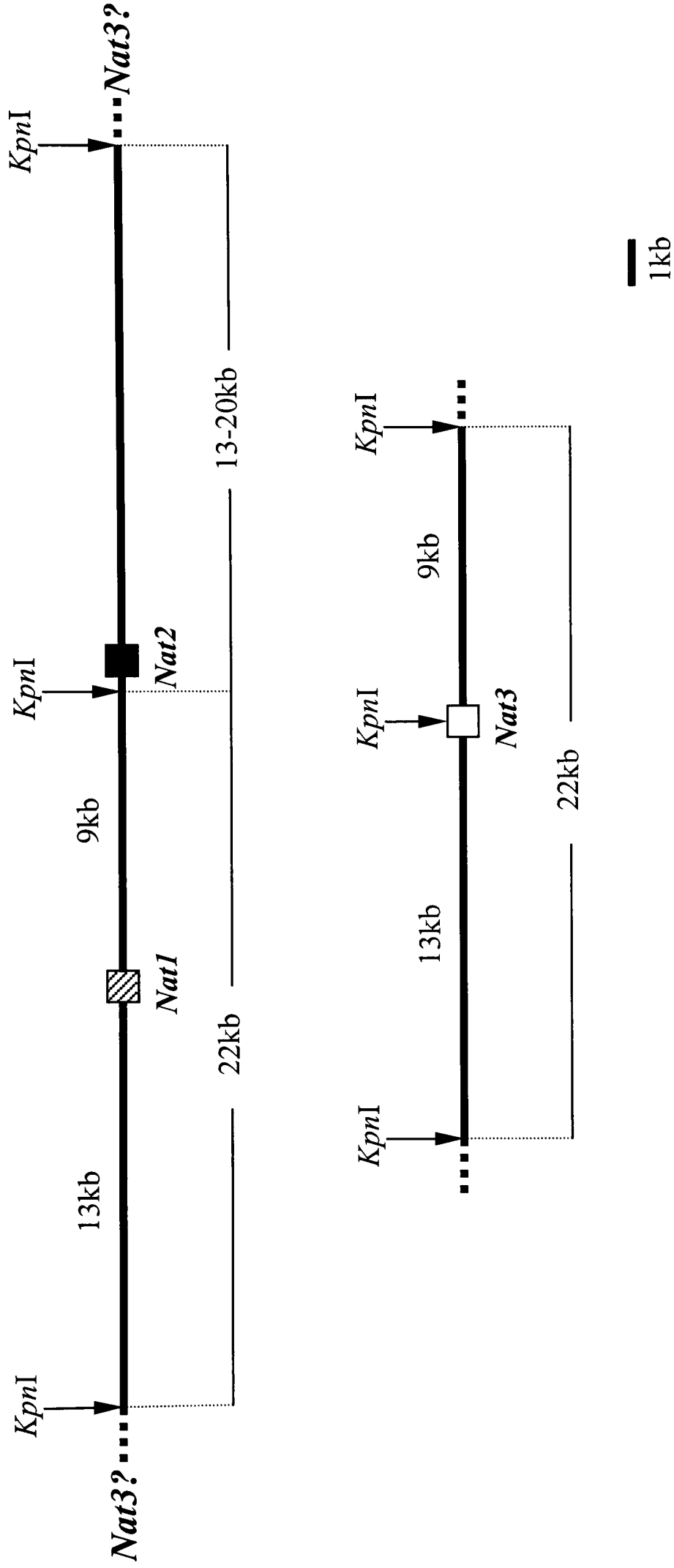


Figure 5.8: Position of the murine *Nat* genes relative to their closest *KpnI* restriction sites. The diagram was based on information from partial and complete digestion of the 511-O17 mouse PAC genomic clone. A *KpnI* site is located approximately 0.5kb upstream of the *Nat2* coding region, separating it from *Nat1*, while another *KpnI* site is located at nucleotide position 476 of the *Nat3* coding region.

5.3.5 Screening of *Nat*-positive mouse PAC genomic clones for chromosome 8 specific markers

On the basis of previous linkage analysis (Mattano et al, 1988), the *Nat1* and *Nat2* genes have been reported by the Mouse Genome Database (MGD) to map at 31cM from the centromere of mouse chromosome 8 (MGI:109201). To accurately place the *Nat* genes on the genetic and physical maps of murine chromosome 8, the three *Nat*-positive mouse PAC genomic clones (section 5.3.1) were screened for a series of polymorphic D8Mit markers (section 5.2.4), previously mapped to the 31cM region (MGD, <http://www.informatics.jax.org/searches/linkmap.cgi>). DNA from PAC clones 584-N8, 393-K20 and 511-O17 was used as template for amplification of ten SSR markers (table 5.4). The PCR conditions and the size of the expected products are indicated in table 5.4. None of the markers was present on any of the three PAC clones, as indicated by lack of amplification product (figure 5.9). The genomic region covered by the three PAC clones was probably too small to contain any of the selected markers.

Table 5.4: Simple sequence repeat (SSR) markers used for screening the *Nat*-positive mouse PAC genomic clones. The optimal conditions used for amplification of each marker are presented. The indicated size of the amplification product is that of the most common allele, among different mouse strains.

Marker	Annealing temperature (°C)	MgCl ₂ concentration (mM)	Product size (bp)
D8Mit54	57	1.5	245
D8Mit68	57	1.5	140
D8Mit69	57	3.0	154
D8Mit100	57	1.5	109
D8Mit128	57	3.0	146
D8Mit144	57	2.0	148
D8Mit160	57	1.5	147
D8Mit231	57	1.5	137
D8Mit260	57	1.5	97
D8Mit358	57	1.5	120

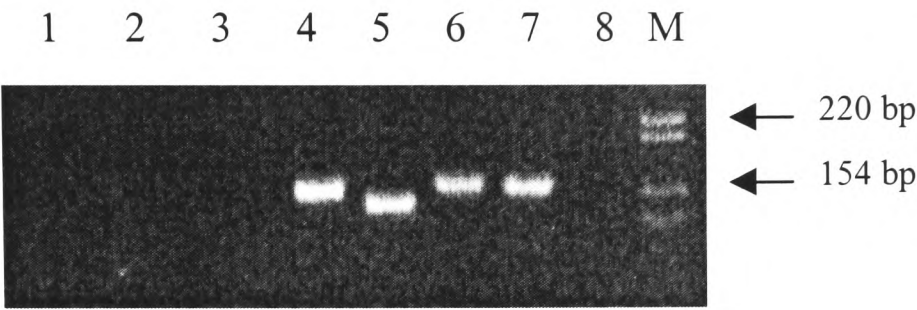


Figure 5.9: Amplification of SSR marker D8Mit69 from the three *Nat*-positive mouse PAC genomic clones. DNA from PAC clones 393-K20 (lane 1), 511-O17 (lane 2) and 584-N8 (lane 3) was used as template for PCR amplification. Genomic DNA from Balb/c (lane 4), C57Bl/6J (lane 5), A/J (lane 6) and 129/Ola (lane 7) mouse strains was used as positive control. Lane 8 was loaded with the reaction negative control (no template) and lane M with 1µg of 1kb ladder. Products were analysed on a 3% metaphor gel. Marker D8Mit69 was not present on any of the three PAC clones (lanes 1 to 3). Amplification from mouse genomic DNA of different strains provided different size products, indicating polymorphism between the strains. Amplification from Balb/c (lane 4), A/J (lane 6) and 129/Ola (lane 7) mouse strains provided a 154bp product, while the product from C57Bl/6J (lane 5) was 144bp in size.

5.3.6 Screening of chromosome 8 specific YAC contigs for murine *Nat* genes

In order to place the *Nat* loci close to established markers of the 31cM region of mouse chromosome 8, four YAC contigs (WC8.14, WC8.16, WC8.17 and WC8.18) spanning the region 29.5-32.8cM of chromosome 8 (section 5.2.3; figure 5.1) were screened for presence of the *Nat* genes, in collaboration with S Boukouvala (Department of Pharmacology, University of Oxford, UK). PCR was carried out with primers Mus12 and Mus13, to investigate the presence of the *Nat1* gene on any of the eighty YAC clones comprising the four YAC contigs. Amplification of SSR markers specific for each contig was also performed, as positive control of amplification from the YAC clones. These were markers D8Mit6, D8Mit205, D8Mit160 and D8Mit68, specific for contigs WC8.14, WC8.16, WC8.17 and WC8.18, respectively (section 5.2.3; figure 5.1). None of the YAC clones was positive for the *Nat1* locus (figure 5.10). Interestingly, markers D8Mit128 and D8Mit160, specific for the WC8.17 contig, as well as markers D8Mit69, D8Mit100, D8Mit68, D8Mit260 and D8Mit231, specific for the WC8.18 contig, were among those used for screening of the *Nat*-

positive PAC clones (section 5.3.5). These data indicated that the *Nat* gene cluster may not be located in region 29.5-32.8cM of mouse chromosome 8, as predicted (Mattano et al, 1988; MGD, <http://www.informatics.jax.org>), although the genomic coverage of the region by the four YAC contigs is not complete.

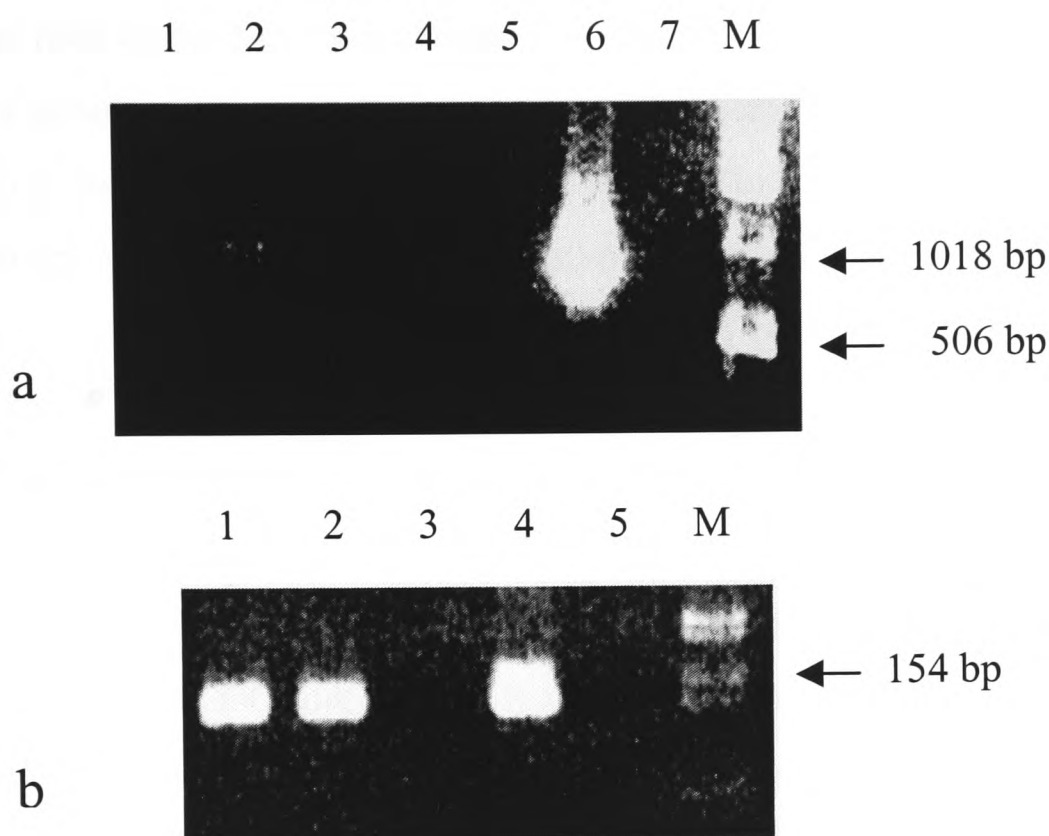


Figure 5.10: Screening of mouse chromosome 8-specific YAC contigs for presence of the *Nat* genes.

YAC clones from contig WC8.16 were used as template for *Nat1*-specific amplification with primers Mus12 and Mus13 (a). Amplification from selected clones was also performed with primers specific for marker D8Mit205, as positive control (b).

(a) Lanes 1 to 5 were loaded with PCR products from clones 456-F2, 187-A3, 462-E11, 148-B9 and 328-G2, respectively. Lanes 6 and 7 were loaded with the reaction positive (C57Bl/6J mouse genomic DNA as template) and negative (no template) controls, respectively. Lane M was loaded with 1µg of 1kb ladder. One clone of the contig (402-D10) was not available from the HGMP Resource Centre. Products were run on a 1% agarose gel. None of the clones was carrying the *Nat1* gene, as indicated by the absence of amplification product.

(b) Lanes 1 and 2 were loaded with PCR products from clones 456-F2 and 462-E11, which are positive for marker D8Mit205. Lane 3 was loaded with product from clone 328-G2, which does not contain this marker. Lanes 4 and 5 were loaded with the reaction positive (C57Bl/6J mouse genomic DNA as template) and negative (no template) controls, respectively. Lane M was loaded with 1µg 1kb ladder. Products were run on a 3% agarose gel. The 146bp PCR product indicated presence of the D8Mit205 marker on clones 456-F2 and 462-E11 of the WC8.16 contig, as expected.

5.3.7 Screening of the WI/MIT mouse YAC genomic library

Screening of the YAC contigs described above (section 5.3.6) did not provide clones carrying any of the *Nat* genes. An alternative approach would be to screen *Nat*-positive YAC clones for established markers, allowing positioning of the *Nat* genes on the physical map of murine chromosome 8. Compared with PACs, YACs provide a much greater genomic coverage per single clone (sections 5.2.1 and 5.2.2), therefore, a given locus of interest would be more likely to co-localise with an established chromosomal marker on the same clone, when YAC cloning systems are used. To isolate *Nat*-positive YAC clones, the WI/MIT mouse YAC genomic library was screened by PCR specific for the coding region of each one of the *Nat* genes, according to the scheme described in section 5.2.2. Amplification was carried out using primers Mus12/ Mus13 for *Nat1*, mNAT2 1/ mNAT2 910 for *Nat2* and Mus12/ Mus15 for *Nat3* (table 2.1). Of the fifty primary pools, eight (MY4, MY5, MY10, MY13, MY18, MY40, MY43 and MY44) were identified as positive for at least one of the *Nat* genes (figure 5.11a). Secondary screening was then performed for positive primary pools MY4, MY5, MY10, MY18 and MY43, while primary pools MY13, MY40 and MY44 were not screened further. Following secondary screening (figure 5.11b), all candidate clones were individually checked for presence of each one of the *Nat*-genes by PCR. Of the thirty-nine clones examined, four clones (107-A2, 110-G4, 156-A8 and 435-B7) were confirmed to be positive *Nat*. Each clone contained all three *Nat* genes (figure 5.12a-c; table 5.5). This further indicated that the *Nat* cluster occupies a relatively short region on murine chromosome 8. Complete screening of primary pools MY13, MY40 and MY44 may provide additional *Nat*-positive YAC genomic clones.

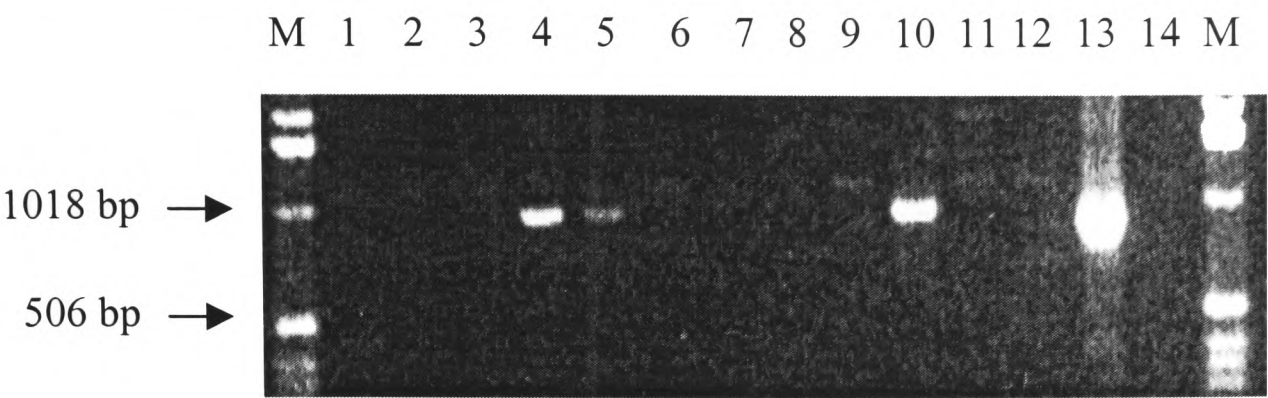


Figure 5.11a: Screening of the WI/MIT mouse YAC genomic library for *Nat* genes. The *Nat3*-specific amplification product from twelve of the fifty primary pools of the library is shown here. PCR was performed using primers Mus12 and Mus15, and 10µl of each product were loaded on a 1.5% agarose gel. Lanes 1 to 12 were loaded with PCR products from primary pools MY1 to MY12. Lanes 13 and 14 were the positive and negative controls for the reaction (1µg mouse genomic DNA and no template, respectively). Lanes M were loaded with 1µg of 1kb ladder. A band 908bp in size was produced from primary pools MY4 (lane 4), MY5 (lane 5) and MY10 (lane10), as well as from the genomic DNA used as positive control (lane 13).

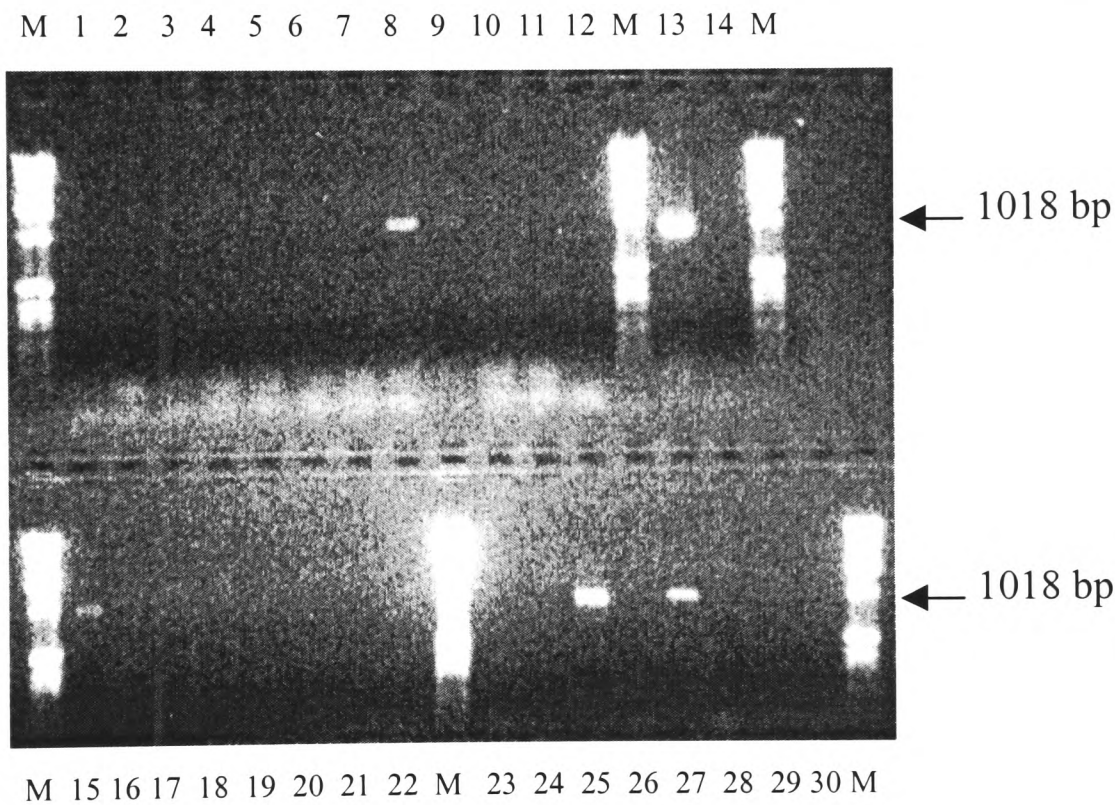


Figure 5.11b: Screening of the WI/MIT mouse YAC genomic library for *Nat* genes. Pools of the eight plates corresponding to the *Nat*-positive primary pool MY10 were screened for *Nat3* by PCR, using primers Mus12 and Mus15. Products (10µl) were loaded on a 1.5% agarose gel. Lanes 1 to 12 were loaded with PCR products from the twelve column pools of all eight plates. Lanes 15 to 22 were loaded with PCR products from the eight row pools of all eight plates. Lanes 23 to 30 were loaded with PCR product from each one of the eight plate pools. Lanes 13 and 14 were the positive and negative controls for the reaction (1µg mouse genomic DNA and no template, respectively). Lanes M were loaded with 1µg of 1kb ladder. A band 908bp in size was produced from column pool 8 (lane 8), row pool A (lane 15) and plate pools 156 and 159 (lanes 25 and 27), as well as from the genomic DNA used as positive control (lane13). The results indicate that clones 156-A8 and 159-A8 potentially contain the *Nat3* gene.

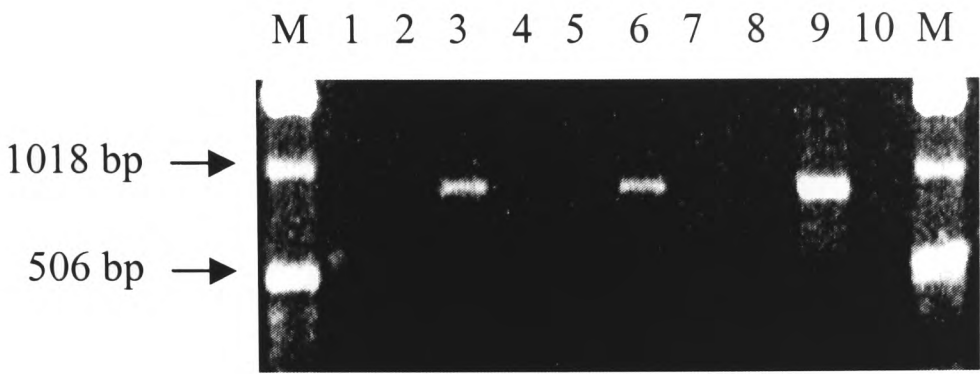


Figure 5.12a: Identification of mouse YAC genomic clones carrying the *Nat* genes. The candidate *Nat*-positive clones were checked for presence of the *Nat1* gene by PCR, using primers Mus12 and Mus13. Amplification products (10 μ l) were loaded on a 1% agarose gel. Lanes 1 to 8 were loaded with PCR products from clones 115-G8, 154-A8, 156-A8, 159-A8, 161-A8, 435-B7, 437-H9 and 439-B7. Lanes 9 and 10 were the positive and negative controls for the reaction (1 μ g mouse genomic DNA and no template, respectively). Lanes M were loaded with 1 μ g of 1kb ladder. A band of 908bp was produced from clones 156-A8 (lane 3) and 435-B7 (lane 6), as well as from the genomic DNA used as positive control (lane 9). The results indicate that clones 156-A8 and 435-B7 contain the *Nat1* gene.

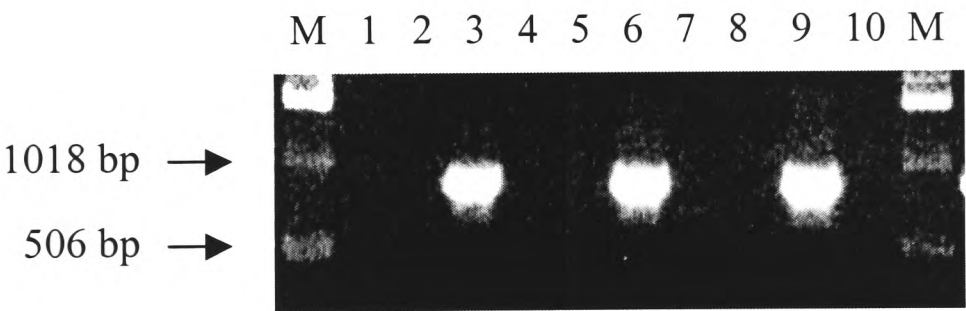


Figure 5.12b: Identification of mouse YAC genomic clones carrying the *Nat* genes. The candidate *Nat*-positive clones were checked for presence of the *Nat3* gene by PCR, using primers Mus12 and Mus15. Amplification products (10 μ l) were loaded on a 1% agarose gel. Lanes 1 to 8 were loaded with PCR products from clones 115-G8, 154-A8, 156-A8, 159-A8, 161-A8, 435-B7, 437-H9 and 439-B7. Lanes 9 and 10 were the positive and negative controls for the reaction (1 μ g mouse genomic DNA and no template, respectively). Lanes M were loaded with 1 μ g of 1kb ladder. A band of 908bp was produced from clones 156-A8 (lane 3) and 435-B7 (lane 6), as well as from the genomic DNA used as positive control (lane 9). The results indicate that clones 156-A8 and 435-B7 contain the *Nat3* coding region.

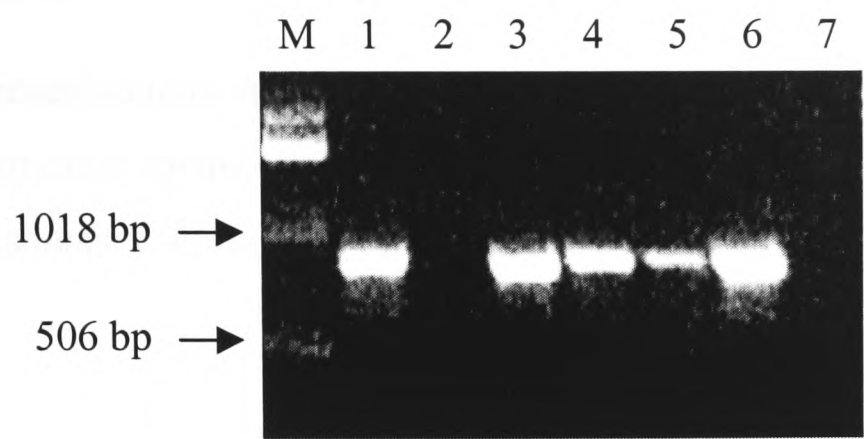


Figure 5.12c: Identification of mouse YAC genomic clones carrying the *Nat* genes. The four clones identified as positive for *Nat1* and *Nat3* were also screened for the *Nat2* gene by PCR, using primers mNAT2 1 and mNAT2 910. Amplification products (10µl) were loaded on a 1% agarose gel. Lanes 1 to 5 were loaded with PCR products from YAC clones 107-A2, 107-E2, 110-G4, 156-A8 and 435-B7. Lanes 6 and 7 were the positive and negative controls for the reaction (1µg mouse genomic DNA and no template, respectively). Lanes M were loaded with 1µg of 1kb ladder. A band 910bp in size was produced from clones 107-A2 (lane 1), 110-G4 (lane 3), 156-A8 (lane 4) and 435-B7 (lane 5), as well as from the genomic DNA used as positive control (lane 6). The results indicate that clones 107-A2, 110-G4, 156-A8 and 435-B7 contain the *Nat2* coding region. Clone 107-E2 (lane 2) was randomly selected from the library, to serve as negative control, and was not expected to provide amplification product.

Table 5.5: Summary of the content of the *Nat*-positive clones isolated from the WI/MIT mouse YAC genomic library. All positive YAC clones were identified to contain the entire *Nat* gene cluster.

Coding region amplification	YAC clone			
	107-A2	110-G4	156-A8	435-B7
<i>Nat1</i>	+	+	+	+
<i>Nat2</i>	+	+	+	+
<i>Nat3</i>	+	+	+	+

5.4 Discussion

The work presented here aimed to provide information about the cytogenetic and physical position of the murine *Nat* genes. This was achieved using mapping resources (databases, libraries and contigs), developed by the Mouse Genome Project (MGP).

Screening of a mouse PAC genomic library and identification of *Nat*-positive clones established co-localisation of *Nat3* on the same chromosomal region with the other two *Nat* genes in the mouse. All three genes were detected on the same PAC clone, but *Nat3* was also found separately. *Nat1* and *Nat2* were always found together, as would be expected from the fact that the distance between them is only 9.4kb (section 4.4.3.3). These results indicated that *Nat3* may be located further away than *Nat1* and *Nat2* are from each other, however, this distance does not exceed the size of the PAC clone containing all three *Nat* genes (511-O17), estimated to 130kb by PFGE (section 5.3.4; figure 5.7).

Preliminary attempts were made to identify the exact position of *Nat3* in relation to the other two genes, by partial digestion of PAC clone 511-O17 with restriction enzymes of relatively low cutting frequency. Following initial restriction analysis, *KpnI* was selected as the most appropriate enzyme for a number of reasons (section 5.3.4). First, it separates *Nat1* from *Nat2*, making possible the isolation of *Nat3* and one of the other *Nat* genes on the same restriction fragment. This would indicate whether *Nat3* is closer to *Nat1* or *Nat2*. Second, it cleaves within the coding region of *Nat3*, therefore, if it was isolated on the same restriction fragment with another *Nat* gene, this would directly provide the exact physical distance between the two genes. This would also establish orientation of *Nat3* relative to *Nat1* and *Nat2*, as it would be easy to distinguish between the *KpnI* fragment carrying the 5' end of *Nat3* and the fragment carrying the 3' end, using suitable probes. The *KpnI* partial digestions described here did not suggest presence of *Nat3* on the same restriction fragment as *Nat1* or *Nat2*, but the minimum distance between them was estimated to 22kb. Although complete digestion with *KpnI* provided larger restriction fragments than with other enzymes used, it appears that more than one *KpnI* sites separate *Nat3* from

the other two genes. Further investigation of the 511-O17 PAC clone will be useful for a more extensive study of the *Nat* genomic region in mice.

The similarities in the organisation of the genes encoding for NAT between the mouse and human are notable. In both species, there are three loci arranged in a single cluster. Although the distances between them appear to be much shorter in the mouse than in the human, the organisation is similar, with two loci located close to each other and the third lying further apart. Human *NAT1* has been localised 170-360kb away from *NAT2*, while *NAT2* and the pseudogene *NATP* are closer to each other, within 140kb (Matas et al, 1997). Similarly, murine *Nat1* and *Nat2* are only 9.4kb away from each other, with *Nat3* located within 22-130kb. This conservation in the genomic organisation of the two clusters, emphasises the strong evolutionary relationship between human and mouse NATs, also evident at the level of nucleotide sequence of the genes and amino acid sequence of their protein products (Vatsis et al, 1995; table 4.1).

Close linkage of the *Nat* genes was also demonstrated by FISH analysis, using the identified *Nat*-positive PAC clones as probes. These experiments also confirmed localisation of the *Nat* genes on murine chromosome 8 and provided their exact cytogenetic position on the middle portion of chromosome 8, in region B3.1-B3.3 (Fakis et al, 2000; MGD accession number J:58758). Although genetic mapping has been very extensive in mice, very few genes have been cytogenetically assigned to specific chromosomal regions of the mouse genome, compared to human. Of the 20,000 genetic markers available to date for the mouse, only 1,300 have been mapped to a specific cytogenetic band, according to the Mouse Genome Database (<http://www.informatics.jax.org>). This is largely due to the nature of the mouse chromosomes, which are all acrocentric and very similar in size, making karyotypic analysis difficult. Recently, chromosome sorting by flow cytometry, followed by labelling with degenerate oligonucleotide-primed PCR (DOP-PCR) has allowed development of very sensitive whole chromosome-specific probes. In species like the mouse, dog, pig and sheep, which have high degree of similarity between their chromosomes, these chromosome-specific “paints” have been extensively used in various cytogenetic studies (Ferguson-Smith, 1997). Cross-species FISH analysis (“Zoo-FISH”) using these paints has allowed identification of chromosomal regions

of conserved synteny between different species and has been particularly useful for comparative mapping. However, this technique allows only identification of large syntenic segments and cannot detect small rearrangements within them (Scalzi and Hozier, 1998; Grutzner et al, 1999).

Genetic mapping of the murine *Nat* genes has been based on the study of recombinant inbred mouse strains generated by crossing the fast acetylating C57Bl/6J with the slow acetylating A/J inbred strain. Linkage has been established between *Nat2* and the *Es1* gene, located at 43cM from the centromere of mouse chromosome 8, placing the two genes approximately 12cM from each other (Mattano et al, 1988). Although it was not established whether *Nat2* lies on the centromeric or the telomeric side of *Es1*, there are several clues that suggest localisation of the *Nat* genes on the centromeric side of *Es1*, at 31cM from the centromere. Linkage between *Nat2* and the *Emv2* locus, located at 67cM from the centromere, was very weak, indicating a large genetic distance between the two loci (Mattano et al, 1988). Therefore, *Nat2* is likely to be located at 31cM from the centromere of chromosome 8, i.e. 12 and 36cM on the centromeric side of *Es1* and *Emv2*, respectively. This is further suggested by cytogenetic localisation of the *Nat* genes to bands B3.1-3.3, on the middle portion of chromosome 8, established in the present study (section 5.3.3). Although integrated cytogenetic and genetic mapping data have been limited for the mouse, chromosome 8 bands B3.1-3.3 have been linked to the 33cM region, based on several chromosomal markers (2000 Chromosome 8 Committee Report available from the Mouse Genome Database, <http://www.informatics.jac.org/ccr>). Moreover, conserved synteny has been observed between the 33cM region of mouse chromosome 8 and region 8p22-p21 in humans, as indicated by presence of several homologous loci. These include the genes for lipoprotein lipase (*Lpl*), solute carrier family 18 member 1 (*Slc18a1*) and lysosomal H⁺ transporting ATPase beta isoform 1 (*Atp6b1*) (Oxford Grid, <http://www.informatics.jax.org/searches/oxfordgrid.cgi>).

