

**Molecular Evolution of the
Mouse Major Histocompatibility Complex
-The Detection of Low Frequency Gene Conversion Events-**

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Jann Thorsten Martinsohn
aus Heidelberg

Köln 2001

Berichterstatter: Prof. Dr. J.C. Howard
Prof. Dr. B. Kemper

Tag der mündlichen Prüfung: 06.07.2001

Vorwort

Die Grundgedanken der in diesem Buche entwickelten Theorie wurden von mir am 14. September 1927 auf dem 5. internationalen Kongreß für Vererbungswissenschaft in Berlin vorgetragen. Die Theorie fand keine Zustimmung und stieß auf allgemeine Nichtbeachtung, die soweit ging, daß in den Berichten über den Kongreß z. B. in den "Naturwissenschaften" (Berichterstatter Nachtsheim) und in der "Nature" (Berichterstatter R.R. Gates) mein Vortrag, obwohl er in einer allgemeinen Sitzung stattgefunden hatte, überhaupt nicht erwähnt wurde. Daß dies für mich kein Grund sein konnte, auf die damals angekündigte weitere Ausarbeitung der Theorie zu verzichten, ist selbstverständlich.

Diese Ausarbeitung hat leider wesentlich länger gedauert, als ich vermutet hatte und als mir selber lieb war. Erhebliche Belastung mit Berufsgeschäften trug das ihre dazu bei. Vor allem aber waren es die in der Sache selbst liegenden Schwierigkeiten, die mich nur langsam vorankommen ließen. Besonders die sehr zahlreichen und umfangreichen Berechnungen und Formelaufstellungen, die notwendig durchgeführt werden mussten, waren für mich als vollständigen Laien in mathematischen Dingen eine recht mühevoll und zeitraubende Abgelegenheit.

Wenn ich die Arbeit an der weiteren Ausgestaltung der Theorie nunmehr vorläufig abschließe und sie in der vorliegenden Gestalt veröffentliche, so geschieht das gewiß nicht, weil ich glaubte, etwas Abgeschlossenes vorlegen zu können. Niemand weiß besser als ich was alles noch fehlt, und wieviel zur weiteren Begründung und Vertiefung noch zu geschehen hätte. Da es indessen bei einer Arbeit wie der vorliegenden wohl unmöglich ist, ein Stadium zu erreichen, auf dem man sich mit gutem Gewissen sagen könnte, man sei nun endgültig fertig, so muß man sich einmal zum Abschluß entschließen, so schwer dieser Entschluß auch fallen mag.

Hamburg, im Jahr 1930
Institut für allgemeine Botanik

Aus
Die Konversion der Gene
-Eine vererbungstheoretische Untersuchung-
von
Dr. Hans Winkler
(Jena, Verlag Gustav Fischer, 1930)

TABLE OF CONTENTS

ABSTRACT	1
ZUSAMMENFASSUNG	2
1. INTRODUCTION	3
1.1. GENE CONVERSION AS A USEFUL DIVERSIFYING MACHINE IN MHC GENES	3
1.1.1. <i>The MHC</i>	3
1.1.2. <i>Origin and maintenance of diversity of MHC alleles</i>	5
Origin.....	5
Maintenance.....	6
1.1.3. <i>Gene conversion as a useful diversifying machine in MHC genes</i>	7
1.1.4. <i>The MHC running alongside the red queen?</i>	10
1.1.5. <i>The gene conversion hypothesis of MHC evolution</i>	10
1.2. EVIDENCE FOR GENE CONVERSION IN THE MOUSE MHC	11
1.2.1. <i>Characteristics</i>	11
1.2.2. <i>Frequencies</i>	13
1.2.3. <i>Interlocus and interallelic conversions</i>	13
1.2.4. <i>Acceptors and donors</i>	14
1.2.5. <i>Multiple mutant births</i>	14
1.2.6. <i>Sex bias</i>	16
1.3. IN VITRO ANALYSIS OF RECOMBINATORIAL EVENTS IN MAMMALS	23
1.3.1. <i>Intrachromosomal recombination</i>	24
1.3.2. <i>Regulation of homologous recombination in mammalian genomes</i>	25
1.4. IN VITRO ANALYSIS OF GENE CONVERSION EVENTS IN MHC GENES	26
1.4.1. <i>PCR-based analysis of gene conversion events in the mouse class II gene Ab</i>	26
1.4.2. <i>PCR-based analysis of gene conversion events in the human class II gene DPB1</i>	27
1.4.3. <i>H2 class I gene conversion studies in transgenic yeast</i>	28
1.4.4. <i>Gene conversion in tandem gene constructs in the germline of transgenic mice</i>	28
1.5. THE PRESENT PROJECT	29

2. MATERIALS AND METHODS	31
2.1. GENERALITIES.....	31
2.2. MOLECULAR BIOLOGY METHODS.....	31
2.2.1. Buffer exchange of DNA samples using self assembled Sephadex G-25 spin columns	31
2.2.2. DNA-Sequencing.....	31
2.2.3. Site directed mutagenesis using the "QuikChange™ XL Site-Directed Mutagenesis" protocol.....	32
2.2.4. Electrophoretic Mobility Shift Assay (EMSA).....	32
2.2.5. Capillary Electrophoresis (CE)	33
2.2.6. Radioactive labelling of probes.....	33
5'-end labelling (forward reaction) of [OPN/OPC]	33
Labelling of pW310 through fill in.....	34
Probes for Southern blotting	34
2.2.7. Synthetic oligonucleotides.....	34
Annealing of oligonucleotides	35
2.2.8. Plasmids and constructs.....	35
2.2.9. Polymerase Chain Reaction	36
LNC-PCR on cells (optimised protocol).....	36
18S-PCR.....	37
PCR on mouse tail DNA.....	37
Pfu-buffers	37
2.2.10. Southern blot analysis	37
Probes used	38
2.2.11. Sequence Specific Oligonucleotide Probing (SSOP).....	38
Non-radioactive labelling of oligonucleotides	38
The oligonucleotides in use for the <i>bg</i> -specific SSOP assay	38
Blotting procedure	39
Hybridisation and detection	39
Evaluation	40
2.2.12. Isolation of genomic DNA	40
Isolation of genomic DNA out of ES cells (E14.1; LNC28.28)	40
Isolation of genomic DNA out of mouse tails	40
2.2.13. M450 ^{I_Z} - <i>lacOp</i> rescue: standard protocol.....	41
Arming M450 Dynabeads® with α - β -galactosidase antibody	41
The M450 ^{I_Z} - <i>lacOp</i> rescue protocol.....	41
Evaluation of the M450 ^{I_Z} - <i>lacOp</i> rescue by southern blotting and densitometry	42

2.3. BIOCHEMICAL METHODS	42
2.3.1. <i>SDS-Polyacrylamide Gel electrophoresis for the separation of proteins.....</i>	42
2.3.2. <i>Coomassiestaining of SDS-Polyacrylamide Proteingels.....</i>	42
2.3.4. <i>Spectrophometric Determination of Protein Concentration</i>	43
2.3.5. <i>Purification of Proteins.....</i>	43
Charging of NHS-activated HiTrap™ affinity columns with annealed oligonucleotides	43
Preparation of LacIZ raw extract	44
Purification of Pfu using the Uno™ Q1 polishing column	44
2.4. CELLULAR BIOLOGY METHODS	44
2.4.1. <i>Culture of ES-cells</i>	44
2.4.2. <i>Freezing and storage of ES-cells</i>	45
2.4.3. <i>Preparation of spleen cell suspension.....</i>	45
2.4.4. <i>Cytofluorometric analysis.....</i>	45
2.5. MAINTENANCE AND CHARACTERIZATION OF 129^{svLNC}-MICE.....	46
2.5.1. <i>Purifying breeding of 129^{kbLNC}-mice: monitoring by PCR.....</i>	47
2.5.2. <i>Expression of K^{blNC} on the cell surface in comparison to Kb</i>	48
2.5.3. <i>Determining the sex of cells: SRY-PCR</i>	49

3. RESULTS51

3.1. THE M450^{IZ}-LACOP RESCUE: SPECIFIC EXTRACTION OF *E. COLI* LAC-OPERATOR CONTAINING DNA-FRAGMENTS WITH M450^{IZ} DYNABEADS®51

3.1.1. Affinity based purification of the fusion protein LacIZ 52

3.1.2. Extraction of target sequence from pools of fragmented plasmid DNA 58

3.1.3. Target extraction out of pools of genomic DNA: High Copy Number 62

Uptake of DNA by M450^{IZ} beads 62

Specific elution of *lac-operator* containing DNA fragments with IPTG 72

3.1.4. Target extraction out of pools of genomic DNA: single copy gene extraction 78

3.1.5. Target extraction out of pools of cells: Isolation of the *Kb-LNC* transgene from LNC28.28 cells 80

3.2. DETECTION OF *BG*-LIKE MUTATIONS WITH *K^{bLNC}*-SPECIFIC POLYMERASE CHAIN REACTION COUPLED TO SEQUENCE SPECIFIC OLIGONUCLEOTIDE PROBING.....83

3.2.1. The *bg*-target region 84

3.2.2. LNC-PCR on cells 86

Isolation and performance of the *Pyrococcus furiosus* DNA Polymerase (“Pfu”) 86

Test of HEPES buffer for Pfu based PCR amplification 99

LNC-PCR on cells: final protocol 101

3.2.3. Development of a *bg* specific SSOP-dot blot assay 103

Optimisation of the oligonucleotides BGOT and BGOC 104

SSOP on PCR-products: Triton X-100 effect 106

SSOP: signal intensity dependence on exposure time in the CDP-Star™ detection system 108

3.2.4. Design and cloning of the evaluation plasmid pBbgMCB 109

Assembly of the insert bgMCB 112

Cloning of the plasmid pBbgMCB 115

Test of pBbgMCB in the LNC-PCR system 115

3.2.5. Evaluation of the power of the *K^{bLNC}*-PCR/SSOP based *bg* mutation detection assay 118

Detection of pBbgMCB diluted into pools of 50 cells 119

Detection of pBbgMCB diluted into pools of 100 cells 127

The $k \geq 2$ problem 134

Comparison of the *bg*-mutation detection oligonucleotides BGOT and BGOC 136

An estimate of the LNC-PCR efficiency 142

Following the fate of rare specimen in PCRs with computer simulations 147

4. DISCUSSION	149
4.1. VIEWS ON A MULTIPLE COMPONENT GENE CONVERSION DETECTOR.....	150
4.1.1. <i>Enrichment of target sequences through the M450^{I_Z}-lacOp rescue</i>	150
4.1.2. <i>Enrichment and detection of mutations through the K^{bLNC}-PCR coupled to a bg specific Sequence Specific Oligonucleotide Probing Assay.....</i>	152
4.1.3. <i>The mutation detection efficiency of the LNC-PCR/bgSSOP assay</i>	155
4.1.4. <i>Hunting for gene conversion events in germline MHC genes of Mice: The perspectives in the year 2001</i>	156
4.1.5. <i>Features of mutation detection methods presently available</i>	158
Physical methods of mutation detection	158
Chemical cleavage of mismatch	159
Enzymatic detection of mutations.....	160
DNA amplification based detection of mutants	161
DNA chip technology	164
The royal road to mutation detection: Sequencing	165
4.2. AND STILL NOT ALL CONVERTED YET!	167
5. REFERENCES.....	169
ERKLÄRUNG.....	I
CURRICULUM VITAE.....	II
ACKNOWLEDGEMENTS.....	III

ABSTRACT

The patchwork-like pattern of variation which emerges from the visual inspection of MHC alignments can best be explained by gene conversion. The linear array of substitutions which characterise alleles and homologues of MHC genes have individually arisen *de novo* by point mutation. Microrecombination of such point mutational variation by gene conversion is an efficient generator of new peptide-binding grooves on which natural selection may act. This is the gene conversion hypothesis of MHC evolution.

The best evidence for the occurrence of gene conversion in MHC genes comes from the documented instances of spontaneous class I gene mutations in the laboratory mouse. Based on screening programs involving over 150 000 mice an overall frequency of germ line class I gene conversion events of about 10^{-4} is estimated.

The spontaneous histocompatibility mutants found in the mouse excited much interest, and their subsequent characterisation at the amino acid and nucleotide levels has stimulated speculations about the nature and evolutionary relevance of the mutational processes operating on MHC genes, as expressed in the gene conversion hypothesis of MHC evolution. However, no general experimental approach has been described for direct molecular analysis of mutational events in different MHC loci so far, and thus most of the original questions remain unanswered.

A major obstacle for the analysis of gene conversion events in class I genes of the mouse is their low frequency. The purpose of the work presented here was to establish and characterise a mutation detection system for the *H2 K^b* gene. The ultimate goal was to get hold of a reliable high throughput gene conversion detector, which would allow us to systematically analyse mutations occurring in class I mouse germline genes with respect to underlying molecular mechanisms.

We describe a target-gene enrichment step, called M450^{Iz}-*lacOp* rescue, for the extraction of *E.coli lac-operator* tagged DNA sequences. This step serves to specifically enrich for DNA fragments which would subsequently be analysed for mutations. After optimisation of the assay for our purposes, we could show that, starting from pools of transgenic cells containing the *H2 K^b* gene in an *lacOp* tagged form, this assay effectively reduces the abundance of contaminating DNA-sequences. In principle this assay can precede any other component of a mutation detection assay.

Furthermore we describe a mutation detection assay which consists of a highly specific PCR-step performed on cells which is followed by a sequence specific oligonucleotide probing assay (SSOP) for the detection of a particular *H2 K^b* mutation (“*bg*”). We examined the power of this assay with respect to the detection of mutations occurring at a frequency of 1/10 000, using model experiments. We showed that the PCR step on cells works with high efficiency and reliability. The same is true for the SSOP assay. However assuming a binomial distribution of these mutations we found that the assay works 5 times below the expectation at a 100% detection efficiency. Using computer simulations for PCRs on pools of molecules containing an abundant specimen (the “wildtype molecule”) and a rare specimen (the “mutant molecule”) we demonstrated that with decreasing PCR efficiency the probability of losing the rare specimen increases. For our calculated PCR efficiency of 0.7 this can under certain conditions already account for 50% loss of the mutants.

We conclude that our proposed mutation detection assay is feasible, but in its present form very labour intensive. It holds the potential to be adapted to modern PCR-based high throughput screening methods using fluorescently labelled tags.

ZUSAMMENFASSUNG

Die Sequenzmuster in den Allelvarianten von Haupthistokompatibilitätsgenen (MHC), lassen als das Ergebnis von Genkonversionsereignissen interpretieren. Durch den Transfer von Sequenzfragmenten zwischen MHC Genen könnten effektiv neue peptidbindende Taschen geformt werden, auf die dann natürliche Selektion einwirkt. Dies ist die Genkonversionshypothese der MHC Evolution.

Der überzeugendste Hinweis, dass sich zwischen MHC-Genen Genkonversion abspielt, kommt von dokumentierten Fällen spontaner Mutationen in MHC Klasse I Genen der Maus. Basierend auf Untersuchungsreihen, kann man Genkonversionsereignissen in Keimbahn Klasse I Genen eine Häufigkeit von etwa 10^{-4} zugrunde legen.

Die Maus MHC Mutanten erzeugten ein großes Interesse, und ihre genauere Untersuchung auf Peptid- und DNA-Ebene, führten zu zahlreichen Hypothesen über die Eigenschaften und die evolutionäre Relevanz von Genkonversion in MHC Genen. Jedoch wurde bis zum heutigen Tage kein experimenteller Ansatz beschrieben, der es erlaubt hätte, diese seltenen Ereignisse systematisch, und für beliebige MHC Gene zu untersuchen.

Die größte Schwierigkeit, die mit der Analyse von Genkonversionserignissen in MHC Genen verbunden ist, stellt die niedrige Frequenz mit der sie auftreten dar. Das hier beschriebene Projekt wurde durchgeführt, um ein Mutationsdetektionsprotokoll für das *H2 K^b* Gen zu etablieren und charakterisieren. Unser Ziel war es, über einen zuverlässigen Genkonversionsdetektor zu verfügen, der es erlauben würde, Genkonversionsereignisse, die in MHC Genen der Mauskeimbahn auftreten, systematisch, und mit hohem Durchsatz zu untersuchen. Dadurch sollten Fragen bezüglich der molekularen Mechanismen solcher Ereignisse beantwortet werden können.

Wir beschreiben in dieser Arbeit einen spezifischen Genanreicherungsschritt, der dazu dient, DNA Fragmente (in unserem Fall von MHC Genen) aufzureinigen, die mit der *E.coli lac-operatorsequenz* versehen sind. Diese Fragmente können dann bezüglich Mutationen untersucht werden. Nach der Optimierung des Protokolls für unsere Zwecke, konnten wir zeigen, dass es effektiv die Gegenwart kontaminierender Sequenzen vermindert. Im Prinzip kann das Protokoll jedem Mutationsdetektionsprotokoll vorangestellt werden.

Wir beschreiben weiterhin ein Mutationsdetektionsprotokoll, das aus einem hochspezifischen PCR-Schritt besteht, der direkt auf Aliquots von Zellen ausgeführt wird, und dem sich ein sequenzspezifisches Oligonukleotid Hybridisierungsprotokoll (*Sequence Specific Oligonucleotide Probing* – SSOP) anschließt. Dieses Protokoll dient dazu, eine spezifische Mutation aufzuspüren. Wir untersuchten die Effizienz dieses kombinierten Mutationsdetektionsassays in Modellexperimenten. Dabei wurde gezeigt, dass der PCR-Schritt sehr effektiv ist. Dies gilt auch für das SSOP Protokoll. Einschränkend muß jedoch gesagt werden, dass wir, unter der Annahme einer Binomialverteilung der Mutationen, zeigen konnten, dass das Assay um den Faktor fünf unterhalb der Erwartung für eine 100%ige Effizienz arbeitet. Wir verwendeten Computersimulationen, um PCRs zu simulieren, in denen häufige (Wildtyp) und seltene (mutante) DNA-Moleküle amplifiziert werden. Dabei konnte gezeigt werden, dass mit abnehmender PCR-Effizienz die Wahrscheinlichkeit zunimmt, die seltenen Moleküle zu verlieren. Für unsere berechnete PCR Effizienz von 0,7 kann das unter Umständen für den Verlust von 50% der Mutanten verantwortlich sein.

Wir kommen zu dem Schluß., dass unser Detektionsassay durchführbar, aber in seiner gegenwärtigen Form sehr arbeitsaufwendig ist. Es beinhaltet jedoch das Potential, in Verbindung mit PCR Detektionsmethoden die Fluoreszenzprimer verwenden, effektiv eingesetzt werden zu werden.

1. INTRODUCTION

1.1. Gene conversion as a useful diversifying machine in MHC genes

1.1.1. The MHC

The major histocompatibility complex (MHC) is a set of genes present in all vertebrates studied so far, which have both immunological and non-immunological functions (Trowsdale 1995; Gruen and Weissmann 1997). The discovery is connected to early tumor transplantation studies in mice by Peter Gorer and George Snell. They realised that there must be a genetic basis for a tissue to be recognised as foreign by an animal of a different genetic background. Consequently, they re-named this genetic region as the histocompatibility 2 (*H2*) genes of the mouse (Snell 1986).

The MHC is classically divided into the class I, II, and III regions, but this subdivision has meanwhile been expanded. The telomeric region to the classical MHC complex in humans is now called the class Ib region; a class IV region has also been suggested locating at the telomeric end of the class III region (Gruen and Weissmann 1997). The MHC in humans - Human Leukocyte Antigens (HLA)- is located on chromosome 6p21.31 and covers a region of about 3.6 Mbp depending on the haplotype (MHC seq. consortium 1999). In Mice the *H2* lies on chromosome 17 and spans about 4Mbp ((Trowsdale 1995); for a comparison of mouse and man MHC genetic organisation see: NCBI Human-Mouse homology map; www.ncbi.nlm.nih.gov/Homology/human6.html).

MHC class I and II molecules are glycoproteins which serve as cell surface recognition elements. While MHC *class I* molecules are expressed on virtually all somatic cells, MHC *class II* molecules are expressed on the surface of antigen presenting cells (APCs) (Natarajan *et al.* 1999; Nelson and Fremont 1999). MHC class I molecules sample peptides generated within the cell and signal the cell's physiological state to effector cells of the immune system, both CD8+-cytotoxic T-lymphocytes and natural killer (NK) cells (Natarajan *et al.* 1999). In addition, molecules structurally related to MHC-I, collectively known as MHC-Ib, are more specialized and, in some cases, interact with more limited subsets of lymphoid cells (Fischer Lindahl 1997; Gobin and van den Elsen 2000; Soloski *et al.* 2000). The peptides presented by MHC class II molecules derive from proteins that are either internalised from the extracellular space by fluid phase or receptor-mediated endocytosis, or

from proteins that are endogenously expressed in the secretory pathway (Busch *et al.* 2000). Both MHC class I and II molecules form a fundamental part of the vertebrate immune system in that the binding and presentation of peptides can initiate, regulate and perpetuate an immune response.

The peptide binding grooves of MHC class I and class II molecules provide a variable array of structural pockets, designed to bind the side-chains of antigenic peptides. Individual amino acid polymorphisms characteristic of specific MHC alleles determine the specificity of peptide binding. This provides a direct structural explanation for the ability of genetically different individuals to bind different antigens. Allele specific peptide motifs are known for several *class I* and *class II* molecules found in humans and mice (Buus 1999; Sette and Sidney 1999).

The genes of the MHC class I and II genes show an impressive degree of polymorphism. This is not only reflected in the high number of different alleles, but also in their even distribution within populations (Parham and Ohta 1996). Analysis of sequence alignments has revealed distinct patterns of polymorphism in different parts of the gene. As summarized by Parham and co-workers for *HLA* and since confirmed in further studies most of the variation in class I molecules is confined to the second and third exons encoding the $\alpha 1$ and $\alpha 2$ domains, respectively, which constitute the peptide binding site (Parham *et al.* 1989). The structural analysis of MHC molecules in mouse and man has shown that most of the hypervariable positions are involved in the binding of peptide, confirming the notion that the observed polymorphism is indeed tightly linked to the capability of binding different antigens (Bjorkman and Parham 1990). The distribution of polymorphic positions along with a high observed ratio of coding to non-coding substitutions (K_a/K_s) in positions determining the structure of the peptide binding cleft (Hughes and Nei 1989; Hughes and Yeager 1998) demonstrates that natural selection plays an important role in determining the identity of individual substituents found in natural MHC sequence arrays. In contrast introns have been homogenized relative to exons over evolutionary time, suggesting that balancing selection acts to maintain diversity in the latter, in contrast to the former (Hughes and Nei 1989; Hughes and Yeager 1998).

1.1.2. Origin and maintenance of diversity of MHC alleles

Origin

Polymorphism in MHC alleles could be explained by the generation and accumulation of point mutations. An attractive possibility is the existence of a hypermutation process directed towards MHC genes (discussed in (Klein 1978)). However studies based on statistical methods and computer simulations provided no evidence for increased mutation rates within the divergent regions of MHC genes (Lundberg and McDevitt 1992; Satta *et al.* 1993). Alternatively, mutations could accumulate at a slower rate over a long evolutionary time, as proposed by Jan Klein in the transspecies hypothesis of MHC polymorphism. According to this hypothesis, most polymorphism in MHC alleles exists prior to speciation events and is passed on to new species (Klein 1987). A large founding population is assumed, containing a high number of alleles. Support for the transspecific nature of MHC polymorphism has come from a series of studies which showed that alleles can be more similar to alleles found at the orthologous locus in a related species than they are to alleles at the same locus within the species (McConnell *et al.* 1988; Wakeland *et al.* 1990; Klein *et al.* 1993; Lienert and Parham 1996). Meanwhile, data has accumulated which indicates that allelic lineages are indeed old, predating speciation (Hickson and Cann 1997; Bergström *et al.* 1998). It seems, however, that alleles within a lineage are much more recent, which would agree with more conservative estimations of founder size.

Further evidence supporting the idea of old allelic lineages that then accumulate diversifying mutations after speciation comes from the finding of new HLA alleles in Amerindian tribal populations from South America (Parham and Ohta 1996; Cadavid and Watkins 1997). The new alleles are found predominantly in tribal populations in South America and appear to result from microrecombinational events involving two alleles that were present in the founding population. That microrecombination (gene conversion – see below) is possibly involved in the evolution of MHC sequences also emerges from the visual inspection of the alignments is the “patchworklike” patterns found in the alleles of MHC genes (Parham *et al.* 1989). Perhaps the most striking example of this is the *HLA-DPBI* locus (Moonsamy, Suraj *et al.* 1992; Zangenberg, Huang *et al.* 1995) in which almost all of the variation found in the allele set is confined to discrete short regions which appear reassorted in a one-armed bandit fashion to form new alleles. The patchwork characteristic of polymorphic variation is found generally in HLA genes and has been repeatedly identified and commented upon in this context (Parham and Lawlor 1991; Gaur and Nepom 1996) as

well as in alignments of MHC sequences from other species (rodent and feline: (Wakeland *et al.* 1990; Yuhki and O'Brien 1994)). Small segmental exchanges, i.e. gene conversion have been invoked as the mechanism generating this pattern as a more plausible explanation than the occurrence of multiple recurrent mutations.