As part of the work presented here, physical mapping of the *Nat* genes in relation to established chromosome 8-specific markers was attempted. The first approach was to screen the *Nat*-positive PAC clones for a series of SSR markers mapping to the 31cM region of mouse chromosome 8. None of these markers was found on the three PAC

clones, suggesting that either the *Nat* genes are not located exactly where predicted or that the selected markers were not as close to the *Nat* genes as to co-localise with them on the same PAC clone. A second approach was to screen a number of YAC contigs covering the 29.5cM to 32.8cM region of mouse chromosome 8 for the *Nat* locus. These contigs contained several chromosomal markers (table 5.1), including the SSR markers used for screening of the PAC clones. None of the YAC clones comprising the contigs was positive for the *Nat* genes. It is possible that the *Nat* cluster could be located within the 1.1cM gap spanning the region from 31.7 to 32.8cM between contigs WC8.16 and WC8.17 (figure 5.1). Alternatively, the *Nat* genes could be located in the regions flanking the four contigs, most likely towards the 33cM region where the *Lpl* gene has been localised. Screening of contigs available for the 33cM region is, therefore, likely to prove useful.

To generate additional tools for physical mapping of the *Nat* genes, a number of *Nat*-positive YAC clones were isolated by screening of the WI/MIT genomic library (Haldi et al, 1996), currently being used for construction of a physical map for the mouse genome (Nusbaum et al, 1999). In the future, the identified clones can be used to localise the *Nat* genes in relation to established chromosomal markers. Only a portion of the library (five out of eight positive primary pools) was screened to completion, due to the very large number of potentially positive clones generated following secondary screening. The four clones identified as positive for the *Nat* genes have not previously been mapped by the Whitehead Institute/MIT Centre for Genome Research and are not part of any of the established contigs of chromosome 8 (MGD, <http://www.informatics.jax.org>). Only a subset of the mouse WI/MIT YAC genomic library (24,000 clones) has been used for construction of the physical map and this is probably not the same as the portion used here.

About a thousand loci have been mapped on murine chromosome 8 to date. Of those, 226 are established genes and another 76 are ESTs of unknown function. Synteny between mouse and human has been investigated for extended regions of mouse chromosome 8, and nearly 100 syntenic markers have been identified to date. However, there has been very little work on comparative mapping between the mouse and other mammalian species (2000 Chromosome 8 Committee Report, <http://www.informatics.jax.org/ccr>). The largest syntenic group has been identified for

mouse chromosome 8 and human chromosome 16, with a total of 52 syntenic genes between the two species. At least 25 genes have been identified as syntenic between mouse and human chromosome 8. Although not comprising a single syntenic region, most of these genes are located on the centromeric half of murine chromosome 8, with their human counterparts located in region 8p12-q21. (Oxford Grid, <http://www.informatics.jax.org/searches/oxfordgrid.cgi>; 1999 Chromosome 8 Committee Report, <http://www.informatics.jax.org/ccr>).

In contrast to genetic mapping, which depends on the availability of polymorphic markers for linkage analysis, comparative mapping mostly relies upon conserved markers between different species. These “anchor loci” are genes showing very little polymorphism between different species, the map position of which has normally been determined by at least two mapping methods (e.g. genetic and physical or genetic and cytogenetic mapping). These loci provide the basis for integrating different types of mapping data from the same or different species and are particularly useful for comparative mapping between the mouse and human. Most of the mapping data currently available for the mouse are in the form of genetic maps, while in humans (where linkage analysis is more difficult) cytogenetic mapping has been much more extensive (Andersson et al, 1996; Carver and Stubbs, 1997; O’Brien et al, 1997; Serikawa et al, 1999). According to the report of the HUGO Committee on Comparative Mapping (Andersson et al, 1996), the human and murine genes for lipoprotein lipase (*LPL* and *Lpl* respectively), plasminogen activator (*PLAT* and *Plat*) and DNA-directed polymerase beta (*POLB* and *Polb*) are anchor loci, establishing synteny between regions of human and mouse chromosome 8. Once their physical mapping is completed, the mouse *Nat* genes can serve as ideal anchored loci for comparative mapping, as well as for integrating genetic and cytogenetic mapping data for mouse chromosome 8.

In addition to conserved chromosome 8 markers, the *Nat* loci represent a useful polymorphic marker for linkage analysis. *Nat2* polymorphism has been detected among established lines of inbred mouse strains, such as the C57Bl/6J and the A/J, and can be easily identified both at the phenotypic (based on pABA N-acetylation activity assays) and genotypic (as a *Tru9I* polymorphism) level. Study of the genomic region surrounding the *Nat* locus in mice may prove particularly interesting, in view

of the involvement of the syntenic region 8p22 in carcinogenesis (Chapter 6). Moreover, the *Nat2* polymorphism has been linked to susceptibility to teratogen-induced cleft palate (CP) and cleft lip with or without cleft palate (CL/P) among the fast acetylating C57Bl/6J (resistant to facial clefting) and the slow acetylating A/J (susceptible to facial clefting) mouse strains. Study of recombinant inbred mouse strains has associated the slow *Nat2* phenotype with 6-aminonicotinamide (6-AN) and glucocorticoid-induced CP. Interestingly, slow acetylator congenic strains carrying the *Nat* region of the A/J strain on a C57Bl/6J genetic background became susceptible to CP, when exposed to these teratogens. At the same time, fast acetylating congenic strains carrying the *Nat* region of the C57Bl/6J strain on an A/J genetic background became resistant to the effects of these compounds (Karolyi et al, 1988; Karolyi et al, 1990). These findings were later supported by a genome-wide search for murine loci affecting susceptibility to teratogen-induced facial clefting, although other genetic loci were also shown to be involved (Diehl and Erickson, 1997). Linkage between the *Nat2* polymorphism and phenytoin-induced CL/P has also been suggested, but it is the fast acetylator phenotype that appears to be associated with higher predisposition (Karolyi et al, 1987; Diehl and Erickson, 1997). It will be of interest to investigate whether polymorphism of the *NAT* genes in humans is associated with different predisposition to this kind of developmental disorders.

It appears that a single gene in the *Nat* region is responsible for variable susceptibility to teratogen-induced facial clefting in mice, but it is unclear whether this might be *Nat2* or another gene in close proximity. The NAT enzymes have not been demonstrated to be directly involved in the metabolism of teratogens, like phenytoin and 6-AN (Karolyi et al, 1990). However, recent research has suggested that expression of *Nat2* may be regulated by a glucocorticoid response element, located upstream of the coding region (Estrada-Rodgers et al, 1998a). Moreover, a role in development has recently been suggested for murine NAT2, via its possible implication in the metabolism of folate (Payton et al, 1999b).

Mapping of the region marked by the *Nat* genes on murine chromosome 8 will be useful for identification of neighbouring genes playing key roles in processes leading to developmental disorders or carcinogenesis. The work presented here extended previous knowledge about the localisation and organisation of the *Nat* cluster in the mouse genome and generated useful tools for physical mapping of the *Nat* region. It is anticipated that this work will contribute towards integration of different types of mapping data for the *Nat* region of murine chromosome 8 and that it will be useful for comparative studies between the mouse and other mammalian species.

Chapter 6: Arylamine N-Acetyltransferase – A Study in Healthy Colon, Breast and Cancerous Bladder Material

6.1 Introduction

6.1.1 NAT in colon cancer

People are exposed to carcinogens through their diet. Red meat cooked at high temperatures is known to contain heterocyclic amines, produced from the pyrolysis of protein, and these compounds have been demonstrated to be potent mutagens and carcinogens (Sugimura, 1985). In laboratory animals, administration of heterocyclic amines, such as PhIP, IQ and MeIQx [2-amino-3,8-dimethylimidazo- (4,5-f)-quinoxaline], in their diet induced tumours of the intestine. The currently accepted mechanism for colon carcinogenesis due to dietary heterocyclic amines involves the NAT enzymes (Lang et al, 1994; Kaderlik and Kadlubar 1995; Hanna, 1996). It is believed that the first step towards bioactivation of ingested heterocyclic amines is N-oxidation catalysed by the hepatic P450 enzyme CYP1A2, producing N-hydroxy heterocyclic amines. These may be further metabolised by UDP-glucuronosyl transferase, leading to detoxification and excretion. Heterocyclic amines are not good substrates for N-acetylation by NAT and cannot be detoxified via this metabolic route. However, they can undergo O-acetylation by both human NAT enzymes, especially NAT2, as shown *in vitro* using recombinant NAT1 and NAT2 enzymes expressed in COS-1 cells (Minchin et al, 1993). This activating reaction may occur in the colon mucosa, but it may also occur in the liver with subsequent transport of the N-acetoxy derivatives to the colon through the circulation. In either case, the

formation of N-acetoxy heterocyclic amine compounds, which are suspected to be the ultimate carcinogens, may lead to DNA adducts formation and the initiation of carcinogenesis (Kaderlik and Kadlubar, 1995).

Both NAT1 and NAT2 isoenzymes are present in human colon. The mRNA for both proteins has been detected in human colorectal cytosols by RT-PCR (Ilett et al, 1994). In the same study, N-acetylation of pABA and SMZ has been detected in colorectal cytosols, indicating the presence of functional NAT1 and NAT2, respectively, in this tissue. This was further confirmed using hybridisation histochemistry with specific NAT1 and NAT2 riboprobes that detected mRNA transcripts from both genes (Windmill et al, 1997; Windmill et al, 2000). In these studies, neither RT-PCR nor hybridisation histochemistry were able to quantify the detected NAT1 and NAT2 mRNA and therefore it was not possible to predict the relative amount of NAT1 and NAT2 protein in colon. The presence of both NAT isoenzymes in human colon suggests that they may both have a role in the local activation of carcinogens (Ilett et al, 1994; Hickman et al, 1998).

A currently proposed mechanism of colon carcinogenesis suggests that individuals with the fast acetylation phenotype may be more susceptible to colon cancer induced by heterocyclic amines, due to more efficient activation of these carcinogens (Lang et al, 1994; Kaderlik and Kadlubar 1995; Hanna, 1996). Consequently, the NAT polymorphism in colorectal cancer has been investigated on the hypothesis that fast acetylation is a susceptibility factor, in contrast to bladder cancer where slow NAT2 activity confers an increased risk, as will be described later (section 6.1.2). A number of studies have reported evidence in support of this assumption (Ilett et al, 1987; Wohlleb et al, 1990; Minchin et al, 1993). A study conducted in the Portuguese population suggested association between the homozygous fast *NAT2**4/*4 genotype and higher risk for colorectal cancer, indicating that very high acetylation activity might be a predisposing factor for the disease (Gil and Lechner, 1998). However, other studies have failed to detect statistically significant differences in the frequency of *NAT2* alleles between the colorectal cancer and the control groups (Ladero et al, 1991; Rodriguez et al, 1993; Probst-Hensch et al, 1995; Hubbard et al, 1997; Lee et al, 1998; Katoh et al, 2000). Therefore, it remains controversial whether any particular

NAT2 acetylation phenotype may confer increased risk for chemically induced carcinogenesis in colon.

The involvement of NAT1 polymorphism in colon carcinogenesis is also unclear. The *NAT1*10* allele, which carries a single base substitution in the polyadenylation signal, has been controversially associated with higher acetylation activity by some researchers (section 1.3.3.2). It has been reported that the *NAT1*10* allele is linked to increased risk of colorectal cancer and that it is more frequently found in patients with higher stage tumours (Bell et al, 1995b). The risk for colorectal cancer conferred by the *NAT1*10* allele was estimated to increase by up to 1.9-fold. The same study reported the possibility of a combined effect of *NAT1* and *NAT2* rapid acetylation genotypes on the risk for colon cancer. However, the involvement of *NAT1*10*, either on its own or by interaction with a fast *NAT2* genotype, in colorectal cancer was later challenged by another study using a larger population (Probst-Hensch et al, 1996). A number of other studies have also showed that there is no connection between *NAT1*10* or any other *NAT1* allele and this type of cancer (Hubbard et al, 1998; Lin et al, 1998). Moreover, no difference in the distribution of *NAT1* genotypes was evident between patients at different stages of the disease. (Hubbard et al, 1998). Katoh et al (2000) recently looked at colorectal and gastric adenocarcinoma in relation to both *NAT* genes and found that *NAT1*10* allele was not a significant risk factor in the colorectal form of the disease. In addition, the small increase they observed in the frequency of the *NAT1*10* allele among the group of gastric adenocarcinoma patients, compared with controls, did not reach statistical significance. Therefore, as in the case of NAT2 polymorphism, the role of NAT1 polymorphism in colon carcinogenesis remains obscure.

6.1.2 NAT in bladder cancer

Bladder cancer has long been characterised as a malignancy with strong occupational background. More than a century ago, it was noted that the bladder cancer incidence among workers in the dye, rubber, leather and other types of industry was much higher than in the general population (for a historical review, see Weber and Hein 1985). Following these early observations, it was established that these workers developed bladder cancer because of their exposure to certain chemicals and this was

one of the first occupational hazards ever described (Case et al, 1954). The chemicals responsible for bladder carcinogenesis in these high-risk groups were identified to be various aromatic amines, such as benzidine and β -naphthylamine. As a result of this research, the production of these two compounds was banned in Britain. Several decades later, the incidence of bladder cancers due to industrial exposure to arylamines was still estimated at 8-20% (Wallace, 1988).

Cigarette smoke contains arylamines, like 4-aminobiphenyl, and it has been suggested that smoking could be another risk factor for bladder carcinogenesis. This was supported by observations that arylamine-DNA adducts were among the most frequent types of adducts found in biopsies from bladder cancer patients (Okkels et al, 1997). NATs are responsible for the detoxification of carcinogenic arylamines via N-acetylation, therefore competing with metabolic activation pathways which involve N-oxidation by cytochrome P450 (Kaderlik and Kadlubar, 1995).

The NAT polymorphism has been identified to affect predisposition to bladder cancer by epidemiological studies. By measuring the rate of dapsone acetylation, in order to determine the NAT2 acetylator phenotype of individuals, Cartwright and his colleagues showed significantly increased susceptibility of slow acetylators to bladder cancer caused by aromatic amines (Cartwright et al, 1982). At the time of that study, methods for determining the NAT2 genotype of individuals were not available. Following the introduction of genotyping methods, this connection between slow NAT2 acetylation and bladder cancer risk was verified (Risch et al, 1995). It was concluded that homozygous slow acetylators account for 71.0% of occupationally exposed bladder cancer patients and 65.4% of patients without previous occupational exposure. The percentage of homozygous slow acetylators in the healthy control group was only 44.1%. No increase was observed in the frequency of any particular slow allele in the bladder cancer population compared with the control group. Therefore, it was concluded that the slow acetylation phenotype, rather than any particular *NAT2* genotype, contributes to the bladder cancer predisposition. More recently, another study, also based on genotyping, confirmed these results (Filiadis et al, 1999).

The effect of the NAT1 polymorphism in bladder cancer is less straightforward. To date, the relevant studies report conflicting results. Okkels et al (1997) reported that bladder cancer risk was not associated with any particular *NAT1* genotype. On the other hand, increased risk for the disease was predicted for individuals with the *NAT1*10* allele, especially if they were smokers and slow NAT2 acetylators. There was evidence for a gene-dosage effect, so that *NAT1*10/*10* homozygotes were at higher risk than *NAT1*10* heterozygotes (Taylor et al, 1998). Similar results have been obtained by other groups of researchers working with the Taiwanese and Japanese population. Although the *NAT1*10* allele was more prevalent among bladder cancer cases, this difference was not considered as statistically significant. However, individuals with combined *NAT1*10* and slow *NAT2* genotypes were found to be at considerable risk for the disease (Hsieh et al, 1999; Katoh et al, 1999).

Only NAT1 protein has been detected in bladder tissue and urothelial cell lines using Western blotting, immunohistochemistry and enzymatic activity assays (Coroneos and Sim 1991; Stacey et al, 1996; Stanley et al, 1996; Stacey et al, 1998). However, both NAT1 and NAT2 transcripts have been detected by RT-PCR in bladder of normal individuals and a urothelial cell line immortalised with SV40 (Kloth et al, 1994). In view of the fact that NAT2 protein is absent in bladder tissue, the significance of this result is unclear. Because of the intronless nature of the NAT genes, this finding could be due to genomic DNA contamination of the cDNA template used for amplification. A control without enzyme in the reverse transcription reaction would indicate such contamination, if present, but was not included in the study of Kloth et al (1994).

The chromosomal region of the *NAT* loci is interesting because it is often deleted in bladder cancer cells (Knowles et al, 1993; Knowles, 1999). Studies have also shown that there is loss of heterozygosity (LOH) at the *NAT2* and the neighbouring loci in tumor cells from bladder cancer patients (Schnakenberg et al, 1998; Thygesen et al, 1999). Genetic instability of the region has also been observed in colorectal cancer, although the *NAT* genes were not included in the area of deletion (Hubbard et al, 1997). These LOH studies and information generated by deletion mapping have indicated the presence of two tumour suppressor genes in the chromosomal region of the *NAT* loci (Fujiwara et al, 1993; Farrington et al, 1996; Knowles, 1999). In that

context, the *NAT* genes could be very useful markers for the study of bladder carcinogenesis (Stacey et al, 1996; Stacey et al, 1999).

6.1.3 NAT in breast cancer

Breast cancer is the most common form of cancer among women. Its incidence has increased throughout the world during the last 20 years, although a parallel decrease has been observed in mortality rates, probably due to earlier diagnosis and better awareness. The highest incidence rates of the disease have been observed in California, USA, while Algeria, India, Korea and Thailand are among the countries with the lowest incidence (Parkin et al, 1997).

There seems to be a strong link between some forms of breast cancer and the individual's genetic background, as was established by studying certain familial forms of the disease. Mutations at the *BRCA1* and *BRCA2* genes have been associated with inherited predisposition to breast cancer, especially among younger women, although inheritance probably accounts for only up to 10% of all cases (King, 1992; Bishop, 1994). Another risk factor for breast cancer is prolonged exposure to high levels of estrogens. Events like early menarche and late menopause that increase exposure of a woman to estrogens are suspected risk factors (Madigan et al, 1995). Environmental chemicals, like dichlorodiphenyltrichloroethane (DDT) and benzo[a]pyrene, that bind to estrogen receptors and mimic their hormonal effect may also be involved (Davis et al, 1993). However, it seems likely that there are additional pathways leading to breast carcinogenesis that are independent of the hormonal route. Since the established risk factors account for approximately only one third of all breast cancer cases, there must be more, as yet unidentified, susceptibility factors. Environmental pollutants, ionising radiation, diet and tobacco smoking have long been suspected as potential risk factors that may contribute to breast carcinogenesis (Welp et al, 1998).

Many researchers have shown interest in the role that the metabolism of carcinogens might play in the aetiology of breast cancer. In contrast to other types of cancer, epidemiological evidence does not point to any particular environmental or occupational cause for breast cancer, but cigarette smoking and consumption of red meat have been suggested as candidates (reviewed in Josephy and Coomber, 1998;

Welp et al, 1998). Potential breast carcinogens in humans include polycyclic aromatic hydrocarbons, aromatic amines and heterocyclic aromatic amines. These compounds are constituents of cigarette smoke and cooked red meat and are known to be mammary carcinogens in rodents (Nagao et al, 1994). NATs are involved in the metabolism of tobacco and dietary carcinogens, therefore it has been postulated that polymorphisms of the *NAT1* and *NAT2* genes may be linked to variable predisposition to breast cancer.

Studies on the association between the NAT2 acetylator phenotype and breast cancer have been conflicting. It has been suggested by some researchers that fast acetylation may be associated with higher risk to the disease, but this has not been confirmed by others (reviewed in Hein et al, 2000a). Furthermore, it has been observed that the acetylation phenotype of some patients may change during chemotherapy, so NAT2 phenotyping is not the most appropriate way of determining the potential risk associated with the NAT2 polymorphism. Studies using *NAT2* genotyping to predict the phenotype have generally failed to reveal any significant association between disease status and the NAT2 polymorphism (Agundez et al, 1995; Ambrosone et al, 1995; Hunter et al, 1997; Millikan et al, 1998). However, one of these studies has suggested that there may be a relationship between the wild type (*NAT2*4*) allele, responsible for fast acetylation, and the lobular type of breast cancer, which is a relatively rare form of the disease (Agundez et al, 1995). It has also been suggested that the *NAT2* genotype may modify the effects of cigarette smoking and that slow acetylators may be more susceptible to developing breast cancer, due to higher sensitivity to inhaled tobacco carcinogens. Based on evidence that smoking may be involved in breast cancer, the association between the NAT2 acetylation phenotype and breast cancer has been investigated in relation to the smoking history of patients (Ambrosone et al, 1995). Current and past smoking habits appeared to increase the risk for breast cancer among slow acetylators. This increase was more evident in post-menopausal women who smoked, but not in women before menopause (Ambrosone et al, 1996). A similar finding was reported for Chinese women in Taiwan (Huang et al, 1999). These results suggest that slow NAT2 activity may increase susceptibility to breast cancer through decreased capacity of detoxifying the carcinogens present in tobacco smoke via N-acetylation. On the other hand, Hunter et al (1997) failed to find significant difference in the risk for breast cancer between slow acetylator smokers

and non-smokers. In contradiction with the above, a study has reported that it is the fast acetylators, rather than the slow, who may be at higher risk among post-menopausal women with smoking habits (Millikan et al, 1998). In view of these contrasting results, the role of NAT2 polymorphism in relation to smoking habits remains unclear.

The role of *NAT2* polymorphism in breast cancer has also been investigated in relation to diet, in particular to meat consumption. Ambrosone et al (1998) reported that the *NAT2* genotype was not a risk factor for breast cancer and that meat intake did not contribute to that risk. These results were later independently confirmed by another study (Gertig et al, 1999). Therefore, it appears that there is no interaction between any particular *NAT2* genotype and meat intake, considered to be a modest risk factor for breast cancer. However, it is possible that the *NAT2* polymorphism may play a role in breast carcinogenesis via a combined effect with other polymorphic enzymes involved in the metabolism of dietary carcinogenic compounds.

NAT1 is the only form of N-acetyltransferase that has been detected in human breast tissue (Sadrieh et al, 1996). However, the role of NAT1 polymorphism in breast cancer has not been studied extensively. Millikan et al (1998) reported that previous history of smoking may increase the risk of post-menopausal women who are carriers of the *NAT1*10* allele, but failed to find any association between the *NAT1* genotype and breast cancer independently of the smoking habits of the individuals. According to a more recent study, designed to detect eleven *NAT1* alleles, *NAT1*10* was associated with a slightly elevated risk of breast cancer and this was influenced by the smoking habits of the individuals. Moreover, there was substantial over-representation of the *NAT1*11* allele in the group of breast cancer patients, especially among women who smoked and consumed high levels of well done red meat (Zheng et al, 1999). In contrast to the above, a recent report that re-assessed the results of Millikan et al (1998) and further examined the association between of *NAT1*10* and *NAT1*11* alleles and breast cancer has failed to reproduce these findings (Millikan, 2000).

6.2 NAT1 and NAT2 in healthy human intestine

6.2.1 Study design and samples

In view of the potential role of NAT in cancer of the gut, it is important to establish its role in healthy tissue. A study was carried out to investigate the expression of NAT in the intestine, in collaboration with researchers from the Department of Pharmaceutics, University of Washington in Seattle, USA. Four samples of healthy intestinal tissue were genotyped for *NAT1* and *NAT2*, as will be described in this chapter. They were also analysed for NAT1 and NAT2 expression by the collaborators in Seattle using enzymatic assays, Western blotting and immunohistochemistry, as described in Hickman et al, 1998.

The intestines were removed in Seattle from four adult organ donors who had died of road accidents or cerebrovascular incidents, causes unrelated to gastric diseases. The four samples were designated 11/5, 18/5, 22/3 and 24/3. Following dissection, they were rinsed with ice cold 0.9% NaCl to remove particulate material and then cut into eight portions of 30cm each, from the duodenum to the rectum. They were frozen in liquid nitrogen and stored at -70°C, except during their transport to Oxford on dry ice.

6.2.2 Results

For genotyping the four intestinal samples for the *NAT1* and *NAT2* genes, genomic DNA was extracted from homogenised pieces of tissue using proteinase K treatment and phenol:chloroform:isoamyl alcohol (24:24:1 v/v/v) extraction, as described in Materials and Methods (section 2.2.3). The concentration and purity of the extracted genomic DNA was estimated by measuring the absorbance of the preparation at 260nm and 280nm (section 2.2.4). Approximately 0.5-1.0µg of gDNA was used as template in PCR amplifications (section 2.2.7), as part of the following genotyping procedures.

6.2.2.1 NAT1 genotyping

The presence of four of the most common *NAT1* alleles, *NAT1**4, *NAT1**3, *NAT1**10 and *NAT1**11, was examined. Genotyping was performed by a combination of PCR and restriction fragment length polymorphism (PCR-RFLP) or by allele-specific PCR (AS-PCR), as described in section 3.2.1. A 344bp product was obtained from all samples, following amplification with primer pair N769 and N1113 (figure 6.1a). Digestion of the amplification products with *Mbo*II produced a pattern indicative of the *NAT1**4/*4 genotype for samples 11/5, 22/3 and 24/3 (figure 6.1b and table 6.1). The fourth sample, 18/5, could either be a *NAT1**4/*10 or a *NAT1**4/*3 heterozygote, but the *Mbo*II digestion is unable to distinguish between the *NAT1**10 and *NAT1**3 alleles. AS-PCR was, therefore, performed to determine the second allele of sample 18/5, using primers N769 and N1110b. This amplification provided a 340bp product, which confirmed the presence of a *NAT1**10 allele (figure 6.1c). Therefore, sample 18/5 was designated as heterozygous *NAT1**4/*10 (table 6.1).

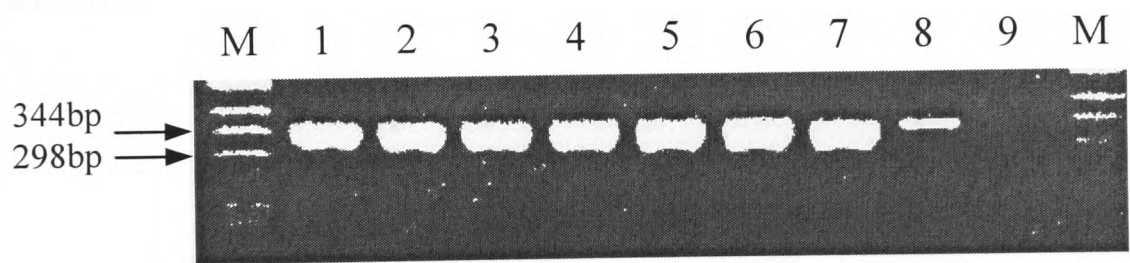


Figure 6.1a: *NAT1* genotyping of human intestine samples. PCR amplification products (10µl) using primer pair N769 and N1113 were loaded on a 2% Metaphor gel. Lanes 1 to 7 were loaded with PCR products from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5) and 24/3 (lanes 6 and 7). Lane 8 is the positive control (sample of known *NAT1**4/*4 genotype) and lane 9 is the negative control (no DNA template in the reaction mix). Lanes M were loaded with 1µg 1kb DNA ladder, as molecular weight marker. All samples provided a 344bp product, except for the negative control (lane 9).

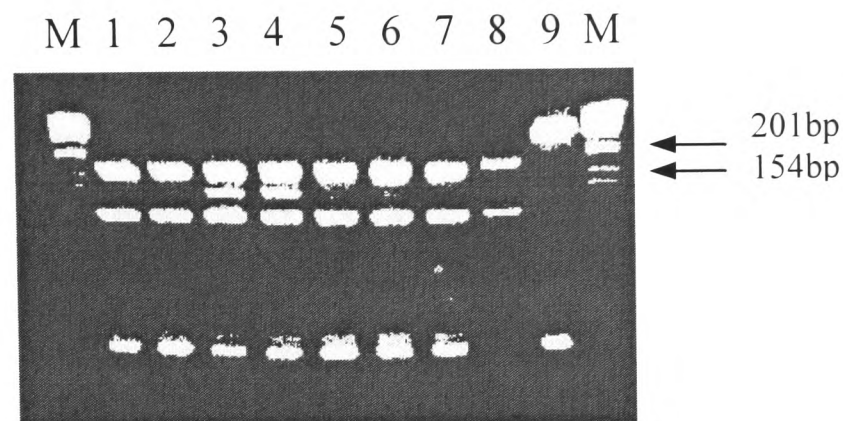


Figure 6.1b: *NAT1* genotyping of human intestine samples. The PCR amplification products (17µl), shown in figure 6.1a, were digested with *Mbo*II at 37°C for approximately 3 hours. A control digestion was included, without restriction enzyme in the reaction mixture. Restriction fragments were separated on a 4% Metaphor gel. Lanes 1 to 7 are *Mbo*II digests from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5) and 24/3 (lanes 6 and 7). Lane 8 was loaded with a sample of known *NAT1**4/*4 genotype, for comparison. Lane 9 was loaded with the undigested control and lanes M with 1µg 1kb DNA ladder, as molecular weight marker. Samples 11/5, 22/3, 24/3, and the positive control, produced four bands 176, 125, 24 and 19 bp in size, indicative of the *NAT1**4/*4 genotype. Sample 18/5 produced five bands 176, 144, 125, 24, 19 bp in size, indicative of a *NAT1**4/*3 or *NAT1**4/*10 heterozygote. Note that under these agarose gel electrophoresis separation conditions, the smallest bands of the digests (24 and 19bp in size) can not be distinguished from the primers used for PCR amplification.

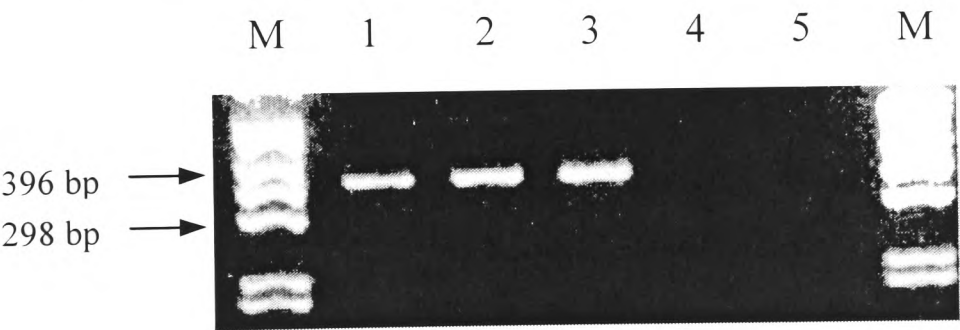


Figure 6.1c: *NAT1* genotyping of human intestine samples. Genomic DNA from sample 18/5 and from two samples of known genotype was amplified by AS-PCR for the *NAT1**10 allele. The products of AS-PCR (10µl) were loaded on a 2% Metaphor gel. Lanes 1 and 2 were loaded with PCR products from sample 18/5. Lanes 3 and 4 were loaded with PCR products from samples of known *NAT1* genotype (*NAT1**4/*10 and *NAT1**4/*4 respectively), for comparison. Lane 5 was loaded with PCR negative control (no DNA template in the reaction mix). Lanes M were loaded with 1µg 1kb DNA ladder, as molecular weight marker. The presence of a 340bp band in lanes 1, 2 and 3 confirmed the presence of a *NAT1**10 allele in sample 18/5 and the *NAT1**4/*10 positive control. The *NAT1**4/*4 control (lane 4) did not provide PCR product, as expected.

6.2.2.2 NAT2 genotyping

The four intestine samples were also examined for the presence of the most common NAT2 alleles in the Caucasian population. The genotyping method used is able to detect alleles NAT2*4, NAT2*5A, NAT2*5B, NAT2*5C, NAT2*6A, and NAT2*7B. Genotyping was performed by PCR-RFLP, as described in section 3.2.2. Two PCR amplifications were performed from genomic DNA, one with primer pair NAT-Hu7 and NAT-Hu8 (figure 6.2a) and the other with primers NAT-Hu14 and NAT-Hu16 (figure 6.2b), producing a 840bp and a 1000bp fragment respectively. The products of the former PCR were subsequently digested with *KpnI* and *TaqI* (figure 6.2c), while the products of the latter PCR were digested with *BamHI* and *DdeI* (figure 6.2d). The combination of digestion patterns was used to determine the NAT2 genotypes, as described in section 3.2.2. To check specificity of amplification, the PCR product with primers NAT-Hu14 and NAT-Hu16 was also digested with *HindIII* (figure 6.2g).

Sample 18/5 was homozygous for the wild type allele (NAT2*4/*4), responsible for fast acetylation phenotype, while samples 24/3 and 22/3 carried only slow alleles. Sample 24/3 was homozygous NAT2*5B/*5B, while sample 22/3 was a NAT2*5B/*6A heterozygote. Based on the genotyping method used here, sample 11/5 could be either a NAT2*4/*5B or a NAT2*5A/*5C heterozygote. However, NAT2 phenotyping carried out by colleagues (Hickman et al, 1998) demonstrated sample 11/5 to be a fast acetylator. This confirmed that the genotype of sample 11/5 is NAT2*4/*5B. The NAT1 and NAT2 genotypes are summarised in table 6.1.

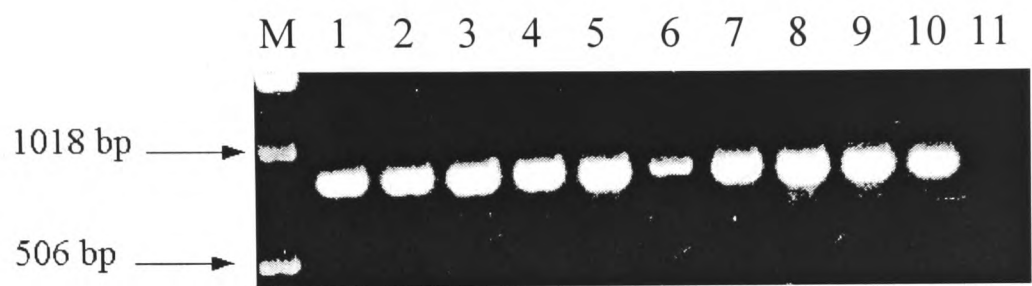


Figure 6.2a: *NAT2* genotyping of human intestine samples. PCR amplification products (5µl), using primer pair NAT-Hu7 and NAT-Hu8, were loaded on a 1% agarose gel. Lanes 1 to 9 were loaded with PCR products from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5 and 6) and 24/3 (lanes 7, 8 and 9). Lane 10 is the positive control (sample of known *NAT2**4/*5*B* genotype) and lane 11 is the negative control (no DNA template in the reaction mix). Lane M was loaded with 1µg 1kb DNA ladder, as molecular weight marker. All samples provided the expected 840bp product, except for the negative control (lane 11).

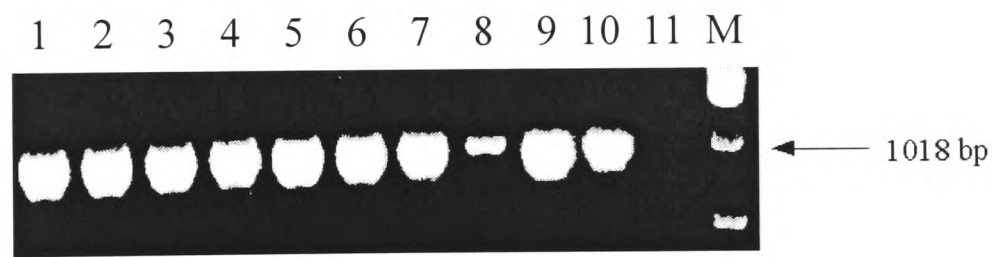


Figure 6.2b: *NAT2* genotyping of human intestine samples. PCR amplification products (5µl), using primer pair NAT-Hu14 and NAT-Hu16, were loaded on a 1% agarose gel. Lanes 1 to 9 were loaded with PCR products from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5 and 6) and 24/3 (lanes 7, 8 and 9). Lane 10 is the positive control (sample of known *NAT2**4/*5*B* genotype) and lane 11 is the negative control (no DNA template in the reaction mix). Lane M was loaded with 1µg 1kb DNA ladder, as molecular weight marker. All samples gave the expected 1000bp product, except for the negative control (lane 11).



Figure 6.2c: *NAT2* genotyping of human intestine samples. The PCR amplification products (10µl), shown in figure 6.2a, were digested with *KpnI* at 37°C for approximately 3 hours. A control digestion was included, without restriction enzyme in the reaction mixture. Restriction fragments were separated on a 1% agarose gel. Lanes 1 to 7 are *KpnI* digests from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5) and 24/3 (lanes 6 and 7). Lane 8 was loaded with a sample of known genotype (*NAT2**4/*5*B*), for comparison. Lane 9 was loaded with the undigested control (840bp) and lanes M with 1µg 1kb DNA ladder, as molecular weight marker. Sample 18/5 produced two bands 450 and 390bp in size, while sample 24/3 was not digested with *KpnI* and produced a single band 840bp in size. Samples 11/5 and 22/3 produced three bands 840, 450 and 390bp in size.

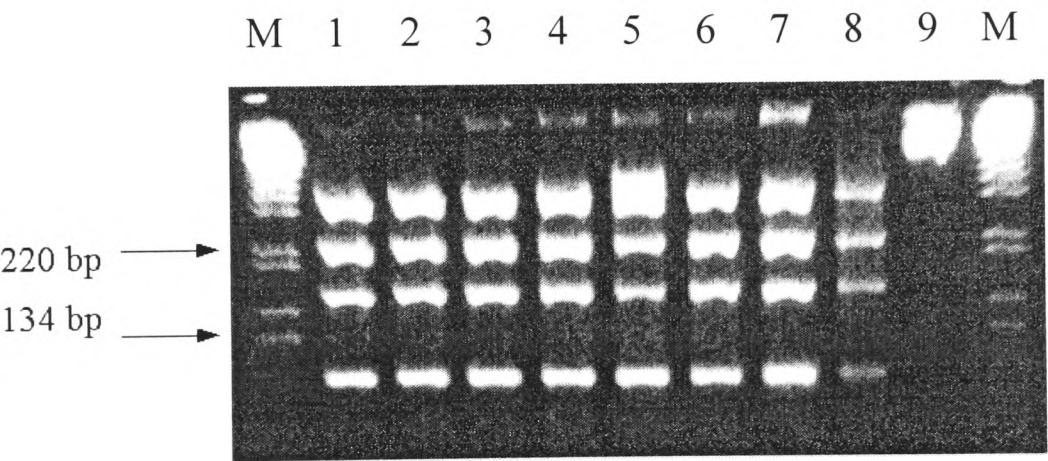


Figure 6.2d: *NAT2* genotyping of human intestine samples. The PCR amplification products (10µl), shown in figure 6.2a were digested with *TaqI* at 65°C for around 3 hours. A control digestion was included, without restriction enzyme in the reaction mixture. Restriction fragments were separated on a 3% Metaphor gel. Lanes 1 to 7 are *TaqI* digests from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5) and 24/3 (lanes 6 and 7). Lane 8 was loaded with a sample of known genotype (*NAT2**4/*5*B*), for comparison. Lane 9 was loaded with the undigested control (840bp) and lanes M with 1µg 1kb DNA ladder, as molecular weight marker. Samples 11/5, 18/5 and 24/3 produced four bands 330, 230, 170 and 110bp in size, while sample 22/3 produced five bands 400, 330, 230, 170 and 110bp in size.

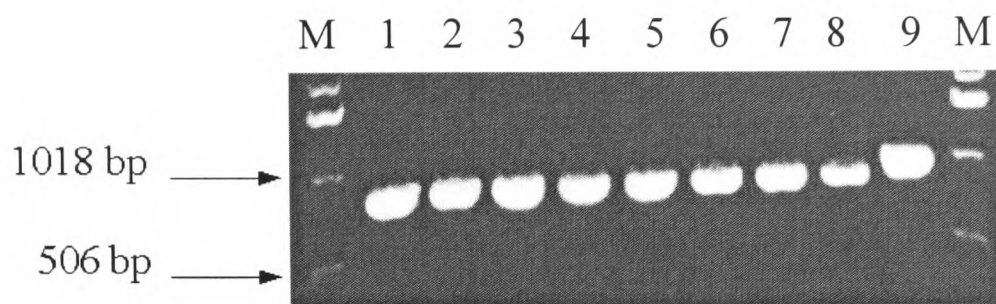


Figure 6.2e: *NAT2* genotyping of human intestine samples. The PCR amplification products (10µl), shown in figure 6.2b, were digested with *Bam*HI at 37°C for approximately 3 hours. A control digestion was included, without restriction enzyme in the reaction mixture. Restriction fragments were separated on a 1% agarose gel. Lanes 1 to 7 are *Bam*HI digests from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5) and 24/3 (lanes 6 and 7). Lane 8 was loaded with a sample of known genotype (*NAT2**4/*5*B*), for comparison. Lane 9 was loaded with the undigested control (1000bp) and lanes M with 1µg 1kb DNA ladder, as molecular weight marker. All samples produced two bands 850 and 150bp in size, of which only the largest one (850bp) was visible.

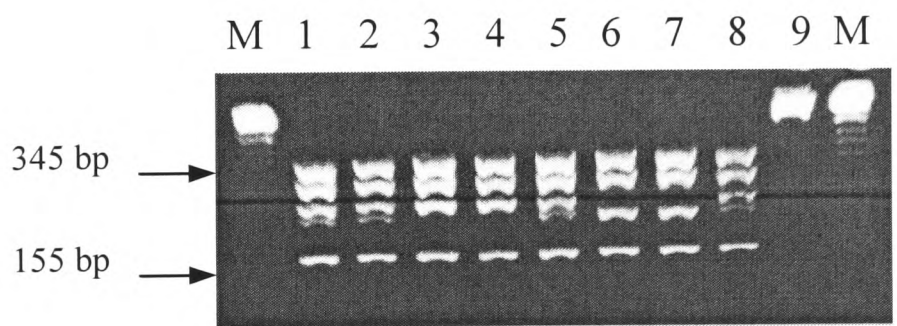


Figure 6.2f: *NAT2* genotyping of human intestine samples. The PCR amplification products (10µl) shown in figure 6.2b were digested with *Dde*I at 37°C for approximately 3 hours. A control digestion was included, without restriction enzyme in the reaction mixture. Restriction fragments were separated on a 3% Metaphor gel. Lanes 1 to 7 are *Dde*I digests from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5) and 24/3 (lanes 6 and 7). Lane 8 was loaded with a sample of known genotype (*NAT2**4/*5*B*), for comparison. Lane 9 was loaded with the undigested control (1000bp) and lanes M with 1µg 1kb DNA ladder, as molecular weight marker. Samples 11/5 and 22/3 produced five bands 345, 280, 220, 200 and 155bp in size. Sample 18/5 produced four bands 345, 280, 220 and 155bp in size, while sample 24/3 produced four bands 345, 280, 200 and 155bp in size.

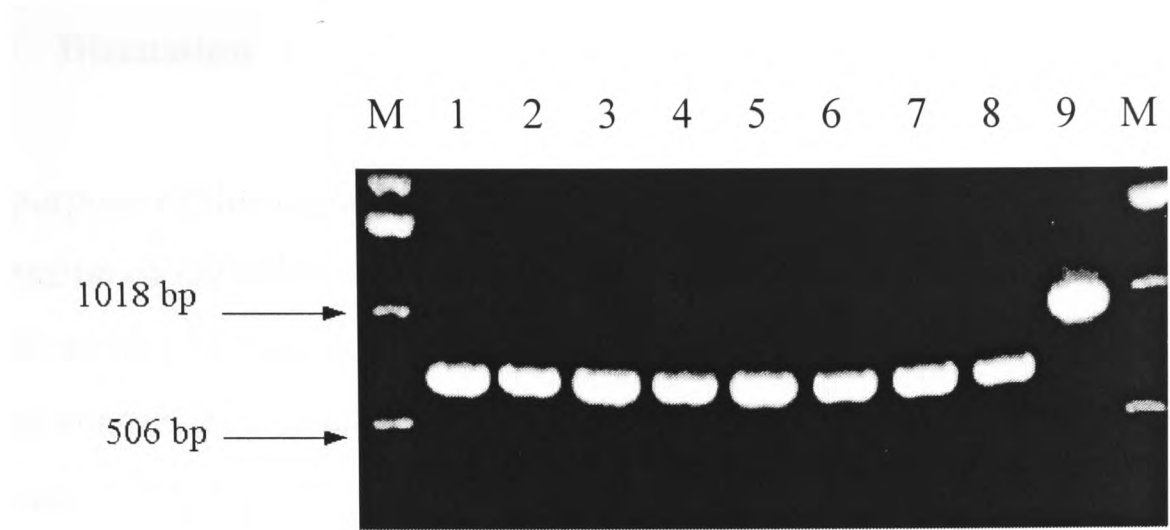


Figure 6.2g: *NAT2* genotyping of human intestine samples. The PCR amplification products (10µl), shown in figure 6.2b, were digested with *Hind*II at 37°C for approximately 3 hours. A control digestion was included, without restriction enzyme in the reaction mixture and PCR product as template. Restriction fragments were separated on a 1% agarose gel. Lanes 1 to 7 are *Hind*II digests from samples 11/5 (lanes 1 and 2), 18/5 (lanes 3 and 4), 22/3 (lane 5) and 24/3 (lanes 6 and 7). Lane 8 was loaded with a sample of known genotype (*NAT2**4/*5B), for comparison. Lane 9 was loaded with the undigested control (1000bp) and lanes M with 1µg 1kb DNA ladder, as molecular weight marker. All samples produced two bands 700 and 300bp in size, of which only the largest is visible, indicating that the digested PCR product has derived from the *NAT2* gene.

Table 6.1: The *NAT1* and *NAT2* genotypes and the predicted NAT2 acetylator status of the four intestine samples. ‘F’ and ‘S’ indicate fast and slow alleles respectively. The NAT2 phenotype was predicted on the basis of the genotype.

Sample	<i>NAT1</i> genotype	<i>NAT2</i> genotype	Predicted NAT2 phenotype
11/5	1*4/1*4	2*4/2*5B	F S
18/5	1*4/1*10 [†]	2*4/2*4	F F
22/3	1*4/1*4	2*5B/2*6A	S S
24/3	1*4/1*4	2*5B/2*5B	S S

[†] Following the discovery of the *NAT1**14 allele, the genotype of this sample was determined to be *NAT1**4/*14 by Dr V Smelt (section 6.2.3; Hickman et al, 1998).

6.2.3 Discussion

The purpose of this collaborative study (Hickman et al, 1998) was to investigate expression of the NAT isoenzymes in healthy human intestine, in order to improve understanding of their potential role in cancer and other diseases. Other studies have also examined expression of NATs in the intestine. Jenne (1965) was the first to demonstrate that samples of intestinal mucosa were able to perform N-acetylation of isoniazid, which was later identified as a NAT2-specific substrate, at levels comparable with the liver. Expression of both NAT1 and NAT2 was later demonstrated by a number of studies (Ilett et al, 1994; Windmill et al, 1997; Windmill et al, 2000), indicating that both NAT isoenzymes may be involved in colon carcinogenesis. The aim of the work, part of which is presented here, was to use a combination of techniques to study NAT in different parts of the intestine, in a more extensive and detailed manner than previously reported.

Expression of NATs was examined along the full length, as well as in specific histological regions of the intestine, by collaborators. Although both NAT1 and NAT2 were present in all sections of the intestine, the distribution of the enzymes was not uniform. Maximum NAT1 activity for pABA was measured at the region between the small and the large intestine, except for sample 18/5 which showed a gradual increase in NAT1 activity from the duodenum to the rectum. Interestingly, sample 18/5 had the lowest NAT1 activity (table 6.2). The amount of NAT1 protein along the length of the intestine, as determined by Western blotting, mirrored the levels of NAT1 activity with pABA. NAT2 activity also varied along the length of the intestine, being highest at the duodenum and showing a decrease towards the more distal sections. This change was more clearly observed for the two samples that were shown here to carry at least one copy of the fast *NAT2*4* allele (11/5 and 28/5), but it was less pronounced for the samples with the slow acetylator phenotype (samples 22/3 and 24/3).

Hickman et al (1998) also examined the cellular distribution of NAT protein in different sections of the intestine, by immunohistochemistry. In the small intestine, NAT was found in the cytoplasm of epithelial cells along the whole length of

intestinal villi, from the tip to the crypts. NAT protein was restricted to the villi and was not found in the underlying muscle layer. The situation was slightly different in the colon, where cells of the crypts did not appear to express NAT.

The genotyping of these four intestinal samples (section 6.2.2) allowed investigation of the relationship between the NAT genotype and the acetylator phenotype. N-acetyltransferase activity values obtained by Hickman et al (1998) from the most proximal section of the intestine for pABA and SMZ are presented in table 6.2. The observed NAT2 phenotype was in agreement with that predicted from genotyping (table 6.1). Samples 22/3 and 24/3 were identified as carrying two slow *NAT2* alleles and had low SMZ activity, as expected. Sample 18/5 was a homozygote for the *NAT2*4* allele, which is known to be responsible for fast acetylation, and had the highest NAT2 activity of all. Based on the genotyping method used in this and previous studies (Hickman and Sim, 1991; Hickman et al, 1992; Smelt et al, 1998), the genotype of sample 11/5 could be either *NAT2*4/*5B* or *NAT2*5A/*5C*. However, this sample showed NAT2 activity with SMZ indicative of a fast acetylator heterozygote. Activity was lower than that of a homozygous fast acetylator (e.g. sample 18/5), but significantly higher than that expected for slow acetylators (e.g. samples 22/3 and 24/3). Therefore, it was concluded that sample 11/5 was heterozygous for the fast *NAT2*4* allele and its genotype was *NAT2*4/*5B*. Although the number of the samples was very small, the trimodal distribution typically observed for NAT2 activity (fast, intermediate and slow) was evident.

Table 6.2: NAT1 and NAT2 activities of the four intestine samples, measured with pABA and SMZ respectively (adapted from Hickman et al, 1998). Standard deviation is provided in parentheses. All assays were carried out using tissue from the most proximal section of the intestines, but they were representative of all sections. ‘F’ represents the fast and ‘S’ the slow acetylator forms of the NAT2 isoenzyme.

Sample	pABA activity (nmol/min/mg protein)	SMZ activity (nmol/min/mg protein)	NAT2 phenotype (as predicted from the genotype)
11/5	6.1 (0.2)	0.85 (0.01)	F S
18/5	2.1 (0.1)	1.18 (0.03)	F F
22/3	6.9 (0.3)	0.22 (0.02)	S S
24/3	7.5 (0.1)	0.23 (0.01)	S S

NAT1 activity in these four intestine samples (table 6.2) showed significant inter-individual variation (3.6-fold). This was mainly due to sample 18/5, which showed considerably lower NAT1 activity, compared with the rest of the samples (Hickman et al, 1998). As described in this Chapter, this sample was originally genotyped as *NAT1**4/*10, but following the discovery of the *NAT1**14 allele (Hughes et al, 1998; Payton and Sim, 1998), its NAT1 genotype was re-examined and identified to be a *NAT1**4/*14 heterozygote (Hickman et al, 1998). This is in agreement with the original observations that the *NAT1**14 allele, which carries a point mutation causing an Arg¹⁸⁷→Gln¹⁸⁷ amino acid substitution (table 1.4), may be associated with slow NAT1 phenotype (Hughes et al, 1998; Payton and Sim, 1998). The remaining three intestine samples were all *NAT1**4/*4 homozygotes and showed limited variation in activity (table 6.2), so that the effect of other *NAT1* alleles on acetylation activity could not be evaluated.

Both NAT1 and NAT2 enzymes were detected in human intestine (Hickman et al, 1998). However, the NAT1 specific activity was significantly higher than that of NAT2. This is in agreement with other studies (Ilett et al, 1994) and suggests that NAT1 may be more important than NAT2 in the local metabolism of colon carcinogens. Prevalence of NAT1 in the intestinal tissue may explain the lack of clear correlation between the NAT2 genotype and susceptibility to colorectal cancer, as reported by many researchers (section 6.1.1). However, it is more difficult to assess the effect of *NAT1* genotype on susceptibility to colon carcinogenesis, due to the unclear genotype to phenotype relationship for NAT1 (section 1.3.3.2).

Animal models have been useful for studying the mechanisms of chemically induced colon carcinogenesis. NAT expression in the intestine has been examined for a number of common laboratory species. Similarly to humans, rabbits possess two NAT isoforms, and they exhibit NAT2 polymorphism (section 1.4). Reeves et al (1991) have demonstrated by Western blotting that both NAT isoenzymes are present in rabbit duodenum tissue and suggested that NAT2 is the predominant form. The distribution of NAT2 in the gastrointestinal tract was later investigated in rapid acetylators rabbits by activity assays with SMZ and Western blotting. It was demonstrated that NAT2 is predominantly expressed in the proximal sections of the

small intestine, especially in the duodenum and jejunum, but its expression decreases along the length of the gastrointestinal tract, so that the NAT1 is the major form expressed in colon (Reeves et al, 1992). Furthermore, the role of both NAT isoforms in the activation of colon carcinogens has been studied in the rabbit and it has been suggested that the NAT2 polymorphism may affect risk for carcinogenesis in the colon (Ilett et al, 1991).

Studies of NAT expression in model rodents have shown that NATs are present in the intestine of the hamster, rat and mouse. Both NAT1 and NAT2 were found in hamster intestinal tissue, although their distribution was not uniform along its length (Ware and Svensson, 1996). Rat intestinal tissue also produces NAT1 and NAT2, but the activity levels for both isoenzymes do not change throughout the gastrointestinal tract (Ware and Svensson, 1996).

The mouse is a popular model for studying NAT, due to its very well characterised NAT2 polymorphism. Moreover, it has been established that mouse NAT2 is similar to human NAT1 in terms of specificity towards some important substrates like pABA and pABGlu (section 1.4.1). In the mouse, NAT2 has been detected in the whole length of the intestine, while NAT1 was only found in the small intestine (Ware and Svensson, 1996). Immunohistochemistry has also confirmed the presence of murine NAT2 in the small intestine (Stanley et al, 1997). However, the role of mouse NAT2 in the bioactivation of colon carcinogens, as measured by the rate of DNA adduct formation, has not been assessed yet. As it is difficult to draw conclusions about the implication of any human *NAT1* genotype in the process of chemically induced carcinogenesis in the human colon, using the well-characterised mouse NAT2 polymorphism as a model for studying human NAT1 polymorphism could be of great potential.

6.3 NAT1 and NAT2 in human bladder cancer material

6.3.1 Study design and samples

Arylamine N-acetyltransferases, and the human *NAT2* polymorphism in particular, have been strongly linked to bladder cancer based on epidemiological evidence (section 6.1.2). In view of the work that reports loss of *NAT* in bladder cancer, as well as in order to investigate the role of the *NAT* locus in the initiation and progression of bladder cancer at the molecular level, the presence of *NAT* genes was investigated in cancerous bladder material. In collaboration with Dr Michael Stacey (Department of Pharmacology, University of Oxford), FISH was carried out to examine the number of copies of each one of the two *NAT* genes in interphase nuclei of tumour cells from a population of bladder cancer patients. The work described in this section represents part of that collaborative study (Stacey et al, 1999).

Barbotage samples were obtained from bladder cancer outpatients of the Urology Department in Churchill Hospital, Oxford, UK. By inserting an endoscope into the bladder, a jet of saline solution was directed towards the location of the tumour and the dislodged cells were collected using a syringe. These barbotage samples were stored at 4°C, then transported to the Department of Pharmacology and used for the preparation of slides for FISH within 16h. Interphase spreads from barbotage samples were prepared as described in Materials and Methods (section 2.3.3) and stored at -20°C. For most samples, information about the grade and stage of the tumour was available.

In the absence of healthy bladder cells from the individuals participating in the study, the bladder carcinoma cell line RT112 and blood lymphocytes from healthy individuals were used as controls for hybridisation, as described in section 3.4.2. Hybridisation efficiency was considered acceptable when at least 95% of the control nuclei showed the expected number of *NAT* and chromosome 8 specific signals. This provided an estimate of the number of hybridisation artifacts expected when using barbotage cells.

6.3.2 Results

The presence of the *NAT1* gene in bladder cancer material was investigated using the recently developed *NAT1* cosmid clone, described in section 3.3.1, as probe. For FISH analysis of the *NAT2* gene, a previously described *NAT2*-specific cosmid (Franke et al, 1994) was used as probe, together with the *NAT2*-specific cosmid probe described in section 3.3.1.

Six samples of cancerous material were investigated for loss of the *NAT1* and *NAT2* genes. The number of examined nuclei for each sample depended on the number and quality of cells available from each patient, as well as the amount of informative hybridisations. This number varied nearly 4.5-fold, but was more than 80 nuclei per sample, on average. Most of the cells examined were normal, showing a diploid number of *NAT*-specific signals and a pair of chromosome 8-specific signals (figure 6.3). However, many cells showed loss of one of the *NAT*-specific signals, usually with retention of both chromosome 8-specific signals (figure 6.4). In a few cells, more than the normal number of *NAT*-specific and chromosome 8-specific signals was observed (figure 6.5), but this was not a very common finding. These results are summarised in table 6.3.

Sample number	Number of <i>NAT1</i> signals per cell (%)					Number of <i>NAT2</i> signals per cell (%)					Number of 8 centromere signals per cell (%)					Tumour grade	Tumour stage
	Number of examined nuclei	1	2	3	4	Number of examined nuclei	1	2	3	4	Number of examined nuclei	1	2	3	4		
4	100	38	62			59	59	41			59	8	92			G2	T2
12	124	40	49	11		100	37	63			100		100			N/A	N/A
47	36		100			28		100			28		100			?	T2
49	100	24	76			100	30	69	1		100	6	82	8	4	G2	T2
50	100	10	84		6	100	12	88			100	2	98			N/A	N/A
52	88	5	93	1	1	73	1	96		3	73		99		1	G2	?

N/A: Data not available, usually due to treatment of the tumour by diathermy.
 ‘?’: Tumour grade or stage could not be determined, although material was available.

Table 6.3: Tumour samples investigated for the loss of *NAT1* and *NAT2* genes. The number of *NAT1*, *NAT2* and 8 centromere signals per cell is presented as a percentage of the total number of examined nuclei. If available, the grade and stage of the tumour are also indicated.

On the basis of how frequently gene loss was observed, the samples were classified in groups of those that showed *NAT* loss in more than 30%, between 10% and 30%, or less than 10% of their examined nuclei. When grouped according to loss of either *NAT1* or *NAT2* genes in a similar way, it became evident that all samples fell into the same category for both types of deletion. This indicated that in samples where loss of one *NAT* locus was observed, the other *NAT* locus was also missing, and suggested that both *NAT* loci may be lost together in bladder cancer tumours (table 6.4).

Table 6.4: Samples with *NAT1* or *NAT2* gene deletion classified according to their observed frequency of loss.

The two genes appear to be deleted together in these six bladder cancer samples.

	Samples showing loss of <i>NAT1</i> or <i>NAT2</i> genes	
Categories	<i>NAT1</i>	<i>NAT2</i>
more than 30%	4, 12	4, 12
10-30%	49, 50	49, 50
less than 10%	47, 52	47, 52

In few cases of bladder tumours, gain of the *NAT* loci was observed. This uncommon finding was seen for both *NAT* genes, but most commonly in *NAT1* hybridisations (table 6.3). In contrast to *NAT* deletion, gain of one *NAT* gene did not always reflect gain of the other. For example, sample 52 showed infrequent multiplication of both *NAT1* and *NAT2*, but this was not evident for samples 12 and 50 which had gained copies of *NAT1* only, or sample 49 which had only extra *NAT2* copies (table 6.3). In this small sample, there appeared to be no obvious correlation between a particular grade or stage of tumour and gain or loss of the *NAT* signals (table 6.3).

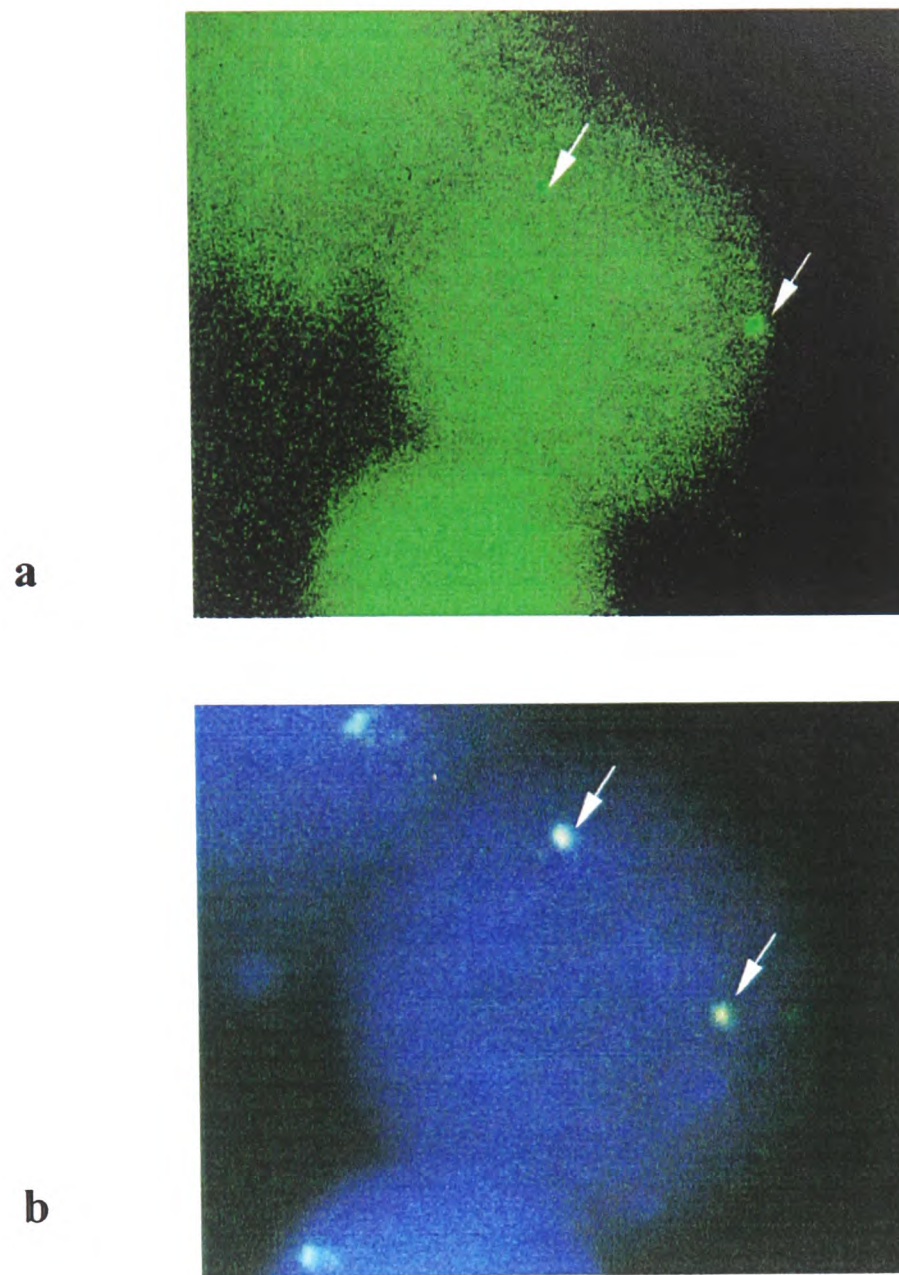


Figure 6.3: Fluorescence *in situ* hybridisation with *NAT1* cosmid probe on bladder cancer cells from barbotage fluid: An example of cell with diploid number of *NAT1* genes.

Approximately 100ng of biotin-labelled *NAT1* cosmid, 10ng of digoxigenin labelled chromosome 8 centromeric probe and 2 μ g human COT-1 DNA were hybridised to the slides at 37°C overnight. Following washes, the chromosome 8 centromeric probe was detected through rhodamine-conjugated anti-digoxigenin Fab fragments. Detection of the *NAT1* signal was achieved through FITC- conjugated avidin and amplified with biotinylated goat anti-avidin antibodies and a second layer of FITC-conjugated avidin.

(a) Two *NAT1* signals were detected (arrows), indicating that the cell shown here had the normal (diploid) number of *NAT1* genes. FITC signals were visualised under the fluorescent microscope, using the green filter.

(b) The same cell also showed two signals for chromosome 8 centromere (arrows), indicating that it contains two copies of chromosome 8. Rhodamine signals were visualised on DAPI-stained nuclei, using the red and blue filters under the fluorescence microscope.

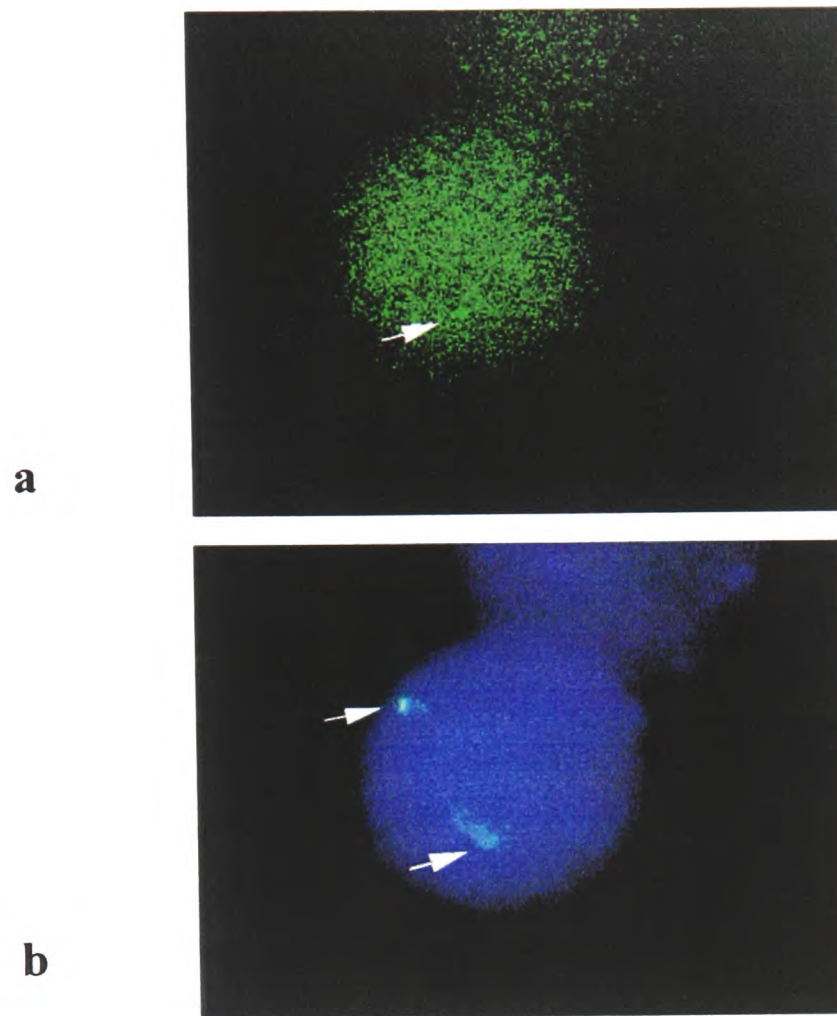


Figure 6.4: Fluorescence *in situ* hybridisation with *NAT1* cosmid probe on bladder cancer cells from barbotage fluid: An example of cell with loss of one copy of the *NAT1* gene.

Hybridisation and signal detection were performed as described in figure 6.3.