Maintenance

Pathogen driven selection can favor genetic diversity of the MHC through both heterozygote advantage (overdominance) and frequency-dependent selection (Potts and Wakeland 1993). Following the frequency-dependent selection hypothesis selection is thought to favour rare MHC genotypes, since pathogens are more likely to have developed mechanisms to evade the MHC-dependent immunity encoded by common MHC genotypes. In the overdominance model all heterozygotes are favored over all homozygotes as proposed by Flaherty (Flaherty 1988). Several molecular models of pathogen-driven selection have been presented which were separated into pathogen evasion models (e.g. molecular mimicry) and host-pathogen interactions (e.g. heterozygote advantage) (Potts and Slev 1995).

Direct MHC associations with specific infectious diseases have been difficult to demonstrate, probably due to the complexity of host-pathogen interactions. The best known ones are Marek's disease in chickens (Briles *et al.* 1977; Briles *et al.* 1983) ({Briles, 1977 #96}, parasitic infestations in Soay sheep (Paterson, Wilson *et al.* 1998), and malaria in humans (Hill *et al.* 1991). Only two examples of heterozygote advantage in human infectious diseases have been reported to date: one for a specific genotype in human immunodeficiency virus (HIV) infection (Carrington *et al.* 1999) and another in hepatitis B virus (HBV) infection (Thursz *et al.* 1997; Thio, Carrington *et al.* 1999).

MHC diversity could also be favored by non-pathogen driven mechanisms. An example is selection through inbreeding depression which acts indirectly by favouring MHC-based disassortative mating (mating preferences; - the MHC is exploited to discriminate against genetic similarity at highly polymorphic loci to avoid inbreeding) (Potts *et al.* 1991; Ober *et al.* 1997). Progeny derived from MHC-dissimilar parents would possess increased fitness because of reduced levels of inbreeding depression and increased resistance to infectious diseases due to increased MHC heterozygosity. Furthermore reproductive mechanisms are discussed with respect to MHC diversity. This could be achieved through mate selection (Penn and Potts 1998), selective fertilization (Ho *et al.* 1994; Wedekind *et al.* 1996), and selective abortion (Hamilton and Hellstrom 1978). The mechanisms involved are

unknown but the plant self-incompatibility system (Haring *et al.* 1990) and the invertebrate allorecognition system (Scofield *et al.* 1982) deserve attention in this respect.

1.1.3. Gene conversion as a useful diversifying machine in MHC genes

Gene conversion describes a process in which gene B acts as a sequence donor and remains unaffected, while gene A receives a block of B sequence and thus turns into a variant allele. The term gene conversion looks back on a long history. In the 30s Hans Winkler, being deeply disappointed about the indifference that the scientific community and the journal Nature conferred to his work, decided to publish an essay, in which he summarised and analysed cases of non-Mendelian segregation and deduced a theory on “the conversion of genes (*“Die Konversion der Gene”*)”, which was based on a physiological process (Winkler 1930). Until today many models were developed around gene conversion, like the copy-choice model and the model of Holliday ((Holliday 1964) and reviewed in (Stahl 1994)).

Mechanistically, the phenomenon has best been explained by the double-strand break repair (DSBR) model of (Szostack *et al.* 1983). This model originally proposed that conversion is initiated by a DSB and results from the repair of a double-strand gap (figure 1-1). DSB formation is followed by exonucleolytic degradation at both the 5' and, to a lesser extent, the 3' termini, creating a double-strand gap with a 3' overhang on either side. One of the free ends then invades the homologous region in the donor duplex and acts as a primer for DNA synthesis, so that a D-loop is displaced in the donor and subsequently enlarged. Eventually, the D-loop anneals to complementary sequences on the 3' overhang to the other side of the gap, priming a second round of DNA synthesis from this end. Once the gap is filled on both strands, ligation ensues resulting in the formation of two Holliday junctions. Resolution of these junctions in the same plane results in the noncrossover configuration of flanking markers, whereas resolution in opposite orientations yields a crossover. Cleavage of the two Holliday junctions independently of each other is predicted to result in half of all gene conversions being associated with crossovers. Later experimental observations forced a revision of the model (Sun *et al.* 1991). It is proposed that after DSB formation, only the 5' ends are resected, leaving intact 3' tails. Annealing of complementary sequences occurs as before, only in this case long tracts of hDNA (heteroduplex DNA) are generated beginning at the site of the DSB and extending in both directions (figure 1-1 B)). Thus, conversion is postulated to result mainly from mismatch correction of hDNA rather than from gap repair. Recently, it has been suggested that the recombination intermediate in

these models could also be resolved via the action of a topoisomerase (Schwacha and Kleckner 1995) or through cleavage of a single Holliday junction (Gilbertson and Stahl 1996) (figure 1-2), in which case only noncrossover conversions will be generated.

Under the right conditions, gene conversion has the power to act as a recombination machine, exploiting the combinatorial possibilities implicit in a system of homologous sequences each marked by distinctive substitutions introduced by point mutation (Holland 1992; Stemmer 1995). Gene conversion acts in diversifying mode when it creates a new allele by transferring a segment of differential sequence from a homologous gene. The homologous-sequence donor may be a tandem (or unlinked) gene, or it may be an allele, and the process may occur mitotically or meiotically. The essential feature of the diversifying mode is that it creates recombinant patchwork genes, which have segments of recipient and segments of donor sequence. In diversifying mode, one end of the conversion tract must lie inside a segment differing between the two converting sequences. If the conversion tract includes the whole differential region between the two genes, then conversion acts only in homogenisation mode and converts the recipient sequence into that of the donor.

The paradigm of diversifying gene conversion is the generation of the chicken primary immunoglobulin (Ig) repertoire. Chicken Ig variable regions are assembled from unique functional V (variable) and J (joining) segments both for the light (L) and the heavy (H) chains (Reynaud *et al.* 1985; Reynaud *et al.* 1987; Reynaud *et al.* 1989). This limited potential for combinatorial diversity is compensated for by extensive somatic diversification through gene conversion in the bursa of Fabricius (reviewed by (Weill and Reynaud 1987; McCormack *et al.* 1991; Thompson 1992; Bezzubova and Buerstedde 1994; Ratcliffe and Jacobsen 1994)). In this case, gene conversion is recruited as a locus-specific mechanism operating in a tissue-regulated manner during a brief period of extensive mitotic proliferation in B-cell development. The recombining sequences are 75-92% identical, only limited regions of perfect pairing are required and the resulting conversion tracts are short. Somatic gene conversion also appears to play a role in the generation of the primary antibody repertoire in rabbit (Becker and Knight 1990), cattle (Parng *et al.* 1996), and pig (Butler *et al.* 1996).

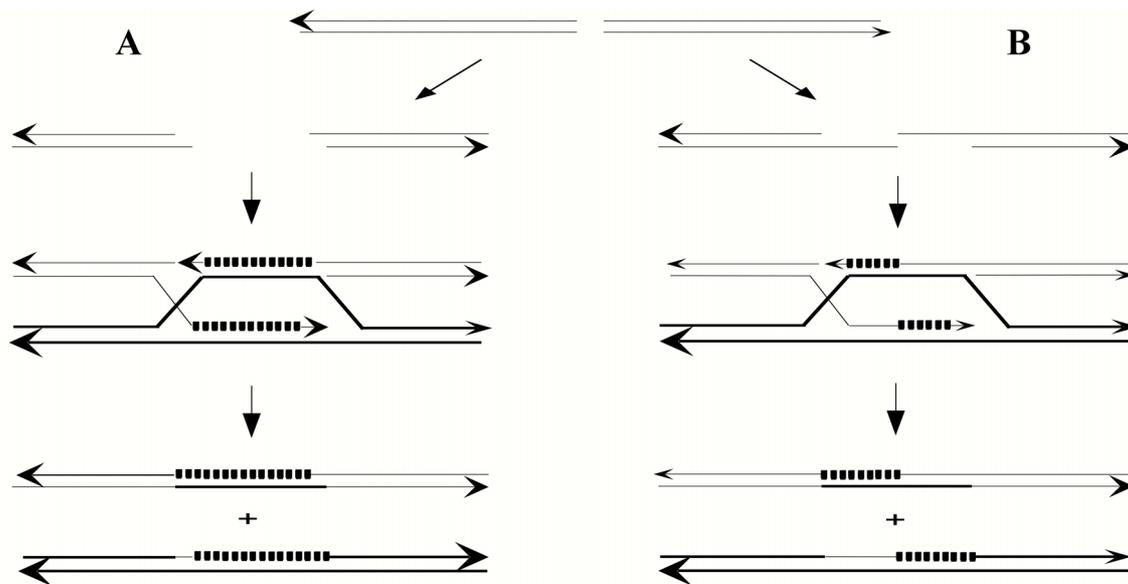


Figure 1-1. Double-strand gap and DSB repair models. Thin and thick lines represent the two strands of homologous duplexes, with 3'-ends marked by arrowheads and repair synthesis by rows of boxes. (A) Original gap repair model as described by Szostak et al. (1983). A DSB is expanded into a double-strand gap. Free ends invade a homologous duplex and prime repair synthesis producing a double Holliday junction structure. Resolution of Holliday junctions can produce either crossover or noncrossover products; only the latter are shown. Also not shown is Holliday junction branch migration, which can produce hDNA adjacent to the repaired gap and lead to conversion via mismatch repair (see Szostak et al. 1983 for details). (B) Modified DSB repair model as described by Sun et al. (Sun et al. 1991). A DSB is processed to two 3'-extensions, which invade and prime repair synthesis, again producing two Holliday junctions. This model predicts conversion as a consequence of strand switching between alleles coupled with repair synthesis (thick line paired with boxed line) and mismatch repair of hDNA (thick line paired with thin line). Again, only noncrossover products arising without branch migration are shown. (From Nickoloff and Hoekstra 1998).

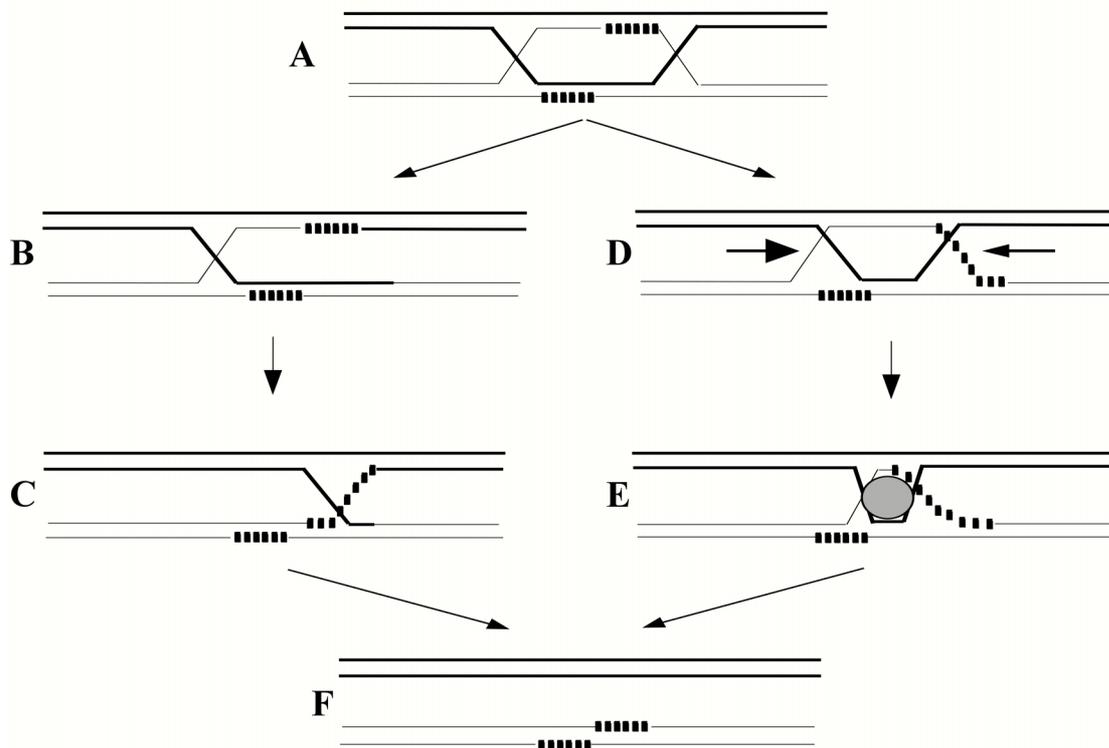


Figure 1-2. Two alternative DSB repair resolutions. The intermediate (A) is the same as in the DSB repair model (Figure x pathway B). Boxed lines indicate newly synthesised DNA. This intermediate may be resolved by cutting only one Holliday junction (B and C) or by cutting neither Holliday junction (D and E). In the former case, junction cleavage leaves single-strand nicks on both molecules (B). The remaining Holliday junction is free to move, and eventually slides to the site of the nicks, resulting in resolution to give (after ligation) the products shown (F). In the latter case, the Holliday junctions converge (D), presumably with the assistance of a topoisomerase to relieve the supercoiling that will occur when they do. A topoisomerase acts on the final intertwinings (E) to resolve the two molecules, giving the products shown (F). (From Gilbertson and Stahl 1996).

1.1.4. The MHC running alongside the red queen?

It was J.B.S. Haldane who first raised the notion that competition between pathogens and their hosts can lead to stable polymorphisms (Haldane 1949). He also pointed out that it would be advantageous for a species if genes concerned in disease resistance would be particularly mutable provided that this mutation process is highly regulated. Examples of regulated and targeted mechanisms creating variability (e.g. chromosomal rearrangement, gene conversion, hot spots of mutagenesis) are found in the vertebrate immune system, and they often find their counterparts in pathogens (Weill and Reynaud 1996). By restricting the generation of variability to specialised cells and certain regions within the genome (as is the case for B and T cells in the mammalian immune system) the host diminishes the cost associated with the generation of genetic variability, which is conform with Haldane's propositions. Evolution can be regarded as an interplay of genetic variation and phenotypic selection and mutation is generally recognized as the ultimate source of evolutionary change. However its relative importance in adaptive processes has been underestimated. Recently the beneficial effects of raised mutation rates due to distinct mutator mechanisms in Bacteria in response to stress were extensively discussed (Taddei *et al.* 1997; Radman *et al.* 1999). Non-minimal mutation rates could be linked to the intensity of selective pressure, a trait which would go along with the red queen hypothesis as proposed by Van Valen (Valen 1973). It is tempting to consider that similar processes apply to MHC genes, but the issue has to remain highly speculative for the moment.

Gene conversion is certainly an attractive candidate as a process introducing variability in a highly efficient way with regards to pathogen-challenge response into MHC genes. This leads us to the gene conversion hypothesis of MHC evolution.

1.1.5. The gene conversion hypothesis of MHC evolution

Patterns of allelic variation between MHC genes are inconsistent with simple mutational divergence from a series of ancestral sequences by point substitution in independent lineages (see also 1.1.2. - Origin and maintenance of diversity of MHC alleles). The patchwork-like pattern of variation which emerges from the visual inspection of MHC alignments can best be explained by gene conversion. The linear array of substitutions which characterise alleles and homologues of MHC genes have individually arisen *de novo* by point mutation. Microrecombination of such point mutational variation by gene conversion is an

efficient generator of new peptide-binding grooves on which natural selection may act. This is the gene conversion hypothesis of MHC evolution.

We have distinguished two forms of the gene conversion hypothesis, which we have called *strong* and *weak* on the basis of the adaptive significance attributed to the gene conversion process in MHC genes (Martinsohn *et al.* 1999). The *weak* form accepts that observed singular events resembling gene conversions, as well as historical evidence based on sequence alignments, may record genuine gene conversions, but makes no claims about the special significance of the process; the *strong* form asserts that diversification of MHC alleles by gene conversion is a powerful force for optimising sequence variability and pattern. In the *weak* form, gene conversion is an inevitable genetic process affecting homologous genes, whether tandem or allelic, with no special significance for the evolution of MHC genes; in the *strong* form the existence of adaptations is anticipated, structural and mechanistic, which facilitate gene conversions in the right place on the right genes and in the right tissues. The *weak* form makes no claims for a specific connection between this potential mutator and pathogen variation; in the *strong* model, gene conversion functions as an adaptive mutator under selection, functionally related to the unpredictability of infectious pathogen selection.

1.2. Evidence for gene conversion in the mouse MHC

The most compelling evidence for the occurrence of gene conversion in MHC genes comes from the documented instances of spontaneous class I gene mutation in the laboratory mouse. Over 40 mutational events have been recorded in the MHC, both in the renowned screening programmes in mice, mainly carried out in the laboratories of I.K. Egorov, D.W. Bailey, H.I. Kohn, and R.W. Melvold, and in occasional independent studies. Altogether over 150 000 mice were screened by skin grafting for the expression of anomalous transplantation specificities ((Bailey and Usama 1960; Bailey and Kohn 1965; Bailey 1966) and reviewed in (Klein 1978; Nathenson *et al.* 1986; Melvold *et al.* 1990; Melvold *et al.* 1997; Pease *et al.* 1991; Martinsohn *et al.* 1999). The *in vivo* mutants documented so far are listed in table 1-1.

1.2.1. Characteristics

The first numbered MHC mutation from the screen, originally known as *H21*, subsequently *H2ba* and finally *H2-K^{bm1}*, is exemplary for the whole set (figure 1-3). The mutation occurred in the third exon, the $\alpha 2$ domain of the class I heavy-chain gene *K^b*, and involves the replacement of the wildtype sequence over a stretch of 13 consecutive

nucleotides with a homologous segment containing 7 novel residues. These residues can be found in another class I gene, $Q10^b$, mapping about 1 megabase (Mb) telomeric to K^b (Mellor, Weiss et al. 1983), which, elsewhere, differs extensively in sequence from K^b . The three key features of K^{bm1} are:

(1) the mutation appears to involve sequence transfer from an identifiable donor in the same genome;

(2) the mutation cannot be explained by a simple crossover event;

(3) the tract length of the apparent double crossover is very short (in the case of $bm1$, 13 bp minimum, 50 bp maximum to the flanking markers on both sides of the mutational event).

These properties were immediately recognised as carrying the stamp of the gene conversion process, already proposed to be active in MHC genes on other grounds (Evans *et al.* 1982; Gachelin *et al.* 1982; Mellor *et al.* 1983). The amino acid substitutions in $H2 K^{bm1}$ result in radical structural alterations to the peptide-binding site (I. Wilson, -as cited in (Martinson *et al.* 1999)). These in turn cause the reported “gain-and-loss” immunological specificity of the mutant; that is, the mutant molecule no longer expresses all the T-cell epitopes of the parental K^b , and in addition, expresses novel epitopes of its own. $H2 K^{bm1}$ is a major histocompatibility antigen (Widmer and Macdonald 1980) and a novel restriction specificity.

All mutations in $H2$ class I genes that have been reported occurred in exon 2 or 3, which encode the peptide binding groove of MHC molecules. This might well be due to the fact that the detection assay relies on the presence of strong rejection antigens. This seems to be confirmed by $bm29$, which was detected by serological anomaly (Horton *et al.* 1991), and which does not affect the peptide binding cleft, and does not induce graft rejection in the appropriate tests.

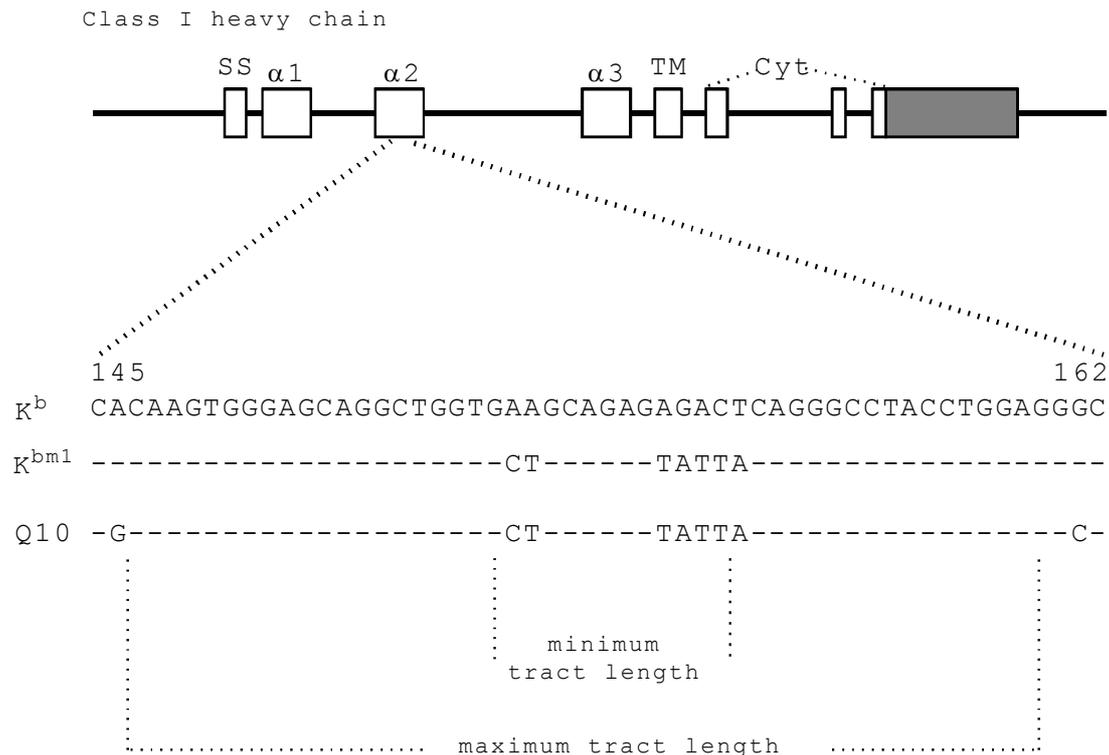


Figure 1-3. Alignment of the nucleotide sequences of K^b , K^{bm1} and Q10 genes from codon 145 to 162. (adapted in part from Geliebter and Nathenson (Geliebter and Nathenson 1987))

1.2.2. Frequencies

40 gene conversion mutants detected among approximately 150 000 screened mice give an overall frequency of $2-3 \times 10^{-4}$ per gene per generation. Correction for the number of fully documented conversions reduces this figure by roughly half. At first sight this seems to be a high mutation rate compared to other examined loci (Schlager and Dickie 1971; Schlager 1972; Russel and Russel 1992). But in fact most of the excess frequency in MHC gene mutations can be attributed to the K^b allele alone. Within the group of K^b mutations, a further bias is visible since 9 of altogether 22 K^b mutants belong to the “ bg ”-group (table 1-1). The other $H2$ genes examined show mutation rates which may be as much as an order of magnitude lower and therefore in the range of frequencies observed for non-MHC genes (see discussion in (Martinsohn *et al.* 1999)). However in general frequency estimates based on the mouse screen are probably too low due to the requirement of immunogenicity in the assay.

1.2.3. Interlocus and interallelic conversions

All of the mice used in the parental generation during the skin grafting screen were inbred, so that interallelic exchanges could not have been detected. The HLA allele sets show a high degree of locus-specific differentiation and the predominant mode of exchange appears to be interallelic (Huang *et al.* 1995; Parham *et al.* 1995; Zangenberg *et al.* 1995). However,

the human allele set is viewed only after selection and may be relatively uninformative about the relative frequencies of primary events. Thus the dichotomy between interlocus exchanges in mouse and interallelic exchanges in human may well be more apparent than real. The underlying issue, however, concerns the nature of the cellular mechanisms involved in the two classes of exchange: interallelic exchanges are probably meiotic, while interlocus exchanges may as easily be mitotic as meiotic events.

1.2.4. Acceptors and donors

All the experimentally determined class I mutations were in classical genes, while both classical and non-classical genes appear to have participated as donors. This bias can also be attributed to the conditions set by the assay (skin grafting: strong transplantation reaction) and therefore there is no support for the view that the class I gene conversion mutations show nonclassical to classical directionality. The striking predominance of $Q4^b$ in the dataset as a potential donor gene has often been noted and attributed to the large tract of sequence identity between this gene and K^b surrounding the two base pair (bp) disparities associated with the “*bg*” group of mutations (Geliebter *et al.* 1986). Some mutations differ from the wild type at a single coding base, and could therefore reasonably be attributed to untemplated single point mutation. However, in all these cases, potential donor genes carrying the appropriate sequence are available in the known class I gene set for the donor haplotype. For the two mutations *bm3* and *bm23*, no one donor gene in the parental genome is known which accounts for all the substitutions. Extrachromosomal cDNA intermediates have been invoked to account for these complex mutation clusters (Pease *et al.* 1993).