(a) Only one *NAT1* signal was detected in this cell (arrow), indicating deletion of the second copy of the *NAT1* gene. The FITC signal was visualised under the fluorescent microscope, using the green filter.

(b) The same cell showed two signals for chromosome 8 centromere (arrows), indicating that it contains two copies of chromosome 8. Rhodamine signals were visualised on DAPI-stained nuclei using the red and blue filters under the fluorescence microscope.

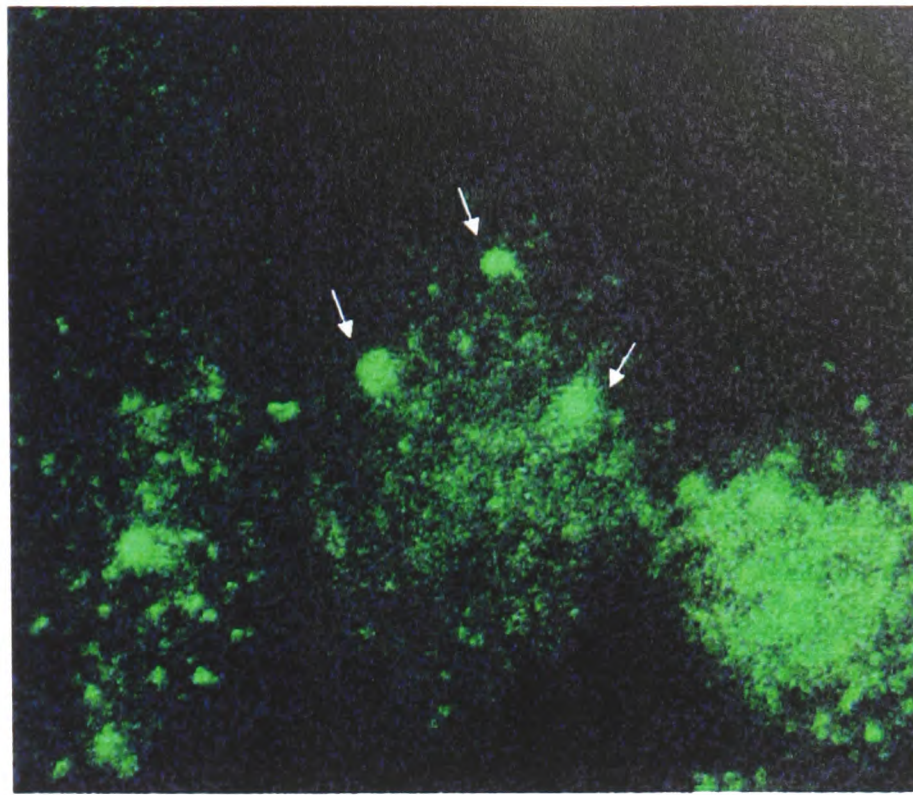


Figure 6.5: Fluorescence *in situ* hybridisation with *NAT1* cosmid probe on bladder cancer cells from barbotage fluid: An example of cells with a triploid number of *NAT1* genes.

Hybridisation was performed as described in figure 6.3. FITC signals were visualised under the fluorescence microscope, using the green filter.

Three *NAT1* signals were detected (arrows), indicating gain of an additional copy of the *NAT1* gene.

6.3.3 Discussion

The aim of the collaborative study (Stacey et al, 1999), part of which is described here, has been to examine the chromosomal abnormalities associated with chromosomal region 8p22, where the human *NAT* genes are localised, in bladder cancer. FISH was used to analyse the presence of *NAT1* and *NAT2* in bladder cancer cells obtained from barbotage washes.

Other methods have also been used for investigation of the genetic events involved in bladder carcinogenesis. LOH studies use polymorphic microsatellite or RFLP markers to detect deletions of genomic material in cancer cells. This approach has been undertaken successfully in the past for studying bladder cancer material and has allowed construction of an allelotype for this type of cancer (Knowles et al, 1994). The well-characterised polymorphism of the *NAT2* gene has been used to study LOH at the *NAT2* locus in bladder cancers with deletion of 8p (Knowles et al, 1993; Schnakenberg et al, 1998; Thygesen et al, 1999). FISH analysis eliminates the need for established polymorphic markers in the region of interest and this is an advantage over LOH methodologies. FISH also allows detection of chromosomal aberrations other than deletions, such as duplications and translocations of the examined region. On the other hand, FISH requires availability of cloned DNA fragments of the genomic region of interest, in the form of cosmid, BAC, PAC, or YAC clones, to be used as probes. FISH is a very powerful technique and, with selection of appropriate DNA probes, it can detect most of the chromosomal aberrations involved in carcinogenesis.

Two other methods have been used for the identification of chromosomal regions involved in carcinogenesis, but these have not been extensively applied to bladder cancer. Comparative genomic hybridisation (CGH) can detect variations of chromosomal copy number, but does not reveal translocations, inversions and other aberrations that do not change the number of copies. The use of CGH in bladder cancer has confirmed findings obtained with other methods and has identified a few more chromosomal regions of interest. However, its application in this disease has been limited (Kallioniemi et al, 1995). Representational difference analysis (RDA) is

suitable for the detection of loss or amplification of chromosomal regions in tumours, but has not been applied to bladder cancer (Lisitsyn et al, 1995).

For FISH analysis with the *NAT1*- and *NAT2*-specific cosmid probes (section 3.3.1), bladder cancer cells were extracted from barbotage washes of patients. Exfoliated bladder cells extracted from urine have been used in the past for FISH analysis of the *NAT* locus, but the quality and number of cells recovered from that source was very low (Stacey et al, 1996). In this respect, the barbotage material used in the present study has been a much more reliable source of cells. To avoid problems related with selection and subsequent enrichment of particular cell types from barbotage fluid during *in vitro* culture, slides were prepared directly from barbotage washes without previous culture. As a result, metaphase spreads could not be prepared and interphase nuclei were used instead.

Samples of healthy bladder cells from the individuals who provided the barbotage washes were not available for use as controls. Therefore, the efficiency of hybridisation was assessed using the bladder carcinoma cell line RT112, as well as cultured lymphocytes from the blood of healthy individuals, as described in section 3.4.2. It should be noted, however, that these control experiments were performed on cultured cells and were, therefore, expected to produce better quality hybridisation signals, compared with uncultured barbotage cells. Given these limitations, loss of the *NAT* signal in less than 10% of the examined nuclei (table 6.4) was attributed to possible hybridisation artifacts, while loss of *NAT* signal in more than 10% of the nuclei (table 6.4) was considered to represent true loss of the 8p22 region. The gain of signal observed for some of the bladder cancer samples (table 6.3) was never observed in control hybridisations with cultured cells and therefore was unlikely to be a hybridisation artifact. In conclusion, the results obtained here show that the *NAT1* and *NAT2* cosmid probes can be reliably used for FISH analysis of tumour cells for chromosomal aberrations involving the *NAT* region on 8p22. However, differences in the quality of individual cell preparations should be taken into account during interpretation of the results.

To allow estimation of the frequency of *NAT* loss or gain, a large number of nuclei should be examined. More than one hundred nuclei were scored for each sample in

similar studies (Thygesen et al, 1999). In this study, the number of examined cells was typically around one hundred and, with the exception of sample 47, at least fifty-nine (table 6.3). In the case of sample 47, cells consistently showed the normal diploid number of *NAT1* and *NAT2* signals, so scoring additional nuclei was not considered necessary. The number of examined cells largely depended on the number and integrity of the cells extracted from each barbotage wash, the hybridisation efficiency and the number of informative nuclei.

In most cases, information about the stage and grade of bladder tumour was available, so it was possible to investigate the correlation between aggressiveness of the tumour and the frequency of *NAT* deletion. However, two of the patients had previously been treated by diathermy for bladder cancer at the papillary and non-invasive tumour stage (Ta), so the relevant data were unavailable. Sample 4 (table 6.3), which showed invasion of the superficial muscle layer (tumour stage T2) and medium degree of cell differentiation (tumour grade G2), had high occurrence of *NAT1* and *NAT2* deletion (38% and 59% respectively). Unfortunately, a similar observation was not possible for sample 12, which also showed *NAT* deletion in more than 30% of the examined nuclei, due to lack of information about the grade and stage of the tumour. However, another tumour classified as G2/T2 (sample 49) also had relatively high frequency of *NAT1* and *NAT2* deletion (24% and 30% respectively). Taking into account the results from the entire bladder cancer population ($n=52$) used in this collaborative project and described in Stacey et al (1999), no obvious correlation could be observed between the grade or stage of bladder cancer and the frequency of *NAT* deletion in the barbotage samples examined.

It has been possible to examine cells from bladder cancer samples for both *NAT* genes (table 6.4). The results indicated that the two *NAT* loci tend to be deleted at similar frequencies in individual samples. The *NAT1* and *NAT2* genes are located relatively close to each other within a distance of 170-360kb on chromosome 8 (Matas et al, 1997). Therefore, the critical region on chromosome 8 appears to be longer than the distance between the two *NAT* genes. This is an important finding in view of other studies reporting loss of chromosomal regions in bladder tumours, with chromosomes 9, 11p, 17p, 8p, 4p and 13q most commonly affected (Sandberg, 1992; Sandberg and Berger, 1994; Marano et al, 2000). The loss of fragments of chromosome 9 is very

widespread in bladder cancer and is the most common abnormality in low stage and grade papillary tumours. As a consequence, it has been considered as a potential initiating event for this malignant disease. Fine deletion mapping using a large number of bladder cancer cases has identified three distinct regions of loss associated with bladder carcinogenesis (Knowles, 1999 for review). LOH of chromosome 11p is also common and has been associated with higher grade of bladder cancer. Region 11p15, an area also deleted in breast and testicular cancers, has been defined as harbouring a tumour suppressor gene (Knowles, 1999 for review). Deletions of chromosome 4p have been detected in several types of cancer, including nearly a quarter of the bladder tumours examined. A short genomic region on 4p16.3 appears to be involved in the majority of chromosome 4 deletions, but the identity of the involved gene remains unknown (Bell et al, 1996).

A number of studies have implicated the 8p region in bladder cancer. A LOH study has shown that this region is commonly deleted in tumours, especially those of high grade and stage (Knowles et al, 1993). The region of the *NAT* loci on 8p22 has been considered as particularly important for this type of cancer (Tackle and Knowles, 1996; Wagner et al, 1997; Schnakenberg et al, 1998), although 8p12 has also been implicated (Knowles, 1999 for review). These findings have also been confirmed by a recent study that used the *NAT2* gene as polymorphic marker for LOH study, in combination with FISH analysis at the *NAT2* locus (Thygesen et al, 1999). Two genes, encoding for DNA polymerase beta (*POLB*) and the beta isoform of the catalytic subunit of protein phosphatase 2A (*PPP2CB*), previously mapped on region 8p12-p11.1 have been considered as candidate tumour suppressor genes. However, mutation analysis of these genes in bladder tumours and bladder cancer cell lines did not reveal any polymorphisms associated with the disease and it was concluded that they should not be the targets for deletion of 8p (Eydmann and Knowles, 1997).

The region on 8p22 involved in bladder cancer may also be important for other types of cancer. Detailed LOH and deletion mapping has identified 8p22 as a region of common loss in colorectal carcinomas (Fujiwara et al, 1993; Farrington et al, 1996) and in prostate cancer (Macoska et al, 1995; Crundwell et al, 1996). These studies have also suggested the presence of at least two tumour suppressor genes (TSGs) in the region close to the *NAT* loci. Although the identity of the TSGs in the 8p22 region

remains unknown, their presence is further supported by monochromosome transfer experiments. Introduction of a healthy chromosome 8 into colorectal cancer cell lines that lacked the 8p22 region suppressed the tumour phenotype in some cases (Gustafson et al, 1996). Specific LOH for *NAT2* has not been observed in colorectal cancer material (Hubbard et al, 1997). On the basis of their involvement in the metabolism of arylamine carcinogens, the *NAT* genes have been demonstrated to play an important role in the process of bladder carcinogenesis (section 6.1.2), but are unlikely to act as TSGs themselves. *NAT2*, in particular, is not consistently missing in tumours that show 8p deletions (Schnakenberg et al, 1998).

The results presented here are a representative part of a larger study (Stacey et al, 1999) on the involvement of 8p22 chromosomal region in bladder cancer. With the use of cosmid probes and FISH, the presence of both *NAT1* and *NAT2* genes was examined in bladder tumour cells. There is strong evidence that this region on the small arm of chromosome 8 has a role in bladder cancer and may harbour TSGs. This study supports the existing hypothesis and makes further contributions towards the identification of the critical region on 8p22, by making use of the two closely linked *NAT* genes, the mapping position of which has previously been determined in relation with each other and with other genetic markers on 8p22 (Matas et al, 1997). Although the frequency of chromosomal abnormalities involving the *NAT* region could not be correlated with a particular grade or stage of tumour, and despite the fact that the *NAT* genes themselves are unlikely to act as TSGs in bladder or other types of cancer, they should be useful markers for studying the role of the 8p22 region in human carcinogenesis.

6.4 NAT1 and NAT2 in healthy breast tissue

6.4.1 Materials and study design

To investigate the possible role of the NAT isoenzymes in breast carcinogenesis, it is important to understand their action in healthy breast tissue. In collaboration with Dr J.A. Williams (Institute of Cancer Research, Sutton, UK), a study was conducted to investigate expression of NAT1 and NAT2 in breast, as well as the contribution of different alleles of these isoenzymes to the bioactivation of suspected breast carcinogens, such as 2-amino-3-methylimidazo[4,5-f] quinoline (IQ). The aim of the work presented here was to examine expression of the two NAT isoenzymes in healthy human breast, using enzymatic activity assays and Western blotting.

Eighteen healthy tissue samples were obtained from women of Caucasian ethnic origin following operation for reduction mammoplasty. Cytosols were prepared from tissue homogenates by the Sutton group, as described in Adams et al, 1985, then frozen and stored at -70°C. Cytosols were transported to Oxford on dry ice. They were labelled as HB1-HB19 (HB11 was not available). For some experiments, a pool of all eighteen samples was prepared by mixing an equal volume from each sample.

For immunological detection of NAT1 protein with Western and dot blot analysis, rabbit primary anti-serum #184 (Ab184), raised against the C-terminal peptide (CVPKHGDRFFTI) of human NAT1, was used. NAT2 protein was detected using rabbit primary anti-serum #4768 (Ab4768), raised against the C-terminal peptide (CVPKPGDGSLTI) of human NAT2 (Stanley et al, 1996).

6.4.2 Results

6.4.2.1 Measurement of protein concentration in breast cytosols

The concentration of total cytosolic protein in the eighteen breast samples was measured as described in section 2.4.1. The protein concentration of a human placental cytosol sample (provided by Dr V. Smelt), used as positive control in all subsequent activity assays, was also measured. The average protein concentration of the breast cytosols was $10.95 \pm 1.63 \text{ mg/ml}$ (mean $\pm$ standard deviation). The protein concentrations of all samples, including the placental control, are listed in table 6.5.

Table 6.5: Protein concentration of the human breast cytosols. The concentration of the eighteen human breast cytosol samples (HB1-HB10 and HB12-HB19) and of the control placental cytosol sample (PS4) is indicated.

Sample Name	Protein Concentration (mg/ml)
HB1	10.98
HB2	10.55
HB3	9.27
HB4	9.23
HB5	12.1
HB6	10.82
HB7	10.32
HB8	11.05
HB9	11.75
HB10	10.42
HB12	9.68
HB13	10.4
HB14	12.85
HB15	9.32
HB16	9.2
HB17	11.22
HB18	15.89
HB19	12.04
PS4	87.23

6.4.2.2 NAT1 enzymatic activity in human breast cytosols

All eighteen breast cytosolic samples were assayed for NAT activity, as described in section 2.5. NAT1 enzymatic activity was measured using pABA as substrate at a final concentration of 0.2mM (20nmol per reaction). All samples were assayed over 60min and it was determined that the acetylation rate was linear for the initial 10min, for the amount of total cytosolic protein used (550µg on average). Reaction rates were measured over this linear period. The reactions were performed in triplicate and the mean absorbance and standard deviation (*s*) were calculated. Typically, the calculated coefficient of variation (c.v.) for the measured absorbance of each triplicate was lower than 10%, while the maximum acceptable c.v. value for each triplicate was 20%. Statistical analysis of data was performed as described in Kirkwood, 1988.

The initial rate of NAT1 acetylation was expressed in nmoles of N-acetylated pABA per min per mg of total cytosolic protein (nmol/min/mg). The NAT1 specific activity with pABA for each individual sample is shown in table 6.6 and figure 6.6. The specific activity of the control placental sample (PS4) was 1.42nmol/min/mg, in accordance to the expected 1.38nmol/min/mg previously reported for this sample (Smelt et al, 1998).

Table 6.6: NAT1 specific activity for pABA of the eighteen human breast cytosols and the control placental cytosol PS4.
The relative NAT1 specific activity is expressed as percentage of the mean specific activity. The *NAT1* genotype of these samples (determined by the Sutton group) is also indicated.

Sample	NAT1 specific activity with pABA (nmol/min/mg)	Relative NAT1 specific activity (% of mean [*])	<i>NAT1</i> genotype (from Williams et al, 2000)
HB1	0.47	47.5	1*10/1*10
HB2	0.35	35.3	1*4/1*4
HB3	0.52	52.5	1*4/1*10
HB4	0.98	98.9	1*4/1*3
HB5	0.39	39.4	1*4/1*4
HB6	1.57	158.6	1*4/1*4
HB7	1.31	132.3	1*4/1*10
HB8	0.85	85.8	1*4/1*14
HB9	0.75	75.7	1*4/1*4
HB10	1.53	154.5	1*4/1*4
HB12	1.45	146.5	1*4/1*4
HB13	0.90	90.9	1*4/1*4
HB14	0.84	84.8	1*4/1*4
HB15	1.14	115.1	1*4/1*10
HB16	2.22	224.2	1*4/1*10
HB17	0.83	83.8	1*4/1*10
HB18	0.50	50.5	1*4/1*4
HB19	1.28	129.3	1*4/1*4
PS4	1.42	Not applicable	Not applicable

^{*}Mean is the average specific activity of all eighteen samples.

NAT1 activity with pABA as substrate was evident in all eighteen samples assayed. The mean specific activity of the cytosol samples was 0.99±0.5nmol/min/mg (mean±s). The calculated NAT1 specific activity was normally distributed and ranged from 0.35nmol/min/mg (HB2) to 2.22nmol/min/mg (HB16), showing a 6.34-fold interindividual variation (figure 6.7).

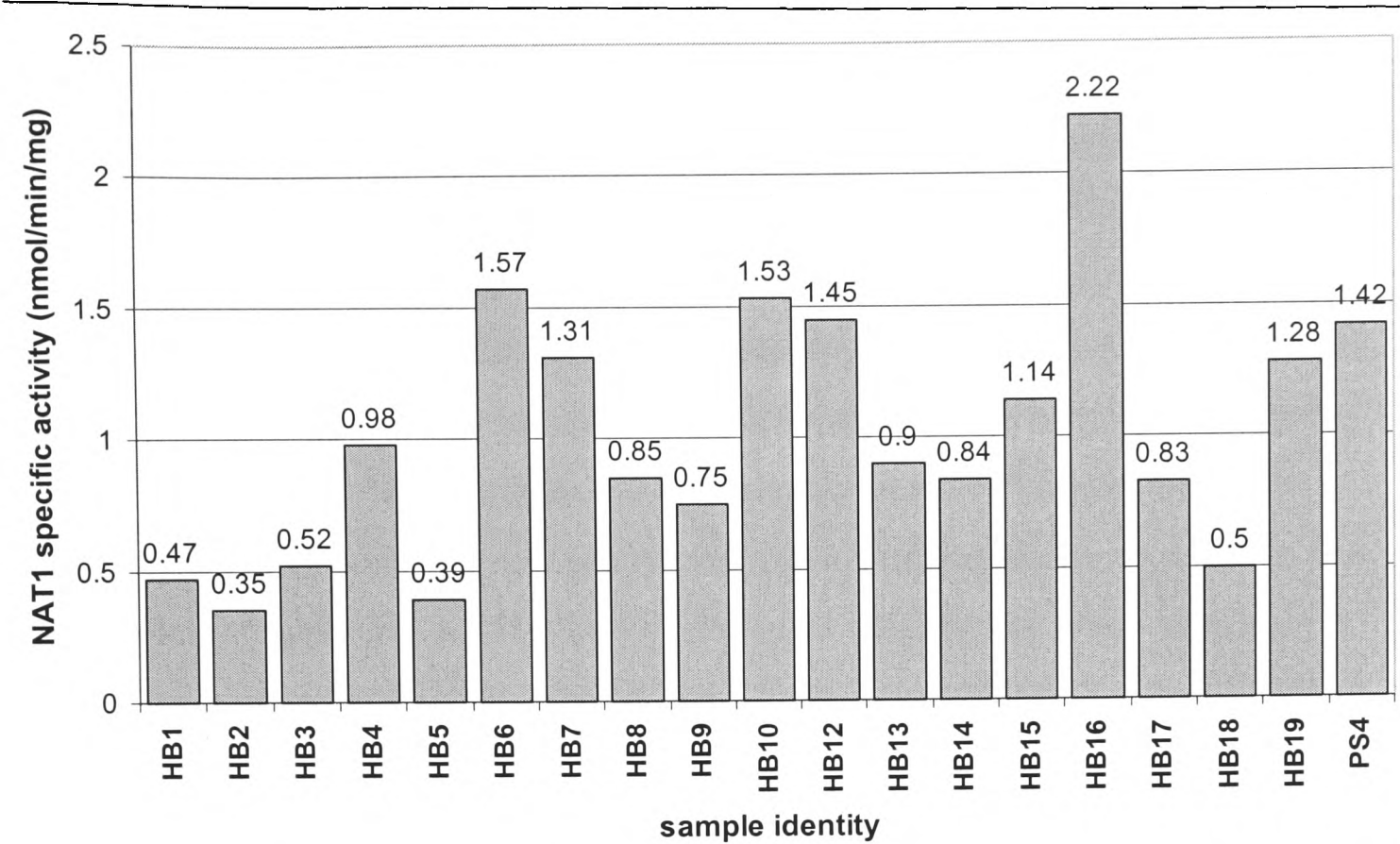


Figure 6.6: Diagram showing the NAT1 specific activity (nmol/min/mg) measured for the eighteen breast cytosol samples, following incubation with pABA for 10min. The mean specific activity and standard deviation for the eighteen samples was 0.99 ± 0.5 nmol/min/mg. The specific activity of the placental cytosolic sample PS4, used as positive control, is also shown.

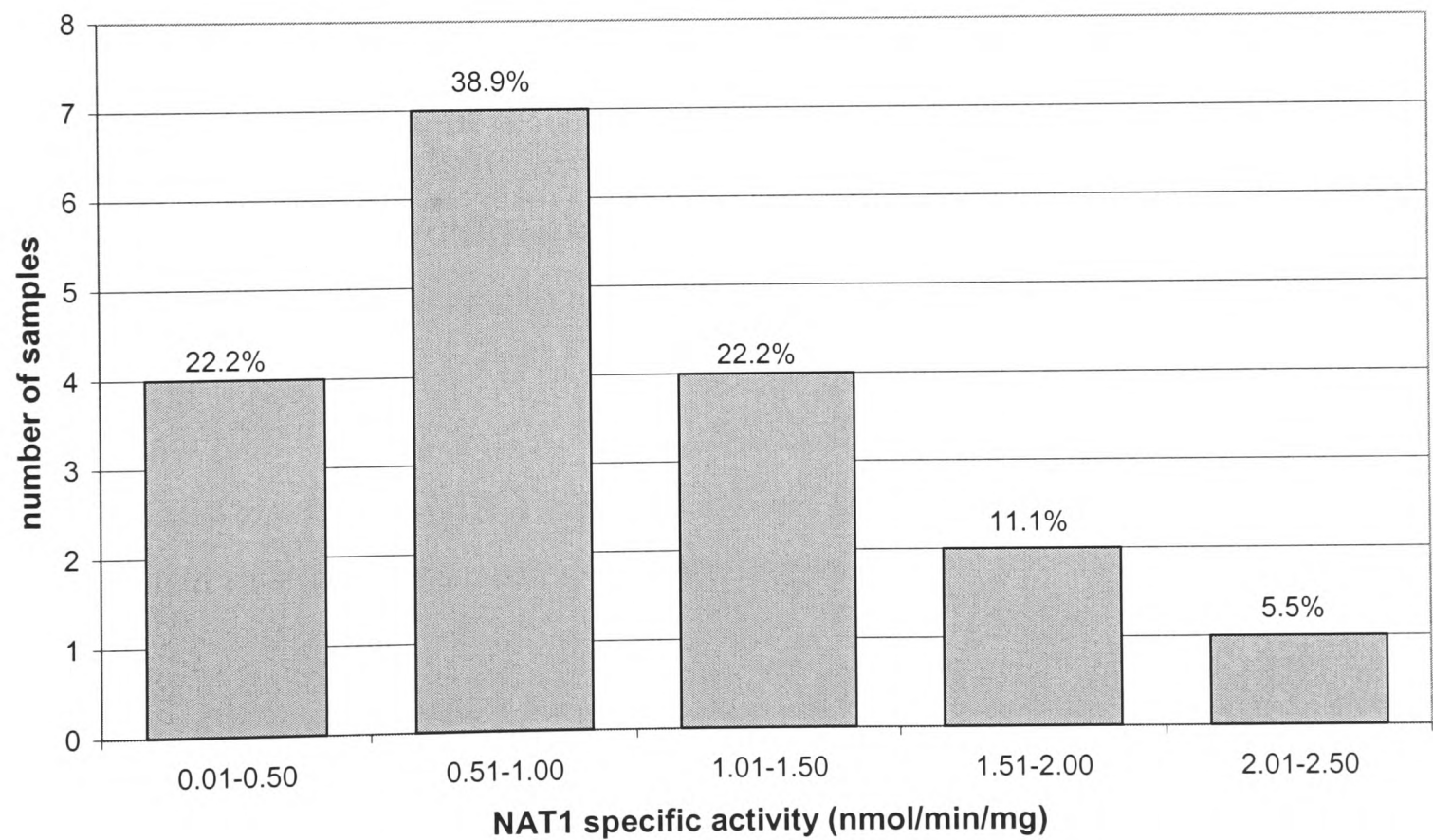


Figure 6.7: Histogram of frequency distribution for the eighteen breast cytosolic samples, based on their calculated NAT1 specific activity. Five intervals with a width of 0.5nmol/min/mg were defined. The height of each bar represents the number of samples with specific activity falling into the corresponding interval. Frequencies are presented at the top of each bar, as percentage of the total number of samples. The distribution appears symmetric around the mean (0.99nmol/min/mg), which indicates normal distribution of the values.

Genotyping of the eighteen samples was performed by the Sutton group (table 6.6; Williams et al, 2000). The distribution of the relative NAT1 specific activity calculated for each sample (expressed as percentage of the mean specific activity for all samples, table 6.6) in relation to its *NAT1* genotype was examined and the results are illustrated in figure 6.8. Genotypes were classified either as homozygous for the wild type allele *NAT1**4 (wt/wt), or as heterozygous for *NAT1**4 and one mutant allele (wt/mt), or as homozygous for any combination of mutant alleles (mt/mt). The mutant alleles found were *NAT1**10, *NAT1**3 and *NAT1**14. Of all samples, ten (55.5%) were genotyped as *NAT1**4/*4, five (27.7%) as *NAT1**4/*10, one (5.5%) as *NAT1**4/*3, one (5.5%) as *NAT1**4/*14 and one (5.5%) as *NAT1**10/*10 (table 6.6). The relative NAT1 specific activity of the homozygous *NAT1**4 samples ranged from 35.3% to 158.6% (samples HB2 and HB6 respectively), with $96.6 \pm 47.7\%$ (mean $\pm$ s). There was a 4.5-fold interindividual variation in the relative NAT1 specific activity of the *NAT1**4 homozygotes. The relative NAT1 specific activity of the heterozygous *NAT1**4 samples ranged from 52.5 to 224.2% (samples HB3 and HB16 respectively), with $113.2 \pm 55.1\%$ (mean $\pm$ s). The interindividual variation for the group of *NAT1**4 heterozygotes was 4.3-fold. Overall, the mean relative specific activity of the group of heterozygotes was higher than that of the group of homozygotes for the *NAT1**4 allele by 16.7%. However, this difference was not statistically significant ($p > 0.5$), as assessed using the t-test for small samples ($n_1=10$, $n_2=7$) with comparable *s* values. Sample HB4 carrying the *NAT1**3 allele had relative specific activity equal to 98.9% and sample HB8 carrying the *NAT1**14 allele had relative specific activity equal to 85.8%. Only one sample (HB1) was homozygous for a mutant allele (*NAT1**10/*10) and this had relative specific activity 47.8% (figure 6.8).

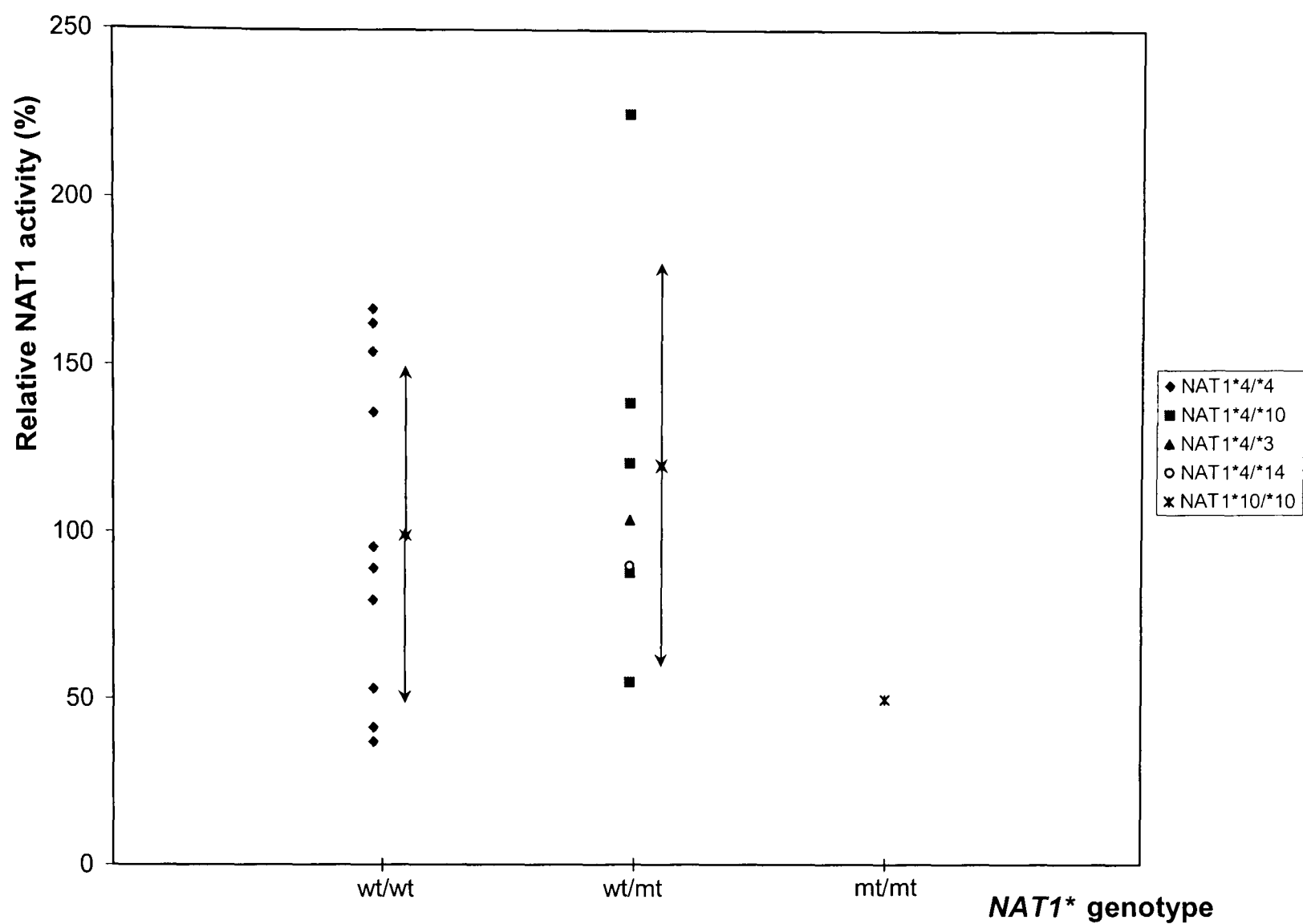


Figure 6.8: Comparison of the breast cytosolic NAT1 activity for pABA with the *NAT1* genotype of the eighteen individuals. Samples were classified as homozygous for the wild type *NAT1**4 allele (wt/wt), as heterozygous (wt/mt), or homozygous (mt/mt) for mutant alleles. NAT1 activity was expressed as percentage of specific activity relative to the mean of all samples (0.99nmol/min/mg). The error bars indicate the mean relative NAT1 specific activity for each group of samples and the corresponding standard deviation (96.56±47.73% for the wt/wt group and 113.24±55.07% for the wt/mt group).

A similar analysis was performed in relation to the number of *NAT1**10 alleles in the genotype of each individual (figure 6.9). Of the eighteen samples assayed for NAT1, twelve (66.6%) did not have any *NAT1**10 alleles, five (27.8%) were heterozygous *NAT1**4/*10 (therefore contained one *NAT1**10 allele), and only one (5.5%) had two *NAT1**10 alleles. The relative NAT1 activity for the non-*NAT1**10 samples ranged from 35.3% to 158.6% (samples HB2 and HB6 respectively), with 95.8±43.3% (mean±s). The relative NAT1 activity for the heterozygous *NAT1**10 alleles ranged

from 52.5 to 224.2% (samples HB3 and HB16 respectively), with $121.6 \pm 64.9\%$ (mean $\pm$ s). The interindividual variation for the non-*NAT1*10* group and the group of the *NAT1*10* heterozygotes was the same as that of the wt/wt (4.5-fold) and the wt/mt (4.3-fold) groups of figure 6.8, respectively. The mean relative NAT1 activity for the group of *NAT1*10* heterozygotes was higher than that of the non-*NAT1*10* group by 25.8%. However, this difference was not statistically significant ($p > 0.2$) using the t-test for small samples ($n_1=12$, $n_2=5$) with comparable s values.