1.2.5. Multiple mutant births

In three of the reported *H2* mutants, *bm6*, *bm9*, and *bm23*, the parents produced clusters of identical mutant offspring (Melvold *et al.* 1982; Egorov and Egorov 1984), suggesting gonadal chimerism as a result of a mitotic mutational event during germ cell expansion (Geliebter *et al.* 1986; Pease *et al.* 1991). In two cases (*bm9* and *bm23*), the mutation originated in the female germ line; in *bm6* it could not be determined. The history and genealogy of these mutations is documented ((Egorov and Egorov 1984) and personal communication; (Melvold *et al.* 1982) and personal communication and for detailed genealogy see tables 2 to 4 in (Martinson *et al.* 1999)). Before considering the gonadal chimerism interpretation further, it is important to exclude the trivial possibility that one of the parents of each of the three affected litters was in fact already heterozygous for the

mutation. Although the parents of each mating were tested before establishment of the pair, none of the pairs generating multiple mutant births were retested after the mutant births were detected, allowing the possibility of a husbandry error. The overall numbers for the three multiple mutant families were: *bm6*, 3 mutant out of 14 progeny tested; *bm9*, 5 mutant out of 20 progeny tested; *bm23*, 4 mutant out of 6 progeny tested. None of these values is inconsistent with a heterozygous parent. If husbandry error is excluded, however, the frequency of mutants is extraordinarily high. The mutants are not monozygotic siblings, since in each case both sexes were represented and in the case of *bm23* appeared in two litters. For *bm9* and *bm23*, the mutation must have occurred in the female germ line. On the face of it, these mutants must have been derived from gonads chimeric to the level of 20–50%. Such high values imply genetic events occurring at a time when the germ cell primordium contains only two to four cells. The number of founder primordial germ cells appears to be higher, possibly considerably higher, than this (Ginsburg *et al.* 1990). Such levels of chimerism in the germ cell pool therefore seem to imply high levels of chimerism outside the germ cell pool in the mother. It is not clear whether this would have been detected by the screening of the parents: wild-type grafts would be accepted, while grafts from either parent to wild type would be accepted to the extent that they contained wild-type cells. A curious aspect of the data is the clustering of the mutant births into single litters in the case of *bm6* (in sum 5 litters recorded) and *bm9* (in sum 3 litters recorded). This might indicate that the overall frequency of the mutant germ cells in the respective parental animals was after all not high. A possibility is that the mutations indeed occurred among the primordial germ cell pool, and during further mitotic division, a cluster of closely neighbored mutant cells stayed together until fertilization. Recent findings indeed indicate that in mouse ovaries, the germ cells sit together in clonal clusters of up to 16 cells. These go through mitotic cycles in synchrony and are interconnected by cytoplasmic processes (Pepling and Spradling 1998; Pepling *et al.* 1999). If the germ cells in these synchronized clusters are also ovulated at the same time, which is not yet known, single litters containing more than one individual carrying a rare mutation could be accounted for. Analysis of the representation in litters of genotypes from female gonadal chimeras established experimentally through embryonic stem cell technology will confirm or refute this possibility. Despite the complexities detailed above, the multiple mutant births have always been assumed to represent the result of mutation during mitotic expansion of germ cells (Klein 1978; Geliebter *et al.* 1986; Pease *et al.* 1991). Consistent with this view, recent experimental work has shown that mitotic germ cells in the male gonad are competent to effect interlocus gene conversion at a significant rate (Murti *et al.* 1992), and surprisingly

gene conversion in *trans* between the MHC class II $A\beta$ and $E\beta$ loci has also been claimed for premeiotic male germ cells (Hogstrand and Bohme 1997). Thus the multiple mutant births appear to have focused attention on mitotic recombination processes in the MHC, a feature which brings them into comparison with the highly regulated gene conversion processes occurring in immunoglobulin genes during the mitotic expansion of pre-B cells and the formation of the B-cell repertoire in chickens and rabbits (see above).

1.2.6. Sex bias

For the mutants available to them at the time, (Loh and Baltimore 1984) observed that the mutation must have occurred in the female germ line in all 16/16 cases where the haplotype of origin could be determined. table 1-1 includes two further mutants, *dm4* and *dm6*, both single base exchanges which are not excluded as gene conversions in which the mutations arose on the paternal chromosome. Both *dm4* and *dm6* were induced by mutagens. Melvold and coworkers (1997) have pointed out that the observed female bias may arise from the combination of two other biases, namely the excess mutation rate in K^b discussed above and the fact that in this screening program, fivefold more animals were screened in which the maternal haplotype was *b* than vice versa. Experimental investigations into gene conversion processes in mice have focused for technical reasons on events occurring in the male germ line (Murti *et al.* 1992; Hogstrand and Bohme 1994). In summary, it is not yet clear if and to what extent gene conversion in the mouse MHC occurs preferentially in the female germ line.

The *H2* mutants found by *in vivo* screening have been extremely influential in guiding the current general acceptance that a process leading to gene conversion-like events in the MHC occurs in the mammalian germ line (Pittman and Schimenti 1998). Nevertheless, much still remains unknown or unclear. Frequency, specificity, directionality, all need extensive further analysis. Are the events meiotic or mitotic? Do they occur in *trans* as well as in *cis*? Is there really a sex bias involved? While these mutants were invaluable for the stimulation of certain hypothesis, at the same time they did not provide any conclusive answer. This is mainly due to the fact that despite the impressive number of animals screened in the combined skin grafting approach, the sample size is still too small to gain enough data allowing conclusive statements. All the issues which rose with the detection of the *in vivo H2* mutants finally flow into the question whether gene conversion in mouse MHC genes is a regulated adaptive machine, or merely the expression of inevitable genomic processes read

out in a system in which unit events are relatively easy to detect. - It is unlikely that further progress can be made in this discussion until a high-throughput general method is developed for determining mutations directly in MHC genes in tissues. The next section describes some recent progress in this direction.

Table 1-1. Summary of recorded <i>in vivo</i> H2 mutants (taken from Martinsohn et al. 1999)										
name	other names ⁱ	locus	donor genes ⁱⁱ	original mutant ⁱⁱⁱ [sex]	H2	gain or loss ^{iv}	position altered AA	altered nucleotides/ consecutive nucleotides	maximal extension ^v	reference {m}= mutant {s}= sequence ^{vi}
bm1	H2 ^{ba} ; Hz1	K ^b	Q10 ^b	(B6By X CBy)F1 [f]	bx	GL	152 E→A 155 R→Y 156 L→Y	7/13	50	[1] {m} [2][3][4][5] {s}
bm2 ^{vii}	H2 ^{bb} ; Hz49	K ^b		(B6By X CBy)F1 [f]	bx	GL	155 R→Y 156 L→Y			[1][7] {m}; [6] {s}
bm3 ^{viii}	H2 ^{bd} ; M505	K ^b	complex ^{ix} (Q4 ^b +Q10 ^b)	B6JY	b	GL	77 D→S 89 K→A	4/38	extension into intron 2	[7] {m} [9][10] {s}
bm4	H2 ^{bf} ; Hz170	K ^b	T5	(B6By X CBy)F1	bx	GL	162 G→D 163 T→P 165 V→L 173 K→E 174 N→L	6/37	62	[7][11] {m} [9] {s}
bm5 (,bg ^{cc})	H2 ^{bg1}	K ^b	Q4 ^b or Q7 ^b -Q9 ^b	B6Kh [f]	b	GL	116 Y→F	1/1	36	[7][12][13][20] {m} [14][15] {s}
bm6 (,bg ^{cc})	H2 ^{bg2}	K ^b	Q4 ^b	B6Kh [2f, 1m]	b	GL	116 Y→F 121 C→R	2/15	95	[12][13] {m} [16][17][19] {s}
bm7 (,bg ^{cc})	H2 ^{bg3}	K ^b	Q4 ^b	(B6Kh X CKh)F1 [m]	bx	GL	116 Y→F 121 C→R			[12][18] {m} [19] {s}
bm8	H2 ^{bh}	K ^b	?Q4 ^b ^x	B6Kh [m]	b	GL	22 Y→F 23 M→I 24 E→S 26 silent (GGC→GGT) 30 D→N	7/24	95	[20] {m} [21] {s}
bm9 (,bg ^{cc})	H2 ^{bi}	K ^b	Q4 ^b	(B6Kh X CKh)F1 [1f,4m]	bx	GL	116 Y→F 121 C→R	2/15	95	[12][18] {m} [17][19] {s}
bm10	H2 ^{bj}	K ^b	K1	(B6Kh X CKh)F1 [m]	bx	GL	163 T→A 165 V→M 167 W→S 173 K→E 174 N→L	6/35	≥79 ^{xi}	[12][18] {m} [9] {s}
bm11	H2 ^{bk}	K ^b	D ^b	(B6Kh X CKh)F1 [m]	bx	GL	77 D→S 80 T→N	3/11	45	[12][18] {m} [9] {s}
bm12	H2 ^{bm}	Ab ^b	Eb ^b	(B6Kh X CKh)F1 [f]	bx	GL	68 I→F 71 R→Q 72 T→K	3/15	44	[22] {m} [23][24][25] {s}
bm13	H2 ^{bn}	D ^b	K ^b	(B6Kh X CKh)F1 [m]	bx	GL	114 L→Q 116 F→Y 118 silent(TAT→TAC) 119 E→D	4/17	36	[18] {m} [26] {s}
bm14	H2 ^{bo} ; zw42	D ^b	many possible ^{xii}	B6.C-H(z-w42)	b	GL	70 Q→H	1/1		[27] {m} [26] {s}
bm15 ^{xiii}	M513			(B10/Sm X B10.D2) F1; [f]	see [28]	L				[28] {m}
bm16 ^{xiv} (,bg ^{cc})		K ^b		B6Kh	b		116 Y→F			[12][14] {m} [14][15] {s}
bm17 (,bg ^{cc})		K ^b		B6Kh [m]	b	GL	116 Y→F 121 C→R			[12][13] {m} [15] {s}
bm18 (,bg ^{cc}) ^{xv}		K ^b		(B6Kh X CKh)F1 [f]	bx	GL				[12][29] {m}
bm19		K ^b		(B6Kh X CKh)F1 [f]	bx	GL				[12][29] {m}
bm20 (,bg ^{cc}) ^{xvi}		K ^b		(B6Kh X CKh)F1 [f]	bx	GL				[12][29] {m}
bm21		K ^b		(B6Kh X CKh)F1 [f]	bx	GL				[12][29] {m}

bm22		K ^b		B6Kh [m]	b	GL					[12][29] {m}
bm23	M568	K ^b	complex ^{xxvii} (Q10b+ either Q4b or Db)	(B10SnEg X B10.D2/nSnEg)F1 [3f; 1m]	see [30]	GL	63 (GAG→GAA) 75 R→H 77 D→S	silent	5/43		[30] {m} [10] {s}
bm24 ^{xxviii}		D ^b	K ^b	B10ScNCr	b	GL	63 (GAA→GAG) 70 Q→N 73 W→S 77 S→D 80 N→T	silent	8/51	93	[31] {m} [26] {s}
bm25		K ^b		B10.RIII(71NS)F1	b						[53] {m}
bm26 ^{xxix}		K ^b		H2 ^b /H2 ^{bxxx}	b	L					[32] {m}
bm27 ^{xxxi}		H2b		H2 ^b /H2 ^{bxxii}	b						[32] {m}
bm28		D ^b	Multiple ^{xxiii}	BL6By	b		97 Q→W 99 S→Y 103 (TTG→CTG)	silent	4/19		[33] {m} [34] {s}
bm29 ^{xxxiv}		K ^b	many possible ^{xxv}	((C3H.CAS3(R4) X B6)F1 X (C3H.CAS 3(R4) X B6)F1)F1	see [35]		89 K→A		2/2		[35] {m; s}
KB-98 („bg ^{6c})		K ^b		B6Kh	b		116 Y→F				[15] {m; s}
KB-96		H2K		B6Kh	b	GL					[7] {m}
KB-97				B6Kh [m]	b	GL					[7] {m}
KH-170		H2K ^b		(B6Kh x CKh)F1 [m]	bx d	GL					[7] {m}
KH-171		H2K ^b		(B6Kh x CKh)F1 [f]	bx d	GL					[7] {m}
dm1 ^{xxxvi}	H2 ^{da} ; M504	D ^d /L ^d ^{xxxvii} hybrid gene		(B10/Sm X B10.D2) F1 [f]	bx d	GL					[8] {m} [44] [45] {s}
dm2 ^{xxxviii}	H2 ^{db}	L ^d deleted		CKh	d	L	deletion mutant				[46] {m} [47] {s}
dm3	H2 ^{dc}	H2 ^d ^{xxxiv}		(B6Kh X CKh)F1 [f]	bx d	L					[13][29] {s}
dm4 ^{xxx}	H2 ^{dd}	K ^d	Ld ^{xxxix}	(B6Kh X CKh)F1 [m]	bx d	GL	114 Q→E		1/1		[48] {m} [49] {s}
dm5		K ^d	T3 ^d T18 ^d ^{xxxii}	(CKh X B6Kh)F1 [f]	dx b	GL	158 A→T		1/1		[50] {m} [49] {s}
dm6 ^{xxxiii}		D ^d	many possible	(B6Kh X CB y)F1 [f]	bx d	L	133 W→R		1/1		[51] {m} [52] {s}
dm7		H2d		(B10.D2/nSnEg X B 10SnEg)F1	dx b	GL					[32] {m}
KH-16 6A		H2d ^{xxxiv}		(B6 X C)F1 [f]	bx d	L					[29] {m}
KH-188 ^{xxxv}		H2d ^{xxxvi}		(B6 X C)F1 [f]	bx d	L					[29] {m}
KH-162A		K ^d ↔ L-A ^d		(CKh x B6Kh)F1 [f]	dx b	GL					[7] {m}
fm1 ^{xxxvii}	H2 ^{fa} ; M506	K ^f	D ^f	(A.CA/SnKIEg X A. Sn)F1 [f]	fx a	GL	94 silent (ACA→ACG) 95 L→I 97 V→R		4/9	71	[28] {m} [41] {s}
fm2	H2 ^{fb} ; FB	D ^f	K ^f	(B10.M X B10.RIII[71NS])F1	fx r	GL	62 P→R		1/1	47	[42] {m} [43] {s}
km1	H2 ^{ka} M523	K ^k	pH2III	CBA/CaLacSto [f]	k	GL	152 D→A		1/1	≥72 ^{xxxviii}	[36] {m} [37][38] {s}
km2 ^{xxxix}		K ^k	D ^k	C3HfB/HeN	k	GL	95 F→I 96 silent (CAA→CAG) 98 M→L 99 Y→S		4/14	45	[39][40] {m} [39] {s}
K ^{bm1} - revertant ^{xxxx}				(B6.C-H2 ^{bm1} X B6.C-H2 ^{bm18} /H2 ^b)F1	bm1x bm18/b						[13] {m}

- [1] = Bailey, D. W. and H. I. Kohn (1965). "Inherited histocompatibility changes in progeny of irradiated and unirradiated inbred mice." Genet Res **6**(3): 330-40.
- [2] = Weiss, E. H., A. Mellor, et al. (1983). "The structure of a mutant H2 gene suggests that the generation of polymorphism in H2 genes may occur by gene conversion-like events." Nature **301**(5902): 671-4.
- [3] = Mellor, A. L., E. H. Weiss, et al. (1983). "A potential donor gene for the bml gene conversion event in the C57BL mouse." Nature **306**(5945): 792-5.
- [4] = Schulze, D. H., L. R. Pease, et al. (1983). "Comparison of the cloned H2K^{bml} variant gene with the H2K^b gene shows a cluster of seven nucleotide differences." Proc Natl Acad Sci U S A **80**(7): 2007-11.
- [5] = Miyada, C. G., C. Klofolt, et al. (1985). "Evidence that polymorphism in the murine major histocompatibility complex may be generated by the assortment of subgene sequences." Proc Natl Acad Sci U S A **82**(9): 2890-4.
- [6] = Nairn, R., K. Yamaga, et al. (1980). "Biochemistry of the gene products from murine MHC mutants." Annu Rev Genet **14**: 241-77.
- [7] = Kohn, H. I., J. Klein, et al. (1978). "The first H2 mutant workshop." Immunogenetics **7**: 179-294.
- [9] = Geliebter, J. and S. G. Nathenson (1988). "Microrecombinations generate sequence diversity in the murine major histocompatibility: Analysis of the K^{bm3}, K^{bm4}, K^{bm10} and K^{bm11} mutants." Molecular and Cellular Biology **8**: 4342-4352.
- [10] = Pease, L. R., R. M. Horton, et al. (1993). "Unusual mutation clusters provide insight into class I gene conversion mechanisms." Mol Cell Biol **13**(7): 4374-81.
- [11] = Bailey, D.W. and Cherry, M. (1975). „A new H2 mutation.“ Forty-sixth Annual Report, The Jackson Laboratory, Bar Harbor, Maine **46**: 1975
- [12] = Melvold, R. W., H. I. Kohn, et al. (1982). "History and genealogy of the H2K^b mutants from the C57BL/6Kh colony." Immunogenetics **15**(2): 177-85.
- [13] = Melvold, R. W. and H. I. Kohn (1990). Spontaneous frequency of H2 mutations. Mutant and Transgenic Mice in Major Histocompatibility Complex Research. C. David and I. Egorov. New York, Springer: 1-13.
- [14] = Yamaga, K. M., G. M. Pfaffenbach, et al. (1983). "Biochemical studies of H2K antigens from a group of related mutants. I. Identification of a shared mutation in B6-H2^{bm5} and B6-H2^{bm16}." Immunogenetics **17**(1): 19-29.
- [15] = Pfaffenbach, G. M., R. W. Melvold, et al. (1990). "Biochemical analysis of related, independently arising histocompatibility mutants: bm17 and KB-98 enlarge the "bg series" of H- 2K^b mutants." Biochem Genet **28**(7-8): 433-41.
- [16] = Geliebter, J., R. A. Zeff, et al. (1986). "Interaction between K^b and Q4 gene sequences generates the Kbm6 mutation." Mol Cell Biol **6**(2): 645-52.
- [17] = Geliebter, J., R. A. Zeff, et al. (1986). "Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: K^{bm9} and K^{bm6}." Proc Natl Acad Sci U S A **83**(10): 3371-5.
- [18] = Melvold, R. W. and H. I. Kohn (1976). "Eight new histocompatibility mutations associated with the H2 complex." Immunogenetics **3**: 185-191.
- [19] = Yamaga, K. M., G. M. Pfaffenbach, et al. (1983). "Biochemical studies of H2K antigens from a group of related mutants. II. Identification of a shared mutation in B6-H2^{bm6}, B6.C-H2^{bm7}, and B6.C-H2^{bm9}." Immunogenetics **17**(1): 31-41.
- [20] = Melief, C. J. M., R. S. Schwartz, et al. (1975). "Dermal histocompatibility and in vitro lymphocyte reactions of three new H2 mutants." Immunogenetics **2**: 337-348.
- [21] = Pfaffenbach, G. M., H. Uehara, et al. (1991). "Analysis of the H2K^{bm8} mutant: correlation of structure with function." Mol Immunol **28**(7): 697-701.
- [22] = McKenzie, I. F., G. M. Morgan, et al. (1979). "B6.C-H2^{bm12}. A new H2 mutation in the I region in the mouse." J Exp Med **150**(6): 1323-38.
- [23] = Mengle-Gaw, L., S. Conner, et al. (1984). "Gene conversion between murine class II major histocompatibility complex loci. Functional and molecular evidence from the bm12 mutant." J Exp Med **160**(4): 1184-94.
- [24] = McIntyre, K. R. and J. G. Seidman (1984). "Nucleotide sequence of mutant I-A beta bm12 gene is evidence for genetic exchange between mouse immune response genes." Nature **308**(5959): 551-3.
- [25] = Widera, G. and R. A. Flavell (1984). "The nucleotide sequence of the murine I-E beta b immune response gene: evidence for gene conversion events in class II genes of the major histocompatibility complex." Embo J **3**(6): 1221-5.
- [26] = Hemmi, S., J. Geliebter, et al. (1988). "3 Spontaneous H2^{db} Mutants Are Generated By Genetic Micro-Recombination (Gene Conversion) Events - Impact On the H2-Restricted Immune Responsiveness." Journal of Experimental Medicine **168**(6): 2319-2335.

- [27] = Bailey, D.W. and Cherry, M. (1975). „A new mutation at the D end of the H2 complex.“ Jackson Laboratory Annual Report 47:53
- [28] = Egorov, I. K. and Z. K. Blandova (1972). “Histocompatibility mutations in mice: chemical induction and linkage with the H2 locus.” Genetic Research **19**: 133-143.
- [29] = Melvold, R. W., K. Wang, et al. (1997). “Histocompatibility gene mutation rates in the mouse: a 25-year review.” Immunogenetics **47**(1): 44-54.
- [30] = Egorov, O. S. and I. K. Egorov (1984). “H2^{bm23}, a new K^b mutant similar to, but not identical with H2^{bm3}.” Immunogenetics **20**: 83-87.
- [31] = Melvold, R.W., and C. Waltenbaugh (1986). Mouse News Letter **74**:75
- [32] = Egorov, I. K. and O. S. Egorov (1988). “Detection of new MHC mutations in mice by skin grafting, tumor transplantation and monoclonal antibodies: a comparison.” Genetics **118**(2): 287-98.
- [33] = Melvold, R.W. (1989). Mouse News Letter **84**:58
- [34] = Yun, T. J., R. W. Melvold, et al. (1997). “A complex major histocompatibility complex D locus variant generated by an unusual recombination mechanism in mice.” Proc Natl Acad Sci U S A **94**(4): 1384-9.
- [35] = Horton, R. M., B. E. Loveland, et al. (1991). “Characterization of the spontaneous mutant H2K^{bm29} indicates that gene conversion in H2 occurs at a higher frequency than detected by skin grafting.” J Immunol **147**(9): 3180-4.
- [36] = Blandova, Z., Y. A. Mnatsakanyan, et al. (1975). “Study of H2 mutations in mice. VI. M523, a new K end mutant.” Immunogenetics **2**: 291-295.
- [37] = Martinko, J. M., J. C. Solheim, et al. (1987). “The H2K^{km1} mutation: nucleotide sequence and comparative analysis.” Mol Immunol **24**(2): 197-200.
- [38] = Martinko, J. M., J. C. Solheim, et al. (1988). “The H2K^{km1} mutation: a single nucleotide substitution is responsible for multiple functional differences in a class I MHC molecule.” Mol Immunol **25**(3): 267-74.
- [39] = Vogel, J. M., A. C. Davis, et al. (1988). “Molecular characterization of the C3HfB/HeN H2K^{km2} mutation. Implications for the molecular basis of alloreactivity.” J Exp Med **168**(5): 1781-800.
- [40] = Vogel, J. M., M. McMillan, et al. (1989). “Evidence that derivation of the adenocarcinoma LT85 predated establishment of the H2^{km2} mutation in a C3Hf colony of mice.” J Immunogenet **16**(4-5): 363-71.
- [41] = Horton, R. M., W. H. Hildebrand, et al. (1990). “Structural analysis of H2K^f and H2K^{fm1} by using H2K locus-specific sequences.” J Immunol **145**(6): 1782-7.
- [42] = Mobraaten, L.E. and D.W. Bailey (1973) Mouse News Letter **48**:17
- [43] = Hildebrand, W. H., R. M. Horton, et al. (1992). “Nucleotide sequence analysis of H2D^f and the spontaneous in vivo H-2D^{fm2} mutation.” Mol Immunol **29**(1): 61-9.
- [44] = Burnside, S. S., P. Hunt, et al. (1984). “A molecular hybrid of the H2D^d and H2L^d genes expressed in the dm1 mutant.” Proc Natl Acad Sci U S A **81**(16): 5204-8.
- [45] = Sun, Y. H., R. S. Goodenow, et al. (1985). “Molecular basis of the dm1 mutation in the major histocompatibility complex of the mouse: a D/L hybrid gene.” J Exp Med **162**(5): 1588-602.
- [46] = Blanden, R. V., I. F. McKenzie, et al. (1977). “Cytotoxic T-cell response to Ectromelia virus-infected cells. Different H2 requirements for triggering precursor T-cell induction or lysis by effector T cells defined by the BALB/c-H2^{db} mutation.” J Exp Med **146**(3): 869-80.
- [47] = Rubocki, R. J., T. H. Hansen, et al. (1986). “Molecular studies of murine mutant BALB/c-H2^{dm2} define a deletion of several class I genes including the entire H2L^d gene.” Proc Natl Acad Sci U S A **83**(24): 9606-10.
- [48] = Morgan, G., H. Dellos, et al. (1981). “Study of an H2K^d mutant strain: C.B6-H2dm4.” Immunogenetics **12**(5-6): 555-60.
- [49] = Pullen, J.K., M.D. Tallquist et al. (1994). „Recognition of a single amino acid change on the surface of a major transplantation antigen is in the context of self peptide“ J Immunol **152**: 3445-3452.
- [50] = Melvold, R. W., M. S. Sandrin, et al. (1982). “The C.B6-H2^{dm5} mutant.” Immunogenetics **15**(3): 331-2.
- [51] = Melvold, R. W., P. M. Stuart, et al. (1983). “A new loss mutant in the H2^d haplotype.” Transplant Proc **15**(4): 2045-7.
- [52] = Rubocki, R. J., J. M. Connolly, et al. (1991). “Mutation at amino acid position 133 of H2D^d prevents beta 2m association and immune recognition but not surface expression.” J Immunol **146**(7): 2352-7.
- [53] = Blandova, Z. (1989). “H2^{bm25} in the strain B10.RIII(71NS).” Mouse News Letter **84**: 58.

ⁱ Lists former designations of mutants as to facilitate orientation in original literature.

ⁱⁱ Proposed donor genes based on sequence identity at homologous positions to the mutated locus.

ⁱⁱⁱ B6=C57BL/6; C=BALB/c; B10=C57BL/10

^{iv} Skin graft reaction. G=Gain; L=Loss

^v The number of consecutive, identical nucleotides between the mutant recipient gene and the donor gene in the region of the substitution.

- ^{vi} Aminoacid or nucleotide sequence.
- ^{vii} It might be that $bm1$ and $bm2$ do not represent independent mutants [6].
- ^{viii} The $bm3$ mutation was detected by a complicated skin grafting-scheme in the Yurlovo subline of the C57BL/6J strain [7]. This mutant was detected only when it already segregated in a colony of inbred mice. The precise generation that gave rise to the mutation could not be determined [10]. Additionally to the nucleotide substitutions leading to the aminoacid exchanges a single base substitution was found in the second intron [10].
- ^{ix} See reference [10].
- ^x Genomic Southern analysis was performed with spleen DNA from parental B6 and $H2K^{bm8}$ DNA. Hybridization was performed with an oligonucleotide specific for the K^{bm8} mutation. Two bands appeared, one corresponding to $Q4^b$ and the other one remained unidentified. It can therefore not be excluded that an uncharacterized MHC class I gene or another gene served as a donor in the K^{bm8} mutation [21].
- ^{xi} The K^{bm10} and $K1$ genes are identical in at least 79 consecutive nucleotides, beginning from codon 157 and extending into intron 3. Because the K^{bm10} sequence was determined by mRNA sequencing, the intron sequence was not obtainable, and the maximum recombination between K^b and $K1$ is not known [9].
- ^{xii} Hemmi et al. (1988) proposed several possible donors: K^b ; $Q4$; $Q7$; $Q8$; $Q9$; $Q10$; $T3^b$; ($T1^b$ - $T5^b$) [26].
- ^{xiii} A histocompatibility mosaic: six grafts of the (B10 X B10.D2) F1 female no. 513 were rejected by B10 mice; of grafts on B10.D2 mice four were rejected on days 12 or 13 but two survived over 40 days. The mutant carriers were found among (BC1 X BC1)F1 progeny [28].
- ^{xiv} Mutant was found segregating in the B6Kh colony in the Netherlands when tests were performed for homogeneity, and a coisogenic line has been developed (C.J.M. Melief, et al., unpublished results as cited in [14]. See also [12]).
- ^{xv} The mutant was lost prior to structural analysis but was included into the „bg“-group based on skin graft compatibility [12][29].
- ^{xvi} The mutant was lost prior to structural analysis but was included into the „bg“-group based on skin graft compatibility [12][13][29].
- ^{xvii} See reference [10].
- ^{xviii} Already segregating in stock when identified [13].
- ^{xix} Detected by tumour transplantation [32].
- ^{xx} Starting with B10 mice a complex breeding pattern was employed involving B10.D2 mice during the syngeneic tumour transplantation protocol, which was used to detect H2 mutants [32].
- ^{xxi} detected by tumour transplantation [32].
- ^{xxii} Starting with B10 mice a complex breeding pattern was employed involving B10.D2 mice during the syngeneic tumour transplantation protocol, which was used to detect H2 mutants [32].
- ^{xxiii} From cosmid hybridization analysis and comparison to nucleotide sequences, it was predicted that $Q5^b$; $Q6^b$; $Q7^b$; $Q8^b$; $Q9^b$; $Q10^b$ and $T12^b$ could have served as potential donors for codons 97 and 99, and $Q2^b$ as well as $T5^b$ for codon 103. As for K^{bm3} and K^{bm23} a complex gene conversion event involving two donors is proposed [34].
- ^{xxiv} This mutant was detected by the loss of affinity of a monoclonal antibody instead of skin grafting. The mutant appeared to be identical to the parental K^b by skin grafting and serological analysis [35].
- ^{xxv} The potential donor gene in this gene conversion-like event cannot be identified because several genes present in the parent such as K^k , D^b , $Q2^k$, $Q4^k$ and $Q4^b$ contain the same nucleotide configuration as K^{bm29} at position 89.
- ^{xxvi} Induced mutant. Arose as an offspring from a B10.D2 father which had been treated with DES [44].
- ^{xxvii} The $Dm1$ gene is a hybrid of the 5' end of the D^d gene and the 3' end of the L^d gene, with the region in between the two genes deleted [46]. Presumably represents an unequal crossover.
- ^{xxviii} Already segregating in stock when identified [29]. Presumably represents an unequal crossover.
- ^{xxix} Mapped to H2 complex but not assigned to a specific locus, as the mutant got lost due to breeding difficulties. The K , $I-A$, and D loci were excluded by complementation testing and the L locus is proposed to be most likely involved in this event [29].
- ^{xxx} Male offspring of an irradiated BALB/c male [49].
- ^{xxxi} Single nucleotide substitution. L^d contains same sequence at homologous position [50].
- ^{xxxii} Single nucleotide substitution. $T3^d$ and $T18^d$ contain same sequence at homologous position [50].
- ^{xxxiii} Father had been treated with ENU [52]. The $H2^{dm6}$ mouse strain got lost after 10 generations of backcrossing onto the BALB/c background [53].
- ^{xxxiv} Due to breeding failure no further mapping data within the $H2^d$ haplotype was obtained [29].
- ^{xxxv} X-ray treatment applied to BALB/c father [29].
- ^{xxxvi} Preliminary data suggested that $H2L$ was involved, but the mutant was lost before further testing [29].
- ^{xxxvii} The Asn male of the parents was treated with DES and mated to A.Ca females at the age of 60-70 days [28].
- ^{xxxviii} The 3' boundary of the potential gene donation was not accurately defined because the sequence of $pH2III$ is identical to K^k at every codon 3' to the mutation site through codon 160, beyond which the $pH2III$ sequence was unknown [37].
- ^{xxxix} This mutant was found by syngeneic tumour transplantation. It was already segregating in the colony when detected [39][40].
- ^{xxxx} An apparent reverse mutation of K^{bm1} found in the course of complementation testing to map the $H2^{bm18}$ mutation. A female who should have been K^{bm1}/K^b had the phenotype K^b/K^b , by direct skin grafting and by the fact that all 13 of her progeny inherited an apparently normal $H2^b$ haplotype from her [13].

1.3. In vitro analysis of recombinatorial events in mammals

The study of meiotic recombination mechanisms in mammals faces several constraints (reviewed in (Pittman and Schimenti 1998)). A major obstacle is the low frequency of events. Thousands of animals have to be screened before recombinations are found, and to obtain statistically significant numbers becomes a gigantic endeavour. Such an approach was used for mouse MHC genes, as described above. On the other hand, the formation of male and female gametes is preceded by extensive phases of mitotic expansion. When the analysis focuses on progeny or sperm pools, mitotic recombination events in primordial germ cells or gonadal cells cannot be distinguished from meiotic events, as is the case for the MHC mutations uncovered in laboratory mice. Finally, segregation patterns in individual meiosis cannot be followed as it is not possible to recover all the products of a meiosis in the mammalian germ line. This is in contrast to fungi in which gene conversion was first examined (Orr-Weaver and Szostak 1985). As a consequence gene conversion becomes indistinguishable from a double crossover, which has forced cautious descriptions of the process under the designation of gene conversion-like event or microrecombination.

Because mammalian meiosis is not easily accessible for examination, most studies of mammalian recombination have looked at mitosis in cultured cells. The strategy most commonly employed for the study of homologous recombination involves monitoring the reconstruction, via recombination between two artificial defective sequences, of a gene encoding a selectable marker. This approach has been used more extensively on extrachromosomal substrates but intrachromosomal recombination and gene targeting have also been assessed (reviewed in (Bollag *et al.* 1989)). More recent studies often rely on the expression of rare-cutting endonucleases to introduce DSBs into very specific locations, thus inducing homologous recombination above spontaneous levels (reviewed in (Jasin 1996)). In this case, recombinational repair of DSBs is used as a magnifying glass to look at homologous recombination.

Together these and other studies have indicated that, contrary to yeast which mainly repairs DSBs through conservative homologous recombination, mammalian cells have a very efficient nonhomologous end-joining (NHEJ) repair pathway which appears to be preferentially used. Moreover, homologous recombination in mammals can be either

conservative or nonconservative. The former relates to the DSBR model of yeast recombination, whereas the latter is best explained by the single-strand annealing (SSA) method. SSA involves degradation of DNA ends by a single-strand exonuclease to expose complementary single strands in two homologous regions, which then anneal to form an apparent crossover product. This model is mainly used to explain certain gene targeting and intrachromosomal events, as it would lead to accentric and/or dicentric products if it were to occur between unlinked chromosomal repeats. The next paragraphs briefly review studies designed to look specifically at mitotic intrachromosomal homologous recombination through conservative mechanisms, bearing in mind the laboratory mouse MHC mutants discussed above are hypothesised to result from segmental exchanges between paralogous loci in *cis* occurring during the mitotic phase of primordial germ cell expansion.

1.3.1. Intrachromosomal recombination

Rates of spontaneous intrachromosomal recombination (ICR) typically fall in the range of 10^{-6} - 10^{-5} upon correction for copy number. Interestingly, this value is only one order of magnitude below estimates of overall conversion rates in MHC genes. Line-to-line variation in such transgenic studies is assumed to derive from position effects reflecting properties specific to the chromosomal region in which the recombination substrate integrates such as chromatin structure and transcription activity (Bollag *et al.* 1989). Homologous recombination between tandemly repeated chromosomal sequences can be stimulated by a DSB within the region of homology from tenfold to several orders of magnitude (Brenneman *et al.* 1996; Liang *et al.* 1996; Sargent *et al.* 1997; Taghian and Nickoloff 1997; Donoho *et al.* 1998; Liang *et al.* 1998).

Relative amounts of homologous and nonhomologous recombination in the repair of DSBs within repeated sequences are somewhat controversial; however, in some cases recombinational repair appears to account for a significant fraction of total repair events (Liang *et al.* 1996; Sargent *et al.* 1997; Liang *et al.* 1998). Intrachromosomal homologous recombination proceeds either by SSA or DSBR, though it has been suggested that the latter may predominate (Taghian and Nickoloff 1997). In conservative ICR, gene conversions have been found to predominate relative to reciprocal exchanges (Liskay *et al.* 1984). On the other hand, gene conversion can accompany crossover during ICR, stressing the potential mechanistic relationship between the two processes (Bollag and Liskay 1988).

Intrachromosomal gene conversion involves the transfer of contiguous tracts of information of varying lengths (Liskay and Stachelek 1986). The minimal length of perfect homology required for efficient ICR in mammalian cells appears to be between 134 and 232 bp (Waldman and Liskay 1988). Shorter homology arms seem to be required for conversion events within MHC genes, as documented also for conversions in the variable region of chicken Igs.

1.3.2. Regulation of homologous recombination in mammalian genomes

In complex organisms with complex genomes, the action of homologous recombination is predicted to be prone to deleterious consequences. Thus, it is speculated that homologous repair might be favored only when sister chromatids are available and is replaced by NHEJ once this is not the case (Liang *et al.* 1998). Such a strategy would bypass the potential of allelic recombination for reduction to homozygosity (Moynahan and Jasin 1997) and the risk of major genomic rearrangements from ectopic recombination between repetitive elements. Several levels of control may have been developed to maintain the integrity of mammalian genomes. Sequence divergence suppresses recombination. Repeats on heterologous chromosomes or at distant locations on the same chromosome recombine at reduced frequencies. And when interchromosomal recombination does occur, crossover events are repressed (Richardson *et al.* 1998). The lack of crossovers may be a reflection of the predominance of short tract gene conversion events in which recombination does not extend to distances long enough to permit crossover events (Taghian and Nickoloff 1997; Richardson *et al.* 1998). Gene targeting experiments have detected a decline in recombination frequencies of up to sixfold for 0.8 to 1.2% sequence divergence (Elliott *et al.* 1998). In mouse cells, intra- and extrachromosomal recombination are differentially affected by sequence heterology in that 19% bp mismatch between recombining sequences reduces the former by a factor of more than a 1000, whereas the latter is reduced only by a factor of 3–15 (Waldman and Liskay 1987). ICR rates are determined primarily by the length of the longest stretch of uninterrupted homology rather than by the overall amount of heterology between two sequences. Single-nucleotide heterologies are most probably able to disrupt the minimal efficient recombination target. However, once recombination is initiated within stretches of perfect homology, it can propagate through adjacent regions of heterology, suggesting that sequence divergence exerts its impact on early recombination steps (Waldman and Liskay 1988). In frame with these results, sequence divergence should preclude recombination between MHC genes and in the chicken immunoglobulin loci (see 1.1. - Gene conversion as a

useful diversifying machine in MHC genes). Interestingly, it has been proposed that the introduction of DSBs might alleviate homology length requirements (Elliott *et al.* 1998; Taghian *et al.* 1998). Suppression of homologous recombination by sequence divergence between recombining substrates is at least to some extent mediated by the MMR system, since mutations in MMR genes can partially overcome some of the barrier to recombination between homologous but not identical DNA sequences (de Wind *et al.* 1995). This observation suggests the possibility that down regulation of MMR could be required to elude the vigilance against heterologous recombination.

1.4. In vitro analysis of gene conversion events in MHC genes

The detection of gene conversion mutations in tissues is difficult. From the mouse skin grafting data summarized above, the overall (i.e., sequence nonspecific) mutation rate was about 10^{-4} per gamete. It seems clear that at this frequency a polymerase chain reaction (PCR)-based analytical approach is obligatory for the detection of mutant segments. The nature of the problem itself, however, presents peculiar difficulties for such an approach. The very class of event to be detected, namely recombination between two homologous genomic sequences, is precisely the class of sequence the PCR is best adapted to generate artifactually, through “jumping” or false annealing of primers or partial extension products to homologous targets in the genomic DNA pool (Dieffenbach and Dveksler 1995). Such artifacts are indistinguishable from true recombinational events. Recently, however, three independent PCR-based systems have been developed which circumvent these obstacles and have generated valuable new information about the nature and frequency of conversion-like events in the MHC. All three systems analyse MHC class II genes, which possess more favourable properties than class I for such analyses. We further describe two experimental models involving transgenesis which may suggest future avenues for exploring this problem in the MHC.

1.4.1. PCR-based analysis of gene conversion events in the mouse *class II* gene *Ab*

Only one MHC class II gene conversion event has been identified in the mouse, namely *H2^{bm12}*. In this event, a segment of the second exon of the nonexpressed *E β^b* gene was transferred in *cis* to the expressed *A β^b* gene (see table 1-1) (McKenzie *et al.* 1979). Högstrand and Böhme (1994) designed a clever PCR-based system to detect similarly structured conversions in *trans* between the two expressed genes *E β^d* and *A β^k* in (dXk)F1 heterozygous

males. By placing one primer partially in the putative donor conversion tract and the other in the recipient gene, the authors were able to identify rare *trans* events in F1 cellular DNA with artifact frequencies well controlled by use of a comparable mix of DNA from the two homozygotes. The frequency of gene conversion between $E\beta^d$ and $A\beta^k$ was 2×10^{-6} per DNA molecule for sperm cells and more than two orders of magnitude lower in somatic tissue (liver cells). Sequencing revealed minimal conversion tract lengths of 13 to 104 nucleotides. Since one of the oligonucleotides primed within the presumed donor sequence, only one breakpoint of the recorded events could be inferred. Högstrand and Böhme (1997, 1998) have subsequently exploited the same approach to analyze a number of aspects of the MHC gene conversion process discussed above. Because event frequencies did not alter in spermatogenic cells of different stages, the authors proposed that gene conversion in the MHC in the male germ line may be a premeiotic event. In a modification of the assay to detect *cis* events, the authors suggest that different haplotypes may have different intrinsic event frequencies. However, by losing the key mixed parental DNA control on artifactual events, this result is perhaps of lower significance for the time being. Mutagenic agents known to cause double-strand breaks (γ -irradiation and the chemical etoposide) induced a significant increase in *trans* conversion-like events (Hogstrand and Bohme 1999). Despite evident limitations in this system, the Högstrand and Böhme assay has provided the first novel data concerning gene conversion events in the mouse MHC since the skin grafting assays.

1.4.2. PCR-based analysis of gene conversion events in the human class II gene *DPBI*

The extraordinary combinatorial character of polymorphism in the *DPBI* alleles of human has stimulated two attempts to document novel gene conversion events at this locus in human male germ cells. In one approach, a specific PCR was performed for exon 2 of *HLA-DPBI* on sorted single sperm derived from a heterozygous individual (*DPBI**0301/0401), taking advantage of the six distinct polymorphic regions as markers for gene conversion-like events (Huang *et al.* 1995). Mutations were detected by direct sequencing of PCR product. Of 800 sperm analyzed, one possible gene conversion-like event was detected in which the allele *0301 served as a donor towards the allele *0401. The tract length was at least 61 bases and the event was accompanied by a single untemplated base pair substitution. The single gene conversion-like event generated the *DPBI**02012 allele. All obvious sources of contamination with this allele were excluded. In three further samples, unrepaired heteroduplex DNA was detected, reminiscent of postmeiotic segregation in yeast. For the time being, this peculiar observation stands alone for mammalian genes. In the second approach,

exon 2 of *DPBI* of heterozygous individuals was analyzed by allele-specific PCR in pools of 50 sperm cells followed by differential oligonucleotide typing specific for polymorphic segments of the other allele (Zangenberg *et al.* 1995). Nine presumed gene conversion-like events were recovered from a total of 111 675 haploid genomes. The maximum tract length was in one case 59 bases, in all other cases around 130 bases. Of the nine detected new *DPBI* sequences in this experiment, five were identical to the known *DPBI*2301* allele. Despite admirable controls, the possibility of laboratory contamination could not be conclusively excluded. In two other cases, the variant sequences were characterized by exchanges of five nucleotides spanning four codons at the left end of the amplicon, so reciprocal recombination could not be excluded. The two remaining events created novel alleles through single base pair substitutions. There remains some doubt as to how many of the events recorded in this analysis are indeed novel gene conversions (Howard 1995). However, at face value, the results record an overall rate of approximately 9×10^{-5} per haploid genome. It is not clear whether this is discrepant with the $1/800$ (1.2×10^{-3}) events recorded for the same gene by Huang *et al.* (1995).

1.4.3. *H2 class I* gene conversion studies in transgenic yeast

In an attempt to bring the power of yeast genetics to bear on the MHC gene conversion problem, Wheeler and coworkers (1990) cloned the mouse class I genes *H2-D^d* and *H2-K^k* into the yeast genome. Haploid yeast strains, each of which carried one of the two class I genes, were crossed and a tetrad analysis performed on the resulting sporulated diploids. In 6% of meioses, short tract gene conversion events occurred resembling the mouse mutant class I sequences. In 5% of analyzed meioses, long tract “yeast-like” conversions could be detected. One of the short tract events was accompanied by a nontemplated single base substitution, a phenomenon also noted by Huang and coworkers (1995) (see above), and in some of the previously analyzed mouse mutants (see table 1-1). Furthermore, the recorded gene conversion events were not randomly distributed along the *D^k* and *K^k* genes but clustered within these genes. Both genes acted equally as both donors and recipients. This imaginative experiment awaits further information from the mouse itself before its implications for the mouse system can be fully assessed.

1.4.4. Gene conversion in tandem gene constructs in the germline of transgenic mice

The duplication-divergence model of genome evolution is weakened by the homogenizing power of interlocus gene conversion. In an attempt to assess the significance of

this theoretical problem experimentally, transgenic mice were developed carrying a tandem *lacZ* construct in which both members were disabled such that a short-tract interlocus gene conversion event could correct one of the partners and release enzyme activity detectable in situ (Murti *et al.* 1992). Expression of the recipient gene was driven by a spermatid-specific mouse promoter. Approximately 2% corrected spermatids were recorded, independently of transgene copy number and relative orientation of the recipient and donor gene. Apparently, both meiotic and mitotic events occurred. Although these important results were discussed in the context of the interlocus gene conversion mutations of the mouse MHC, their relevance to this matter is not yet clear. They establish experimentally that the male germ line is “conversion competent” at both mitotic and meiotic divisions, but the high frequency of events suggests that an integrated transgene may not reflect the properties of natural genes. The abnormal property of the tandem transgene which encourages the high frequency of events is unknown. Certainly, if generalized, such conversion event frequencies present severe difficulties for the duplication-divergence model (Murti *et al.* 1992; Schimenti 1999). Despite such reservations, it is clear that related experimental approaches could contribute further to the MHC gene conversion problem and allow systematic analysis of the mechanisms underlying recombinational processes in mammals (Hanneman *et al.* 1997).

1.5. The present project

The spontaneous histocompatibility mutants found in the mouse excited much interest, and their subsequent characterisation at the amino acid and nucleotide levels has stimulated speculations about the nature and evolutionary relevance of the mutational processes operating on MHC genes, as expressed in the gene conversion hypothesis of MHC evolution. However, so far no general experimental approach has been described for direct molecular analysis of mutational events in different MHC loci, and thus most of the original questions remain unanswered.

Such a mutation detector would allow statistically meaningful measurements of mutation rates in wild type and manipulated mice, helping to clarify the status of the purported sex, haplotype, and locus differences. Manipulated mice could be used to investigate the sequence specificity and context dependence of the mutational process as well as the influence of genomic background. Additional information could be obtained on mutation domains, homology requirements, directionality, and donor preference. Finally, it would be of interest to time the occurrence of gene conversion events, distinguishing not only

between mitotic and meiotic processes but also between interlocus and interallelic exchanges. Only such data would enable a decision in favour of weak or strong forms of the gene conversion hypothesis.

The purpose of the work presented here was to establish and characterise a mutation detection system for the *H2 K^b* gene. A target-gene enrichment step, called M450^{IZ}-*lacOp* rescue, is described for the extraction of DNA sequences which will subsequently be examined for mutations. This step serves to suppress the risk of introducing in vitro artefacts during PCR steps of the mutation analysis. Furthermore a mutation detection assay is described which consists of a highly specific PCR-step (LNC-PCR) performed on cells which is followed by a sequence specific oligonucleotide probing assay (SSOP) for the detection of *bg*-like mutations (see table 1-1.). Both the M450^{IZ}-*lacOp* rescue and the LNC-PCR/SSOP assay are evaluated with respect to power and feasibility. Finally our approach will be discussed with respect of value for the detection of rare mutations in germline *H2* genes, and will be compared to other modern mutation detection techniques.

2. MATERIALS AND METHODS

2.1. Generalities

Protocols or techniques which are not described in detail here, are done according to Ausubel or Coligan (Coligan *et al.* 1997; Ausubel *et al.* 1998), or are explained and discussed in results.

Chemicals were bought in p.a. quality from Sigma, Merck, or Roth. Radiochemicals were ordered from Amersham. Enzymes were ordered from New England Biolabs or Roche. The water used for all buffer solutions, and solutions used in assays was deionised and purified water (Seral™). pH-values are given for room temperature unless otherwise specified. As a reference marker for DNA-agarose gels the kb DNA Ladder (indicated in the figure legends as KBL) from Gibco BRL was used unless specified otherwise.

2.2. Molecular Biology Methods

2.2.1. Buffer exchange of DNA samples using self assembled Sephadex G-25 spin columns

To transfer DNA samples from one buffer into another, into 1ml syringes (Becton Dickinson) a glass ball of 1mm diameter (Faust) was placed into the outlet. With the aid of a pipette Sephadex G-25 (DNA Grade; superfine; Pharmacia Biotech) which had formerly been equilibrated with the appropriate buffer, was applied to the syringe. The syringe was centrifuged to get rid of excess liquid. Subsequently the DNA sample was applied and centrifuged again, this time into a 1.5ml reaction tube.