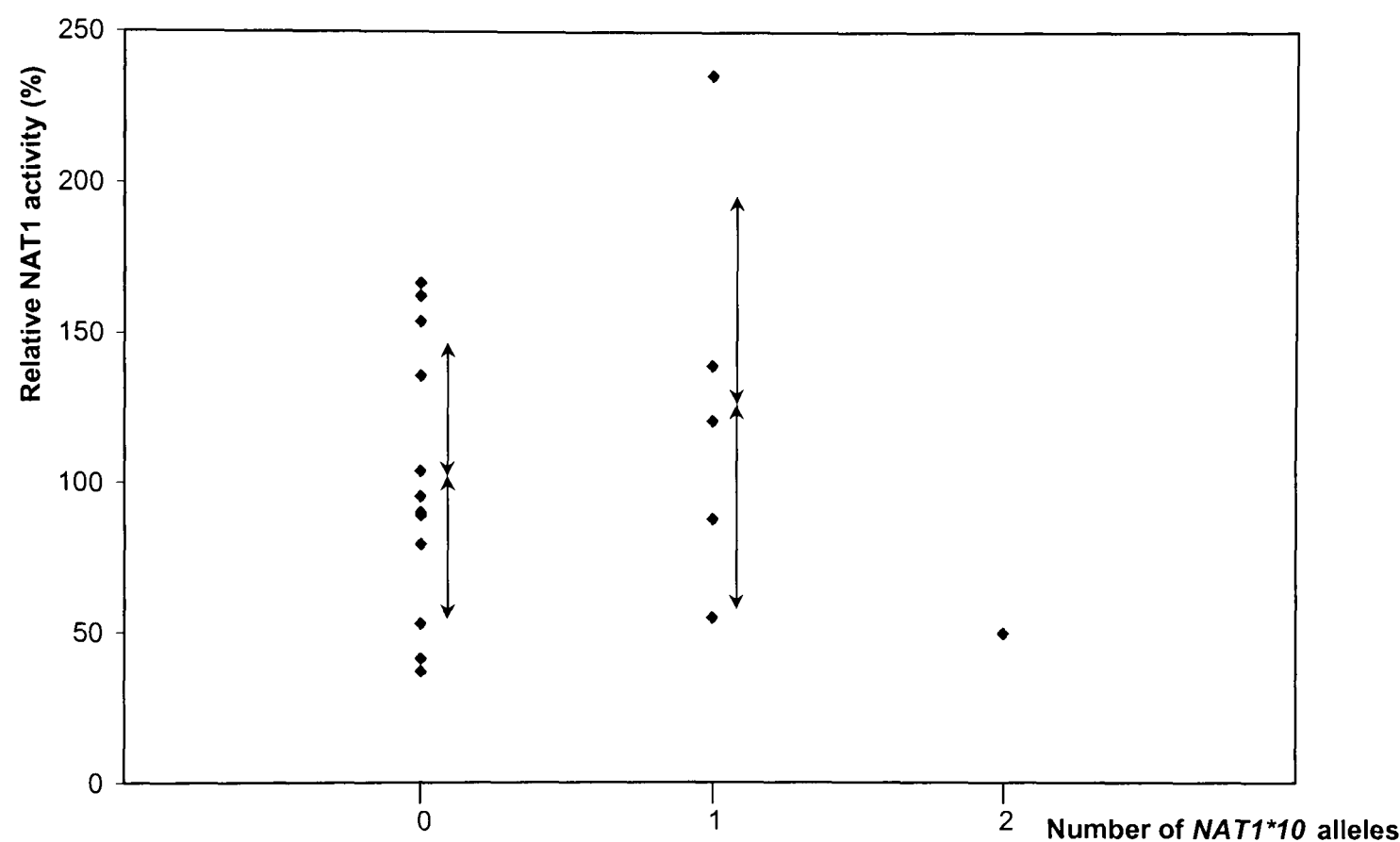


Figure 6.9: Comparison of breast cytosolic NAT1 activity with the number of *NAT1*10* alleles in the genotype of the eighteen individuals. NAT1 activity is expressed as percentage of specific activity relative to the mean of all samples (0.99nmol/min/mg). The error bars indicate the mean relative NAT1 specific activity for each group of samples and the corresponding standard deviation ($95.8 \pm 43.3\%$ for the non-*NAT1*10* group and $121.6 \pm 64.9\%$ for the heterozygous *NAT1*10* group of samples).

6.4.2.3 NAT2 enzymatic activity in human breast cytosols

NAT2 enzymatic activity was measured as described in section 2.5 of materials and methods, using the NAT2-specific substrate sulphamethazine (SMZ) at a final concentration of 0.1mM (10nmol per reaction). Only samples HB1, HB5, HB7, HB8, HB9, HB12 HB14, HB15, HB17, HB18 and HB19 were assayed for NAT2, as they were predicted to have a fast acetylator phenotype, based on their *NAT2* genotype (determined by the Sutton group, Williams et al, 2000). N-acetylation activity was not evident in the breast cytosols assayed, even after 16h of incubation of undiluted sample with 0.1mM SMZ. This indicated lack of NAT2 activity in healthy breast tissue at detectable levels.

6.4.2.4 Detection of NAT protein in healthy human breast tissue

Following detection of NAT1 enzymatic activity in the breast cytosols, the presence of NAT1 protein was examined by Western blotting, using the NAT1-specific antibody Ab184 (section 6.4.1; Stanley et al, 1996). A sample pool and individual samples were used.

A dot blot was initially performed with pooled sample to determine the ideal protein concentration and antibody dilution for detection of NAT1 with antibody Ab184, as described in section 2.4.2. It was determined that optimal detection of NAT1 protein could be achieved using 100µg of total cytosolic protein and an antibody dilution of 1:2000 (figure 6.10). Western blots were subsequently performed under the same conditions. Human breast cytosol samples were separated on denaturing polyacrylamide gels, stained and photographed as described in section 2.4.3 (figure 6.11). Identical gels were blotted and the blots were developed as mentioned in section 2.4.4. Western blots were performed using either the pooled cytosol sample (figure 6.12) or each one of the eighteen samples individually (figure 6.13). A band representing NAT1 protein was not visible on the SDS-PAGE gel (figure 6.11), indicating that NAT1 is not abundant in the breast tissue. Following Western blotting,

however, an NAT1-specific band of around 31-33kDa was visible on the autoradiogram (figures 6.12 and 6.13).

The NAT2-specific antibody Ab4768 (section 6.4.1; Stanley et al, 1996) was used in a similar way to detect NAT2 protein in the breast cytosols. However, NAT2 protein was not detected, which was in agreement with the lack of detectable NAT2 activity, described above (section 6.4.2.3).

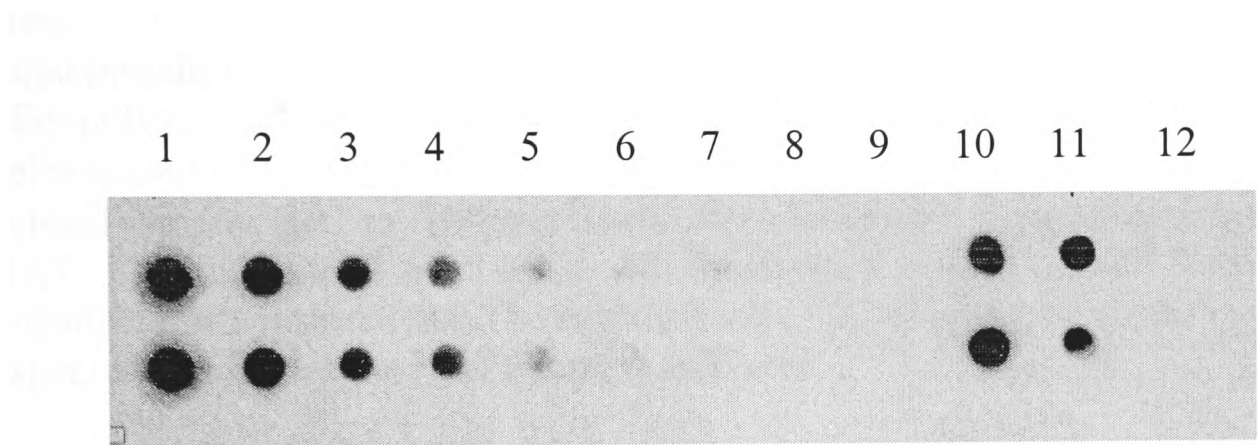


Figure 6.10: Detection of NAT1 protein in human breast cytosols by dot blot. All samples were loaded in duplicate. Slots 1 to 9 were loaded with 1000, 500, 100, 50, 10, 5, 2.5, 1 and 0.5µg of pooled cytosolic protein. Slot 10 contains 1µg of recombinant human NAT1 (positive control), while slot 12 was loaded with TBS only (negative control). Slot 11 was loaded with 10µg of bovine serum albumin (BSA). Following dot blotting, the membrane was cut into identical strips which were incubated in Ab184 anti-serum diluted 1:1000, 1:1500, 1:2000 and 1:3000. Washes and development were performed as described in section 2.4.4. At 1:2000 antibody dilution, shown here, NAT1 protein was clearly detected in 1000, 500, 100 and 50µg of pooled sample (slots 1 to 4) after 1 minute of autoradiographic exposure (section 2.2.12.1.2). The positive control (slot 10) also produced a signal, while the negative control (slot 12) was blank. The signal produced by BSA (slot 11) was due to the method of antibody preparation that involved conjugation of the NAT1 peptide to BSA prior to immunisation (Stanley et al, 1996). The BSA signal did not interfere with subsequent Western blot analysis of the human samples.

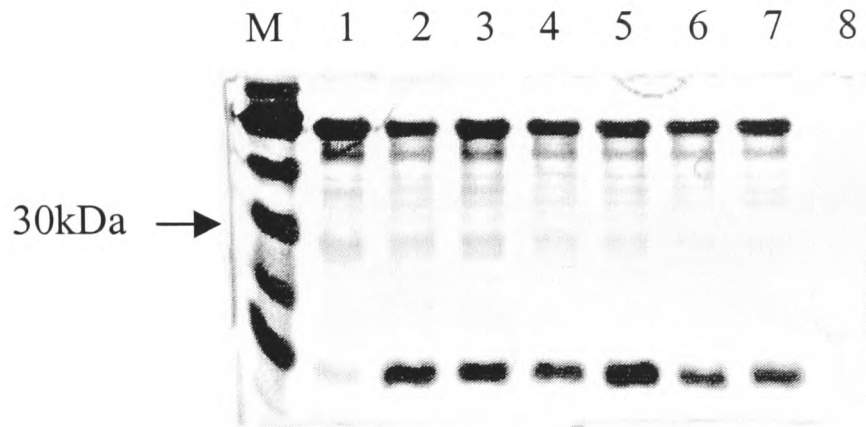


Figure 6.11: Denaturing polyacrylamide gel electrophoresis (SDS-PAGE) of human breast cytosolic samples.

Total protein from breast cytosols (100 μ g) was denatured and separated on a 12% SDS-polyacrylamide gel, as described in section 2.4.3. Following electrophoresis, the gel was stained with Coomassie blue. Lanes 1 to 7 were loaded with 100 μ g protein of cytosol samples HB1 to HB7 and lane 8 was loaded with 10 μ g of human recombinant NAT1 protein. Lane M was loaded with denatured Rainbow marker, according to the manufacturer's instructions. The marker band (30kDa) which is closest to the expected position of the NAT1 band is indicated.

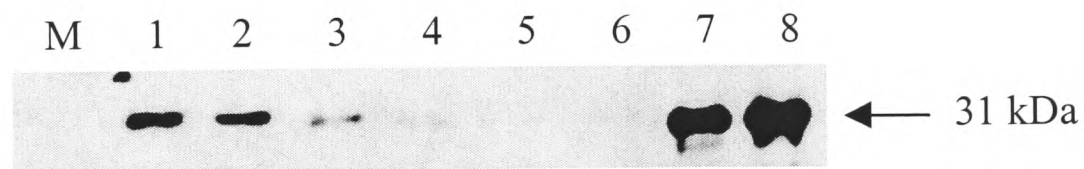


Figure 6.12: Detection of NAT1 protein in the pooled human breast cytosol by Western blotting.

Pooled breast cytosolic protein (150, 100, 50, 25, 10 and 5 μ g) was loaded in lanes 1 to 6 and separated by SDS-PAGE. Lanes 7 and 8 were loaded with 10 and 20 μ g of recombinant human NAT1 protein, as positive control. Lane M was loaded with denatured Rainbow marker, according to the manufacturer's instructions. At 1:2000 dilution of Ab184 and following 5 minutes of autoradiographic exposure, NAT1 protein was detected as a single band of approximately 30kDa in size. A NAT1 signal was produced by 150, 100 and 50 μ g of pooled cytosolic protein (lanes 1, 2 and 3 respectively) and by the positive control (lanes 7 and 8).

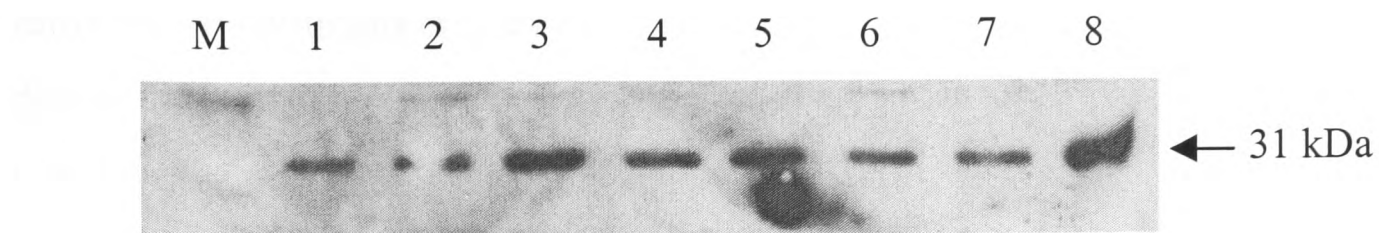


Figure 6.13: Detection of NAT1 protein in samples of human breast cytosol by Western blotting.

Individual samples (HB1-HB7) of total cytosolic protein (100µg) were loaded in lanes 1 to 7 and separated by SDS-PAGE. Lane 8 was loaded with 10µg of recombinant human NAT1 protein, as positive control. Lane M was loaded with denatured Rainbow marker, according to the manufacturer's instructions. At 1:2000 dilution of Ab184 and following 8 minutes of autoradiographic exposure, NAT1 protein was detected as a single band of approximately 30kDa in size. A NAT1 signal was produced by all cytosols used (lanes 1 to 7) and by the positive control (lane 8).

6.4.3 Discussion

In addition to genetic and hormonal factors whose roles are better established, several chemical carcinogens have been suspected to induce breast carcinogenesis. In view of their ability to metabolise carcinogenic compounds, such as polycyclic aromatic hydrocarbons and heterocyclic amines, NATs could be involved in the initiation of this type of cancer (Welp et al, 1998). Epidemiological studies have produced conflicting reports on the link between polymorphisms of the *NAT* genes and risk for breast cancer (section 6.1.3), therefore further investigation is necessary.

Understanding the role of NAT in healthy breast tissue should facilitate assessment of its contribution to the local bio-activation or detoxification of carcinogens. The metabolic pathways leading to chemical carcinogenesis in the breast are currently unclear, but the NAT isoenzymes could be involved. NAT could act as a susceptibility factor by itself, or in combination with other enzymes with similar functions. To address these questions, NAT was studied in healthy breast tissue obtained after operations for reduction mammoplasty. The aim of the collaborative study, part of which is presented here, was to examine NAT1 and NAT2 expression and enzymatic activity in the breast (section 6.4.2), as well as the potential role of

these enzymes in the activation of carcinogens, such as N-hydroxy-IQ and N-hydroxy-PhIP (Williams et al, 2000). It has also been possible to study the correlation between the human NAT1 genotype and phenotype in the small population of samples examined.

Using enzymatic assays and Western blot analysis, the work presented in this Chapter has been able to demonstrate the presence of active NAT1 protein in cytosols prepared from healthy breast tissue. NAT2 activity was not evident using these methods. This is in agreement with a previous study that used p-aminosalicylic acid (pAS) and SMZ as specific substrates for detecting NAT1 and NAT2 activity (Sadrieh et al, 1996). NAT1 expression appears to be ubiquitous in the body, since it has been detected in liver and most of the extrahepatic tissues and cell lines examined to date (section 1.3.4).

As part of the project, Dr J.A. Williams and his colleagues have reported detection of both NAT1 and NAT2 mRNA in breast (Williams et al, 2000). Quantitative RT-PCR showed the presence of transcript for both isoenzymes, with NAT1 mRNA being about 2-3 times more abundant than NAT2 mRNA. Therefore, inability to detect NAT2 activity could be due to very low expression of the enzyme, at amounts below the detection limit of the SMZ assay (1pmol/min/mg). The same group also detected NAT1 protein in the ductal epithelium of breast tissue by immunohistochemistry (Williams et al, 2000). Interestingly, immunohistochemistry was also able to detect NAT2 protein in the ductal epithelium, with the same polyclonal antiserum (Ab4768) used for Western blot analysis (section 6.4.2.4). However, unlike that used here for Western blotting, the antiserum aliquot used for immunohistochemistry had been affinity purified, using the C-terminal peptide of NAT2 (originally used for immunisation of the rabbit) to select for the NAT2-specific subpopulation of antibodies from the polyclonal antiserum. Antiserum Ab4768 has previously been reported to exhibit low binding capacity to the NAT2 protein, based on ELISA and Western blotting (Stanley et al, 1996). Although useful for the detection of NAT2 when present at relatively high levels in the preparation (e.g. when recombinant NAT2 is used), negative results can be expected when the levels of the protein are very low. Therefore, enrichment of the antibody used in immunohistochemistry by affinity purification is likely to have increased sensitivity of the technique to a level

allowing detection of NAT2 in tissue. This further indicates that NAT2 should be present in breast tissue at extremely low amounts.

Genotyping of *NAT1* and *NAT2* was also carried out by the Sutton group (Williams et al, 2000) and it has been possible to examine the effect of the *NAT1* genotype on the N-acetylation phenotype. The majority of samples contained combinations of the wild type *NAT1**4 and the mutant *NAT1**10 alleles, so the study of other genotypes was very limited. The distribution of NAT1 specific activities was similar in the groups of homozygous wild type (*NAT1**4/*4) and heterozygous *NAT1**4/*10 individuals (figure 6.8). There was a difference in the mean activity of the two groups, but this was not statistically significant. One sample (HB4) was heterozygous for *NAT1**4 and the mutant *NAT1**3 allele, and its NAT1 specific activity was very close to the mean value for all samples (table 6.6). The *NAT1**3 allele is known to contain a single point mutation outside the coding region (table 1.4), which does not affect the product of the gene (section 1.3.3.2). Another sample (HB8) was heterozygous for *NAT1**4 and the rare mutant *NAT1**14 allele (table 6.6). The product of this allele carries an amino acid substitution associated with lower NAT1 activity (Hughes et al, 1998a; Payton and Sim, 1998). In this study, sample HB8 had activity below the average, but still within the range of activities measured for the *NAT1**4/*4 group (table 6.6, figure 6.8). Considering the high inter-individual variation in the distribution of NAT1 activity within the *NAT1**4/*4 group (figure 6.8), the NAT1 activity of sample HB8 can be explained. Higher than average NAT1 activity, conferred by the *NAT1**4 allele, could mask the effect of the slow *NAT1**14 allele in this particular sample.

The *NAT1**10 allele has been associated with increased NAT1 activity (Bell et al, 1995a), although this has not been confirmed by other studies (Smelt et al, 1998; Payton and Sim, 1998; Butcher et al, 1998; Bruhn et al, 1999). In this study, there was no clear evidence that the *NAT1**10 allele is responsible for a fast acetylation phenotype. When the samples were grouped according to the number of *NAT1**10 alleles they carried, the distribution of activities was not different between the groups (figure 6.9). The mean activities of groups with no *NAT1**10 and one *NAT1**10 alleles did not differ significantly either. Only one of the *NAT1**4/*10 samples (HB16) showed unusually high NAT1 activity, while the only homozygous *NAT1**10/*10 sample (HB1) was among those with the lowest NAT1 activity observed. Overall,

these data suggest no obvious correlation between the *NAT1**10 allele and higher NAT1 activity, in agreement with the majority of previous studies (section 1.3.3.2). Study of a larger population is required in order to determine the effect of specific *NAT1* alleles on the N-acetylation phenotype of individuals.

Williams and colleagues also investigated the contribution of the NAT1 phenotype to the bio-activation of N-hydroxy-IQ by measuring the levels of DNA adducts (Williams et al, 2000). The formation of DNA adducts was investigated by incubation of calf thymus DNA with cytosolic protein from breast, in the presence of N-hydroxy-IQ. Acetyl coenzyme A was also added as a co-factor for NAT. As shown in figure 6.14, no apparent correlation between the NAT1 specific activity of the eighteen breast cytosols and the levels of IQ-DNA adduct formation could be observed. The calculated correlation coefficient ($R^2=0.0272$) did not reveal a statistically significant association between NAT1 specific activity and the number of IQ-DNA adducts, following two-sided t-test analysis ($p>0.5$). Two of the samples, HB3 and HB10, showed high levels of DNA adducts, compared with the rest of the population. The *NAT1* genotype for these two samples was *NAT1**4/*10 and *NAT1**4/*4, respectively. The former sample had relatively low NAT1 activity, while the latter had the third highest NAT1 activity of all eighteen samples (table 6.6). Exclusion of these two samples from figure 6.14, provided a better correlation ($R^2=0.1727$), but again this was not statistically significant ($0.2>p>0.1$). Therefore, there appears to be no association between NAT1 activity and variation in the levels of adduct formation as a result of bio-activation of heterocyclic amines, such as IQ. However, a more extensive study, using a larger number of samples, is required.

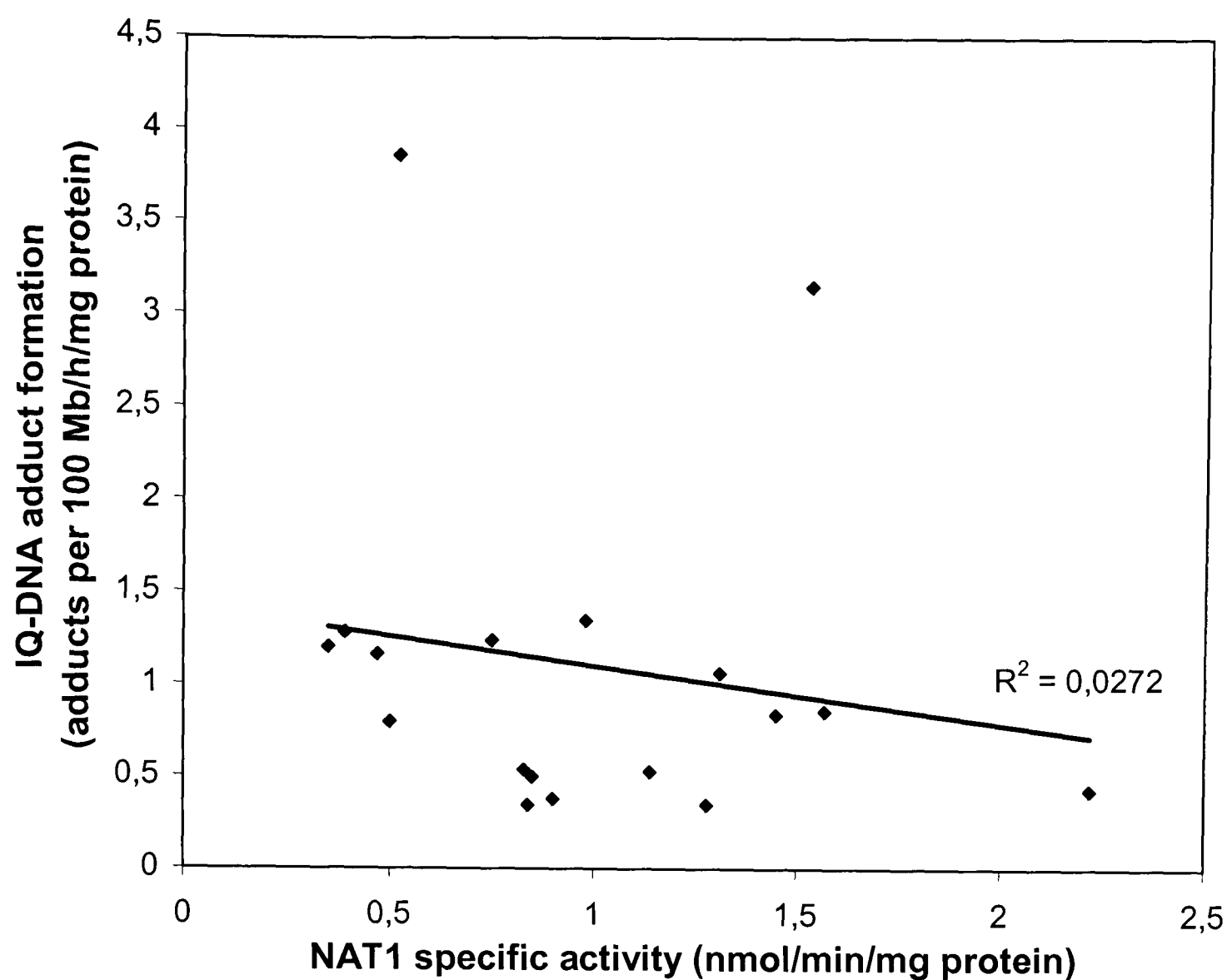


Figure 6.14: Correlation between NAT1 specific activity with pABA (nmol/min/mg) and the number of IQ-DNA adducts (adducts per 100Mb/h/mg), following incubation of calf thymus DNA with breast cytosol and N-hydroxy IQ (Williams et al, 2000). The correlation coefficient (R^2) for the two sets of values is indicated.

In conclusion, the results presented here demonstrate NAT1 protein and enzymatic activity in cytosolic fractions obtained from healthy breast tissue. NAT2 was not detected by Western blotting, or SMZ enzymatic activity assay, although it is likely to be present in breast tissue at very low amounts. There was significant interindividual variation in NAT1 activity, which did not appear to correlate with any particular *NAT1* genotype, in agreement with previous studies. Whether NAT1 has a role in the process of chemically induced carcinogenesis in the breast is not known. However, it seems likely that NAT1 can contribute to the local bio-activation of heterocyclic amines, as suggested by its presence in breast tissue.

Chapter 7: Conclusions

The N-acetylation polymorphism is recognised as a landmark in the fields of Pharmacogenetics and Xenogenetics and has attracted the interest of scientists with diverse expertise, such as pharmacologists, toxicologists, cancer epidemiologists, biochemists, geneticists and molecular biologists (La Du, 1992; Grant, 1993; Sim et al, 1995). In this chapter, the conclusions from the work described in this thesis will be discussed, in the light of current research in the NAT field.

Many epidemiological studies have attempted to identify association between polymorphisms of the NAT isoenzymes and predisposition to various cancers. A strong link has been recognised between the slow NAT2 phenotype and increased predisposition to occupational or smoking-related bladder cancer (Cartwright et al, 1982; Risch et al, 1995), but the results of epidemiological studies have been inconclusive for other types of cancer. Epidemiological research remains an important way for directly assessing the risk for carcinogenesis in human populations (Vainio and McGregor, 1992). However, results are often overinterpreted and fail to take into account the entirety of genetic and environmental factors that contribute towards susceptibility to a particular type of cancer (Sim et al, 1995; Ingelman-Sundberg, 1998). Furthermore, epidemiological studies suffer from publication bias, meaning that work that does not establish a risk factor for a particular condition is less likely to be published than one with positive results.

One of the major problems that epidemiologists often face is collection of data from large populations of affected and control individuals, sufficient to support a statistically significant association between a specific risk factor and the disease. This is particularly difficult in the case of complex diseases, where individuals have to be grouped not only based on the specific phenotype of interest, but also according to

factors that include their ethnic origin, age, sex, social habits, occupational hazards, medical history, etc. When the aim is to assess whether polymorphism in a specific type of xenobiotic metabolising enzyme may affect susceptibility to cancer or the side effects of a drug, then another major problem is to predict how alternative synergistic or competing metabolic pathways might influence the response of different individuals. International collaborative studies are currently in progress, designed to collect and analyse data on xenobiotic metabolising enzyme polymorphisms from large populations of variable ethnic origin (Taioli, 1999). Large-scale epidemiological research is expected to provide valuable information about the combination and distribution of genetic polymorphism in xenobiotic metabolising enzymes among high-risk groups and identify factors modifying the effect of these polymorphisms in different individuals.

Methods for assessing the carcinogenic effect of chemicals are available. Study of the tissue-specific expression profile of xenobiotic metabolising enzymes is of particular importance for evaluating the role of these enzymes and their polymorphisms as protective or predisposing factors, especially in tissues that are highly susceptible to chemical carcinogenesis, like the colon, bladder, lungs and breast. As part of collaborative research (Williams et al, 2000), the work presented here investigated expression of NAT1 and NAT2 in healthy breast tissue, examined the link between NAT1 activity and the genotype, and reported on DNA adduct formation following exposure to established carcinogens (Chapter 6). Measuring the levels of adducts formed between activated mutagenic compounds and cellular DNA or protein is an effective method for assessing levels of exposure to potential carcinogens. The value of the technique lies upon the fact that it can take into account not only the mutagenic potency of the ultimate carcinogen, but also the metabolic efficiency of the individual in detoxifying or bioactivating the procarcinogen upon internalisation. Variability in the levels of *in vitro* adduct formation may reflect variability in individual response to exposure to specific carcinogens and can be correlated to polymorphisms in xenobiotic metabolising enzymes. On the other hand, the levels of DNA adducts are not always informative. DNA adducts can lead to carcinogenesis only if they introduce mutations to critical genes of the cell cycle, and cannot be used to directly predict susceptibility of an individual to any particular type of cancer (Vainio and

McGregor, 1992). The use of the technique is also limited by the fact that its application on a large-scale basis is difficult. Correlation between a specific NAT1 phenotype or genotype and the levels of adducts was not evident, following exposure to established carcinogens metabolised by NAT (Williams et al, 2000). This was not surprising, as most studies attempting evaluation of the phenotypic effect of different *NAT1* alleles have reported high inter-individual variation in NAT1 activity and lack of clear genotype-phenotype relationship, with the exception of only a few alleles, such as the *NAT1*14* and the *NAT1*15* which are associated with a slow NAT1 phenotype (Butcher et al, 1998; Hubbard et al, 1998; Hughes et al, 1998a; Payton et al, 1998; Smelt et al, 1998; Bruhn et al, 1999). High inter-individual variation in NAT1 activity was also observed in the small population described in this study. Expression of active NAT1 could potentially play a role in the local bioactivation of carcinogens in the mammary gland. Variation in N-acetylation activity could, therefore, affect susceptibility to chemically induced breast cancer, although a large case-control study would be necessary to establish that.

Another target tissue for chemical carcinogenesis is the gut. NAT1 is the isoenzyme predominantly expressed in the intestine, although NAT2 is also present at very low amounts (Ilett et al, 1994; Hickman et al, 1998). Variation in the relative levels of expression for the two isoenzymes was observed, along the length of the intestine, and this could be attributed to different combinations of *NAT1* and *NAT2* genotypes, as demonstrated by the collaborative study described here (Chapter 6; Hickman et al, 1998). It appears that NAT1 could be the isoenzyme predominantly responsible for the bioactivation of dietary carcinogens, while NAT2 may not play an important role. This is further supported by the lack of clear relationship between the *NAT2* genotype and susceptibility to colon cancer, reported by many researchers (Probst-Hensch et al, 1995; Hubbard et al, 1998; Katoh et al, 2000). In view of recent research that showed the presence of polymorphic NAT in prokaryotes (Sim et al, 2000; Payton et al, 2000) it would be of interest to examine whether variation in NAT activity of the bacterial flora of the gut would affect the rate of carcinogen bioactivation.

Chromosomal aberrations of genomic regions containing genes which control the cell cycle have also been associated with initiation and progression of cancer. Very often,

inheritance of a defective allele is combined with somatic loss or rearrangement of the second allele to trigger tumourigenesis (Hooper, 1998). Deletion of chromosomal region 8p22 in humans has been implicated in several types of cancer and regions suspected to contain tumour suppressor genes, have been identified. As a well-characterised group of genes located in region 8p22, the human *NAT* genes could serve as genetic markers for detecting chromosomal aberrations of this region in tumours. The work presented here (Chapter 6) established a methodology for easy and reliable screening of material derived from bladder tumours, using FISH analysis with suitable *NAT1*- and *NAT2*-specific cosmid probes. Deletion of the *NAT* region was a common finding in bladder cancer cells, but increase in the number of copies was also occasionally observed. Although there was no obvious association between aberrations of the *NAT* region and the stage or grade of tumours, the results of this collaborative study highlight the importance of the 8p22 region in bladder carcinogenesis (Stacey et al, 1998; Stacey et al, 1999; Thygesen et al, 1999).