2.2.2. DNA-Sequencing

Plasmid DNA was sequenced using the *ABI Prism*[®] *BigDye*[™] Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems). The method is based on the dideoxy-chain termination reaction (Sanger *et al.* 1977), using fluorescently labelled dideoxynucleotides. In short 1µg plasmid DNA went together with 3.2pmol primer and 6µl *BigDye*[™] terminator ready reaction mix (ABI) in a total volume of 20µl into cycle sequencing (96°C, 5min.; 25X[96°C, 30s; 50°C, 15s; 60°C, 4min.]; 4°C, ∞). After EtOH precipitation

(Maniatis *et al.* 1989) the sequencing reactions were pelleted, dried, and handed over to Rita Lange who analysed them on an automated sequencer (ABI 373A).

2.2.3. Site directed mutagenesis using the "QuikChange™ XL Site-Directed Mutagenesis" protocol

Site directed mutagenesis was carried out using the the "QuikChange site-directed mutagenesis" protocol proposed by Stratagene (www.stratagene.com/mutagenesis/quikchngxl.htm). Miniprep plasmid DNA can be used for the assay. The basic procedure starts with a supercoiled, dsDNA vector with an insert of interest and two synthetic oligonucleotide primers containing the desired mutation. The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by Pfu DNA polymerase. On incorporation of the oligonucleotide primers, a mutated plasmid containing staggered nicks is generated. After temperature cycling, the product is treated with *Dpn* I. The *Dpn* I is used to digest the parental DNA template and select for the synthesized DNA containing mutations. Since DNA isolated from most *E.coli* strains is *dam* methylated, it is susceptible to *Dpn* I digestion, which is specific for methylated and hemimethylated DNA. The nicked vector DNA incorporating the desired mutations is then transformed into *E.coli*.

In short in a 50µl Volume 25ng plasmid DNA were mixed with 2.5U Turbo Pfu (Stratagene); 16pmol of each mutagenesis primer (for design see Stratagene instructions); and 0.2mM dNTP in 1X Pfu buffer (Stratagene). The cycling was done as follows: 95°C, 30"; 12X[95°C, 30"; 55°C, 1'; 68°C, 8'40"]; 4°C,∞. To the sample 10U of *Dpn* I (New England Biolabs) were added, and the sample was incubated for 1h at 37°C. 1µl of the PCR product was used for transformation of *E.coli*.

2.2.4. Electrophoretic Mobility Shift Assay (EMSA)

In principle the protocol proposed by Fried and Crothers was applied (Fried *et al.* 1981). A 6% polyacrylamide gel (running buffer 1XTBE) was prerun at 220V for 1h. Per sample (20µl) 1000cpm of the radiolabelled target sequence were mixed carefully with ca. 5ng LacIZ and 20ng dIdC (Roche) in REB (Spassky *et al.* 1984), and incubated for 30 minutes at RT, before loading onto the gel. The gel was run at 220V (34mA) for about 3hours, under cooling conditions. It was transferred onto 3MM paper (Whatman) and dried under vacuum at 80°C before exposure on a phosphoimager screen .

2.2.5. Capillary Electrophoresis (CE)

A CE-system set up by Kirstin Hebenbrock in our laboratory was used (details are described in (Martinson 1996), and for general information see (Landers 1993)). In short 25cm long CS-fused-silica capillaries (inner diameter: 50 μ m; CS-Chromatographie Service) were employed. As a matrix hydroxypropylmethylcellulose (HPMC) dissolved in 90mM TB was utilized. For loading of the DNA-samples electrokinetic migration was used (5000V (200V/cm; the current was monitored and held lower than 30 μ A to avoid boiling of the matrix, - time of loading and details as indicated in respective runs). For loading purposes the DNA had been transferred into deionised water. Running conditions were as follows: 200V/cm (the current was monitored and held lower than 30 μ A to avoid boiling of the matrix); 50°C; 1%HPMC; 0.05 μ g/ml of the intercalating dye Thiazole Orange (Polysciences Inc.). Data acquisition was done through an A/D-change module at 20Hz (Model 600; ABI). Data processing: Model 600 software (ABI). At the anode and cathode platinum wires were used, which were integrated into 0.2ml reaction tubes for sample injection purposes. Decontamination of the platinum wires was achieved by incubation in 50% HNO₃ pH1, followed by rinsing with 0.1M HNO₂ pH1 (this solution is not stable and has to be prepared freshly). Upon this treatment the amount of amplifiable DNA decreases within an hour to less than 10⁻⁴ (under detection limit) of the original amount (Kirstin Hebenbrock, unpublished data). The effect is based on the destruction of the DNA backbone in strong acids and deamination of dCTP to dUTP.

2.2.6. Radioactive labelling of probes

5'-end labelling (forward reaction) of [OPN/OPC]

End labelling was applied for the annealed and ligated oligonucleotides [OPN/OPC] (see below). Ca. 1 μ g of DNA (400pmol) were mixed in 1XT4 polynucleotide kinase buffer (New England Biolabs) with 250 μ Ci [γ -³²P]ATP (specific activity \geq 3000 Ci/mmol), and 20U T4 polynucleotide kinase. The reaction was incubated at 37°C for 60 minutes, and stopped with 1 μ l 0.5M EDTA. Extraction with phenol/chloroform and precipitation with EtOH was applied. The samples were counted in a scintillation counter (Beckman LS 5000TD), dissolved in TE to a final concentration of 50 000cpm/ μ l, and stored at 4°C.

Labelling of pW310 through fill in

10µg of pW310 (see below) were digested with *EcoR* I and *Hind* III in a volume of 50µl. Subsequently 100µCi [α -³²P]ATP and 10U Klenow fragment (New England Biolabs) were added. The reaction was incubated at room temperature for 60 minutes, and stopped by addition of 2µl 500mM EDTA. Extraction with phenol/chloroform and precipitation with EtOH was applied, and the *lacOp* carrying fragment isolated by recovery from a polyacrylamide gel by crush and soak (Maniatis *et al.* 1989).

Probes for Southern blotting

Radioactive labelling of probes for Southern blot analysis was done with the multiprime DNA labelling system from Amersham, according to the manufacturers instructions.

2.2.7. Synthetic oligonucleotides

In table 2 all oligonucleotides employed during this study are listed, along with the application they were used for.

designation	Sequence	Specificity	use
LNCV	5'-CAA AGC GCG CAA TTA ACC CTC ACT A	K ^{blNC}	PCR
LNCK	5'-GCC CAC ACT CGC TGA GGT ATT TCG T	K ^{blNC}	PCR
NLNCV	5'-CGC AAT TAA CCC TCA CTA AAG	K ^{blNC}	PCR
NLNCK	5'-GGT ACA TGG AAG TCG GCT AC	K ^{blNC}	PCR
Kb5	5'-GGA AGT CTC CTT ACC TGA TAA	H2 K ^b (intron 3)	PCR
Kb6	5'-GAC TGA AAC AAA TGA TAA TAG G	H2 K ^b (intron 3)	PCR
18S5'	5'-AAA CCA ACC CGG TGA GCT CCC TC	Mouse <i>18S rDNA</i> genes	PCR
18S3'	3'-AAT TAC AGG GCC TCG AAA GAG TCC	Mouse <i>18S rDNA</i> genes	PCR
ND	5'-TGA CTT GGG GTC GGA CTG GCG C	H2 D ^b	PCR
GC40D	5'-CGC CCG CCG CGC CCC GCG CCC GCC CCG CCG CCC CCG CCC GGT AGG CCT TGT AAT GCT CTG CAG	H2 D ^b	PCR
AK	5'-TGA AGT GGG GTC GGA CGG GCG A	H2 K ^b	PCR
GC40K	5'-GCG CCG CCGCGC CCC GCG CCC GCC CCG CCG CCC CCG CCC GGT AGG CCC TGA GTC TCT CTG CTT	H2 K ^b	PCR
Sry-pA	5'-TCT TAA ACT CTG AAG AAG AGA C	mouse <i>sry</i>	PCR
Sry-pB	5'-GTC TTG CCT GTA TGT GAT GG	mouse <i>sry</i>	PCR
Kb-oligo	5'-GTA CCA GCA GTA CGC CTA CG	H2 K ^b	SSOP
Q4-oligo	5'-CTA CAG AGA AAA GCA AGG CC	H2 Q4	SSOP
KBO58	5'-TGA AGC AGA GAG ACT CAG G	H2 K ^b	SSOP
BGOC	5'-ACG ACG GCC GCG ATT ACA	H2 K ^{bg}	SSOP
BGOT	5'-TAC CAG CAG TTC GCC TAC	H2 K ^{bg}	SSOP
OPN	5'-[6TFA-Aminolinker]-AAG CTT GTT TCT TGG AAT TGT GAG CGG ATA ACA ATT TAC GAG	contains <i>E.coli</i> <i>lac-operator</i>	NHS
OPC	5'-AGC TTC TCG TAA ATT GTT ATC CGC TCA CAA TTC CAA GAA ACA	contains <i>E.coli</i> <i>lac-operator</i>	NHS

designation	Sequence	Specificity	use
Nlado	5' -GTA GCC GAG CAG GCT CCT CAG GTC C	mismatch in exon II of <i>H2 K^b</i>	SDM
OLIX	5' -CTA CAA CCA GAG CAA GGG C	<i>H2 K^b</i>	SEQ
KStsts	5' -GAG AAT TTC CCT GAG GTA AC	pBbgMCB	SEQ
KStsta	5' -GGT TTC CCG ACT GGA AAG C	pBbgMCB	SEQ

Table 2. Oligonucleotides used in this study. PCR: Polymerase Chain reaction. SSOP: Sequence specific oligonucleotide probing. NHS: Affinity based NHS-column protein purification. SDM: site directed mutagenesis. SEQ: sequencing.

Annealing of oligonucleotides

Typically 1 to 2nmol of each oligonucleotide were annealed in a volume of 100µl 10mM Tris/HCl pH8.0. The reaction tube was placed in a 250ml water bath which was heated up to 75°C, and subsequently cooled down *ad. lib.* to room temperature.

Note: Based on the experiments performed by Dr. Ana Sousa in our laboratory in which she has shown that prolonged exposure to heat induces severe damage to DNA (Sousa 2000), I would no longer use the annealing protocol presented here. In a new protocol the oligonucleotides would be very briefly exposed to an elevated temperature in a PCR cycler, and the solution would be cooled down under controlled conditions. Also the oligonucleotide-solution would be buffered with Hepes instead of Tris (see also Results in 3.2.2.).

2.2.8. Plasmids and constructs

Listed are the plasmids and constructs used during this study.

PW310

Plasmid containing the ideal *lac-operator* sequence (O^{id}) on an *EcoR I/Hind III* fragment (Eismann 1989; Fickert *et al.* 1992). This fragment was used for electrophoretic mobility shift assays.

pF1

Vector expressing LacIZ constitutively (cloned and kindly provided by Brigitte von Wilcken-Bergmann). The fusionprotein LacIZ is identical to a fusion of Lac-repressor and β-galactosidase examined formerly by Brake *et al.* (Brake *et al.* 1978). The construct is based on pWB100 (Lehming *et al.* 1987).

pF4

As pF1 but containing a promoter which is about four times stronger than that used in pF1 (Brigitte von Wilcken-Bergmann, personal communication).

pBKb

Plasmid pBluescript KS (+) (Stratagene) containing the first three exons of the *H2 K^b* gene (Rita Lange, personal communication).

pBQ4

Plasmid pBluescript KS (+) (Stratagene) containing the first three exons of the *H2 Q4* gene cloned into the *Hind* III site of the multiple cloning site (Rita Lange, personal communication and (Martinsohn 1996)).

8-8PST

Contains a *Nru* I/*Pst* I fragment including the first three exons of *H2 K^b* cloned into the multiple cloning site of pBR322 (Gibco BRL) (Rita Lange, personal communication and (Martinsohn 1996)).

ND

Contains the first three exons of the *H2 D^b* gene (*Bam* HI/*Nco* I fragment) cloned into the multiple cloning site (*Bam* HI/*Hind* III) of pSP65neo (Rita Lange, personal communication)

pSP72

Standard cloning vector (Promega).

pHbapr-1

Mammalian expression vector containing 3 kilobases of the human beta-actin gene (Gunning *et al.* 1987).

2.2.9. Polymerase Chain Reaction

LNC-PCR on cells (optimised protocol)

The PCRs were performed in 96-well plates (Biozyme). Per well 50 or 100 cells were aliquoted in a volume of 5µl in PBS. The cells were lysed in 30µl volume in 1XPfu buffer (self made-see below) and 70µg/ml ProteinaseK (Merck) at 50°C for 1h after which inactivation of the Proteinase K was ensured by heating to 95°C for 10 minutes. The plates were briefly spun and 30µl PCR-mix was added (1XPfu buffer; 0.5µl Promega Pfu (ca. 1U);

5pmol LNCV/LNCK; 0.2mM dNTP). The PCR program: 95°C, 3min.; 30X[95°C, 30s; 64°C; 30s; 75°C, 90s]; 75°C, 10min.; 4°C, ∞.

18S-PCR

The amplification on 18S rDNA genes (Alberts *et al.* 1994) was employed as a standard for enrichment of the M450^{IZ}-lacOp rescue (see 3.1.5). An amplicon of 320 bp is generated.

The amplification was performed in 50µl volume with 2.5U Taq (produced in this laboratory), 1XTaq buffer (Gibco BRL), 2.5mM MgCl₂, 5pmol 18S5'/18S3'; 0.2mM dNTP for 35 cycles (95°C, 3min.; 35X[95°C, 30s; 67°C; 30s; 75°C, 60s]; 75°C, 10min.; 4°C, ∞).

PCR on mouse tail DNA

After the isolation of genomic DNA from mouse tails (see below), the DNA was diluted 1:100 in 10mM Tris/HCl pH8.0 and 5µl were amplified in a volume of 50µl (5U Taq (produced in this laboratory); 1X Taq buffer (GibcoBRL); 4.5mM MgCl₂; 5pmol Kb5 and Kb6; 95°C; 1min.; 35X[95°C; 30s; 55°C; 30s; 75°C; 1min]; 75°C, 10min; 4°C).

Pfu-buffers

In the course of optimisation of the LNC-PCR several Pfu's were used along with the Pfu buffers provided by the respective producer (Promega and Amersham). Additionally for the LNC-PCR on cells the following 10XPfu buffer was used, containing reduced amounts of detergent: 200mM Tris/HCl pH8.8; 100mM KCl; 100mM (NH₄)₂SO₄ 20mM MgSO₄; 0.25% Triton X-100 (Sigma); 1mg/ml BSA (Sigma).

2.2.10. Southern blot analysis

Southern blot analysis (capillary blotting onto nylon membranes), used in some of the evaluation experiments for the M450^{IZ}-lacOp rescue, was done following standard procedures (Ausubel *et al.* 1998). Details are explained in the presentation of the actual experiments.

Probes used

neo probe: *Bam* HI/*Sal* I fragment of the plasmid pEF β gal (Lisbeth Guethlein, unpublished data). Specific for a part of the neomycin resistance gene.

amp probe: *Bgl* I/*Sca* I fragment of the plasmid pSP72 (Promega). Contains a part of the β -lactamase gene.

The probes were labelled radioactively using the multiprime DNA labelling system from Amersham. Detection was carried out by exposing the filters on phosphoimager screens (BAS 1000 Bio-imaging Analyzer system (exposure times as indicated; Fujix).

2.2.11. Sequence Specific Oligonucleotide Probing (SSOP)

Non-radioactive labelling of oligonucleotides

The Gene Images 3'-End Labelling System (Amersham) uses Terminal Deoxynucleotidyl Transferase to tail oligonucleotides from any source with fluorescein-dUTP. The application results in a tail of 6 to 8 nucleotides. There is no need to purify the probe before hybridization. Detection relies on an anti-fluorescein antibody conjugated to alkaline phosphatase (AP). AP catalysed breakdown of the dioxetane substrate. The CDP-Star detection reagent produces a chemiluminescent reaction that is captured on film. Light output is rapid and continues for up to 5 days.

100pmol oligonucleotide were incubated with 1Xcacodylate buffer, 1 μ l fluorescein-dUTP, and 24U Terminal Deoxynucleotidyl Transferase in a total volume of 80 μ l, for 90 minutes at 37°C. After incubation the sample was immediately stored at -20°C.

The oligonucleotides in use for the *bg*-specific SSOP assay

designation	sequence	specificity	optimised temperature (hs-wash)
Kb-oligo	5' -GTA CCA GCA GTA CGC CTA CG	<i>H2 K^b</i>	58°C
KBO58	5' -TGA AGC AGA GAG ACT CAG G	<i>H2 K^b</i>	58°C
BGOC	5' -ACG ACG GCC GCG ATT ACA	<i>H2 K^{bg}</i>	58°C
BGOT	5' -TAC CAG CAG TTC GCC TAC	<i>H2 K^{bg}</i>	56°C

Table 2-3 The oligonucleotides employed for the *bg*-specific SSOP. hs-wash: high stringency wash after hybridisation.

Blotting procedure

The positively charged nylon hybridisation transfer membranes (Hybond-N+; Amersham) were shortly soaked in 2XSSC (20XSSC: 0.3M tri-Sodium-Citrate 2-hydrate; 3M NaCl; pH 7 to 8) before the transfer of the sample. The sample was transferred in a volume of 50µl to 100µl onto the membrane, in a 96-well manifold device (Minifold I SRC96; Schleicher & Schuell). Transfer was carried out through vacuum. An incubation in 150µl GS1 (denaturation; 0.5M NaOH; 1.5M NaCl) for 5 minutes followed. Neutralization was ensured by incubation in 150µl GS2 (1.5M NaCl; 0.5M Tris/HCl pH 8.0) for 5 minutes (this step is mainly necessary because the nylon membrane is sensitive to alkalic pH values). The membrane were shortly soaked in 4XSSC and dried at 75°C before further processing.

Hybridisation and detection

Prehybridisation was carried out in Hybridisation buffer (6XSSPE [30XSSPE: 4.5M NaCl; 0.3M NaH₂PO₄•H₂O; 30mM EDTA; the pH was adjusted to 7.4 with NaOH], 10XDenhardt's solution [100XDenhardt's solution: 2.0g Bovine Serum Albumin (Sigma); 2.0g Ficoll™400 (Pharmacia Biotech); 2.0g Polivinylpyrrolidone (Sigma)], 20µg/ml fragmented herringsperm DNA (Roth)) at 42°C for 60 minutes. Hybridisation was done for 10 to 12 hours at 42°C in hybridisation buffer to which 30pmol of fluorescein-dUTP labelled oligonucleotide were added.

Washing was carried out three times for 10 minutes in 2XSSPE/0.1%SDS. The first two washing steps were carried out at room temperature, the third washing step (high stringency washing step) was performed at the T_m (melting temperature) of the oligonucleotide.

Blocking was carried out by incubation of the filters for 90 minutes at room temperature in buffer A (300mM NaCl; 100mM Tris/HCl pH9.5; the solution should be autoclaved to avoid background effects due to bacterial growth in the solution) 1/10 volume liquid blocking reagent (Amersham).

Anti-fluorescein antibody incubation was carried out at a dilution of 1:5000 of the antibody in buffer A plus 0.5% w/v bovine serum albumine (Sigma) for 90 minutes at room temperature.

Washing was carried 3 times at room temperature out in buffer A containing 0.3% v/v Polyoxyethylene-Sorbitan Monolaureate (Tween 20; Sigma).

Chemiluminescent detection was carried out by incubating the filters with CDP-Star detection reagent for 5 minutes at room temperature. The filters were sealed into plastic foil and exposed to X-ray films (Kodak X-OMat™ AR).

Stripping was carried out by incubating the filters five times in 300ml 0.1% SDS, that had been heated to boiling.

Evaluation

The exposed X-ray films were scanned (Agfa SNAPSCAN 1236), and imported into the software Adobe Photoshop 5.0. They were saved as “TIFF” files and imported into the program Tina (vs2.08). The intensity of the signals was measured [PSL – Photostimulated Luminescence] and the obtained values imported into Excel 2000 (Microsoft). After background subtraction, and division by 1000, the data was used for further evaluation.

2.2.12. Isolation of genomic DNA

Basically the protocol (non-phenol method) of Torres and Kühn was followed (Torres *et al.* 1997).

Isolation of genomic DNA out of ES cells (E14.1; LNC28.28)

10⁵ cells were harvested after trypsinisation and washing with 1XPBS, and pelleted in 1.5ml reaction tubes. The pellet was resuspended in 500µl lysis buffer (100mM Tris/HCl pH8.5, 0.2% SDS, 200mM NaCl, 100µg/ml Proteinase K (freshly added; Merck)). The mixture was incubated at 37°C for several hours and the sample carefully mixed from time to time by flicking the tube, until the solution became homogenous and viscous. The DNA was precipitated with isopropanol (Maniatis *et al.* 1989) and after washing with 70%EtOH the pellet was taken up in 50 to 100µl Tris/HCl pH8.0.

Isolation of genomic DNA out of mouse tails

To about 0.5cm mouse tail 600µl lysis buffer were added (100mM Tris/HCl pH8.5, 0.2% SDS, 200mM NaCl, 400µg/ml Proteinase K (freshly added; Merck)) in a 1.5ml reaction tube. The sample was incubated in a shaking water bath at 55°C for about 10 to 12hours and spun down at 20 000g for 5 minutes. The supernatant was precipitated with isopropanol (Maniatis *et al.* 1989)and after washing with 700µl 70%EtOH the pellet was taken up in 150 to 200µl Tris/HCl pH8.0.

2.2.13. M450^{I_Z}-*lacOp* rescue: standard protocol

The M450^{I_Z}-*lacOp* rescue went through many optimisation steps which are described in this study. What follows is an optimised version for future reference.

Arming M450 Dynabeads[®] with α - β -galactosidase antibody

150 μ l of M450 Dynabeads[®] (sheep antimouse IgG coated; ca. 4×10^8 beads/ml; Dynal) were pelleted down on a magnetic stand. The beads were washed one time with 1 volume of PBS. 200 μ l of 200 μ g/ml α - β -Galactosidase antibody (Promega) in PBS were put onto the pelleted beads, which was followed by incubation at room temperature or 4°C for 90 minutes on a rotator. Finally the beads were washed four times with 200 μ l 1X rescue binding buffer (RBB; 10mM Tris/HCl pH 7.0; 10mM MgAc; 150mM KCl; 5% v/v glycerol).

The M450^{I_Z}-*lacOp* rescue protocol

The protocol described here started from 10^5 LNC28.28 cells, which were lysed (Torres & Kühn, - see above). The genomic DNA was digested and after phenol extraction transferred through Sephadex spin columns into 1Xrescue binding buffer (RBB; 10mM Tris/HCl pH 7.0; 10mM MgAc; 150mM KCl; 5% v/v glycerol). The assay was performed in 200 μ l reaction tubes, in a volume of 50 μ l. After the Sephadex buffer exchange step 1 μ l 0.2% BSA (New England Biolabs) and 1 μ g LacI_Z were added. The sample was carefully mixed and incubated at room temperature for 90 minutes on a rotator (at this stage it is helpful to introduce a tiny bubble into the solution with the aid of a pipette, to ensure gentle but thorough mixing of the sample during rotation). The mixture was added to pelleted M450- α - β -galactosidase armed beads (20 μ l; the supernatant was taken off), and incubation followed as above. The beads were pelleted and the supernatant taken away. The sample was rinsed three times with 1 Volume 1Xbinding buffer. Elution was performed with 20 μ l IPTG-elution buffer (10mM Tris/HCl pH 7.0; 150mM MgAc; 5% v/v glycerol; 0.01% BSA, 10mM IPTG (Roche)), and incubation for 30 minutes as above. The beads were pelleted and the supernatant taken away and transferred through Sephadex spin columns (Sephadex G-25; superfine; DNA-grade; Pharmacia) into 10mM Tris/HCl pH8.0, 0.1mM EDTA. The samples were stored at -20°C.

Evaluation of the M450^{IZ}-*lacOp* rescue by southern blotting and densitometry

In a part of the experiments on the M450^{IZ}-*lacOp* rescue, Southern blotting followed by densitometry measurement was used. Detection was carried out by exposing the filters on phosphoimager screens (BAS 1000 Bio-imaging Analyzer system (exposure times as indicated; Fujix). The resulting files were imported into the program Tina (vs2.08). The intensity of the signals was measured [PSL – Photostimulated Luminescence] and the obtained values imported into Excel 2000 (Microsoft). After background subtraction the data was used for further evaluation. Enrichment factors were calculated as follows: Per sample the signal intensity of the target plasmid NKrQ (i.e. the plasmid that should be enriched for) was divided by the signal intensity of a contaminating sequence (i.e. pSP72, pHbapr1). Each ratio obtained was subsequently divided by the ratio calculated for a standard sample (i.e. a sample consisting of target and contaminant, but did not go through the M450^{IZ}-*lacOp* rescue protocol).

2.3. Biochemical Methods

2.3.1. SDS-Polyacrylamide Gel electrophoresis for the separation of proteins

Applied was one-dimensional discontinuous gel electrophoresis under denaturing conditions (presence of 0.1% SDS; “Laemmli gel method”). This ensures separation of proteins based on their molecular size (Coligan *et al.* 1997). The separating gel (7.5% for LacIZ; 12.5% for Pfu) was overlaid by a 5% stacking gel (preparation as suggested in Current Protocols in Protein Science (Coligan *et al.* 1997)). The protein samples were taken up in 1XSDS sample buffer (60mM Tris/HCl pH6.8; 2.3%w/v SDS; 5% glycerol; 0.1mg/ml bromphenol blue (Sigma)). For the reduction of disulfide bonds β -mercaptoethanol (50 μ l/ml) was added freshly to the samples. The samples were heated up to 95°C for 5 to 10 minutes before loading onto the gels. The gel runs (running buffer: 25mM Tris; 190mM Glycin; 0.1% SDS) were performed at 20mA for about 2hours in Mini-Protean[®] II gel chambers (BioRad).

2.3.2. Coomassiestaining of SDS-Polyacrylamide Proteingels

After the runs the gels were stained in Coomassie staining solution (1.25g Coomassie Brilliant Blue R-250 (Sigma), 500ml methanol; 100ml acetic acid; 400ml H₂O) for about 20 minutes at RT. Distaining was achieved by repeated incubation of the gel in big volumes of

tap water and boiling in a microwave oven (Hervieu 1997). Subsequently the gels were dried under vacuum at 80°C for 2 hours.

2.3.4. Spectrophometric Determination of Protein Concentration

Spectrophometric Determination of Protein Concentration was done in a standard Spectrophotometer (Beckman DU-62), using a 1cm path length quartz cuvette. The absorbance was either measured at 280nm (absorbance of aromatic amino acids in protein solutions) or at 205nm (absorbance by the peptide bond) (Coligan *et al.* 1997). For Pfu the absorbance was also measured at 278nm ($A_{278}=0.74$ for 1mg/ml; (Lu *et al.* 1997)).

2.3.5. Purification of Proteins

The purification of the proteins Pfu and LacIZ was subject of experiments and is as such explained in the results. Short descriptions of protocols which are not explained explicitly in results are given below.

Charging of NHS-activated HiTrap™ affinity columns with annealed oligonucleotides

NHS-activated HiTrap™ affinity columns (Pharmacia Biotech) contain NHS-activated Sepharose which is designed for the covalent coupling of ligands containing primary amino groups. The column bed gel carries 6 atom spacer arms attached to the matrix by epichlorohydrin and activated by N-hydroxy-succinimide. We used these columns to couple DNA fragments which contained the *lac-operator* sequence, and to which a primary amino group had been linked, to the matrix to purify LacIZ (see results). Here a short description of the coupling procedure is given.

In principle the instructions of the manufacturer were followed. In short the columns was activated with an acidic solution (1mM HCl). Afterwards the DNA fragments got loaded in coupling buffer (0.4M NaHCO₃; 0.1M NaCl; pH8.3). After incubation the coupling buffer was replaced by blocking buffer A (0.5M MetOH-Amine; 0.5M NaCl; pH8.3), which was followed by a deactivation step (0.1M Acetate; 0.5M NaCl; pH4.0). This blocking/deactivation cycle was repeated three times. Finally the column was transferred into storage buffer (10mM Tris/HCl pH8.0; 1mM EDTA; 0.3mM NaCl; 0.04% NaN₃), and stored at 4°C.

Preparation of LacIZ raw extract

A 500ml bacterial culture (LB-medium) was inoculated with pF1 or pF4 transformed bacteria of the strain *BMH8117* [*F*⁺ $\Delta(lac-proAB)$, *thi*, *supE*, *gyrA*, λ] (kindly provided by the department Benno Müller-Hill) and grown at 37°C for 8 to 10 hours. The culture was pelleted and dissolved in 1ml/mg raw extract buffer (REB: 40mM Tris/HCl pH8.0; 100mM KCl; 0.1mM DTT; 1mM EDTA; 5% w/v glycerol; (Spassky *et al.* 1984)). The samples were sonicated (Labsonic^o; duty cycle 0.5; 3X30s; 30" pause) and centrifuged for 1h at 20 000g at 4°C. Finally the supernatants were snap frozen on a dry ice/EtOH mixture and stored at – 80°C.

Purification of Pfu using the Uno™ Q1 polishing column

The Uno™ Q1 polishing column (BioRad) is an anion exchange column, for continuous bed ion exchange chromatography. Unlike traditional columns which consist of a bed of packed beads or particles, each UNO column contains a polymer matrix that is completely homogeneous. This matrix is supposed to overcome problems associated with chromatography using beaded supports. It is offered as a late-stage tool to purify and concentrate dilute samples at the same time, subsequently to additional purification steps (see www.bio-rad.com). The column was used with a FPLC™ system (Pharmacia). Precycling and recycling was done according to the manufacturers instructions. The purification of Pfu was done on a sample that had formerly been purified with a phosphocellulose column (see Results). The aim was not primarily to obtain greater purity but to assess whether the specific activity of Pfu could be increased. About 500µg of Pfu were loaded onto the column and eluted with a KCl gradient (50mM Tris/HCl pH8.0; gradient: 0 to 500mM KCl over 30ml; flow rate 0.5ml/min). The eluted material was concentrated and transferred into storage buffer (Amicon; Centriprep-50 columns; storage buffer: 70mM Tris/JHCl pH8.0, 1mM DTT, 1mM DTT; 0.1% v/v IGEPAL; 0.1% v/v Tween20; 0.1mM EDTA). The activity of Pfu after this purification step was measured in appropriate PCR test series (see Results).

2.4. Cellular Biology Methods

2.4.1. Culture of ES-cells

The ES cell line E14.1 (Kuhn *et al.* 1991), and the transgenic ES cell line LNC28.28 (Gaby Vopper and Lisbeth Guethlein, unpublished data) were held in culture. The cells were

grown in 75cm² polystyrene tissue culture flasks (Sarstedt) in a cell incubator at 37°C and 5%CO₂. Medium: ES-cell medium (DMEM (high glucose, without sodium pyruvate; Gibco); 15% Fetal Calf Serum (FCS; Sigma); 1mM sodium pyruvate (Gibco BRL); 100IU/ml penicillin (Gibco BRL); 100µg/ml streptomycin (Gibco BRL); 2mM L-glutamine (Gibco BRL); 1Xnon-essential amino acids (Gibco BRL); 1.2ml/l leukaemia inhibitory factor (LIF (Williams *et al.* 1988); serves for maintenance of pluripotent state of the ES cells; source: supernatant from LIF-transfected CHO cells; kindly provided by department Klaus Rajewsky); 0.1mM β-mercaptoethanol (Merck). The cells were trypsinized and diluted every second to third day.

2.4.2. Freezing and storage of ES-cells

The cells were trypsinized, washed two times in 1XPBS, and taken up at a concentration of 10⁷ cells/ml in FCS (Sigma) containing 10% Dimethylsulfoxid (DMSO; Sigma). The cells were immediately stored at -80°C. Long-term storage was carried out in liquid nitrogen.

2.4.3. Preparation of spleen cell suspension

Mice were sacrificed by cervical dislocation, and the spleen removed. Single cell suspensions were prepared by gentle tearing of spleens in complete RPMI (RPMI 1640 (Gibco BRL), 10% fetal calf serum (Sigma), 50µM β-Mercaptoethanol (Merck), 100U penicillin/streptomycin (Gibco BRL)). Erythrocytes were eliminated by incubating the cell suspensions in hypotonic buffer (150mM NH₄Cl; 1mM KHCO₃; 0.1mM EDTA) for 2minutes at room temperature. Subsequently the suspension was transferred again into complete RPMI until further processing.

2.4.4. Cytofluorometric analysis

In short, cell suspensions were washed two times with DFN (1XDAB (500XDAB: 0.5M CaCl₂; 0.25M MgCl₂); 2% fetal calf serum (Sigma); 0.1%NaN₃), and collected at a final concentration of 5X10⁶ to 10⁷ cells/ml in DFN. About 2X10⁵ cells per well were transferred into round bottomed microtiterplates (96-well; Costar) in a volume of 25µl. The primary antibody was added at an appropriate dilution in DFN in a volume of 50µl. The plate was covered and incubated under gently shaking for 30minutes at room temperature. After adding 125µl of DFN the plates were gently centrifuged, and the supernatant carefully shaken off. This step was repeated three times. FITC labelled rabbit anti-mouse IgG antibody (Dianova)

was added in a volume of 50µl at a dilution of 1:200. Incubation was performed as above, as well as the subsequent washing steps. Finally the samples were resuspended in 200µl DFN and transferred into 1.5ml reaction tubes. Immediately before analysis on the FACScan (Becton Dickinson), 0.2µg/ml propidium iodide (Sigma) was added to the samples. Propidium iodide enters into dead cells (while it is excluded by living cells), and intercalates with DNA. It emits red fluorescence upon excitation, which allows to exclude the dead cells during the analysis (live gating). This is helpful to suppress background, since dead cells take up FITC-labelled antibodies. In table 2 the antibodies used for FACS analysis are listed.

H-2 specific antibodies					
panel	designation	isotype	recognizes	does not recognize	reference
1	H142-23	IgG2b	K ^{b,k,q,r,s}	K ^{d,l} D ^{b,d,l}	Horton et al. (1991)
2	28-14-8S	IgG2a	D ^{b,q} L ^d		Hämmerling et al. (1982)
3	SF1-1.1	IgG2b	K ^u	K ^{b,i,k,p,q,s,u} D ^{b,i,k,p,q,s,u}	PharMingen

Table 2. List of antibodies used for FACS analysis in this study.

2.5. Maintenance and characterization of 129^{svLNC}-mice

As described in the introduction, the major obstacle for the detection and analysis of gene conversion like events in the MHC of mice is their anticipated low frequency of about 10^{-4} (see introduction), combined with the presence of a high number of highly homologous genes on chromosome 17 (Trowsdale 1995). The rare occurrence of the events we are interested in makes a PCR step obligatory in order to lift those mutations to detectable levels. However in the context of our analysis a PCR-based approach represents a double edged sword: The PCR is, due to the character of events we want to analyse, namely the outcome of recombinational events between homologous sequences, prone to produce *in vitro* artefacts, through false annealing of primers and "jumping PCR"(Dieffenbach *et al.* 1995). These artefactual products risk to be indistinguishable from recombinational events that were generated *in vivo*. This is probably why until now only few published experiments deal with the detection of gene conversion like events in the MHC, despite the general interest in this process in the context of the evolution of the MHC and gene families in general (Schimenti 1994). Furthermore, apart from the "historical" and gigantic mutation screen during which more than 150 000 mice were tested for anomalous transplantation specificities (see introduction), the other attempts to detect gene conversion in mouse and human beings, were focused on such events occurring in MHC *class II* genes (Hogstrand *et al.* 1994; Huang *et al.*

1995; Zangenberg *et al.* 1995), a system which shows less complexity than MHC I (Trowsdale 1995) and is therefore experimentally more accessible.

However we felt that the true motivation to look for such events in the MHC of mice derives from the detected mutants in the skin grafting screen, which with the exception of *H2 K^{bm12}* (see table 1) occurred in MHC *class* I genes and within this group mainly within *K^b*. We wanted to assemble a mutation detection assay which would allow us to examine the *K^b* gene in germline derived cells. The first step in this assay should consist of an enrichment step for *K^b*-genes which would enable us to perform PCR in the absence of other MHC *class* I genes sequences which introduce the risk of producing *in vitro* artefacts. This enrichment is based on the highly specific affinity of the *E.coli lac-operator* to Lac repressor, and is presented in detail in section 3.1.

In this laboratory Gaby Vopper and Dr. Lisbeth Guethlein generated embryonic stem cells which carry in the third intron of the *H2 K^b* gene the *E.coli lac-operator* sequence (*K^{bLNC}*, see also Introduction). These cells were generated by homologous replacement techniques (“knock-in”) and served ultimately to create transgenic mice. The *K^{bLNC}*-mice represent an essential part in the whole mutation detection process as we had envisioned it. Primordial germ cells should be isolated from these mice. The *lac-operator* sequence should allow for physical extraction of the marked *K^b* gene from the rest of the genome (see 3.1). This in turn would permit to carry the enriched sequences over into PCR, which subsequently could be followed by other measures to enrich and detect mutations. Naturally this approach is not restricted to primordial germ cells, but can also be applied to any other cell type.

We shortly summarize in this section how the *K^{bLNC}*-mice were developed for use. Their BALB/c component derived from the original cross had to be reduced by backcrossing to 129/sv mice, using a specific PCR which allowed for the identification of *K^{bLNC}* positive mice. Furthermore we tested, whether the expression of *K^{bLNC}* was identical to that of the wildtype *K^b* gene. Finally a PCR was developed which should allow for an efficient and simple way to distinguish whether cells that are examined derive from female or male mice, a feature that is important in the context of the discussion whether gene conversion in mammals shows a sex bias (see introduction).

2.5.1. Purifying breeding of 129^{K^{bLNC}}-mice: monitoring by PCR

After having obtained chimeras the *K^{bLNC}* mice were repeatedly backcrossed to 129/sv mice to eliminate the Balb/c component. They are meanwhile in the 11th generation

and bred to homozygosity for the K^{bLNC} transgene. To identify the K^{bLNC} -gene we used the primer pair Kb5 and Kb6. Both primers lie in the third intron and frame the pBluescript sequence of the K^{bLNC} transgene (see figure 2-1). In the presence of the pBluescript sequence an amplicon of 840bp is created, while in its absence the amplicon has a length of 530bp. In figure 2-2 some examples of such PCR outcomes are depicted.

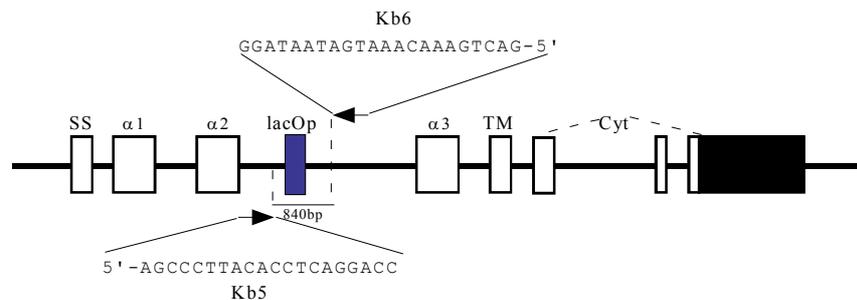


Figure 2-1. Position of primers Kb5 and Kb6 in the Kb and K^{bLNC} gene. Depicted is a scheme of the K^{bLNC} gene and the approximate position of Kb5 and Kb6. SS: signal sequence. $\alpha 1$; $\alpha 2$; $\alpha 3$: first three exons. TM: transmembrane region. Cyt: cytoplasmic tail. *lacOP*: pBluescript derived sequence containing the *E.coli lac-operator* sequence. (Source: Gaby Vopper and Lisbeth Guethlein; unpublished data).

2.5.2. Expression of K^{bLNC} on the cell surface in comparison to Kb

We wanted to confirm that the transgene K^{bLNC} is expressed to the same level as the wildtype K^b gene. For this purpose a heterozygous K^b/K^{bLNC} mouse of the third generation of breeding to 129/sv, was bred to a Balb/c female (K^d/K^d). Spleen cells were isolated from littermates being either K^{bLNC} or K^b positive as well as from the BALB/c mother and stained for K^b -, D^b -, D^d - cell surface expression with $H2$ -specific antibodies (listed in figure 6 D)) and indirect immunofluorescence in the FACScan. Ideally the ratio of expression levels of K^b to K^d should be the same in the littermates. As can be seen in figure 2-3 A) and table 3, the expression of K^{bLNC} is indeed at the same level as that of wildtype K^b .

	K^{bLNC}/K^d Geo Mean of $H2$ expression	K^b/K^d Geo Mean of $H2$ expression
K^{bLNC}	161	-
K^b	-	171
K^d	114	120
ratio (K^{bLNC}/K^d)	1.4	-
ratio Mean (K^b/K^d)	-	1.4

Table 3. Ratio of K^{bLNC}/K^d and K^b/K^d expression in two test animals. Listed are the geometric means (GeoMean) of the H-2 expression levels obtained by FACS analysis, as depicted in figure 6, and the ratios of the K^{bLNC} to K^d , and K^b to K^d expression.

2.5.3. Determining the sex of cells: *SRY*-PCR

A PCR was established which amplifies a 380bp fragment from the male specific *Sry* gene ((Kunieda *et al.* 1992) and for nucleotide sequence Washburn *et al.* (2001) Genbank Accession number AF337050). As can be seen in figure 2-4 this PCR works well. This *Sry*-PCR should facilitate the sex determination of the genital ridges of which primordial germ cells can be isolated. In principle sex determination is possible by visual inspection due to a sex specific morphology, but this needs some experience to be reliable ((Hogan *et al.* 1994) and Anne McLaren (personal communication)).

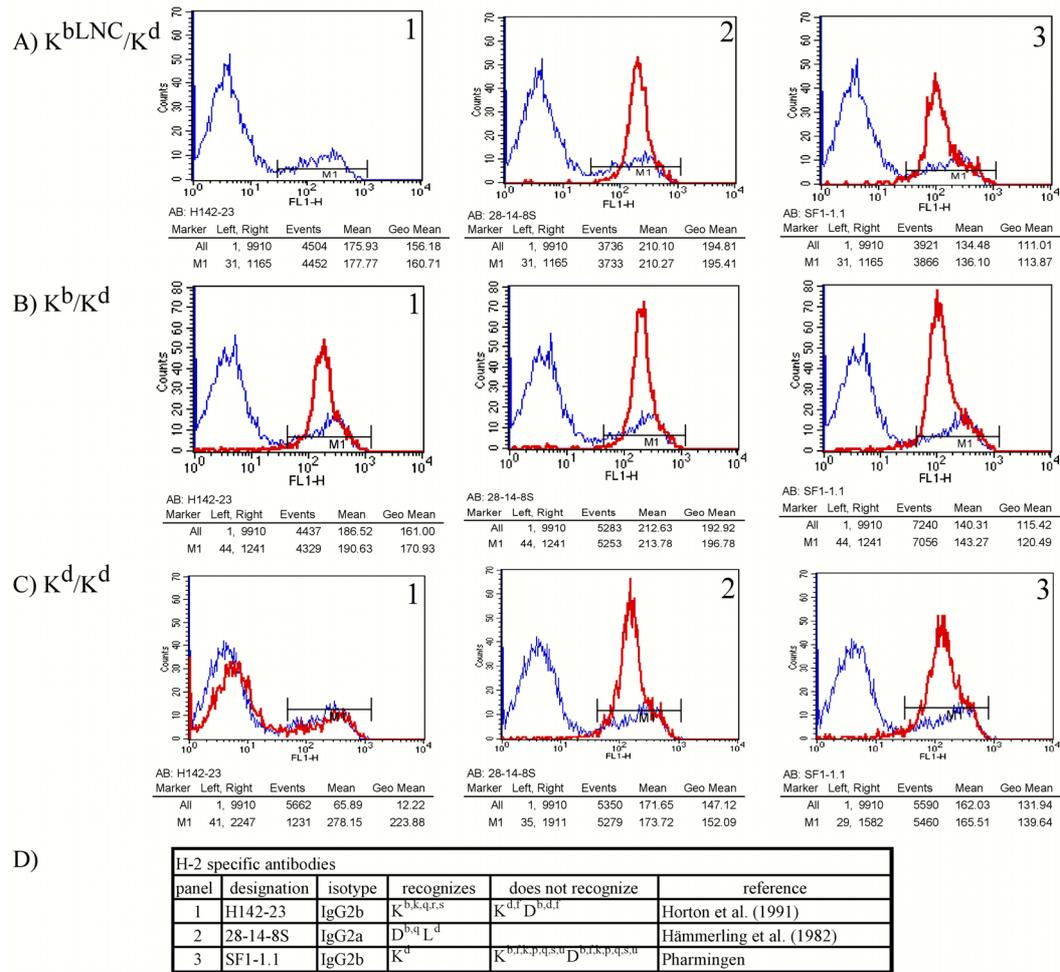


Figure 2-3. Comparative expression of K^{bLNC} and K^b . FACS analysis of surface stainings for K^{bLNC} , K^b , K^d , (in red). Negative control: Myc1-9E10 (in blue; monoclonal mouse antibody against synthetic peptide sequence which is part of the human protein p62c-myc) (Evan et al., 1985). Secondary antibody: FITC coupled rabbit anti-mouse IgG. Analysed were spleen cells of A) a K^{bLNC}/K^d mouse B) a K^b/K^d mouse and C) a BalbC (K^d/K^d) mouse. A) and B) are littermates and offspring from C). D) primary antibodies used and specificities. FL1-H: level of H-2 expression. 1: antibody: H142-23 (Horton et al., 1991). 2: antibody: 28-14-8S (Hämmerling et al., 1982). 3: SF1-1.1. (Pharmingen). The isolation and preparation of spleen cells was done as described above in Material and Methods. The details of the FACS analysis are described above in Materials and Methods.

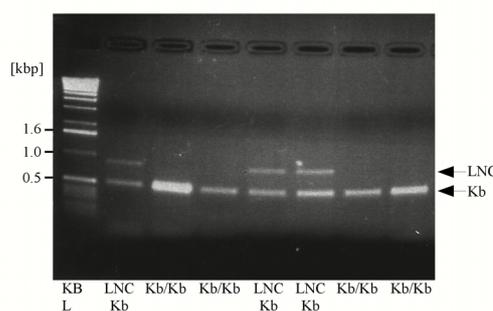


Figure 2-2. Examples of a Kb5/Kb6 based PCR on mouse genomic DNA. Depicted is the picture of an ethidiumbromide stained agarose gel (1%; 1XTAE; 0.01% Ethidiumbromide; loading volume: 25 μ l). PCR on genomic DNA of 7 littermates (all males). The offspring derives from the crossing 129^{SV}KbLNC (F8; female)X129^{SV}. PCR conditions: volume: 50 μ l; 5 μ l of a 1/100 dilution of isolated genomic DNA (see Materials and Methods); 5U Taq; 4.5mM MgCl₂; 5pmol Kb5 and Kb6; 95°C; 1min.; 35X[95°C; 30s; 55°C; 30s; 75°C; 1min]; 75°C, 10min; 4°C. LNC/Kb: heterozygous animals for K^{bLNC} and Kb wildtype. Kb/Kb: homozygous K^b wildtype offspring.

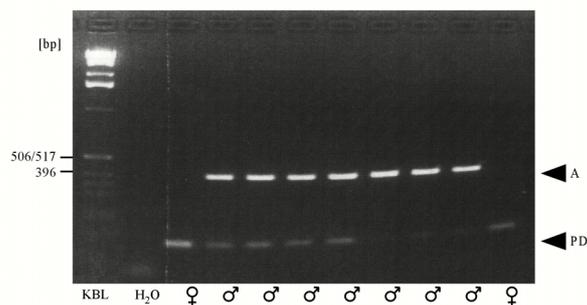


Figure 2-4. Sex determination of mice with a Sry specific PCR. Depicted is the photograph of an ethidiumbromide stained agarose gel (2%; 1XTBE; 0.01% Ethidiumbromide; loading volume: 25 μ l). The genomic DNA of nine mice (129^{SV}-K^b/K^{bLNC}) with known sex were chosen. In all cases the sex was confirmed with the Sry-PCR. PCR conditions: volume: 50 μ l; 5 μ l of a 1/100 dilution of isolated genomic DNA (see Materials and Methods); 5U Taq; 4.5mM MgCl₂; 5pmol Sry-pA: 5'-TCT TAA ACT CTG AAG AAG AGA C; 5pmol Sry-pB: 5'-GTC TTG CCT GTA TGT GAT GG; 95°C; 1min.; 35X[95°C; 30s; 55°C; 30s; 75°C; 1min]; 75°C, 10min; 4°C. A: amplicon. PD: primer dimer.

3. RESULTS

3.1. The M450^{IZ}-*lacOp* rescue: specific extraction of *E.coli lac-operator* containing DNA-fragments with M450^{IZ} Dynabeads[®]

A plasmid rescue protocol, which is based on the high and specific affinity between the *E.coli lac-Operator* sequence and the Lac-repressor protein (Gossen *et al.* 1993), was refined to enable the isolation of a single copy target sequence from genomic DNA, in this case the Mouse *H2* gene, *K^b*.

The assay consists in principle of three steps (see figure 3-1). Genomic or plasmid DNA, after being treated with restriction enzymes (in principle plasmid DNA can go through the assay in circulated form), is incubated with the fusion protein LacIZ (β -Galactosidase-Lac Repressor fusion (Kania *et al.* 1974)). This leads to binding of *lac-operator* containing DNA fragments to the Lac-repressor part of the fusion protein. After this binding step magnetic beads are added, to which anti- β -Galactosidase monoclonal antibodies are coupled. As a result LacIZ binds to the magnetic beads, which allows for the extraction of the target sequence by pelleting down the beads with a magnet. After washing steps the target sequence can be specifically eluted with IPTG. Details of the assay are provided in Materials and Methods, and whenever modifications are applied.