Physical mapping information about the *NAT* cluster in humans is quite extensive (Matas et al, 1997). The search for other genes could, therefore, start from this relatively well-characterised region of 8p22. To facilitate analysis of this genomic region, a number of *NAT*-positive human PAC clones were characterised (Chapter 3). Identification of additional PAC clones containing other established chromosomal markers of the region would also be useful for construction of a more refined physical map around the *NAT* genes. At least two BAC clones covering the *NAT* region on 8p22 have been sequenced as part of the Human Genome Project and the relevant sequencing data are available in a draft format from the EMBL database (accession nos. AC025648 and AC025062). The size of BACs is similar to that of the PAC clones characterised in this study, so the sequencing information published for the *NAT* region should facilitate screening of the PACs for genomic elements that might be of interest. As more extensive genetic resources have become available through the Human Genome Project, in the form of sequence information, cDNA libraries and ESTs, our knowledge about the 8p22 region is expected to expand rapidly.

The mouse is a well established animal model for studying toxicity and genetic disease in humans and its importance is increasing as more information becomes

available via the Mouse Genome Project. Although human epidemiological and *in vitro* studies are extremely useful, animal experimentation remains the most effective way of evaluating the toxic or carcinogenic effect of xenobiotics *in vivo* (Levy et al, 1992). It is generally difficult to assess the impact of a suspected toxic agent, due to variability in the environmental and genetic factors affecting human response, as well as to lack of adequate data about the history of exposure (Vainio and McGregor, 1992). The use of inbred and congenic mouse strains allows direct evaluation of the effect of xenobiotics on living organisms of defined genetic background, under controlled experimental conditions and without the ethical implications associated with human subjects. Transgenic technologies have been a breakthrough in the study of human genetic disease. Mice genetically engineered to lack or overexpress genes encoding for xenobiotic metabolising enzymes can be extremely useful in understanding the role of these enzymes in the bioactivation or detoxification of drugs and carcinogens (Wolf and Henderson, 1998). Most of these transgenic lines have apparently normal phenotypes, indicating that xenobiotic metabolising enzymes either do not have a critical role for survival, or show functional complementation. However, upon exposure to specific compounds, these transgenic animals exhibit differences in susceptibility, compared with the wild-type controls (Nebert and Duffy, 1997; Gonzalez and Kimura, 1999). Although human physiology is in many ways different to that of the mouse, information obtained from laboratory strains can be valuable (Wolf and Henderson, 1998).

The present work has been the basis for the generation of NAT transgenic mouse strains. The 129/Ola plasmid genomic clone (clone A) that was mapped in this study (Chapter 4) has already been used to generate constructs for the production of mice lacking murine *Nat2* or overexpressing human *NAT1*. In view of the recent evidence suggesting involvement of human NAT1 and its functional equivalent mouse NAT2 in folate metabolism (section 1.3.5 for overview), production of these transgenic strains is anticipated to generate valuable information about functions of these isoenzymes, other than in the metabolism of xenobiotics. For example, transgenic strains could be used as models to investigate the possible link recently reported between Alzheimer's disease, associated with low folate levels, and specific *NAT1* alleles in humans (Johnson et al, 2000). Potential involvement of human NAT1

(Smelt et al, 2000) and mouse NAT2 (Stanley et al, 1998; Payton et al, 1999b) in early embryonic development should also become evident by studying these strains.

Furthermore, the transgenic strains will be useful for studying the role of NAT in carcinogenesis. For that purpose, the availability of a murine PAC genomic clone, demonstrated to contain all three *Nat* genes (Chapter 5), will be the starting point for the generation of transgenic strains involving all three mouse *Nat* genes.

The chromosomal region carrying the genes for NAT is of interest in both humans and mice. The present study (Chapter 5) demonstrated that the genes for murine *Nat* are localised in a region of chromosome 8 which is syntenic with the human 8p22 region, where the human *NAT* genes have been localised (Hickman et al, 1994; Franke et al, 1994). Comparative mapping of the region between the two organisms could provide valuable information about the evolution of the human *NAT* genes and their surrounding region, and could help assign mouse genetic orthologues for each one of the human *NAT* genes. More importantly, mice could be used to identify tumour suppressor genes possibly located in the syntenic region marked by the *Nat* locus, and this information could then be used to identify homologous genes on human 8p22. Furthermore, parallel studies in mice and humans could lead to identification of loci responsible for variable susceptibility to cleft lip and palate. In that context, the *NAT1* and *NAT2* polymorphisms in humans, as well as the single *Nat2* polymorphism in mice, could serve as well-characterised SNPs for linkage analysis. The work presented here has significantly contributed towards precise localisation and mapping of the genes for murine *Nat* genes, to complement information already available for their human homologues and is expected to facilitate integration of data useful for comparative mapping of the *NAT* region between the mouse and human.

List of References

- Abdel-Rahman SZ, El-Zein RA, Zwischenberger JB, Au WW (1998). Association of the *NAT1*10* genotype with increased chromosome aberrations and higher lung cancer risk in cigarette smokers. *Mutation Research* **398**: 43-54.
- Abu-Zeid M, Nagata K, Miyata M, Ozawa S, Fukuhara M, Yamazoe Y, Kato R (1991). An arylamine acetyltransferase (AT-1) from Syrian golden hamster liver: Cloning, complete nucleotide sequence and expression in mammalian cells. *Molecular Carcinogenesis* **4**: 81-88.
- Adams DJS, Seilman Z, Amelizad F, Oesch F, Wolf CJ (1985). Identification of human cytochromes P450 analogous to forms induced by phenobarbital and 3 methylcholanthrene in the rat. *Biochemical Journal* **232**: 869-876.
- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185-2195.
- Agundez JAG, Jimenez-Jimenez FJ, Luengo A, Molina JA, Orti-Pareja M, Vazquez A, Ramos F, Duarte J, Coria F, Ladero JM, Alvarez-Cermenio JC, Benitez J (1998a). Slow allotypic variants of the *NAT2* gene and susceptibility to early-onset Parkinson's disease. *Neurology* **51**: 1587-1592.
- Agundez JAG, Ladero JM, Olivera M, Abildua R, Roman JM, Benitez J (1995). Genetic analysis of the arylamine N-acetyltransferase polymorphism in breast cancer patients. *Oncology* **52**: 7-11.
- Agundez JAG, Martinez C, Olivera M, Gallardo L, Ladero JM, Rosado C, Prados J, Rodriguez-Molina J, Resel L, Benitez J (1998b). Expression in human prostate of drug- and carcinogen- metabolising enzymes: Association with prostate cancer risk. *British Journal of Cancer* **78**: 1361-1367.
- Agundez JAG, Martinez C, Olivera M, Ledesma MC, Ladero JM, Benitez J (1994). Molecular analysis of the arylamine N-acetyltransferase polymorphism in a Spanish population. *Clinical Pharmacology and Therapeutics* **56**: 202-209.
- Agundez JAG, Menaya JG, Tejeda R, Lago F, Chavez M, Benitez J (1996a). Genetic analysis of the *NAT2* and *CYP2D6* polymorphisms in white patients with non-insulin-dependent diabetes mellitus. *Pharmacogenetics* **6**: 465-472.
- Agundez JAG, Olivera M, Ladero JM, Rodriguez-Lescure A, Ledesma MC, Diaz-Rubio M, Meyer UA, Benitez J (1996b). Increased risk for hepatocellular carcinoma in *NAT2*-slow acetylators and *CYP2D6*-rapid metabolisers. *Pharmacogenetics* **6**: 501-512.

-
- Agundez JAG, Olivera M, Martinez C, Ladero JM, Benitez J (1996c). Identification and prevalence study of 17 allelic variants of the human NAT2 gene in a white population. *Pharmacogenetics* **6**: 423-428.
 - Ambrosone CB, Freudenheim JL, Graham S, Marshall JR, Vena JE, Brasure JR, Michalek AM, Laughlin R, Nemoto T, Gillenwater KA, Harrington AM, Shields PG (1996). Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms and breast cancer risk. *Journal of the American Medical Association* **276**: 1494-1501.
 - Ambrosone CB, Freudenheim JL, Marshall JR, Graham S, Vena JE, Brasure JR, Michalek AM, Laughlin R, Nemoto T, Shields PG (1995). The association of polymorphic N-acetyltransferase (NAT2) with breast cancer risk. *Annals of the New York Academy of Sciences* **768**: 250-252.
 - Ambrosone CB, Freudenheim JL, Sinha R, Graham S, Marshal JR, Vena JE, Laughlin R, Nemuta T, Shields PG (1998). Breast cancer risk, meat consumption and N-acetyltransferase (NAT2) genetic polymorphisms. *International Journal of Cancer* **75**: 825-830.
 - Anand R, Riley JH, Butler R, Smith JC, Markham AF (1990). A 3.5 genome equivalent multi access YAC library: Construction, characterisation, screening and storage. *Nucleic Acids Research* **18**: 1951-1956.
 - Andersson L, Archibald A, Ashburner M, Audun S, Barendse W, Bitgood J, Bottema C, Broad T, Brown S, Burt D, et al (1996). Comparative genome organization of vertebrates. The First International Workshop on Comparative Genome Organization. *Mammalian Genome* **7**: 717-734.
 - Andres HH, Kolb HJ, Schreiber RJ and Weiss L (1983). Characterisation of the active site, substrate specificity and kinetic properties of acetyl-CoA: arylamine N-acetyltransferase from pigeon liver. *Biochimica et Biophysica Acta* **746**: 193-201.
 - Andres HH, Vogel RJ, Tarr GE, Johnson L, Weber WW (1987). Purification, physicochemical and kinetic properties of liver acetyl-CoA: arylamine N-acetyltransferase from rapid acetylators rabbits. *Molecular Pharmacology* **31**: 446-456.
 - Arbieva ZH, Banerjee K, Kim SY, Edassery SL, Maniatis VS, Horrigan SK, Westbrook CA (2000). High-resolution physical map and transcript identification of a prostate cancer deletion interval on 8p22. *Genome Research* **10**: 244-257.
 - Badawi AF, Hirvonen A, Bell DA, Lang NP, Kadlubar FF (1995). Role of aromatic amine acetyltransferases, NAT1 and NAT2, in carcinogen-DNA adduct formation in urinary bladder. *Cancer Research* **55**: 5230-5237.
 - Bandmann O, Vaughan J, Holmans P, Marsden CD, Wood NW (1997). Association of slow acetylator genotype for N-acetyltransferase 2 with familial Parkinson's disease. *Lancet* **350**: 1136-1139.

- Baranova H, Canis M, Ivaschenko T, Albuisson E, Bothorishvilli R, Baranov V, Malet P, Bruhat MA (1999). Possible involvement of arylamine N-acetyltransferase 2, glutathione S-transferases M1 and T1 genes in the development of endometriosis. *Molecular Human Reproduction* **5**: 636-641.
- Barret JH (1992). Genetic mapping based on radiation hybrid data. *Genomics* **13**: 95-103.
- Battey J, Jordan E, Cox D, Dove W (1999). An action plan for mouse genomics. *Nature Genetics* **21**: 73-75.
- Beisiegel U (1998). Lipoprotein metabolism. *European Heart Journal* **19 Suppl. A**: A20-23.
- Bell DA, Badawi AF, Lang NP, Ilett KF, Hirvonen A (1995a). Polymorphism in the N-acetyltransferase (NAT1) polyadenylation signal: Association of the *NAT1*10* allele with higher N-acetylation activity in bladder and colon tissue. *Cancer Research* **55**: 5226-5229.
- Bell DA, Stephens EA, Castranio T, Umbach DM, Watson M, Deakin M, Elder J, Hendrickse C, Duncan H, Strange R (1995b). Polyadenylation polymorphism in the acetyltransferase 1 gene (*NAT1*) increases risk of colorectal cancer. *Cancer Research* **55**: 3537-3542.
- Bell SM, Zuo J, Meyers RM, Knowles MA (1996). Fluorescence *in situ* hybridisation deletion mapping at 4p16.3 in bladder cancer cell lines refines the localisation of the critical interval to 30kb. *Genes, Chromosomes and Cancer* **17**: 108-117.
- Bellane-Chantelot C, Lacroix B, Ougen P, Billault A, Beauvils S, Bertrand S, Georges I, Gilbert F, Gros I, Lucotte G, Susini L, Codani J-J, Gesnouin P, Pook S, Vaysseix G, Lu-Kuo J, Ried T, Ward D, Chumakov I, Le Paslier D, Barillot E, Cohen D (1992). Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* **70**: 1059-1058.
- Bigler J, Chen C, Potter JD (1997). Determination of human *NAT2* acetylator genotype by oligonucleotide ligation assay. *BioTechniques* **22**: 682-690.
- Bishop DT (1994). BRCA1, BRCA2, BRCA3... a myriad of breast cancer genes. *European Journal of Cancer* **12**: 1738-1739.
- Blake JA, Richardson JE, Davisson MT, Eppig JT and the Mouse Informatics Group (1997). The Mouse Genome Database (MGD). A comprehensive public resource of genetic, phenotypic and genomic data. *Nucleic Acids Research* **25**: 85-91.

-
- Blum M, Demierre A, Grant DM, Heim M, Meyer U (1991). Molecular mechanism of slow acetylation of drugs and carcinogens in humans. *Proceedings of the National Academy of Sciences of the USA* **88**: 5237-5241.
 - Blum M, Grant DM, Demierre A, Meyer UA (1989a). Nucleotide sequence of a full-length cDNA for arylamine N-acetyltransferase from rabbit liver. *Nucleic Acids Research* **17**: 3589.
 - Blum M, Grant DM, Demierre A, Meyer UA (1989b). N-acetylation pharmacogenetics: A gene deletion causes absence of arylamine N-acetyltransferase in liver of slow acetylators rabbits. *Proceedings of the National Academy of Sciences of the USA* **86**: 9554-9557.
 - Blum M, Grant DM, McBride W, Heim M and Meyer UA (1990). Human N-acetyltransferase genes: Isolation, chromosomal localisation and functional expression. *DNA and Cell Biology* **9**: 193-203.
 - Bock KW and Lilienblum W (1994). Roles of uridine diphosphate glucuronosyltransferases in chemical carcinogenesis. In *Conjugation-Deconjugation Reactions in Drug Metabolism and Toxicity: Handbook of Experimental Pharmacology*, edited by Kauffman FC, Springer-Verlag, Berlin, pp. 391-428.
 - Bookstein R, Levy A, MacGrogan D, Lewis TB, Weissenbach J, O'Connell P, Leach AJ (1994). Yeast artificial chromosome and radiation hybrid map of loci in chromosome band 8p22, a common region of allelic loss in multiple human cancers. *Genomics* **24**:317-323.
 - Botstein P, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314-331.
 - Bouchardy C, Mitrunen K, Wikman H, Husgafvel-Pursiainen K, Dayer P, Benhamou S, Hirvonen A (1998). N-acetyltransferase NAT1 and NAT2 genotypes and lung cancer risk. *Pharmacogenetics* **8**: 291-298.
 - Boukouvala S and Sim E (2000). Tissue-specific expression of arylamine N-acetyltransferases in mice. *18th International Congress of Biochemistry and Molecular Biology*, Birmingham, UK.
 - Boukouvala S, Fakis G, Pinter K, Buckle V, Priddle H, Sim E (1999). Expression and localisation of the genes for arylamine N-acetyltransferases in mice. *Genetical Society's 10th Mammalian Genetics and Development Workshop*, London, UK.
 - Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254.

-
- Brady KP, Rowe LB, Her H, Stevens TJ, Eppig J, Sussman DJ, Sikela J, Neier DR (1997). Genetic mapping of 262 loci derived from expressed sequences in the murine interspecific cross using single-strand conformational polymorphism analysis. *Genome Research* **7**: 1085-1093.
 - Bruhn C, Brockmöller J, Cascorbi I, Roots I, Borchert H-H (1999). Correlation between genotype and phenotype of the human arylamine N-acetyltransferase type 1 (NAT1). *Biochemical Pharmacology* **58**: 1759-1764.
 - Buckle VJ and Rack KA (1993). Fluorescent *in situ* hybridisation. In *Human Genetic Disease Analysis: A Practical Approach*, edited by KE Davies, 2nd edition, IRL Press, Oxford, pp 59-80.
 - Burke DT, Carle GF, Olson MV (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**: 806-812.
 - Butcher NJ, Ilett KF, Minchin RF (1998). Functional polymorphism of the arylamine N-acetyltransferase type 1 gene caused by C¹⁹⁰T and G⁵⁶⁰A mutations. *Pharmacogenetics* **8**: 67-72.
 - Caldwell J, Gardner I, Swales N (1995). An introduction to drug disposition: The basic principles of absorption, distribution, metabolism and excretion. *Toxicologic Pathology* **23**: 102-114.
 - Caporaso N, Landi MT, Vineis P (1991). Relevance of metabolic polymorphisms to human carcinogenesis: Evaluation of epidemiologic evidence. *Pharmacogenetics* **1**: 4-19.
 - Cartwright RA, Glashan RW, Rogers HJ, Ahmad RA, Barham-Hall D, Higgins E, Kahn MA (1982). The role of N-acetyltransferase phenotype in bladder carcinogenesis: A pharmacogenetic epidemiological approach to bladder cancer. *Lancet* **2**: 842-846.
 - Carver EA and Stubbs L (1997). Zooming in on the human-mouse comparative map: Genome conservation re-examined on a high-resolution scale. *Genome Research* **7**: 1123-1137.
 - Cascorbi I and Roots I (1999). Pitfalls in N-acetyltransferase 2 genotyping. *Pharmacogenetics* **9**: 123-127.
 - Cascorbi I, Brockmöller J, Bauer S, Reum T, Roots I (1996a). NAT2*12A (803A>G) codes for rapid arylamine N-acetylation in humans. *Pharmacogenetics* **16**: 257-259.

-
- Cascorbi I, Brockmöller J, Mrozekiewicz PM, Bauer S, Loddenkemper R, Roots I (1996b). Homozygous rapid arylamine N-acetyltransferase (NAT2) genotype as a susceptibility factor for lung cancer. *Cancer Research* **56**: 3961-3966.
 - Case RA, Hosker ME, McDonald DB, Pearson JT (1954). Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. *British Journal of Industrial Medicine* **11**: 75-182.
 - Chapman VM and Nadeau JH (1992). The mouse genome: An overview. *Current Opinion in Genetics and Development* **2**: 406-411.
 - Chartier FL, Keer JT, Sutcliffe MJ, Henriques DA, Mileham P, Brown DM (1992). Construction of a mouse yeast artificial chromosome library in a recombinant-deficient strain of yeast. *Nature Genetics* **1**: 132-136.
 - Chen C-J, Yu M-W, Liaw Y-F (1997). Epidemiological characteristics and risk factors of hepatocellular carcinoma. *Journal of Gastroenterology and Hepatology* **12 (Suppl)**: S294-S308.
 - Chinen K, Isomura M, Izawa K, Fujiwara Y, Ohata H, Iwamasa T, Nakamura Y (1996). Isolation of 45 exon-like fragments from 8p22→p21.3, a region that is commonly deleted in hepatocellular, colorectal and non-small cell lung carcinomas. *Cytogenetics and Cell Genetics* **75**: 190-196.
 - Chotai J (1984). On the lod score method in linkage analysis. *Annals of Human Genetics* **48**: 359-378.
 - Chumakov IM, Rigault P, Le Gall I, Bellane-Chantelot C, Billault A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I, et al (1995). A YAC contig map of the human genome. *Nature* **377**: 175-183.
 - Chung JG, Levy GN, Weber WW (1993). Distribution of 2-aminofluorene and p-aminobenzoic acid N-acetyltransferase activity in the tissues of C57BL/6J rapid and B6.A-Nat^S slow acetylator congenic mice. *Drug Metabolism and Disposition* **21**: 1057-1063.
 - Clapper ML and Szarka CE (1998). Glutathione S-transferases — biomarkers of cancer risk and chemopreventive response. *Chemico-Biological Interactions* **111-112**: 377-388.
 - Clarke R, Smith AD, Jobst KA, Refsum H, Sutton L, Ueland PM (1998). Folate, vitamin B12 and serum total homocysteine levels in confirmed Alzheimer disease. *Archives of Neurology* **55**: 1449-1455.

-
- Collins F and Galas D (1993). A new five-year plan for the US Human Genome Program. *Science* **262**: 43-46.
 - Collins FS (1992a). Positional cloning: Let's not call it reverse anymore. *Nature Genetics* **1**: 3-6.
 - Collins FS (1992b). Identifying human disease genes by positional cloning. *The Harvey Lectures* **86**: 149-164.
 - Collins FS, Guyer MS, and Chakravarti A (1997). Variations on a theme: Cataloging human DNA sequence variation. *Science* **278**: 1580-1581.
 - Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L (1998). New goals for the US Human Genome Project: 1998-2003. *Science* **282**: 682-689.
 - Cooper DN and Smidtke J (1984). DNA restriction fragment length polymorphisms and heterozygosity in the human genome. *Human Genetics* **66**: 1-16.
 - Coroneos E and Sim E (1991). N-acetyltransferase in human urothelium and bladder cancer cell lines. *Biochemical Society Transactions* **19**: 129S.
 - Coroneos E and Sim E (1993). Arylamine N-acetyltransferase activity in human cultured cell lines. *Biochemical Journal* **294**: 481-486.
 - Coroneos E, Gordon JW, Kelly SL, Wang PD, Sim E (1991). Drug metabolising N-acetyltransferase activity in human cell lines. *Biochimica et Biophysica Acta* **1073**: 593-599.
 - Cox DR, Burmeister M, Price ER, Kim S, Myers R (1990). Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**: 245-250.
 - Cribb AE, Grant DM, Miller MA, Spielberg SP (1991). Expression of monomorphic arylamine N-acetyltransferase (NAT1) in human leucocytes. *Journal of Pharmacology and Experimental Therapeutics* **259**: 1241-1246.
 - Crundwell MC, Chughtai S, Knowles M, Takle L, Luscombe M, Neoptolemos JP, Morton DG, Phillips SM (1996). Allelic loss on chromosomes 8p, 22q and 18q (DCC) in human prostate cancer. *International Journal of Cancer* **69**: 295-300.
 - Csordas A (1989). A proposal for a possible role of nucleosome positioning in the evolutionary adjustment of introns. *International Journal of Biochemistry* **21**: 455-461.

-
- Curran JL (1997). Human linkage mapping. In *Genome Mapping: A Practical Approach*, edited by PH Dear, Oxford University Press.
 - Czeizel AE and Dudás I (1992). Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *New England Journal of Medicine* **327**: 1832-1835.
 - Dausset J, Cann H, Cohen D, Lathrop M, Lalouel J-M, White R (1990). Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative Genetic Mapping of the Human Genome. *Genomics* **6**: 575-577.
 - Davies KE and Tilghman SM, editors (1990). Genome analysis volume 1: Genetic and physical mapping. Cold Spring Harbor Laboratory Press, USA.
 - Davis DL, Bradlow HL, Wolff M, Woodruff T, Hoel DG, Anton-Culver H (1993). Medical hypothesis: Xenoestrogens as preventable causes of breast cancer. *Environmental Health Perspectives* **101**:372-377.
 - De Leon JH, Martell KJ, Vatsis KP, Weber WW (1995). Slow acetylation in mice is caused by a labile and catalytically impaired mutant N-acetyltransferase (NAT2 9). *Drug Metabolism and Disposition* **23**: 1354-1361.
 - Dear PH, editor (1997). *Genome Mapping: A Practical Approach*. Oxford University Press, New York.
 - Deguchi T, Mashimo M, Suzuki T (1990). Correlations between acetylator phenotypes and genotypes of polymorphic arylamine N-acetyltransferase in human liver. *Journal of Biological Chemistry* **265**: 12757-12760.
 - Delaforge M (1998). Importance of metabolism in pharmacological studies: Possible *in vitro* predictability. *Nuclear Medicine and Biology* **25**: 705-709.
 - Deloménie C, Goodfellow GH, Krishnamoorthy R, Grant DM, Dupret JM (1997). Study of the role of the highly conserved residues Arg⁹ and Arg⁶⁴ in the catalytic function of human N-acetyltransferases NAT1 and NAT2 by site-directed mutagenesis. *Biochemical Journal* **323**: 207-215.
 - Deloménie C, Grant DM, Mathelier-Fusade P, Jacomet C, Leynadier F, Jacqz-Aigrain E, Rozenbaum W, Krishnamoorthy R, Dupret JM (1994). N-acetylation genotype and risk of severe reactions to sulphonamides in AIDS patients. *British Journal of Clinical Pharmacology* **38**: 581-582.
 - Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, Matise TC, McKusick KB, Beckmann JS, et al. (1998). A physical map of 30,000 human genes. *Science* **282**:744-746.

-
- Derewlany LO and Koren G (1994). Biotransformation of carcinogenic arylamines and arylamides by human placenta. *Journal of Laboratory and Clinical Medicine* **124**: 134-141.
 - Derewlany LO, Knie B, Koren G (1994a). Arylamine N-acetyltransferase activity of the human placenta. *Journal of Pharmacology and Experimental Therapeutics* **269**: 756-760.
 - Derewlany LO, Knie B, Koren G (1994b). Human placental transfer and metabolism of p-aminobenzoic acid. *Journal of Pharmacology and Experimental Therapeutics* **269**: 761-765.
 - Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**:152-154.
 - Diehl SR and Erickson RP (1997). Genome scan for teratogen-induced clefting susceptibility loci in the mouse: Evidence of both allelic and locus heterogeneity distinguishing cleft lip and cleft palate. *Proceedings of the National Academy of Sciences of the USA* **94**: 5231-5236.
 - Dietrich WF, Copeland NG, Gilbert DJ, Miller JC, Jenkins NA, Lander ES (1995). Mapping the mouse genome: Current status and future prospects. *Proceedings of the National Academy of Sciences of the USA* **92**: 10849-10853.
 - Dietrich WF, Miller JC, Steen RG, Merchant M, Damron D, Nahf R, Gross A, Joyce DC, Wessel M, Dredge RD, et al (1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics* **7**: 220-245.
 - Dilela AG and Woo SLC (1987). Cloning large segments of genomic DNA using cosmid vectors. *Methods in Enzymology* **152**: 199-212.
 - Doll MA and Hein DW (1995). Cloning, sequencing and expression of NAT1 and NAT2 encoding genes from rapid and slow acetylator inbred rats. *Pharmacogenetics* **5**: 247-251.
 - Doll MA, Fretland AJ, Deitz AC, Hein DW (1995). Determination of human NAT2 acetylator genotype by restriction fragment length polymorphism and allele specific amplification. *Analytical Biochemistry* **231**: 412-420.
 - Doll MA, Jiang W, Deutz AC, Rustan TD, Hein DW (1997). Identification of a novel allele at the human *NAT1* acetyltransferase locus. *Biochemical and Biophysical Research Communications* **233**: 584-591.

-
- Donnis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keih TP, Bowden DW, Smith DR, Lander ES, et al (1987). A genetic linkage map of the human genome. *Cell* **51**: 319-337.
 - Dooley TP (1998). Cloning of the human phenol sulfotransferase gene family: Three genes implicated in the metabolism of catecholamines, thyroid hormones and drugs. *Chemico-Biological Interactions* **109**: 29-41.
 - Dunham I, Shimizu N, Roe BA, Chisoe S, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, et al (1999). The DNA sequence of human chromosome 22. *Nature* **402**: 489-495.
 - Ebisawa T and Deguchi T (1991). Structure and restriction fragment length polymorphism of genes for human liver arylamine N-acetyltransferases. *Biochemical and Biophysical Research Communications* **177**: 1252-1257.
 - Edwards A and Caskey CT (1991). Genetic marker technology. *Current Opinion in Biotechnology* **2**: 818-822.
 - El-Yazigi A, Johansen K, Raines DA, Dossing M (1992). N-acetylation polymorphism and diabetes mellitus among Saudi Arabians. *Journal of Clinical Pharmacology* **32**: 905-910.
 - Emi M, Asaoka H, Matsumoto A, Itakura H, Kurihara Y, Wada Y, Kanamori H, Yazaki Y, Takahashi E, Lepert M, Lalouel J-M, Kodama T, Mukai T (1993a). Structure, organisation and chromosomal mapping of the human macrophage scavenger receptor gene. *Journal of Biological Chemistry* **268**:2120-2125.
 - Emi M, Fugiwara Y, Ohata H, Tsuda H, Hirohashi S, Koike M, Miyaki M, Monden M, Nakamura Y (1993b). Allelic loss at chromosome band 8p21.3-p22 is associated with progression of hepatocellular carcinoma. *Genes Chromosomes and Cancer* **7**: 152-157.
 - Eppig JT, Blake JA, Davisson MT, Richardson JE (1998). Informatics for mouse genetics and genome mapping. *Methods: A Companion to methods in Enzymology* **14**: 179-190.
 - Estrada L, Kanelakis KC, Levy GN, Weber WW (2000). Tissue- and gender-specific expression of N-acetyltransferase 2 (*NAT2**) during development of the outbred mouse strain CD-1. *Drug Metabolism and Disposition* **28**: 139-146.
 - Estrada-Rodgers L, Levy GN, Weber WW (1998a). Characterisation of a hormone response element in the mouse N-acetyltransferase 2 (*Nat2**) promoter. *Gene Expression* **7**: 13-24.
 - Estrada-Rodgers L, Levy GN, Weber WW (1998b). Substrate selectivity of mouse N-acetyltransferase 1, 2, and 3 expressed in COS-1 cells. *Drug Metabolism and Disposition* **26**: 502-505.

-
- Evans DAP and White TA (1964). Human acetylation polymorphism. *Journal of Laboratory and Clinical Medicine* **63**: 394.
 - Evans DAP, Manley KA and McKusick VA (1960). Genetic control of isoniazid metabolism in man. *British Medical Journal* **2**: 4485-4491.
 - Evans WE and Relling MV (1999). Pharmacogenomics: Translating functional genomics into rational therapeutics. *Science* **286**: 487-491.
 - Eydmann ME and Knowles MA (1997). Mutation analysis of 8p genes *POLB* and *PPP2CB* in bladder cancer. *Cancer Genetics and Cytogenetics* **93**: 167-171.
 - Fakis G, Boukouvala S, Buckle V, Payton M, Denning C, Sim E (2000). Chromosomal localisation and mapping of the genes for murine arylamine N-acetyltransferases (NATs), enzymes involved in the metabolism of carcinogens: Identification of a novel upstream non-coding exon for murine *Nat2*. *Cytogenetics and Cell Genetics* **90**: 134-138.
 - Farrington SM, Cunningham C, Boyle SM, Wyllie AH, Dunlop MG (1996). Detailed physical and deletion mapping of 8p with isolation of YAC clones from tumour suppressor loci involved in colorectal cancer. *Oncogene* **12**: 1803-1808.
 - Feinberg AP and Vogelstein B (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**: 6-13.
 - Ferguson-Smith MA (1997). Genetic analysis by chromosome sorting and painting: Phylogenetic and diagnostic applications. *European Journal of Human Genetics* **5**: 253-265.
 - Fidlerova H, Senger G, Kost M, Sanseau P, Sheer D (1994). Two simple procedures for releasing chromatin from routinely fixed cells for fluorescence *in situ* hybridisation. *Cytogenetics and Cell Genetics* **65**: 203-205.
 - Filiadis IF, Georgiou I, Alamanos Y, Kranas V, Giannakopoulos X, Lolis D (1999). Genotypes of N-acetyltransferase-2 and risk of bladder cancer: A case-control study. *Journal of Urology* **161**: 1672-1675.
 - Flammang TJ, Couch LH, Levy GN, Weber WW and Wise CK (1992). DNA adduct levels in congenic rapid and slow acetylator mouse strains following chronic administration of 4-aminobiphenyl. *Carcinogenesis* **13**: 1887-1891.
 - Fleming A and Copp AJ (1998). Embryonic folate metabolism and mouse neural tube defects. *Science* **280**: 2107-2109.

-
- Franke S, Klowitz I, Schnakenberg E, Rommel B, Van de Ven W, Bullerdiek J, Schloot W (1994). Isolation and mapping of a cosmid clone containing the human *NAT2* gene. *Biochemical and Biophysical Research Communications* **199**: 52-55.
 - Fretland AJ, Doll MA, Gray K, Feng Y, Hein D (1997). Cloning, sequencing and recombinant expression of *NAT1*, *NAT2* and *NAT3* derived from the C3H/HeJ (rapid) and A/HeJ (slow) acetylator inbred mouse: Functional characterisation of the activation and deactivation of aromatic amine carcinogens. *Toxicology and Applied Pharmacology* **142**: 360-366.
 - Fujiwara Y, Emi M, Ohata H, Kato Y, Nakajima T, Mori T, Nakamura Y (1993). Evidence for the presence of two tumour suppressor genes on chromosome 8p for colorectal carcinoma. *Cancer Research* **53**: 1172-1174.
 - Fujiwara Y, Ohata H, Kuroki T, Koyama K, Tsuchiya E, Mondn M, Nakamura Y (1995). Isolation of a candidate tumour suppressor gene on chromosome 8p21.3-p22 that is homologous to an extracellular domain on the PDGF receptor beta gene. *Oncogene* **10**:891-895.
 - Fukutome K, Watanabe M, Shiraishi T, Murata M, Uemura H, Kubota Y, Kawanamura J, Ito H, Yatani R (1999). N-acetyltransferase 1 genetic polymorphism influences the risk of prostate cancer development. *Cancer Letters* **136**: 83-87.
 - Gertig DM, Hankinson SE, Hough H, Spiegelman D, Colditz GA, Willett W, Kelsey KT, Hunter DJ (1999). N-acetyltransferase 2 genotypes, meat intake and breast cancer risk. *International Journal of Cancer* **80**: 13-17.
 - Gibson GG and Skett P (1994). Introduction to drug metabolism. 2nd edition, Blackie Academic and Professional, Glasgow, UK.
 - Gil JP and Lechner MC (1998). Increased frequency of wild-type arylamine N-acetyltransferase allele *NAT2*4* homozygotes in Portuguese patients with colorectal cancer. *Carcinogenesis* **19**: 37-41.
 - Glatt H (1997). Sulfation and Sulfotransferases 4: Bioactivation of mutagens by sulfation. *FASEB Journal* **11**: 314-321.
 - Glatt H, Engelke CEH, Pabel U, Teubner W, Jones AL, Coughtrie MWH, Andrae U, Falany CN, Meinel W (2000). Sulfotransferases: Genetics and role in toxicology. *Toxicology Letters* **112-113**: 341-348.
 - Glowinski IB and Weber WW (1982a). Genetic regulation of aromatic amine N-acetylation in inbred mice. *Journal of Biological Chemistry* **257**: 1424-1430.
 - Glowinski IB and Weber WW (1982b). Biochemical characterisation of genetically variant aromatic amine N-acetyltransferases in A/J and C57BL/6J mice. *Journal of Biological Chemistry* **257**: 1431-1437.