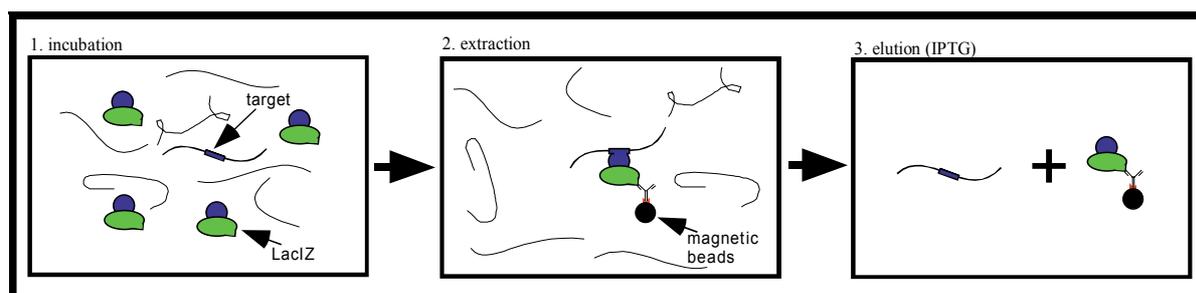


Figure 3-1. The M450^{IZ}-*lacOp* rescue: sketch. 1. incubation: DNA is treated with a restriction enzyme. The resulting fragment pool which contains *lac-operator* carrying target sequences is incubated with the fusion protein LacIZ. 2. LacIZ, to which the target sequences are bound, is coupled to magnetic beads via a anti- β -galactosidase monoclonal antibody. 3. elution: after washing steps the target sequence is specifically eluted with IPTG.

This part of the Result section is concerned with the description of the characterisation and optimisation of the M450^{IZ}-*lacOp* rescue protocol. To begin with, I will present the purification of the fusion protein LacIZ, which holds a central position in the process of specifically enriching for *lac-operator* containing target sequences. Subsequently the process of optimising the M450^{IZ}-*lacOp* rescue towards the isolation of single copy target genes, namely the *K^{bLNC}*-transgene, out of whole genomes, will be described.

3.1.1. Affinity based purification of the fusion protein LacIZ

The fusion protein LacIZ holds an essential place in the M450^{IZ}-*lacOp* rescue, and therefore in the whole scope of the project presented and discussed here. The β -Galactosidase part bridges the LacI/target-DNA complex to the magnetic beads. The LacI part is responsible for the binding of the target *lac-operator* containing sequence. Its properties determine the specificity of target binding and therefore the enrichment power of the assay. Originally this protein was available through Promega. With the onset of our project however, the company discontinued its production. As a consequence we had to purify it ourselves (see also (Martinsohn 1996)). The characterisation of LacI/ β -Galactosidase chimeras goes back to studies of Jürgen Kania and Benno Müller-Hill (Kania *et al.* 1974; Kania 1976; Kania *et al.* 1976; Brake *et al.* 1978). It has a monomeric molecular weight of 150 000kDa, and forms a Tetramer (Kania 1976). Two expression plasmids (pF1 and pF4) were cloned by Brigitte von Wilcken-Bergmann, based on a clone originally characterised by Jürgen Kania (Kania 1976; Brake *et al.* 1978), and kindly provided to us. Due to a different promotor, pF4 expresses LacIZ about four times stronger than pF1 (Brigitte von Wilcken-Bergmann, personal communication). In the context of the work presented here only pF4 was used. Traditionally Lac IZ, like the Lac repressor (LacI) is isolated by a precipitation step with ammonium sulphate followed by purification with phosphocellulose (Kania *et al.* 1974). This was formerly also done in this laboratory (Martinsohn 1996). However due to the sporadic occurrence of activity loss during the purification of LacIZ, we decided to adapt a protocol proposed by C. Larson and G. Verdine for the affinity based purification of the MHC class I transcription factor H2TF1 (Larson *et al.* 1992; Potter *et al.* 1993) for our own needs. Concatemers of *lac-operator* sequence containing oligonucleotides which are coupled to NHS-activated Sepharose (Pharmacia-Amersham), should lead to specific binding of LacIZ-protein out of bacterial raw lysate, and subsequently the bound protein should specifically be eluted with IPTG. This method should not only lead to purification of LacIZ, but also enrich for active material.

Two oligonucleotides, OPC and OPN were designed which, upon annealing, contain the wildtype *lac-operator* sequence (Gilbert *et al.* 1973). Also *Hind* III overhangs are formed, which allow for ligation. The oligonucleotide OPN carries at its 5' end an 6TFA-aminolinker (figure 3-2) which allows binding to N-hydroxysuccinimide (NHS) activated Sepharose (for chemistry see Materials and Methods and Amersham Pharmacia Biotechnology technical notes).

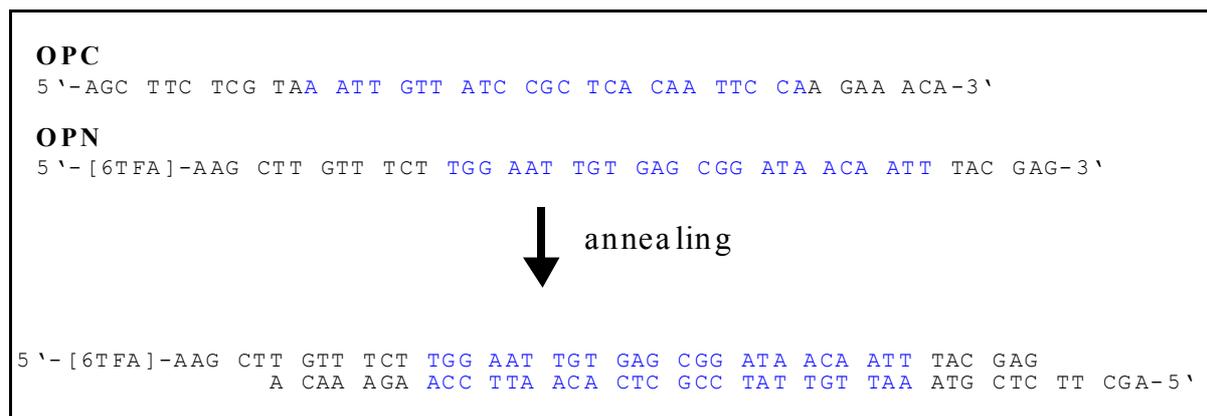


Figure 3-2. The oligonucleotides OPC and OPN. Depicted are the two 42 bp long oligonucleotides OPC and OPN. OPN has a 6TFA-aminolinker attached to its 5'-end ([6TFA]) which allows coupling to NHS-activated sepharose. The wildtype *lac-operator* sequence (Gilbert & Maxam, (1973)) is drawn in blue. Upon annealing overhangs are formed which after ligation lead to *Hind* III sites.

After annealing and phosphorylation of the two oligonucleotides (see figure 3-4 A)) it was tested in an electrophoretic mobility shift assay (EMSA) whether LacIZ binds efficiently to the fragments, and also whether it can subsequently be eluted efficiently with IPTG (figure 3-3). Both binding and IPTG-elution appeared to work efficiently when compared to the positive control pW310.

The [OPN/OPC] fragments were ligated (figure 3-4 B)), and the ligated material was coupled to HiTrap NHS-activated columns (1ml volume) following the instructions of the manufacturer (see Material and Methods and Amersham Pharmacia Biotech technical notes). The HiTrap column has a total ligand density of 10 μ mol per ml gel (Amersham Pharmacia Biotech HiTrapTM affinity columns data file). In sum 90nmol (2.5mg) of [OPN/OPC] were loaded onto the column. The binding efficiency was estimated with ³²P-radioactively labelled tracer (see figure 3-5 and table 3-1). Ca. 20% (18nmol) of the total material bound to the NHS column. This is well in the range of coupling efficiencies reported by others (50%, (Larson *et al.* 1992); 20%, Amersham Pharmacia Biotech technical note Smart system (Biotechnology 1995)). Larson *et al.* had used a total amount of 500nmol concatenated DNA fragments, while in the experiment presented by Pharmacia 9nmol of concatenated DNA was used.

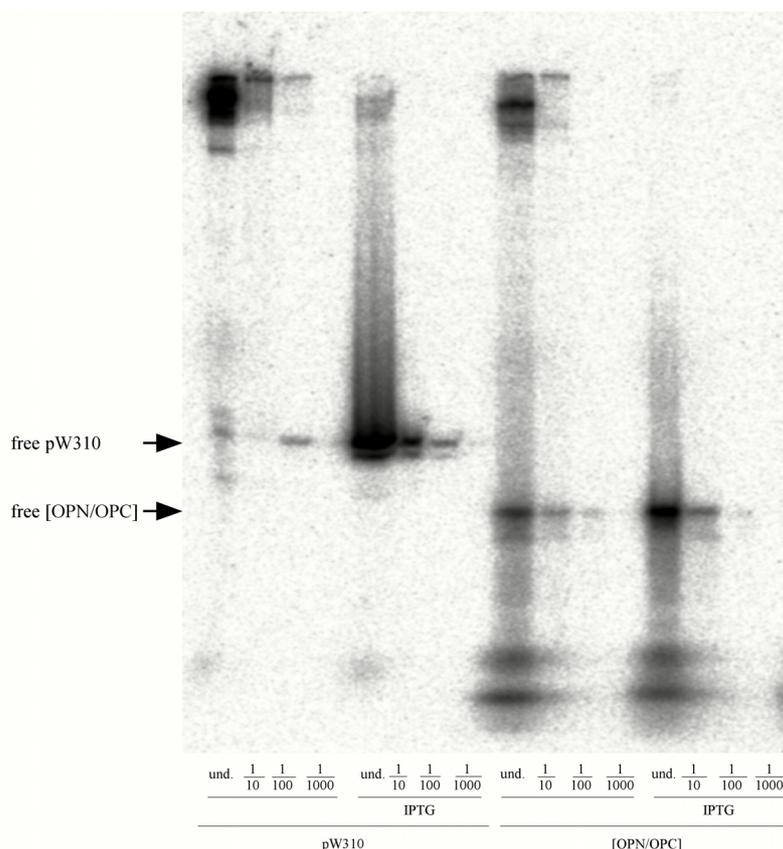


Figure 3-3. Binding to and IPTG-elution from the annealed oligonucleotides OPC/OPN of LacIZ. Depicted is a phosphoimager picture of an electrophoretic mobility shift assay (EMSA). Ca. 250ng of phosphocellulose purified LacIZ (see Martinsohn (1996)) were incubated with 20 000cpm ^{32}P -radioactively labelled fragment pW 310 which contains the ideal *lac-operator* sequence (O^d ; Fickert & Müller-Hill (1992), and see Material and Methods), or 20 000cpm of ^{32}P -radioactively labelled [OPN/OPC] in 100 μl volume, at room temperature (RT) for 30 minutes. The samples were split into two fractions of equal volume, of which one fraction was incubated with 2 μmol IPTG at RT for 30 minutes. 17 μl of each sample (undiluted, 1:10; 1:1000; 1:10 000) was loaded onto a PAGE gel (6%; 1XTBE; Prerun: 220V; 66mA, 258min.; run: 220V; 34mA).

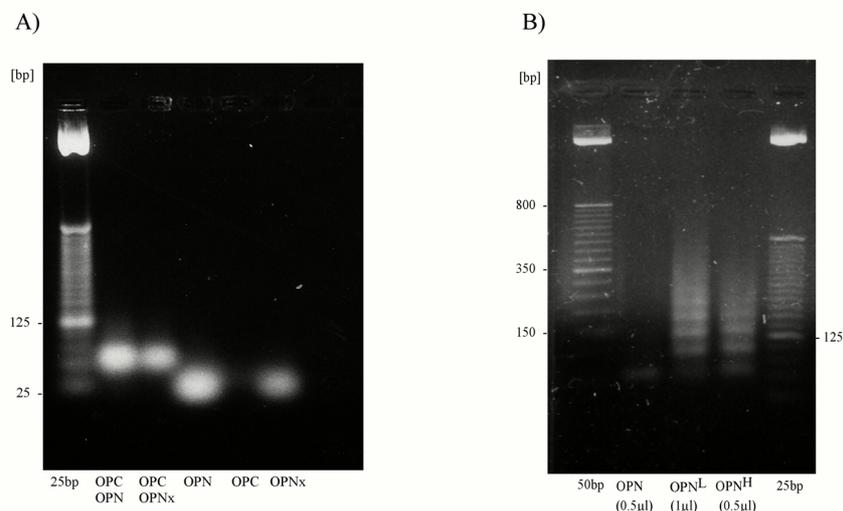


Figure 3-4. Annealing, phosphorylation and ligation of OPC and OPN. Depicted are pictures of ethidiumbromide stained agarose gels (2%; 1XTBE; 0.01% ethidiumbromide). A) annealing: per annealing reaction (ca. 100 μl volume) 1.2nmol of each oligonucleotide was heated to 75°C in a 250ml water bath and subsequently cooled down *ad. lib.* to room temperature. (Phosphorylation: was done with T4 Polynucleotid Kinase following the conditions proposed in Asubel (1998)). 5 μl of each sample is loaded onto the gel. B) ligation: per sample 2.5nmol of the annealed and phosphorylated ds-oligonucleotide [OPN/OPC] were ligated in 100 μl volume in the presence of 80WU (8000 ligation units) T4-Ligase (NEBioLabs), at 16°C for 8h. The samples were phenol extracted and ethanol precipitated, and finally taken up in 50 μl deionised H_2O . The indicated amounts were loaded onto the gel. Note: in A) OPNx is an oligonucleotide as OPN, with the exception that the 6TFA-aminolinker is attached to the T at position 5 (from the 5' end). It was ordered as there was some concern about the ligation efficiency of OPN. As it could be shown that [OPN/OPC] fragments ligate as well as [OPNx/OPC] fragments, OPNx was not used in later experiments. OPN^L: ligated material. OPN^H: *Hind* III digested material. Markers: 25bp and 50bp ladder (Gibco BRL).

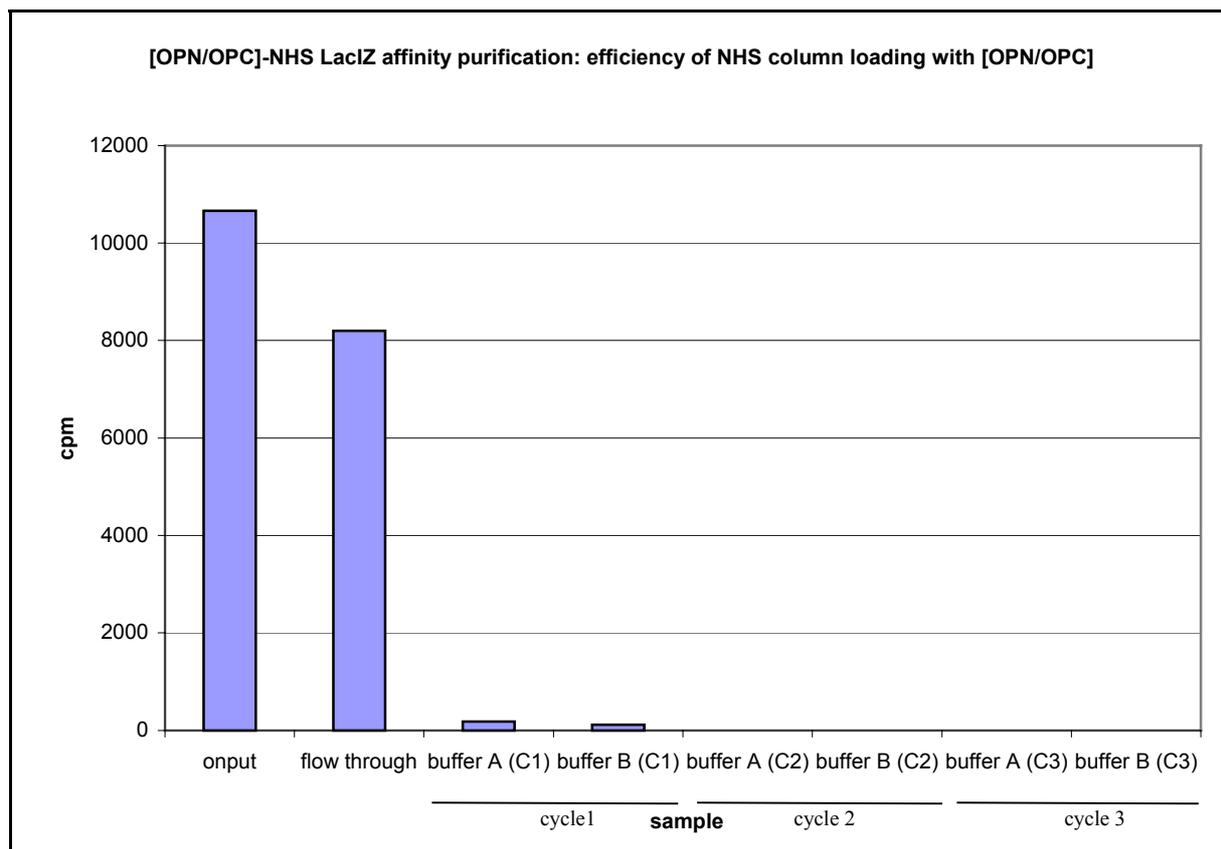


Figure 3-5. Binding efficiency of concatenated [OPN/OPC] fragments to HiTrap NHS-activated columns. Depicted is a graphical evaluation of the scintillation counts of fractions of all steps during the binding of DNA to the HiTrap columns (see Materials and Methods). In short the column was activated with an acidic solution (1mM HCl). Afterwards the [OPN/OPC] fragments were taken up in coupling buffer (0.4M NaHCO₃; 0.1M NaCl; pH8.3) and loaded onto the NHS-column. After incubation the coupling buffer was replaced by blocking buffer A (0.5M MetOH-Amine; 0.5M NaCl; pH8.3), which was followed by a deactivation step (0.1M Acetate; 0.5M NaCl; pH4.0). This blocking/stopping cycle was repeated three times (Cycle 1 to 3). Finally the column was transferred into storage buffer (10mM Tris/HCl pH8.0; 1mM EDTA; 0.3mM NaCl; 0.04% NaN₃), and stored at 4°C. Of each step 1/50th volume was taken off, deposited onto Whatman 3MM paper and after drying counted in a Beckman scintillation counter. Onput: total DNA applied onto the column. Flow Through: Material recovered after the loading steps. The amount of DNA bound to the column was estimated by subtracting the flow through amount from the onput. 90nmol (2.5mg) of OPC/OPN fragments to which 5X10⁵ cpm 32P-labelled fragments had been added were loaded in a total volume of 500 µl onto the column. Note: the exclusion volume of 1ml HiTrap NHS-activated columns was estimated to be 600µl by measurement with a blue dextran solution.

	cpm	[%] of onput
onput	10662	100,00
flow through	8199	76,90
buffer A (C1)	184	1,73
buffer B (C1)	113	1,06
buffer A (C2)	0	0,00
buffer B (C2)	0	0,00
buffer A (C3)	0	0,00
buffer B (C3)	0	0,00
[OPN/OPC] on column	2166	20,32

Table 3-1. Binding efficiency of concatenated OPN/OPC fragments to HiTrap NHS-activated columns. Listed are the number of counts (cpm) per sample examined as described in Figure 3-5.

Next the NHS column which had been loaded with DNA was used to purify LacIZ in a FPLC system (figure 3-6 A) and B)). The purity of the product was estimated to be 90% (as estimated on coomassie stained gels (figure 3-6 D)). The yield was roughly 20%, in sum about 1.4mg (table 3-2). The IPTG elution was very efficient as no remaining protein could be eluted in a high salt buffer washing step (0.5M NaCl, figure3-6C)). Several runs like this were

performed with very similar outcomes. The LacIZ batches showed high activity in the M450IZ-lacOp rescues, and were used throughout the optimisation experiments presented in the following sections. Unfortunately it turned out that using the same [OPN/OPC] coupled NHS-HiTrap column with each run the efficiency of purification in terms of yield declined constantly (see figure 3-7). We suspected that this was due to degradation of the concatenated oligonucleotides being bound to the columns, possibly because of nuclease activity in the LacIZ raw extract. To "dilute" the concatenated oligonucleotide [OPN/OPC] as a target for degradation in later runs the LacIZ raw extract before being loaded onto the NHS-column was mixed with 100µl sonicated herring sperm DNA (10mg/ml). This did not have major effects. Also despite increasing the concentration of EDTA in the NHS-column storage buffer (up to 50mM EDTA) the efficiency loss persisted. This means that the columns have up to now only a limited life time for about four to five runs. However it was shown that in one run enough active LacIZ for about 500 M450^{IZ}-lacOp rescues can be isolated. In one case, in which the NHS-lacOp column purification was applied the batch of purified LacIZ failed to show any activity in electrophoretic mobility shift assays (EMSA). However since the purification as such was successful (as could be shown in the FPLC elution profile and by coomassie stained PAGE analysis), and since the condition for successful purification is the function of the Lac repressor part of LacIZ, the loss of activity must have occurred post purification and is probably due to wrong treatment of the protein.

Fraction (dilution)	OD595	mg/ml (undiluted)	sum volume [ml]	sum protein [mg]
IPTG elution	0.082	0.45	3	1.4

Table 3-2. Yield of affinity purified LacIZ. The Protein amounts were estimated with the BioRad Protein assay. Listed is the measured Od₅₉₅, the concentration of the different fractions, and the sum of protein yielded.

The successful isolation of a high quantity of LacIZ enabled us to further characterise and optimise the M450IZ-lacOp rescue. As discussed in the introduction, transgenic mice have been generated, carrying the *lac-Operator* sequence in the third intron of the *H-2K^b* gene. The efficient specific extraction of this gene, or parts of it, from genomic DNA would simplify any further mutational analysis, as well as greatly diminish the risk of contamination and artefacts. The feasibility and potential of this approach was first explored in a simple plasmid system, pBluescript (Stratagene), which carries the *lac-operator* sequence downstream of the multiple cloning site.