-
- Gonzalez FJ and Kimura S (1999). Role of gene knockout mice in understanding the mechanisms of chemical toxicity and carcinogenesis. *Cancer Letters* **143**: 199-204.
 - Gonzalez FJ and Nebert DW (1990). Evolution of the P450 gene superfamily: Animal-plant “warfare”, molecular drive and human genetic differences in drug oxidation. *Trends in Genetics* **6**: 182-186.
 - Gonzalez MV, Alvarez V, Pello MF, Menendez MJ, Suarez C, Coto E (1998). Genetic polymorphism of N-acetyltransferase-2, glutathione S-transferase-M1, and cytochromes P450IIE1 and P450IID6 in the susceptibility to head and neck cancer. *Journal of Clinical Pathology* **51**: 294-298.
 - Goss SJ and Harris H (1975). New method for mapping genes in human chromosomes. *Nature* **255**: 680-684.
 - Goss SJ and Harris H (1977a). Gene transfer by means of cell fusion I: Statistical mapping of the human X-chromosome by analysis of radiation-induced gene segregation. *Journal of Cell Science* **25**: 17-37.
 - Goss SJ and Harris H (1977b). The mapping of 8 loci on human chromosome 1 by statistical analysis of gene assortment in somatic cell hybrids. *Journal of Cell Science* **25**: 39-57.
 - Grant DM (1993). Molecular genetics of the N-acetyltransferases. *Pharmacogenetics* **3**: 45-50.
 - Grant DM, Blum M, Beer M, Meyer UA (1991). Monomorphic and polymorphic human arylamine N-acetyltransferases: a comparison of liver isozymes and expressed products of two cloned genes. *Molecular Pharmacology* **39**: 184-191.
 - Grant DM, Blum M, Demierre A, Meyer UA (1989a). Nucleotide sequence of an intronless gene for a human arylamine N-acetyltransferase related to polymorphic drug acetylation. *Nucleic Acids Research* **17**: 3978.
 - Grant DM, Hughes NC, Janezic SA, Goodfellow GH, Chen HJ, Gaedigk A, Yu VL, Grewal R (1997). Human acetyltransferase polymorphisms. *Mutation Research* **376**: 61-70.
 - Grant DM, Lottspeich F, Meyer UA (1989b). Evidence for two closely related isozymes of arylamine N-acetyltransferase in human liver. *FEBS Letters* **244**: 203-207.
 - Green MD and Tephly TR (1998). Glucuronidation of amine substrates by purified and expressed UDP-glucuronosyltransferase proteins. *Drug Metabolism and Disposition* **26**: 860-867.

- Grutzner F, Himmelbauer H, Paulsen M, Ropers H-H, Haaf T (1999). Comparative mapping of mouse and rat chromosomes by fluorescence *in situ* hybridisation. *Genomics* **55**: 306-313.
- Gryfe R, Swallow C, Bapat B, Redston M, Gallinger S, Couture J (1997). Molecular biology of colorectal cancer. *Current Problems in Cancer* **21**: 233-300.
- Gustafson CE, Wilson PJ, Lukeis R, Baker E, Woollatt E, Annab L, Hawke L, Barrett JC, Chenevix-Trench G (1996). Functional evidence for colorectal cancer tumor suppressor gene at chromosome 8p22-23 by monochromosome transfer. *Cancer Research* **56**: 5238-5245.
- Guyer MS and Collins FS (1995). How is the Human Genome Project doing and what have we learned so far? *Proceedings of the National Academy of Sciences of the USA* **92**: 10841-10848.
- Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillet D, Muselet D, Prud'Homme J-F, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN (1996). A radiation hybrid map of the human genome. *Human Molecular Genetics* **5**: 339-346.
- Haldi ML, Strickland C, Lim P, VanBerkel V, Chen X-N, Noya D, Korenberg JR, Husain Z, Miller J, Lander ES (1996). A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. *Mammalian Genome* **7**: 767-769.
- Hanna PE (1996). Metabolic activation and detoxification of arylamines. *Current Medicinal Chemistry* **3**: 195-210.
- Harhangi BS, Oostra BA, Heutink P, van Duijn CM, Hofman A, Breteler MMB (1999). N-acetyltransferase-2 polymorphism in Parkinson's disease: The Rotterdam study. *Journal of Neurology, Neurosurgery and Psychiatry* **67**: 518-520.
- Harrison DJ (1998). Life and death in the genesis of the tumour cell. *Toxicology Letters* **102-103**: 115-119.
- Hastings RJ and Franks LM (1981). Chromosome pattern, growth in agar and tumorigenicity in nude mice of four human bladder carcinoma cell lines. *International Journal of Cancer* **27**: 15-21.

-
- Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park H-S, Yoyoda A, Ishii K, Totoki Y, Choi D-K, et al (2000). The DNA sequence of human chromosome 21. *Nature* **405**: 311-319.
 - Hein DW (1988). Acetylator genotype and arylamine-induced carcinogenesis. *Biochimica et Biophysica Acta* **948**: 37-66.
 - Hein DW, Doll MA, Fretland AJ, Leff MA, Webb SJ, Xiao GH, Devanaboyina U-S, Nangju NA, Feng Y (2000a). Molecular genetics and epidemiology of the *NAT1* and *NAT2* acetylation polymorphisms. *Cancer Epidemiology, Biomarkers and Prevention* **9**: 29-42.
 - Hein DW, Grant DM, Sim E (2000b). Update on consensus arylamine N-acetyltransferase gene nomenclature. *Pharmacogenetics* **10**: 291-293.
 - Hein DW, Hirata M, Glowinski IB, Weber WW (1982a). Biochemical evidence for the coexistence of monomorphic and polymorphic N-acetyltransferase activities on a common protein in rabbit liver. *Journal of Pharmacology and Experimental Therapeutics* **220**: 1-7.
 - Hein DW, Rustan TD, Bucher KD, Martin WJ, Furman EJ (1991). Acetylator phenotype-dependent and -independent expression of arylamine N-acetyltransferase isozymes in rapid and slow acetylators in inbred rat liver. *Drug Metabolism and Disposition* **19**: 933-937.
 - Hein DW, Smolen TN, Fox RR, Weber WW (1982b). Identification of genetically homozygous rapid and slow acetylators of drugs and environmental carcinogens among established inbred rabbit strains. *Journal of Pharmacology and Experimental Therapeutics* **223**: 40-44.
 - Hein DW, Trinidad A, Yerokun T, Ferguson RJ, Kirilin WG, Weber WW (1988). Genetic control of acetyl coenzyme A-dependent arylamine N-acetyltransferase, hydrazine N-acetyltransferase, and N-hydroxy-arylamine O-acetyltransferase enzymes in C57BL/6J, A/J, AC57F₁, and the rapid and slow acetylators A.B6 and B6.A congenic inbred mouse. *Drug Metabolism and Disposition* **16**: 341-347.
 - Heiskanen M, Hellsten E, Kollioniemi O-P, Makela TP, Alitalo K, Peltonen L, Palotie A (1995). Visual mapping by Fiber-FISH. *Genomics* **30**: 31-36.
 - Henning S, Cascorbi I, Münchow B, Jahnke V, Roots I (1999). Association of arylamine N-acetyltransferases *NAT1* and *NAT2* genotypes to laryngeal cancer risk. *Pharmacogenetics* **9**: 103-111.
 - Hickman D and Sim E (1991). N-acetyltransferase polymorphism: Comparison of phenotype and genotype in humans. *Biochemical Pharmacology* **42**: 1007-1014.

-
- Hickman D, Pope J, Patil SD, Fakis G, Smelt V, Stanley LA, Payton M, Unadkat JD, Sim E (1998). Expression of arylamine N-acetyltransferase in human intestine. *Gut* **42**: 402-409.
 - Hickman D, Risch A, Buckle V, Spurr NK, Jeremiah SJ, McCarthy A and Sim E (1994). Chromosomal localisation of human genes for arylamine N-acetyltransferase. *Biochemical Journal* **297**: 441-445.
 - Hickman D, Risch A, Camilleri JP, Sim E (1992). Genotyping human polymorphic arylamine N-acetyltransferase: Identification of new slow allotypic variants. *Pharmacogenetics* **2**: 217-226.
 - Hirvonen A (1999). Polymorphisms of xenobiotic-metabolising enzymes and susceptibility to cancer. *Environmental Health Perspectives* **107S**: 37-47.
 - Hooper ML (1998). Tumour suppressor gene mutations in humans and mice: Parallels and contrasts. *EMBO Journal* **17**: 6783-6789.
 - Horn D, Krebsova A, Kunze J, Reis A (2000). Homozygosity mapping in a family with microcephaly, mental retardation and short stature to a Cohen syndrome region on 8q21.3-8q22.1: Redefining a clinical entity. *American Journal of Medical Genetics* **92**: 285-292.
 - Hou S-M, Ryberg D, Fält S, Deverill A, Tefre T, Børresen A-L, Haugen A, Lambert B (2000). GSTM1 and NAT2 polymorphisms in operable and non-operable lung cancer patients. *Carcinogenesis* **21**: 49-54.
 - Hsieh F-I, Pu Y-S, Chern H-D, Hsu L-I, Chiou H-Y, Chen C-J (1999). Genetic polymorphisms of N-acetyltransferase 1 and 2 and risk of cigarette smoking-related bladder cancer. *British Journal of Cancer* **81**: 537-541.
 - Huang C-S, Chern H-D, Shen C-Y, Hsu S-M, Chang K-J (1999). Association between N-acetyltransferase 2 (*NAT2*) genetic polymorphism and development of breast cancer in post-menopausal Chinese women in Taiwan, an area of great increase in breast cancer incidence. *International Journal of Cancer* **82**: 175-177.
 - Hubbard A, Moyes C, Wyllie AH, Smith CAD, Harisson DJ (1998). N-acetyltransferase 1: Two polymorphisms in coding sequence identified in colorectal cancer patients. *British Journal of Cancer* **77**: 913-916.
 - Hubbard AL, Harrison DJ, Moyes C, Wyllie AH, Cunningham C, Manion E, Smith CAD (1997). N-acetyltransferase 2 genotype in colorectal cancer and selective gene retention in cancers with chromosome 8p deletions. *Gut* **41**: 229-234.

-
- Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu S-H, et al (1995). An STS-based map of the human genome. *Science* **270**: 1945-1954.
 - Hughes HB, Biehl JP, Jones AP and Schmidt LH (1954). Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. *American Journal of Tuberculosis* **70**: 266-273.
 - Hughes NC, Janezic SA, McQueen KL, Jewett MAS, Castranio T, Bell DA, Grant D (1998a). Identification and characterisation of variant alleles of human acetyltransferase NAT1 with defective function using p-aminosalicylate as an in-vivo and in-vitro probe. *Pharmacogenetics* **8**: 55-66.
 - Hughes SJ, Glover TW, Zhu XX, Kuick R, Thoraval D, Orringer MB, Beer DG, Hanash S (1998b). A novel amplicon at 8p22-23 results in overexpression of cathepsin B in esophageal adenocarcinoma. *Proceeding of the National Academy of Sciences USA* **95**:12410-12415.
 - Hunter DJ, Hankinson SE, Hough H, Gertig DM, Garcia-Closas M, Spiegelmn D, Manson JAE, Colditz GA, Willet WC, Speizer FE, Kelsey K (1997). A prospective study of NAT2 acetylation genotype, cigarette smoking and risk of breast cancer. *Carcinogenesis* **18**: 2127-2132.
 - Hunter KW, Watson ML, Rochelle J, Ontiveros S, Munroe D, Seldin MF, Housman DE (1993). Single strand conformational polymorphism (SSCP) mapping of the mouse genome: Integration of the SSCP, microsatellite and gene maps of mouse chromosome 1. *Genomics* **18**: 510-519.
 - Idle JR, Armstrong M, Boddy AV, Boustead C, Cholerton S, Cooper J, Daly AK, Ellis J, Gregory W, Hadidi H, et al (1992). The pharmacogenetics of chemical carcinogenesis. *Pharmacogenetics* **2**: 246-258.
 - Ilett KF, David BM, Detchon P, Castleden WM, Kwa R (1987). Acetylator phenotype in colorectal carcinoma. *Cancer Research* **47**: 1466-1469.
 - Ilett KF, Ingram DM, Carpenter DS, Teitel CH, Lang NP, Kadlubar FF, Minchin RF (1994). Expression of monomorphic and polymorphic N-acetyltransferases in human colon. *Biochemical Pharmacology* **47**: 914-917.
 - Ilett KF, Reeves PT, Minchin RF, Kinnear BF, Watson HF, Kadlubar FF (1991). Distribution of acetyltransferase activities in the intestines of rapid and slow acetylator rabbits. *Carcinogenesis* **12**: 1465-1469.
 - Ingelman-Sundberg M (1983). Bioactivation or inactivation of toxic compounds? In *Drug Metabolism and Distribution*, edited by JW Lamble, Elsevier Biomedical Press, Amsterdam, pp. 22-29.

-
- Ingelman-Sundberg M (1998). Functional consequences of polymorphism of xenobiotic metabolising enzymes. *Toxicology Letters* **102-103**: 155-160.
 - Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Baltzer MA, de Jong PJ (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genetics* **6**: 84-89.
 - Ishibe N, Wiencke JK, Zuo Z-F, McMillan A, Spitz M, Kelsey KT (1998). Polymorphisms in the N-acetyltransferase 1 (*NAT1*) gene and lung cancer in a minority population. *Biomarkers* **3**: 219-226.
 - Isomura M, Ikegawa S, Kinjo T, Takeuchi K, Yamane-Tanaka Y, Kitami K, Nakamura Y (1999). Sequence analysis of a total of three megabases of DNA in two regions of chromosome 8p. *DNA Research* **6**: 387-400.
 - Iyer L and Ratain MJ (1998). Pharmacogenetics and Cancer Chemotherapy. *European Journal of Cancer* **34**: 1493-1499.
 - Jencks WP, Gresser M, Valenzuela MS, Huneeus FC (1972). Acetyl coenzyme A: arylamine acetyltransferase. Measurement of the steady state concentration of the acetyl-enzyme intermediate. *Journal of Biological Chemistry* **247**: 3756-3760.
 - Jenne JW (1965). Partial purification and properties of the isoniazid trans-acetylase in human liver. Its relationship to the acetylation of p-aminosalicylic acid. *Journal of Clinical Investigation* **44**: 1992-2002.
 - Johnson N., Troen A, Fernando S, Warden D, Nagy Z, Smith AD, Sim E (2000). Investigation of N-acetyltransferase (*NAT1*) in Alzheimer's disease: Identification of a novel *NAT1* allelic variant. *European Toxicological Society, Eurotox* September 2000, London, UK.
 - Jones PA and Laird PW (1999). Cancer epigenetics comes of age. *Nature Genetics* **21**: 163-167.
 - Josephy PD and Coomber BL (1998). The 1996 Veylien Henderson Award of the Society of Toxicology of Canada. Current concepts: Neutrophils and the activation of carcinogens in the breast and other organs. *Canadian Journal of Physiology and Pharmacology* **76**: 639-700.
 - Jourenkova-Mironova N, Wikman H, Bouchardy C, Mitrinen K, Dayer P, Benhamou S, Hirvonen A (1999). Role of N-acetyltransferase 1 and 2 (*NAT1* and *NAT2*) genotypes in susceptibility to oral/pharyngeal and laryngeal cancers. *Pharmacogenetics* **9**: 533-537.
 - Kaderlik KR and Kadlubar FF (1995). Metabolic polymorphism and carcinogen-DNA adduct formation in human populations. *Pharmacogenetics* **5**: S108-S117.

-
- Kadlubar FF, Butler MA, Kaderlik KR, Chou H-C, Lang NP (1992). Polymorphisms for aromatic amine metabolism in humans: Relevance in human carcinogenesis. *Environmental Health Perspectives* **98**: 69-74.
 - Kallioniemi A, Kallioniemi OP, Citro G, Sauter G, DeVries S, Kerschmann R, Carroll P, Waldman F (1995). Identification of gains and losses of DNA sequences in primary bladder cancer by comparative genomic hybridisation. *Genes, Chromosomes and Cancer* **12**: 213-219.
 - Karolyi J, Erickson RP, Liu S (1988). Genetics of susceptibility to 6-aminonicotinamide-induced cleft palate in the mouse: Studies in congenic and recombinant inbred strains. *Teratology* **37**: 283-287.
 - Karolyi J, Erickson RP, Liu S, Killewald L (1990). Major effects on teratogen-induced facial clefting in mice determined by a single genetic region. *Genetics* **126**: 201-205.
 - Karolyi J, Liu S, Erickson RP (1987). Susceptibility to phenytoin-induced cleft lip with or without cleft palate: Many genes are involved. *Genetical Research* **49**: 43-49.
 - Katoh T, Boissy R, Nagata N, Kitagawa K, Kuroda Y, Itoh H, Kawamoto T, Bell DA (2000). Inherited polymorphism in the N-acetyltransferase 1 (*NAT1*) and 2 (*NAT2*) genes and susceptibility to gastric and colorectal adenocarcinoma. *International Journal of Cancer* **85**: 46-49.
 - Katoh T, Inatomi H, Yang M, Kawamoto T, Matsumoto T, Bell DA (1999). Arylamine N-acetyltransferase 1 (*NAT1*) and 2 (*NAT2*) genes and risk of urothelial transitional cell carcinoma among Japanese. *Pharmacogenetics* **9**: 401-404.
 - Katoh T, Kaneko S, Boissy R, Watson M, Ikemura K, Bell DA (1998). A pilot study testing the association between N-acetyltransferase 1 and 2 and risk of oral squamous cell carcinoma in Japanese people. *Carcinogenesis* **19**: 1803-1807.
 - Kaufmann GR, Wenk M, Taeschner W, Peterli B, Gyr K, Meyer UA, Haefeli WE (1996). N-acetyltransferase 2 polymorphism in patients infected with human immunodeficiency virus. *Clinical Pharmacology and Therapeutics* **60**: 62-67.
 - Kelly SL and Sim E (1991). Expression of N-acetyltransferase in a human monocytic cell line, U937. *Human and Experimental Toxicology* **10**: 33-38.
 - Kelly SL and Sim E (1992). Arylamine N-acetyltransferase in mice. *Biochemical Society Transactions* **20**: 106S.

-
- Kelly SL and Sim E (1994). Arylamine N-acetyltransferase in Balb/c mice: Identification of a novel mouse isoenzyme by cloning and expression *in vitro*. *Biochemical Journal* **302**: 347-353.
 - Kennedy S (1999). The genetics of endometriosis. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **82**: 129-133.
 - Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**: 1073-1080.
 - King M-C (1992). Breast cancer genes: How many, where and who are they? *Nature Genetics* **2**: 89-90.
 - Kirkwood BR (1988). *Essentials of Medical Statistics*. Blackwell Science, Oxford, UK.
 - Kloth MT, Gee RL, Messing EM, Swaminathan S (1994). Expression of N-acetyltransferase (NAT) in cultured human uroepithelial cells. *Carcinogenesis* **15**: 2781-2787.
 - Knowles MA (1999). Identification of novel bladder tumour suppressor genes. *Electrophoresis* **20**: 269-279.
 - Knowles MA, Elder PA, Williamson M, Cairns JP, Shaw ME, Law MG (1994). Allelotype of human bladder cancer. *Cancer Research* **54**: 531-538.
 - Knowles MA, Shaw ME, Proctor AJ (1993). Deletion mapping of chromosome 8 in cancers of the urinary bladder using restriction fragment length polymorphism and microsatellite polymorphisms. *Oncogene* **8**: 1357-1364.
 - Korenberg JR, Chen X-N, Sun Z, Shi Z-Y, Ma S, Vataru E, Yimlamai D, Weissenbach JS, Shizuya H, Simon MI, Gerety SS, Nguyen H, Zemsteva IS, Hui L, Silva J, Wu X, Birren B, Hudson TJ (1999). Human genome anatomy: BACs integrating the genetic and cytogenetic maps bridging genome and biomedicine. *Genome Research* **9**: 994-1001.
 - Kownacki-Brown PA, Wang C, Bailey LB, Toth JP, Gregory JF (1993). Urinary excretion of deuterium-labelled folate and the metabolite p-aminobenzoylglutamate in humans. *Journal of Nutrition* **123**: 1101-1108.
 - Kwok P-Y, Deng Q, Zakeri H, Taylor SL, Nickerson A (1996). Increasing the information content of STS-based genome maps: Identifying polymorphisms in mapped STSs. *Genomics* **31**: 123-126.
 - La Du BN (1992). Overview of pharmacogenetics. In *Pharmacogenetics of Drug Metabolism*, edited by W. Kalow, Pergamon Press, New York, pp. 1-12.

- Labuda D, Krajcinovic M, Richer C, Skoll A, Sinnett H, Yotova V, Sinnett D (1999). Rapid detection of CYP1A1, CYP2D6 and NAT variants by multiplex polymerase chain reaction and allele-specific oligonucleotide assay. *Analytical Biochemistry* **275**: 84-92.
- Ladero JM, Arrojo A, de Salamanca RE, Gomez M, Cano F, Alfonso M (1982). Hepatic acetylator phenotype in diabetes mellitus. *Annals of Clinical Research* **14**: 187-189.
- Ladero JM, Barquero MS, Coria F, Molina JA, Jimenez-Jimenez FJ, Benitez J (1993). Acetylator polymorphism in Alzheimer's disease. *European Journal of Medicine* **2**: 281-283.
- Ladero JM, Gonzalez JF, Benitez J, Vargas E, Fernandez MJ, Baki W, Diaz-Rubio M (1991). Acetylator polymorphism in human colorectal carcinoma. *Cancer Research* **51**: 2098-2100.
- Ladero JM, Jimenez FJ, Benitez J, Fernandez-Gundin MJ, Martinez C, Llerena A, Cobaleda J, Munoz JJ (1989). Acetylator Polymorphism in Parkinson's disease. *European Journal of Clinical Pharmacology* **37**: 391-393.
- Lang NP, Butler MA, Massengill J, Lawson M, Stotts RC, Hauer-Jensen M, Kadlubar FF (1994). Rapid metabolic phenotypes for acetyltransferase and cytochrome P4501A2 and putative exposure to food-borne heterocyclic amines increase the risk for colorectal cancer or polyps. *Cancer Epidemiology, Biomarkers and Prevention* **3**: 675-682.
- Lawrence JB, Villnave CA, Singer RH (1988). Sensitive, high-resolution chromatin and chromosome mapping *in situ* – presence and orientation of 2 closely integrated copies of EBV in a lymphoma line. *Cell* **52**: 51-61.
- Lawrence S, Morton NE, Cox DR (1991). Radiation hybrid mapping. *Proceedings of the National Academy of Sciences of the USA* **88**: 7477-7480.
- Leach RJ, Banga SS, Ben-Othame K, Chugtai S, Clarke R, Daiger SP, Kolehmainen J, Kumar S, Kuo M, Macoska J, Mada N, Naylor SL, Nunes M, O'Connell P, Pebusque M-J, Pekkel V, Porter CJ, Simons CT, Sohocki MM, Trapman J, Wells D, Westbrook C, Wood S (1996). Report of the third international workshop on human chromosome 8 mapping 1996. *Cytogenetics and Cell Genetics* **75**: 71-84.
- Lee BL, Wong D, Benowitz NL, Sullam PM (1993). Altered patterns of drug metabolism in patients with acquired immunodeficiency syndrome. *Clinical Pharmacology and Therapeutics* **3**: 529-535.
- Lee EJD, Zhao B, Seow-Choen F (1998). Relationship between polymorphism of N-acetyltransferase gene and susceptibility to colorectal carcinoma in a Chinese population. *Pharmacogenetics* **8**: 513-517.

- Leff MA, Epstein PN, Doll MA, Fretland AJ, Devanboyina U-S, Rustan TD, Hein DW (1999). Prostate-specific human N-acetyltransferase 2 (NAT2) expression in the mouse. *The Journal of Pharmacology and Experimental Therapeutics* **290**: 182-187.
- Levy GN and Weber WW (1989). 2-Aminofluorene-DNA adduct formation in acetylator congenic mouse lines. *Carcinogenesis* **10**: 705-709.
- Levy GN and Weber WW (1992). 2-Aminofluorene-DAN adducts in mouse urinary bladder: Effect of age, sex and acetylator phenotype. *Carcinogenesis* **13**: 159-164.
- Levy GN, Chung J-G, Weber WW (1994). 2-Aminofluorene metabolism and DNA adduct formation by mononuclear leukocytes from rapid and slow acetylator mouse strains. *Carcinogenesis* **15**: 353-357.
- Levy GN, Martell KJ, DeLeon JH, Weber WW (1992). Metabolic, molecular genetic and toxicological aspects of the acetylation polymorphism in inbred mice. *Pharmacogenetics* **2**: 197-206.
- Levy GN, Martell KJ, Weber WW (1993). Polymorphic N-acetylation of 2-aminofluorene by cell-free colon extracts from inbred mice. *Pharmacogenetics* **3**: 71-76.
- Lichter P, Tang CC, Call K, Hermanson G, Evans G, Housman D Ward DC (1990). High-resolution mapping of human chromosome 11 by *in situ* hybridisation with cosmid clones. *Science* **247**: 64-69.
- Lidksy AS, Robson KJ, Thirumalachary C, Barker PE, Ruddle FH, Woo SL (1984). The PKU locus in man is on chromosome 12. *American Journal of Human Genetics* **36**: 527-533.
- Lin HJ, Han CY, Lin BK, Hardy S (1993). Slow acetylator mutations in the human polymorphic N-acetyltransferase gene in 786 Asians, blacks, Hispanics and whites: application to metabolic epidemiology. *American Journal of Human Genetics* **52**: 827-834.
- Lin HJ, Probst-Hensch NM, Hughes NC, Sakamoto GT, Louie AD, Kau IH, Kin BK, Lee DB, Lin J, Frankl HD, Lee ER, Hardy S, Grant DM, Haile RW (1998). Variants of N-acetyltransferase NAT1 and a case-control study of colorectal adenomas. *Pharmacogenetics* **8**: 269-281.
- Lisitsyn NA, Lisitsina NM, Dalbagni G, Barker P, Sanchez CA, Gnarr J, Linehan WM, Reid BJ, Wigler MH (1995). Comparative genomic analysis of tumours: Detection of DNA losses and amplification. *Proceedings of the National Academy of Sciences of the USA* **92**: 151-155.

-
- Lovett M (1994). Fishing for complements: finding genes by direct selection. *Trends in Genetics* **10**: 352-357.
 - Ludecke HJ, Schmidt O, Nardmann J, von Holtum D, Meinecke P, Muenke M, Hosrthemke B (1999). Genes and chromosomal breakpoints in the Langer-Giedion syndrome region on human chromosome 8. *Human Genetics* **105**: 619-628.
 - Macé K, Bowman ED, Vautravers P, Shields PG, Harris CC, Pfeifer AMA (1998). Characterisation of xenobiotic-metabolising enzyme expression in human bronchial mucosa and peripheral lung tissues. *European Journal of Cancer* **34**: 914-920.
 - Macoska JA, Trybus TM, Benson PD, Sakr WA, Grignon DJ, Wojno KD, Pietruk T, Powell IJ (1995). Evidence for three tumor suppressor gene loci on chromosome 8p in human prostate cancer. *Cancer Research* **55**: 5390-5395.
 - Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN (1995). Proportion of breast cancer cases in the United States explained by well-established risk factors. *Journal of the National Cancer Institute* **87**: 1681-1685.
 - Mallon A-M and Strivens M (1998). DNA sequence analysis and comparative sequencing. *Methods: A Companion to Methods in Enzymology* **14**: 160-178.
 - Mannervik B and Danielson UH (1988). Glutathione transferases-structure and catalytic activity. *CRC Critical Reviews in Biochemistry* **23**: 283-337.
 - Marano A, Pan Y, Li C, Pagliarulo A, Elmberger G, Tribukait B, Ekman P, Bergerheim U (2000). Chromosomal numerical aberrations detected by fluorescence *in situ* hybridisation on bladder washings from patients with bladder cancer. *European Urology* **37**: 358-365.
 - Marra M, Hillier L, Kucaba T, Allen M, Barstead R, Beck C, Blistain A, Bonaldo M, Bowers Y, Bowles L, et al (1999). An encyclopedia of mouse genes. *Nature Genetics* **21**: 191-194.
 - Marshall CJ, Franks LM, Carbonell AW (1977). Markers of neoplastic transformation in epithelial cell lines derived from human carcinomas. *Journal of the National Cancer Institute* **58**: 1743-1751.
 - Martell KJ, Levy GN, Weber WW (1992). Cloned mouse N-acetyltransferases: Enzymatic Properties of Expressed *Nat-1* and *Nat-2* gene products. *Molecular Pharmacology* **42**: 265-272.
 - Martell KJ, Vatsis KP, Weber WW (1991). Molecular genetic basis of rapid and slow acetylation in mice. *Molecular Pharmacology* **40**: 218-227.

-
- Martinez C, Agundez JA, Olivera M, Martin R, Ladero JM, Benitez J (1995). Lung cancer and mutations at the polymorphic *NAT2* gene locus. *Pharmacogenetics* **5**: 207-214.
 - Matas N, Thygesen P, Stacey M, Risch A, Sim E (1997). Mapping AAC1, AAC2 and AACP, the genes for arylamine N-acetyltransferases, carcinogen metabolising enzymes on human chromosome 8p22, a region frequently deleted in tumours. *Cytogenetics and Cell Genetics* **77**: 290-295.
 - Mattano SS and Weber WW (1987). Kinetics of arylamine N-acetyltransferase in tissues from rapid and slow acetylator mice. *Carcinogenesis* **8**: 133-137.
 - Mattano SS, Erickson RP, Nesbitt MN, Weber WW (1988). Linkage of *Nat* and *Es-1* in the mouse and development of strains congenic for N-acetyltransferase. *Journal of Heredity* **79**: 430-433.
 - Mattano SS, Land S, King CM, Weber WW (1989). Purification and biochemical characterisation of hepatic arylamine N-acetyltransferase from rapid and slow acetylator mice: Identity with arylhydroxamic acid N,O-acyltransferase and N-hydroxyarylamine O-acetyltransferase. *Molecular Pharmacology* **35**: 599-609.
 - Mattei MG, Etienne J, Chuat JC, Nguyen VC, Brault D, Bernheim A, Galibert F (1993). Assignment of the human lipoprotein lipase (LPL) gene to chromosome band 8p22. *Cytogenetics and Cell Genetics* **63**:45-46.
 - May DG (1994). Genetic differences in drug disposition. *Journal of Clinical Pharmacology* **34**: 881-897.
 - McCarthy L, Hunter K, Schalkwyk L, Riba L, Anson S, Mott R, Newell W, Bruley C, Bar I, Ramu E, Housman D, Cox R, Lehrach H (1995). Efficient high resolution genetic mapping of mouse interspersed repetitive sequence PCR products, towards integrated genetic and physical mapping of the mouse genome. *Proceedings of the National Academy of Sciences of the USA* **92**: 5302-5306.
 - McCarthy LC (1996). Whole genome radiation hybrid mapping. *Trends in Genetics* **12**: 491-493.
 - McGurk KA, Brierley CH, Burchell B (1998). Drug glucuronidation by human renal UDP-glucuronosyltransferases. *Biochemical Pharmacology* **55**: 1005-1012.
 - McNulty H, McPartlin JM, Weir DG, Scott JM (1993). Folate catabolism is increased during pregnancy in rats. *Journal of Nutrition* **123**: 1089-1093.
 - McPartlin J, Courtney G, McNulty H, Weir D, Scott J (1992). The quantitative analysis of exogenous folate catabolites in human urine. *Analytical Biochemistry* **206**: 256-261.

-
- McPartlin J, Halligan A, Scott JM, Darling M, Weir DG (1993). Accelerated folate breakdown in pregnancy. *Lancet* **341**: 148-149.
 - McPherson JD (1997). Sequence ready - or not? *Genome Research* **7**: 1111-1113.
 - McPherson JD, Apostol B, Wagner-McPherson CB, Hakim S, Del Mastro RG, Aziz N, Baer E, Gonzales G, Krane MC, Markovich R, Masny P, Ortega M, Vu J, Vujicic M, Church DM, Segal A, Grady DL, Moyzis RK, Spence MA, Movett M, Wasmuth JJ (1997). A radiation hybrid map of human chromosome 5 with integration of cytogenetic, genetic and transcript maps. *Genome Research* **7**: 897-909.
 - Meisel M, Schneider T, Siegmund W, Nikschick S, Klebingat KJ, Scherber A (1986). Development of human polymorphic N-acetyltransferase. *Biological Research in Pregnancy and Perinatology* **7**: 74-76.
 - Meisler MH (1996). The role of the laboratory mouse in the Human Genome Project. *American Journal of Human Genetics* **59**: 764-771.
 - Meyer UA (1992). Molecular genetics and the future of pharmacogenetics. In *Pharmacogenetics of Drug Metabolism*, edited by W Kalow, Pergamon Press, New York, pp. 882-888.
 - Millikan RC (2000). *NAT1*10* and *NAT1*11* polymorphisms and breast cancer risk. *Cancer Epidemiology, Biomarkers and Prevention* **9**: 217-219.
 - Millikan RC, Pittman GS, Newman B, Tse C-KJ, Selmin O, Rockhill B, Savitz D, Moorman PG, Bell DA (1998). Cigarette smoking, N-acetyltransferases 1 and 2, and breast cancer risk. *Cancer Epidemiology, Biomarkers and Prevention* **7**: 371-378.
 - Minchin RF (1995). Acetylation of p-amino benzoylglutamate, a folic acid catabolite, by recombinant arylamine N-acetyltransferase and U937 cells. *Biochemical Journal* **307**: 1-3.
 - Minchin RF, Kadlubar FF, Ilett KF (1993). Role of acetylation in colorectal cancer. *Mutation Research* **290**: 35-42.
 - Mitchell BS, Dhami D, Schumacher U (1992). *In situ* hybridisation – a review of methodologies and applications in the biomedical sciences. *Medical Laboratory Sciences* **49**: 107-118.
 - Mitchell MK, Futscher BW, McQueen CA (1999). Developmental expression of N-acetyltransferases in C57Bl/6 mice. *Drug Metabolism and Disposition* **27**: 261-264.
 - Mitchell RS and Bell JC (1957). Clinical implications of isoniazid blood levels in pulmonary tuberculosis. *New England Journal of Medicine* **257**: 1066-1070.

- Monaco AP and Larin Z (1994). YACs, BACs, PACs and MACs: Artificial chromosomes as research tools. *Trends in Biotechnology* **12**: 280-286.
- Moore KJ (1999). Utilisation of mouse models in the discovery of human disease genes. *Drug Discovery Today* **4**: 123-128.
- Morita S, Yano M, Tsujinaka T, Akiyama Y, Taniguchi M, Kaneko K, Miki H, Fuji T, Yoshino K, Kusuoka H, Monden M (1999). Genetic polymorphisms of drug-metabolising enzymes and susceptibility to head-and-neck squamous-cell carcinoma. *International Journal of Cancer* **80**: 685-688.
- Morita S, Yano M, Tsujinaka T, Ogawa A, Taniguchi M, Kaneko K, Shiozaki H, Doki Y, Inoue M, Monden M (1998). Association between genetic polymorphisms of glutathione S-transferase P1 and N-acetyltransferase 2 and susceptibility to squamous-cell carcinoma of the esophagus. *International Journal of Cancer* **79**: 517-520.
- MRC Vitamin Study Research Group (1991). Prevention of neural tube defects: Results of the Medical Research Council vitamin study. *Lancet* **338**: 131-137.
- Mrozikiewicz PM, Drakoulis N, Roots I (1994). Polymorphic arylamine N-acetyltransferase (*NAT2*) genes in children with insulin-dependent diabetes mellitus. *Clinical Pharmacology and Therapeutics* **56**: 626-634.
- Nagao M and Sugimura T (1993). Carcinogenic factors in food with relevance to colon cancer development. *Mutation Research* **290**: 43-51.
- Nagao M, Ushijima T, Wakabayashi K, Ochiai M, Kushida H, Sugimura T, Hasegowa R, Shirai T and Ito N (1994). Dietary carcinogens and mammary carcinogenesis. Induction of rat mammary carcinomas by administration of heterocyclic amines. *Cancer* **74 (Suppl)**: 1063-1069.
- Nebert DW (1997). Polymorphisms in drug metabolising enzymes: What is their critical relevance and why do they exist? *American Journal of Human Genetics* **60**: 265-271.
- Nebert DW and Duffy JJ (1997). How knockout mouse lines will be used to study the role of drug-metabolising enzymes and their receptors during reproduction and development, and in environmental toxicity, cancer and oxidative stress. *Biochemical Pharmacology* **53**: 249-254.
- Nehls M, Messerle M, Sirulnik A, Smith AJH, Boehm T (1997). Two large insert vectors, λ PS and λ KO, facilitate rapid mapping and targeted disruption of mammalian genes. *BioTechniques* **17**: 770-775.
- Nelson DL, Ballabio A, Victoria MF, Poeretti M, Bies RD, Gibbs RA, Maley JA, Chinault AC, Webster TD, Caskey CT (1991). *Alu*-primed polymerase chain

- reaction for regional assignment of 110 yeast artificial chromosome clones associated with a disease locus. *Proceedings of the National Academy of Sciences of the USA* **88**: 6157-6161.
- Nishioka Y (1995). The origin of the common laboratory mice. *Genome* **38**: 1-7.
 - Nusbaum C, Slonim DK, Harris KL, Birren BW, Steen RG, Stein LD, Miller J, Dietrich WF, Nahf R, Wang V, et al (1999). A YAC-based physical map of the mouse genome. *Nature Genetics* **22**: 388-393.
 - Nyberg F, Hou S-M, Hemminki K, Lambert B, Pershagen G (1998). Glutathione S-transferase μ 1 and N-acetyltransferase 2 genetic polymorphisms and exposure to tobacco smoke in nonsmoking and smoking lung cancer patients and population controls. *Cancer Epidemiology, Biomarkers and Prevention* **7**: 875-883.
 - O'Brien SJ, Wienberg J, Lyons LA (1997). Comparative genomics: Lessons from cats. *Trends in Genetics* **13**: 393-399.
 - O'Brown P (1994). Genome scanning methods. *Current Opinion in Genetics and Development* **4**: 366-373.
 - O'Neil WM, Gilfix BM, DiGirolamo A, Tsoukas CM, Wainer I (1997). N-acetylation among HIV-positive patients and patients with AIDS: When is fast, fast and slow, slow? *Clinical Pharmacology and Therapeutics* **62**: 261-271.
 - Oda Y, Yamazaki H and Shimada T (1999). Role of human N-acetyltransferases, NAT1 or NAT2, in genotoxicity of nitroarenes and aromatic amines in *Salmonella typhimurium* NM6001 and NM6002. *Carcinogenesis* **20**: 1079-1083.
 - Ohata H, Emi M, Fujiwara Y, Higashimo K, Nakagawa K, Futagami R, Tsuchiya E, Nakamura Y (1993). Deletion mapping of chromosome 8 in non-small cell lung carcinoma. *Genes Chromosomes and Cancer* **7**: 85-88.
 - Ohsako S and Deguchi T (1990). Cloning and expression of cDNAs for polymorphic and monomorphic arylamine N-acetyltransferases from human liver. *Journal of Biological Chemistry* **265**: 4630-4634.
 - Ohsako S, Ohtomi M, Sakamoto Y, Uyemura K, Deguchi T (1988). Arylamine N-acetyltransferase from chicken liver: Cloning of cDNA and expression in chinese hamster ovary cells. *Journal of Biological Chemistry* **263**: 7534-7538.
 - Okkels H, Sigsgaard T, Wolf H, Autrup H (1997). Arylamine N-acetyltransferase 1 (NAT1) and 2 (NAT2) polymorphisms in susceptibility to bladder cancer: The influence of smoking. *Cancer Epidemiology, Biomarkers and Prevention* **6**: 225-231.

-
- Olshan AF, Weissler MC, Watson MA, Bell DA (2000). GSTM1, GSTT1, GSTP1, CYP1A1, and NAT1 polymorphisms, tobacco use, and the risk of head and neck cancer. *Cancer Epidemiology, Biomarkers and Prevention* **9**: 185-191.
 - Olson M, Hood L, Cantor C, Botstein D (1989). A common language for physical mapping of the human genome. *Science* **245**: 1434-1435.
 - Olson MV (1993). The human genome project. *Proceedings of the National Academy of Sciences of the USA* **90**: 4338-4344.
 - Ott J (1999). Methods for analysis and resources available for genetic mapping. *The Journal of Heredity* **90**: 68-70.
 - Pacifici GM, Bencini C, Rane A (1986). Acetyltransferase in humans: Development and tissue distribution. *Pharmacology* **32**: 283-291.
 - Palotie A, Heiskanen M, Laan M, Horelli-Kuitunen N (1996). High-resolution fluorescence *in situ* hybridisation: A new approach in genome mapping. *Annals of Medicine* **28**: 101-106.
 - Pariente-Khayat A, Pons G, Rey E, Richard M-O, D'Athis P, Moran C, Badoual J, Olive G (1991). Caffeine acetylator phenotyping during maturation in infants. *Pediatric Research* **29**: 492-495.
 - Parkin DM, Whelan S, Ferlay J, Reymond L, Young J (1997) Cancer incidence in five continents; vol VII Lyon: International Agency for Research on Cancer (IACR). IACR scientific publications, no 143.
 - Payton M, Auty R, Delgoda R, Everett M, Sim E (1999a). Cloning and characterisation of arylamine N-acetyltransferase genes from *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*: Increased expression results in isoniazid resistance. *Journal of Bacteriology* **181**: 1343-1347.
 - Payton M, Smelt V, Upton A, Sim E (1999b). A method for genotyping murine arylamine N-acetyltransferase type 2 (NAT2): A gene expressed in preimplantation embryonic stem cells encoding an enzyme acetylating the folate catabolite p-aminobenzoylglutamate. *Biochemical Pharmacology* **58**: 779-785.
 - Payton MA and Sim E (1998). Genotyping human arylamine N-acetyltransferase type 1 (NAT1): The identification of two novel allelic variants. *Biochemical Pharmacology* **55**: 361-366.
 - Payton MA, Mushtaq A, Sinclair JC, Sandy J, Yu T-W, Sim E (2000). Structure, function and characterisation of prokaryotic arylamine N-acetyltransferases (in preparation).

-
- Pierce JC, Sternberg N, Sauer B (1992). A mouse genomic library in the bacteriophage-P1 cloning system-organisation and characterisation. *Mammalian Genome* **3**: 550-558.
 - Price Evans DA (1992). N-acetyltransferase. In *Pharmacogenetics of Drug Metabolism*, edited by W. Kalow, Pergamon Press (New York), pp.95-178.
 - Probst-Hensch NM, Haile RW, Ingles SA, Longnecker MP, Han CY, Lin BK, Lee DB, Sakamoto GT, Frankl HD, Lee ER, Lin HJ (1995). Acetylation polymorphism and prevalence of colorectal adenomas. *Cancer Research* **55**: 2017-2020.
 - Probst-Hensch NM, Haile RW, Li DS, Sakamoto GT, Louie AD, Lin BK, Frankl HD, Lee ER, Lin HJ (1996). Lack of association between the polyadenylation polymorphism in the NAT1 (acetyltransferase 1) gene and colorectal adenomas. *Carcinogenesis* **17**: 2125-2129.
 - Reeves PT, Kinnear BF, Minchin RF, Ilett KF (1991). Immunological evidence for N-acetyltransferase isozymes in the rabbit. *Molecular Pharmacology* **39**: 42-48.
 - Reeves PT, Kinnear BF, Minchin RF, Ilett KF (1992). Distribution of polymorphic N-acetyltransferase in the intestines of rapid acetylators rabbits. *Asia Pacific Journal of Pharmacology* **7**: 219-223.
 - Reeves PT, Minchin RF, Ilett K (1988). Induction of sulfamethazine acetylation by hydrocortisone in the rabbit. *Drug Metabolism and Disposition* **16**: 110-115.
 - Reeves RH (1999). Genetic, physical and biological mapping in mouse. Handouts from *Current Topics in Genome Analysis*, a mini-course organised by the NHGRI and NCBI (www.nhgri.nih.gov/COURSE99/handouts.html).
 - Rhodes M, Straw R, Fernando S, Evans A, Lacey T, Dearlove A, Greyststrong J, Walker J, Watson P, Weston P, et al (1998). A high-resolution microsatellite map of the mouse genome. *Genome Research* **8**: 531-542.
 - Riddle B and Jencks WP (1971). Acetyl-coenzyme A: arylamine N-acetyltransferase. Role of the acetyl-enzyme intermediate and the effects of substituents on the rate. *Journal of Biological Chemistry* **246**: 3250-3258.
 - Ringwald M, Davis GL, Smith AG, Trepanier LE, Begley DA, Richardson JE, Eppig JT (1997). The mouse gene expression database GXD. *Cell and Developmental Biology* **8**: 489-497.
 - Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al (1989). Identification of the cystic fibrosis gene: cloning and characterisation of complementary DNA. *Science* **245**: 1066-1073.

-
- Risch A, Smelt V, Lane D, Stanley L, van der Slot W, Ward A, Sim E (1996). Arylamine N-acetyltransferase in erythrocytes of cystic fibrosis patients. *Pharmacology and Toxicology* **4**: 235-240.
 - Risch A, Wallace DMA, Bathers S, Sim E (1995). Slow acetylation genotype is a susceptibility factor in occupational and smoking related bladder cancer. *Human Molecular Genetics* **4**: 231-236.
 - Rocha L, Garcia C, de Mendoca A, Gil JP, Bishop DT, Lechner MC (1999). N-acetyltransferase (*NAT2*) genotype and susceptibility to sporadic Alzheimer's disease. *Pharmacogenetics* **9**: 9-15.
 - Rodriguez JW, Kirlin WG, Ferguson RJ, Doll MA, Gray K, Rustan TD, Kemp K, Urso P, Hein DW (1993). Human acetylator genotype: Relationship to colorectal cancer incidence and arylamine N-acetyltransferase expression in colon cytosol. *Archives of Toxicology* **67**: 445-452.
 - Roots I, Brockmöller J, Drakoulis N, Loddenkemper R (1992). Mutant genes of cytochrome P-450IID6, glutathione S-transferase class Mu, and arylamine N-acetyltransferase in lung cancer patients. *Clinical Investigator* **70**: 307-319.
 - Rougraff PN and Paxton R (1987). Purification and partial characterisation of chicken liver acetyl coenzyme A: arylamine N-acetyltransferase. *Comparative Biochemistry and Physiology* **86**: 601-606.
 - Sadrieh N, Davis CD, Snyderwine EG (1996). N-acetyltransferase expression and metabolic activation of the food-derived heterocyclic amines in the human mammary gland. *Cancer Research* **56**: 2683-2687.
 - Sambrook J, Fritsch EF and Maniatis T (1989). Molecular Cloning: A Laboratory Manual. 2nd edition. Cold Spring Harbor Press, New York.
 - Sandberg AA (1992). Chromosome changes in early bladder neoplasms. *Journal of Cellular Biochemistry* **16 (Suppl)**: 76-79.
 - Sandberg AA and Berger CS (1994). Review of chromosome studies in urological tumours 2. Cytogenetics and molecular genetics of bladder cancer. *Journal of Urology* **151**: 545-560.
 - Sanders-Haigh L, Anderson WF, Francke U (1980). The beta-globin gene is on the short arm of human chromosome 11. *Nature* **283**: 683-686.
 - Sasaki Y, Ohsako S, Deguchi T (1991). Molecular and genetic analysis of arylamine N-acetyltransferase of rabbit liver. *Journal of Biological Chemistry* **266**: 13243-13250.
 - Scalzi JM and Hozier JC (1998). Comparative genome mapping: Mouse and rat homologies revealed by fluorescence *in situ* hybridisation. *Genomics* **47**: 44-51.

- Schalkwyk LC, Francis F, Lehrach H (1995). Techniques in mammalian genome mapping. *Current Opinion in Biotechnology* **6**: 37-43.
- Schechter I, Conrad DG, Hart I, Berger RC, McKenzie TL, Bleskan J, Patterson D (1994). Localisation of the squalene synthase gene (FDFT1) to human chromosome 8p22-p23.1. *Genomics* **20**: 116-118.
- Schenkman JB and Gibson GG (1983). Status of the cytochrome P-450 cycle. In *Drug Metabolism and Distribution*, edited by JW Lamble, Elsevier Biomedical Press, Amsterdam, pp. 7-11.
- Schnakenberg E, Ehlers C, Feyerabend W, Werdin R, Hubotter R, Dreikorn K, Schloot W (1998). Genotyping of the polymorphic N-acetyltransferase (NAT2) and loss of heterozygosity in bladder cancer patients. *Clinical Genetics* **53**: 396-402.
- Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, et al (1996). A gene map of the human genome. *Science* **274**: 540-546.
- Senger G, Jones TA, Fidlerova H, Sanseau P, Trowsdale F, Duff M, Sheer D (1994). Released chromatin-linearised DNA for high resolution fluorescence *in situ* hybridisation. *Human Molecular Genetics* **3**: 1275-1280.
- Serikawa T, Kuramoto N, Yokoi N, Andoh Y, Cui Z, Kondo Y, Komeda K, Kitada K (1999). Rat genetic map and a rat/mouse/human comparative gene map. *Transplantation Proceedings* **31**: 1537-1540.
- Shepherd NS, Pfong BD, Coulby JN, Ackerman SL, Vaidyanathan G, Sauer RH, Balkenhol TC, Sternberg N (1994). Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proceedings of the National Academy of Sciences of the USA* **91**: 2629-2933.
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992). Cloning and stable maintenance of 300-kilobase pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proceedings of the National Academy of Sciences of the USA* **89**: 8794-8797.
- Shows TB and Brown JA (1975). Human X-linked genes regionally mapped utilising X-autosome translocations and somatic cell hybrids. *Proceedings of the National Academy of Sciences of the USA* **72**: 2125-2129.
- Sim E and Hickman D (1991). Polymorphism in human N-acetyltransferase - the case of the missing allele. *Trends in Pharmacological Sciences* **12**: 211-213.
- Sim E, Hickman D, Coroneos E, Kelly SL (1992). Arylamine N-acetyltransferase. *Biochemical Society Transactions* **20**: 304-309.

- Sim E, Payton M, Noble M, Minchin R (2000). An update on genetic, structural and functional studies of arylamine N-acetyltransferases in eukaryotes and prokaryotes. *Human Molecular Genetics* (in press).
- Sim E, Stanley LA, Risch A, Thygesen P (1995). Xenogenetics in multifactorial disease susceptibility. *Trends in Genetics* **11**: 509-512.
- Sinclair J and Sim E (1997). A fragment consisting of the first 204 amino-terminal amino acids of human arylamine N-acetyltransferase one (NAT1) and the first transacetylation step of catalysis. *Biochemical Pharmacology* **53**: 11-16.
- Sinclair JC, Delgoda R, Noble MEM, Jarmin S, Goh NK, Sim E (1998). Purification, characterisation and crystallisation of an N-hydryarylamine O-acetyltransferase from *Salmonella typhimurium*. *Protein Expression and Purification* **12**: 371-380.
- Sinclair JC, Sandy J, Delgoda R, Sim E and Noble MEM (2000). Structure of arylamine N-acetyltransferase reveals a catalytic triad. *Nature Stuctural Biology* **7**: 560-564.
- Smelt VA, Mardon HJ, Redman CWG, Sim E (1997). Acetylation of arylamines by the placenta. *European Journal of Drug Metabolism and Pharmacokinetics* **22**: 403-408.
- Smelt VA, Mardon HJ, Sim E (1998). Placental expression of arylamine N-acetyltransferases: Evidence for linkage disequilibrium between *NAT1*10* and *NAT2*4* alleles of the two human arylamine N-acetyltransferase loci *NAT1* and *NAT2*. *Pharmacology and Toxicology* **83**:149-157.
- Smelt VA, Upton A, Adjaye J, Payton MA, Boukouvala S, Johnson N, Mardon HJ, Sim E (2000). Expression of arylamine N-acetyltransferases in pre-term placentas and in human pre-implantation embryos. *Human Molecular Genetics* **9**: 1101-1107.
- Smith CAD, Smith G, Wolf CR (1994). Genetic polymorphisms in xenobiotic metabolism. *European Journal of Cancer* **30A**: 1921-1935.
- Smith CAD, Wadelius M, Gough AC, Harrison DJ, Wolf CR, Rane A (1997). A simplified assay for the arylamine N-acetyltransferase 2 polymorphism validated by phenotyping with isoniazid. *Journal of Medical Genetics* **34**: 758-760.
- Smith GE and Summers MD (1980). The bidirectional transfer of DNA and RNA to nitrocellulose or diazbenzylomethoxy-paper. *Analytical Biochemistry* **109**: 123-129.
- Smolen TN, Brewer JA, Weber WW (1993). Testosterone modulation of N-acetylation in mouse kidney. *The Journal of Pharmacology and Experimental Therapeutics* **264**: 854-858.

- Stacey M, Matas N, Drake M, Payton M, Fakis G, Greenland J, Sim E (1999). Arylamine N-acetyltransferase type 2 (NAT2), chromosome 8 aneuploidy, and identification of a novel *NAT1* cosmid clone: An investigation in bladder cancer by interphase FISH. *Genes, Chromosomes and Cancer* **25**: 376-383.
- Stacey M, Payton M, Fakis G, Pope J, Sim E (1998). Characterisation of arylamine N-acetyltransferase in the human urothelial cell line RT112. *Pathogenesis* **1**: 99-105.
- Stacey M, Thygesen P, Stanley LA, Matas N, Risch A, Sim E (1996). Arylamine N-acetyltransferase as a potential biomarker in bladder cancer: Fluorescent in situ hybridisation and immunohistochemistry studies. *Biomarkers* **1**: 55-61.
- Stanley LA, Copp AJ, Pope J, Rolls S, Smelt V, Perry VH, Sim E (1998). Immunochemical detection of arylamine N-acetyltransferase during mouse embryonic development and in adult mouse brain. *Teratology* **58**: 174-182.
- Stanley LA, Coroneos E, Cuff R, Hickman D, Ward A, Sim E (1996). Immunochemical detection of arylamine N-acetyltransferase in normal and neoplastic bladder. *The Journal of Histochemistry and Cytochemistry* **44**: 1059-1067.
- Stanley LA, Mills IG, Sim E (1997). Localisation of polymorphic N-acetyltransferase (NAT2) in tissues of inbred mice. *Pharmacogenetics* **7**: 121-130.
- Steinbrueck T, Read C, Daiger SP, Sadler LA, Weber JL, Donis-Keller H (1992). A comprehensive genetic linkage map of the human genome [Chromosome 8]. *Science* **258**: 67-86.
- Sternberg N (1990). Bacteriophage P1 cloning system for the isolation, amplification and recovery of DNA fragments as large as 100 kilobase pairs. *Proceedings of the National Academy of Sciences of the USA* **87**: 103-107.
- Sternberg N (1992). Cloning high molecular weight DNA fragments by the bacteriophage P1 system. *Trends in Genetics* **8**: 11-16.
- Strange RC, Jones PW, Fryer AA (2000). Glutathione S-transferase: Genetics and role in toxicology. *Toxicology Letters* **112-113**: 357-363.
- Strange RC, Lear JT, Fryer AA (1998). Glutathione S-transferase polymorphisms: Influence on susceptibility to cancer. *Chemico-Biological Interactions* **111-112**: 351-364.
- Sugimura T (1985). Carcinogenicity of mutagenic heterocyclic amines formed during the cooking process. *Mutation Research* **150**: 33-41.

-
- Surh Y-J (1998). Bioactivation of benzylic and allylic alcohols via sulfo-conjugation. *Chemico-Biological Interactions* **109**: 221-235.
 - Szórády I, Santa A, Veress I (1987). Drug acetylator phenotypes in newborn infants. *Biological Research in Pregnancy and Perinatology* **8**: 23-25.
 - Taioli E (1999). International collaborative study on genetic susceptibility to environmental carcinogens. *Cancer Epidemiology, Biomarkers and Prevention* **8**: 727-728.
 - Takle LA and Knowles MA (1996). Deletion mapping implicates two tumour suppressor genes on chromosome 8p in the development of bladder cancer. *Oncogene* **12**: 1083-1087.
 - Taylor JA, Umbach DM, Stephens E, Castranio T, Paulson D, Robertson C, Mohler JL, Bell DA (1998). The role of N-acetylation polymorphisms in smoking-associated bladder cancer: Evidence of a gene-gene-exposure three-way interaction. *Cancer Research* **58**: 3603-3610.
 - Thygesen P, Risch A, Stacey M, Fakis G, Takle L, Knowles M, Sim E (1999). Genes for arylamine N-acetyltransferases in relation to loss of the short arm of chromosome 8 in bladder cancer. *Pharmacogenetics* **9**: 1-8.
 - Trask B, Vandengh G, Landegent J, Dewal NJI, Vanderploeg M (1985). Detection of DNA sequences in nuclei in suspension by in situ hybridisation and dual beam flow cytometry. *Science* **230**: 1401-1403.
 - Trask BJ (1991). Fluorescence *in situ* hybridisation: Applications in cytogenetics and gene mapping. *Trends in Genetics* **7**: 149-154.
 - Trepanier LA, Ray K, Winand NJ, Spielberg SP, Cribb AE (1997). Cytosolic arylamine N-acetyltransferase (NAT) deficiency in the dog and other canids due to an absence of *NAT* genes. *Biochemical Pharmacology* **54**: 73-80.
 - Vainio H and McGregor D (1992). Approaches to the prediction of human cancer risk. *Pharmacogenetics* **2**: 337-343.
 - Van Etten WJ, Steen RG, Nguyen H, Castle AB, Slonim DK, Ge B, Nusbaum C, Schuler GD, Lander ES, Hudson TJ (1999). Radiation hybrid map of the mouse genome. *Nature Genetics* **22**: 384-387.
 - Vatsis KP and Weber WW (1993). Structural heterogeneity of Caucasian N-acetyltransferase at the NAT1 gene locus. *Archives of Biochemistry and Biophysics* **301**: 71-76.
 - Vatsis KP and Weber WW (1994). Human N-acetyltransferases. In *Conjugation-Deconjugation Reactions in Drug Metabolism and Toxicity: Handbook of*

- Experimental Pharmacology*, edited by F.C. Kaufmann, Springer-Verlag, vol. 112, pp. 109-130.
- Vatsis KP, Martell KJ, Weber WW (1991). Diverse point mutations in the human gene for polymorphic N-acetyltransferase. *Proceedings of the National Academy of Sciences of the USA* **88**: 6333-6337.
 - Vatsis KP, Weber WW, Bell DA, Dupret J-M, Price-Evans DA, Grant DM, Hein DW, Lin HJ, Meyer UA, Relling MV, Sim E, Suzuki T, Yamazoe Y (1995). Nomenclature for N-acetyltransferases. *Pharmacogenetics* **5**: 1-17.
 - Vesel ES (1983). Why individuals vary in their response to drugs. In *Drug Metabolism and Distribution*, edited by JW Lamble, Elsevier Biomedical Press, Amsterdam, pp.143-147.
 - Waddington D, Springbett AJ, Burt DW (2000). A chromosome-based model for estimating the number of conserved segments between pairs of species from comparative genetic maps. *Genetics* **154**: 323-332.
 - Wadelius M, Autrup JL, Stubbins MJ, Andersson S-O, Johansson J-E, Wadelius C, Wolf CR, Autrup H, Rane A (1999). Polymorphisms in *NAT2*, *CYP2D6*, *CYP2C19* and *GSTP1* and their association with prostate cancer. *Pharmacogenetics* **9**: 333-340.
 - Wagner U, Budendorf L, Gasser TC, Moch H, Gorog JP, Richter J, Mihatsch MJ, Waldman FM, Sauter G (1997). Chromosome 8p deletions are associated with invasive tumor growth in urinary bladder cancer. *American Journal of Pathology* **151**: 753-759.
 - Wallace DMA (1988). Occupational urothelial cancer. *British Journal of Urology* **61**: 175-182.
 - Wang JC, Radford DM, Holt MS, Helms C, Goate A, Brandt W, Parik M, Phillips NJ, DeSchuyver K, Schuh ME, Fair KL, Ritter JH, Marshall P, Donnis-Keller H (1999). Sequence-ready contig for the 1.4-cM ductal carcinoma in situ loss of heterozygosity region on chromosome 8p22-p23. *Genomics* **60**: 1-11.
 - Wang M-R, Perissel B, Malet P (1995). Cytogenetic characterisation of a bladder carcinoma cell line by fluorescence in situ hybridisation. *Cancer Genetics and Cytogenetics* **82**: 170-174.
 - Ward A, Hickman D, Gordon JW, Sim E (1992). Arylamine N-acetyltransferase in human red blood cells. *Biochemical Pharmacology* **44**: 1099-1104.
 - Ward A, Summers MJ and Sim E (1995). Purification of recombinant human N-acetyltransferase type 1 (NAT1) expressed in *E. coli* and characterisation of its potential role in folate metabolism. *Biochemical Pharmacology* **49**: 1759-1767.

- Ware JA and Svensson CK (1996). Longitudinal distribution of arylamine N-acetyltransferases in the intestine of the hamster, mouse, and rat. *Biochemical Pharmacology* **52**: 1613-1620.
- Watanabe M, Sofuni T, Nohmi T (1992). Involvement of Cys69 residue in the catalytic mechanism of N-hydroxyarylamine O-acetyltransferase of *Salmonella typhimurium*. Sequence similarity at the amino acid level suggests a common catalytic mechanism of acetyltransferase for *S. typhimurium* and higher organisms. *Journal of Biological Chemistry* **267**: 8429-8436.
- Waterston R and Sulston J (1995). The genome of *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the USA* **92**: 10836-10840.
- Watson AP, Wang PD, Sim E (1990). Arylamine N-acetyltransferase from fast (C57BL6) and slow (A/J) N-acetylating strains of mice. *Biochemical Pharmacology* **39**: 647-654.
- Watson JD (1990). The Human Genome Project: Past, present and future. *Science* **248**: 44-49.
- Watson JD and Cookdeegan RM (1991). Origins of the human genome project. *FASEB Journal* **5**: 8-11.
- Weber JL and May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.
- Weber WW and Cohen SN (1967). N-acetylation of drugs: Isolation and properties of N-acetyltransferase from rabbit liver. *Molecular Pharmacology* **3**: 266-273.
- Weber WW and Hein DW (1985) N-acetylation pharmacogenetics. *Pharmacological Reviews* **37**: 25-78.
- Weber WW and Vatsis KP (1993). Individual variability in p-aminobenzoic acid N-acetylation by human N-acetyltransferase (NAT1) of peripheral blood. *Pharmacogenetics* **3**: 209-212.
- Welp EA, Weiderpass E, Boffetta P, Vainio H, Vasama-Neuvonen K, Petralia S, Partanen TJ (1998). Environmental risk factors of breast cancer. *Scandinavian Journal of Work and Environmental Health* **24**: 3-7.
- Wicking C and Williamson B (1991). From linked marker to gene. *Trends in Genetics* **7**: 288-293.
- Williams JA, Stone EM, Fakis G, Johnson N, Meini W, Glatt H, Sim E, Phillips DH (2000). N-acetyltransferases, sulphotransferases and heterocyclic amine activation in the breast. *Cancer Research* (submitted).

- Williams RT (1959). The metabolism and detoxication of drugs, toxic substances and other organic compounds. In *Detoxication Mechanisms*, Champman & Hall, London, pp. 734-740.
- Williamson P, Holt S, Townsend S, Boyd Y (1995). A somatic cell hybrid panel for mouse gene mapping characterised by PCR and FISH. *Mammalian Genome* **6**: 429-432.
- Windle B, Silvas E, Parra I (1995). High resolution microscopic mapping of DNA using multicolour fluorescent hybridisation. *Electrophoresis* **16**: 273-278.
- Windmill KF, Christiansen A, Teusner JT, Bhasker CR, Birkett DJ, Zhu X, McManus ME (1998). Localisation of aryl sulfotransferase expression in human tissues using hybridisation histochemistry and immunohistochemistry. *Chemico-Biological Interactions* **109**: 341-346.
- Windmill KF, Gaedigk A, Hall PDM, Samaratunga H, Grant DM, McManus ME (2000). Localisation of N-acetyltransferases NAT1 and NAT2 in human tissues. *Toxicological Sciences* **54**: 19-29.
- Windmill KF, McKinnon RA, Zhu X, Gaedigk A, Grant DM, McManus ME (1997). The role of xenobiotic metabolising enzymes in arylamine toxicity and carcinogenesis: Functional and localisation studies. *Mutation Research* **376**: 153-160.
- Wohlleb JC, Hunter CF, Blass B, Kadlubar FF, Chu DZJ, Lang NP (1990). Aromatic amine acetyltransferase as a marker for colorectal cancer: Environmental and demographic associations. *International Journal of Cancer* **46**: 22-30.
- Wolf CR, Hendreson CJ (1998). Use of transgenic animals in understanding molecular mechanisms of toxicity. *Journal of Pharmacy and Pharmacology* **50**: 567-574.
- Wong L-L (1998). Cytochrome P450 monooxygenases. *Current Opinion in Chemical Biology* **2**: 263-268.
- Wormhoudt LW, Commandeur JNM, Vermeulen NPE (1999). Genetic polymorphisms of human N-acetyltransferase, cytochrome P450, glutathione-S-transferase and epoxide hydrolase enzymes: Relevance to xenobiotic metabolism and toxicity. *Critical Reviews in Toxicology* **29**: 59-124.
- Yaremko ML, Kutza C, Lyzak J, Mick R Recant WM, Westbrook CA (1996). Loss of heterozygosity from the short arm of chromosome 8 is associated with invasive behavior in breast cancer. *Genes Chromosome and Cancer* **16**: 189-195.

- Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alish R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, Schellnberg GD (1996). Positional cloning of the Werner's syndrome gene. *Science* **272**: 258-262.
- Zenser TV, Lakshmi VM, Davis BB (1998). N-glucuronidation of benzidine and its metabolites. Role in bladder cancer. *Drug Metabolism and Disposition* **26**: 856-859.
- Zheng W, Deitz AC, Campbell DR, Wen W-Q, Cerhan JR, Sellers TA, Folsom AR, Hein DW (1999). N-acetyltransferase 1 genetic polymorphism, cigarette smoking, well-done meat intake and breast cancer risk. *Cancer Epidemiology, Biomarkers and Prevention* **8**: 233-239.
- Zietkiewicz E, Laduba M, Sinnett D, Glorieux FH, Laduba D (1992). Linkage mapping by simultaneous screening of multiple polymorphic loci using Alu oligonucleotide-directed PCR. *Proceedings of the National Academy of Sciences of the USA* **89**: 8448-8451.