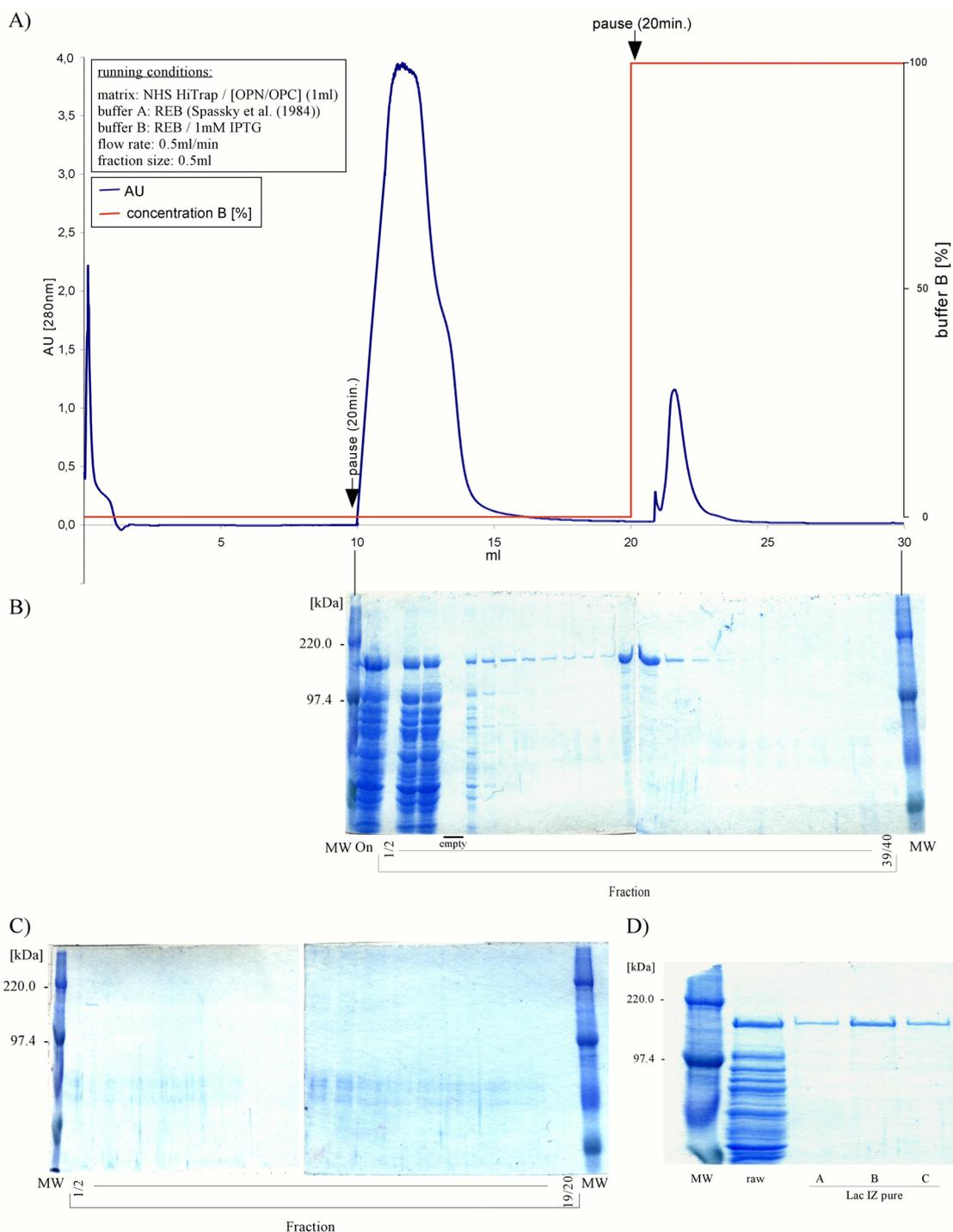


Figure 3-6. NHS-HiTrap column affinity based purification of LacIZ. A) FPLC run: 800 μ l LacIZ raw extract (preparation see Materials and methods) was spun down (10^5g ; 30'; 4°C). To the supernatant 200 μ l REB (Spassky et al., (1984)) and 5 μ l PMSF (10mg/ml; Sigma) were added. The material was immediately loaded onto an equilibrated [OPN/OPC]-NHS-HiTrap column. Running conditions as indicated in the figure. Fraction size 0.5ml. B) Fractions of FPLC LacIZ purification run: Coomassie stained denaturing PAGE; 7.5%PAA; loading volume: For each loaded sample 10 μ l of two fractions (1 and 2; 3 and 4 etc.) where taken up in a total volume of 30 μ l 1Xloading buffer of which 10 μ l were loaded onto the gel. C) Fractions of high salt wash of column. After the purification run the column was washed with high salt buffer (40mM TrisHCl pH8.0; 50mM KCl; 50mM MgCl₂; 0.1mM DTT; 500mM NaCl). Depicted is as in B) a Coomassie stained denaturing PAGE; 7.5%PAA; loading volume: 10 μ l (10 μ l of two following fractions where taken up in a total volume of 30 μ l 1Xloading buffer). After the runs the [OPN/OPC]-NHS-column was transferred into storage buffer (10mM Tris HCl pH8.0; 300mM NaCl; 1mM EDTA (in later runs 50mM EDTA, see text); 0.04% NaN₃). D) Affinity purified LacIZ. Examples of stored LacIZ after purification. Depicted is a Coomassie stained denaturing PAGE. 7.5% PAA; 5 μ l of each sample were taken up in 50 μ l 1Xloading buffer of which 15 μ l were loaded onto the gel. Shown are examples of three independent runs (A; B; C) in comparison to raw extract. Sample C belongs to the FPLC run presented in A) to C). Storage: The fractions which belonged to the peak of purified LacIZ in the respective FPLC runs were pooled and dialysed three times 8h against 2l REB (Spassky et al. (1992)). Subsequently the samples were concentrated with Centricon-50 spin columns (Amicon Millipore), and mixed carefully with 1 Volume 2XLacIZ storage buffer (100mM Tris/HCl pH8.0; 200mM NaCl; 2mM DTT; 80% glycerol). Aliquots (40 μ l) were snap frozen in liquid nitrogen and stored at -80°C. MW: Molecular weight marker (Rainbow (Amersham))

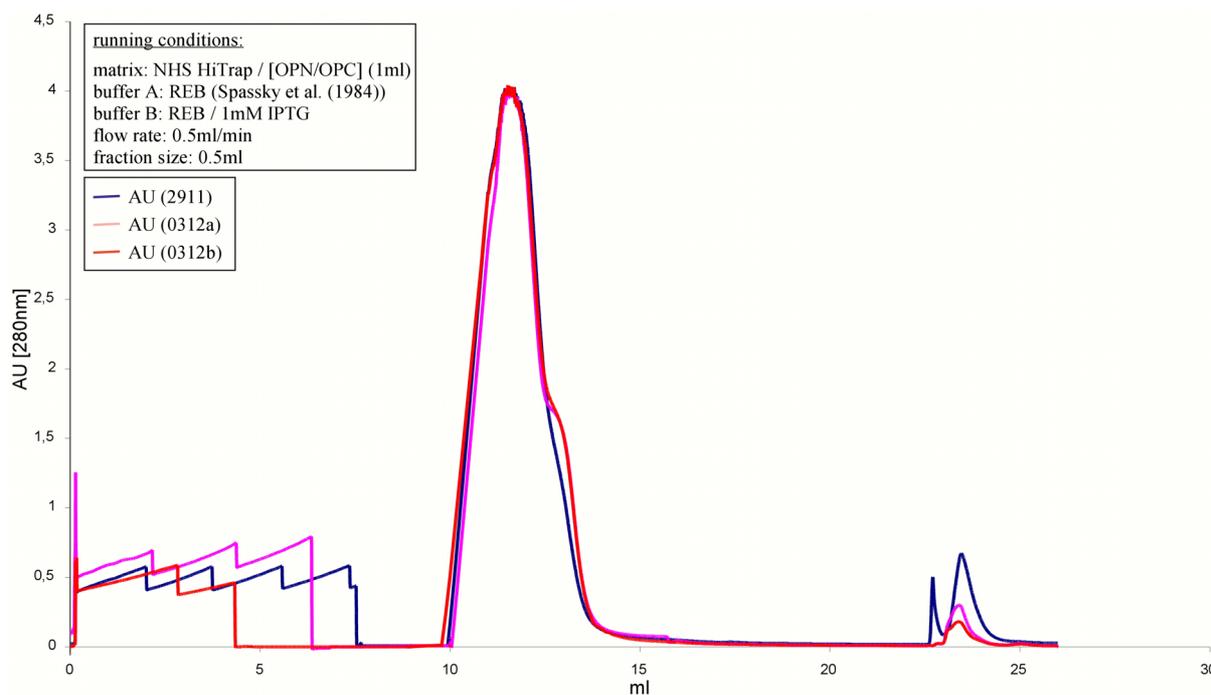


Figure 3-7. Loss of efficiency of [OPN/OPC] coupled HiTrap activated NHS columns. Depicted are the FPLC profiles of three independent [OPN/OPC]-NHS column LacZ purification runs, using the same column. The running conditions were as in Figure 3-6. The running order was: 1: 2911; 2: 0312a; 3: 0312b. 0312a and 0312b were two runs following each others immediately.

3.1.2. Extraction of target sequence from pools of fragmented plasmid DNA

The plasmid pBluescriptKS(+) (Stratagene) was digested with the restriction enzyme *HaeIII* (BioLabs). Fourteen fragments are generated, of which one fragment of 242bp length carries the *E.coli lac-operator* sequence (“242^{Op}”).

The pool of digested DNA went through one cycle of the M450^{Iz}-*lacOp* rescue. Subsequently samples which either derived from the *Hae III* treated material that went into the assay (onput), or samples of the elution step, were analysed with the aid of capillary electrophoresis (figure 3-8). The peak areas, which were obtained through signal integration (Analysis: Model600; ABI), are listed in table 3-3. Apart from the signal of 242^{Op} only one further signal of unknown origin is visible after considerable magnification of the CE run (figure 3-8 B)). It was used to calculate an enrichment factor with reference to the target sequence 242^{Op} (table 3-3). The enrichment factor is greater than 2000. An estimation of an enrichment factor with respect to the whole population of *Hae III* fragments is difficult. If fragment 1 is used as a reference, a signal which cannot be ascribed unambiguously to any of *Hae III* generated fragments, thus chosen randomly, the enrichment factor lies in the range of 70 000 (table 3-3 C)). For fragment 3, which is chosen as fragment 1 it is about 2000.

However due to the shape of these signals as well as to their running behaviour it is likely that fragment 1 to 4 are rather salt signals than DNA (Martinsohn 1996). The final yield of 242^{Op} was about 60% with respect to the onput (see figure 3-8 and table 3-3).

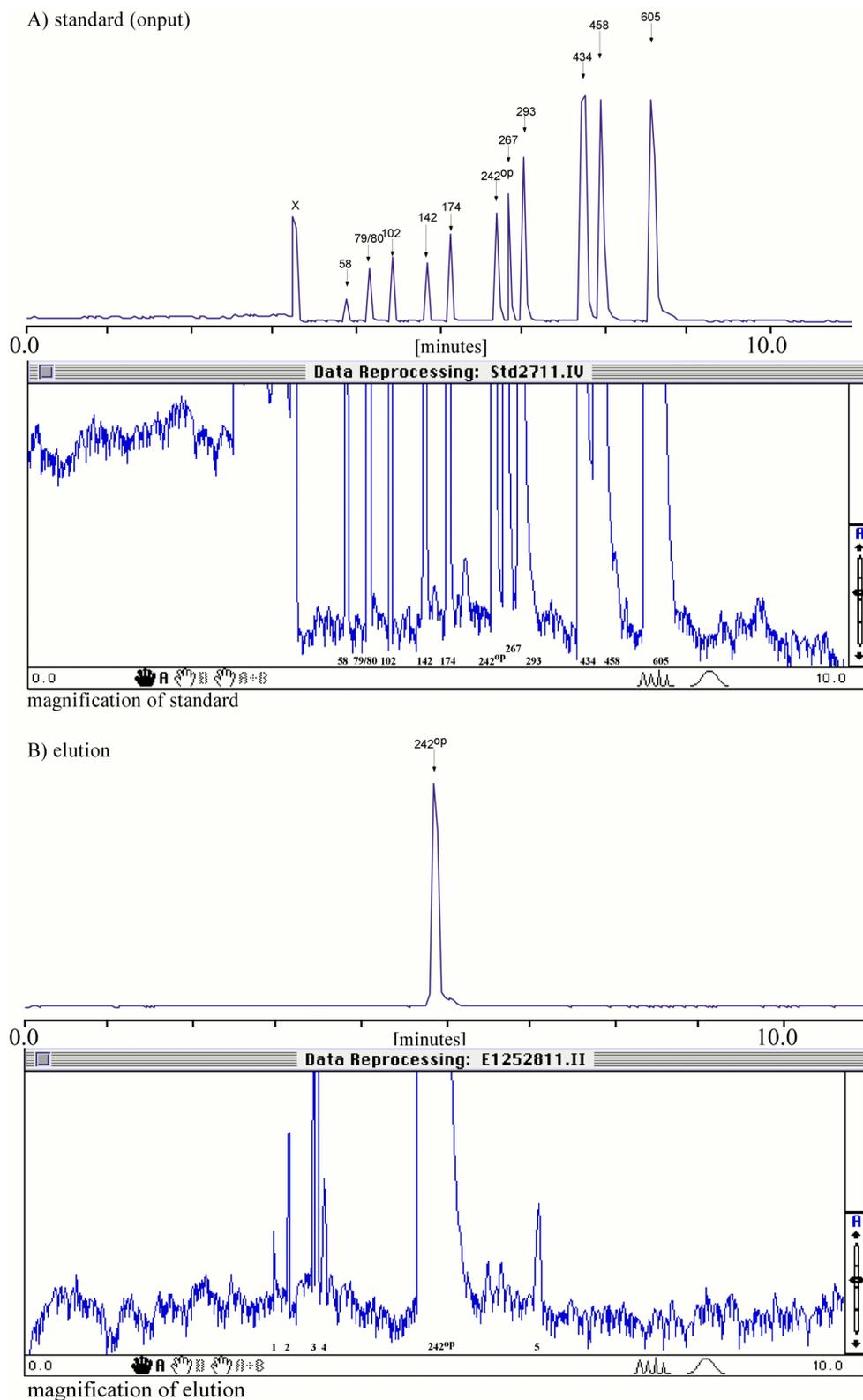


Figure 3-8. Extraction of 242^{OP} from a *Hae* III restriction digest of pBluescript. A): CE-run of the pBluescriptKS(+) fragments generated by a *Hae*III digest, which went into the M450^{IZ}-*lacOp* rescue (reference run; output). B): IPTG-elution of the lac-operator carrying target sequence 242^{OP}, after treatment with one cycle of the M450^{IZ}-*lacOp* rescue. The numbers above the respective signals indicate the fragment length in basepairs. CE-run: 1% HPMC; [0.05µg/ml] TO; 1XTBE; 200V/cm; Injection of output (1:5 dilution; 6ng/µl) pBluescript-*Hae* III in H₂O, 3s, 5kV (200V/cm). Injection of eluted material: transferred into H₂O via Sephadex G-25 columns (Pharmacia; DNA-grade; see Material and Methods); undiluted; 30s, 5kV (200V/cm). M450^{IZ}-*lacOp* rescue: (1) binding: 50µl volume; ca. 2X10⁷ M450IZ-Dynabeads®; 20mM Tris/HCl pH7.6, 2mM EDTA, 5% glycerol; 90min; RT; (2) washing: 125mM NaCl, 20mM Tris/HCl pH 7.6, 2mM EDTA, 5% glycerol, 1 volume; 2 X 5min.; (3) elution: 125mM NaCl, 20mM Tris/HCl pH 7.6, 2mM EDTA, 8.4mM IPTG; 30min; RT. For processing purposes of the CE-data the software Model600 (ABI) was used. X: signal representing fragments of 11bp and 18bp length, and a loading effect.

A) before elution (1:5 dilution; injection 1/10 time of elution)

Fragment [bp]	Time [min.]	Height[μ V]	Area [μ V*s]	Area%	242 ^{Op} /X
58	3,880	11969	7232	0,536	10,7
79/80	4,155	29613	36053	2,674	2,1
102	4,410	36132	24386	1,809	3,1
142	4,835	33040	28855	2,14	2,6
174	5,114	48498	50654	3,757	1,5
242 ^{Op}	5,675	61042	76431	5,67	1,0
267	5,809	71479	101826	7,554	0,8
293	5,990	92575	147408	10,935	0,5
434	6,716	126997	244063	18,106	0,3
458	6,919	124267	276007	20,476	0,3
605	7,532	124510	355037	26,338	0,2
Total Area			1347952	99,995	

B) after elution

Fragment	Time [min.]	Height(μ V)	Area [mV*s]	Area%	242 ^{Op} /X
1	3.222	1745	127	0,005	17329,1
2	3.520	3521	1390	0,062	1583,3
3	3.560	6789	4037	0,182	545,1
4	3.646	1254	2444	0,110	900,4
242 ^{Op}	4.855	945248	2200806	99,485	1,0
458	6.269	1148	3378	0,152	651,5
Total Area			2212182	99,996	

C) enrichment

Fragment	1	2	3	4	242 ^{Op}	458
Enrichment	71 500	5800	1960	3250	1	2352

Table 3-3. Enrichment of target sequence after one cycle of the M450^{Iz}-lacOp rescue. All areas beneath the signal peaks were integrated. Subsequently for each run the quotient of [Area 242Op] to [area fragment x] was calculated. The ratio of this quotient for the fragment of 458bp length before A) and after B) IPTG-elution leads to the enrichment factor with respect to fragment x (C)). Fragments 1 to 4 in B) could not be assigned to any of the fragments which are generated by a *HaeIII* digest of pBluescript. They were included into the calculation of enrichment factors by building the ratio of their respective % area-values in the elution run to that of 458 ([% area 458]/[% area x]), and multiplication of this value with the enrichment factor for fragment 458 (([% area 458]/[% area x])*[2352]). Data processing was carried out with the software program Model600 (ABI).

The approach described above is to my knowledge the first attempt to estimate directly the enrichment power of the M450^{Iz}-lacOp rescue. Beforehand indirect measurements such as plasmid rescue followed by transformation studies were employed (Gossen *et al.* 1993). The result is impressive, but the pBluescript based evaluation system is of course very simple with respect to the objective of extracting a specific gene, or parts of it, out of pools of genomic DNA. Therefore a series of experiments was designed and executed with the intention to assess the power of the target rescue assay under more challenging conditions, namely in the context of genomic DNA. The experiments should allow to optimise the assay, and finally lead to the ultimate goal of isolating single copy target genes from genomic DNA.

3.1.3. Target extraction out of pools of genomic DNA: High Copy Number

In the pBluescript^{HaeIII} system described above the fragment 242^{Op} represents 8% of the total mass of DNA. As discussed in the introduction we are aiming at the isolation of single copy genes or parts thereof from the whole mammalian genome. The mouse genome has a size of 2.7×10^9 bp per haploid genome (Ausubel *et al.* 1998) DNA fragments which are supposed to be isolated from the Mouse genome, will, for practical reasons, have a length of about 0.5 to 4kb. Fragments of this size represent 0.00002 to 0.00016% of the total DNA, which in turn is equivalent to a 50 000 fold dilution of target sequence in comparison to the pBluescript^{HaeIII} system. It was therefore necessary to test whether the M450^{Iz}-*lacOp* rescue also works in a satisfactory manner under such more demanding conditions.

Accordingly DNA mixtures were prepared in which genomic DNA was contaminated with *lacOp*-containing plasmid at a ratio 10^3 copies per genome (That is, still 10^3 fold more abundant than single-copy genes). The genomic DNA, derived from chicken erythrocytes (*Gallus domesticus*), was digested with the restriction enzymes *EcoR* I and *BamH* I. The linearised plasmid NKrQ, which carries the *lac-operator* sequence (see Material and Methods) served as the target sequence. In order to evaluate the enrichment factor, and specificity of the M450^{Iz}-*lacOp* rescue, under the respectively given conditions, plasmids were used which do not carry the *lac-operator* sequence (pHbapr-1 and pSP72; see also Material and Methods). The M450^{Iz}-*lacOp* rescue was monitored by analysis of samples which derived from different steps of the procedure (onput, washing, elution). The analysis was done by Southern blotting using probes for the plasmids.

Uptake of DNA by M450^{Iz} beads

Formally the M450^{Iz}-*lacOp* rescue can be separated into three steps which determine the amount of target sequence that can be recovered after being carried through the assay (yield) and purity with respect to the contaminating DNA fragments present at the beginning (enrichment). These steps are (1) uptake of DNA during incubation with LacIz and the M450 Dynabeads^o, (2) discharging contaminating DNA during rinsing steps, and finally (3) specific elution of the target sequence. It must be stressed that our major concern is enrichment of target sequence and only to a lesser extent yield. Loss of material can be effectively compensated for by PCR. None of these steps has previously been analysed in any detail before (Gossen *et al.* 1993; Gossen *et al.* 1994; Boerrigter *et al.* 1995; Gossen *et al.* 1995; Dolle *et al.* 1996).

The first experiments were designed to look at the specificity and quantity of DNA uptake. For this purpose the starting DNA (restricted chicken genomic DNA and plasmid DNA) was mixed with M450^{I_Z} beads, and compared to the material that could subsequently be recovered by non-specific elution from the beads with high molar salt buffer (1M NaCl) after being processed with the M450^{I_Z}-*lacOp* rescue.

Plasmids used in the assays

All plasmids used in the following experiments were linearised. In figure 3-9 the location of the plasmids used in the following experiments are depicted.

Quality and quantity of DNA uptake in time dependence

For the rescue experiments using pBluescript (see above) incubation times of 90 minutes were employed consistently for the incubation of the target sequence containing DNA pools with M450^{I_Z}. It was not clear to what extent in a more complex system containing genomic DNA and target sequence at low frequencies duration of incubation plays a role. Therefore DNA mixtures containing digested genomic chicken DNA as well as the plasmids NKrQ (*lacOp*⁺), and pH β apr1 (*lacOp*⁻), were incubated for different time intervals with the fusion protein LacIZ. After the final elution with 1M NaCl, the supernatants (i.e. the material that was taken off the beads after incubation), was analysed in a comparative way to the eluted samples (figure 3-10 and for graphical evaluation figure 3-11). No clear time-dependence of uptake could be observed. Additionally only modest enrichment for the target NKrQ to reference plasmid pH β apr1 could be obtained. The high salt elution samples (1M NaCl) show a substantial non-specific uptake of DNA by the M450^{I_Z}-Dynabeads[®]. On average more than 15% of the starting material of pH β apr1 was bound onto the M450^{I_Z}-Dynabeads[®] (figure 3-12), which could only be eluted after treatment with 1M NaCl. As can be seen in figure 3-10 in the sample of NKrQ low molecular DNA fragments are present (here designated as “LMWF”). As they behave independently from the intact plasmid NKrQ during the M450^{I_Z}-*lacOp* rescue they were included into the evaluation. As shown in figure 3-11 these low molecular weight fragments (size: 1 to 1.5kbp) are more efficiently discharged than pH β apr1 (size: 10.5kbp (Gunning et al. (1987))). Already in experiments in which the *Hae* III digest of pBluescript was used, it could be shown, that DNA fragments show an increasing tendency to bind non-specifically to M450^{I_Z}-Dynabeads[®] with increasing length, even though the data was not conclusive and the main cause of non-specific binding of DNA at that point seemed to be inactive LacIZ protein (Martinsohn 1996).

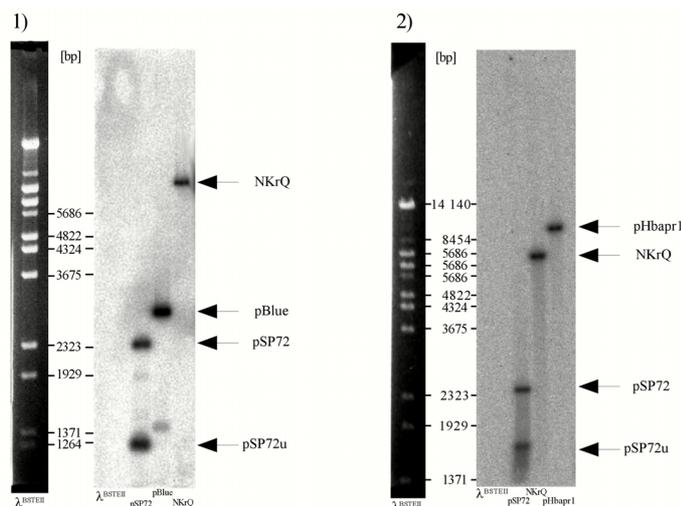


Figure 3-9. Localisation of plasmids being used in the evaluation of the M450IZ-*lacOp* rescue. Depicted are images of the λ BstEII digest (NEBioLabs) loaded onto an agarose gel (0.6%; 1XTBE; 0.01% Ethidiumbromide), and the corresponding Southernblots with the plasmids pSP72 (Promega, l=2462bp), pSP72u, pBluescript (pBlue; Stratagene; l=2961bp), NKrQ (Materials and Methods; l=7.5kb), and pHbaprl (Gunning et al., 1987; l=10.5kb).

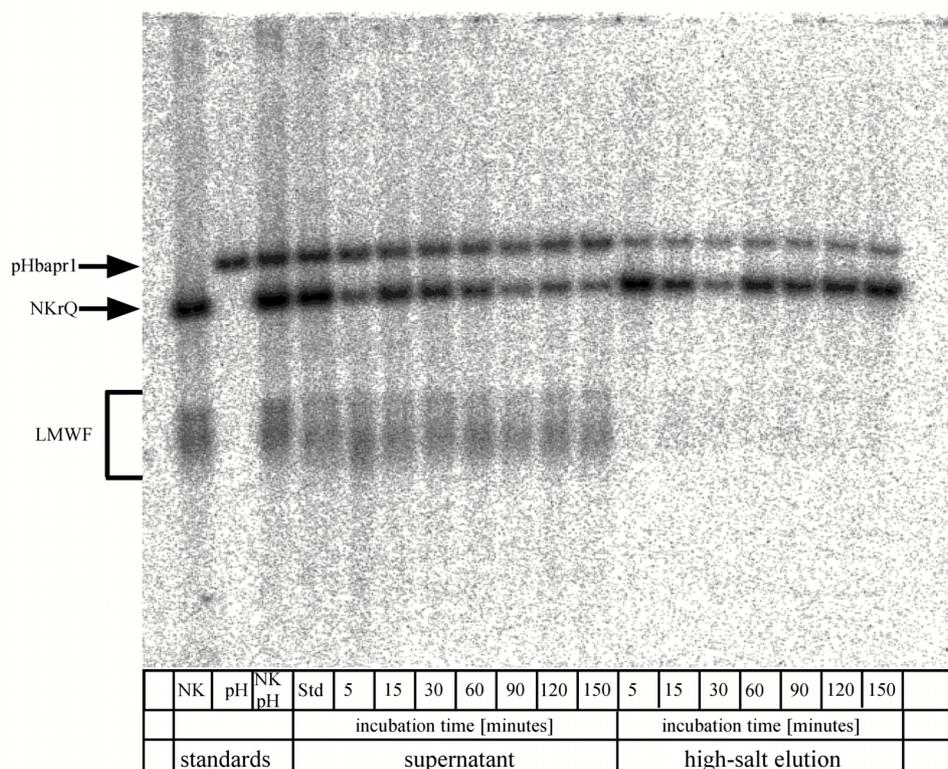


Figure 3-10. M450IZ-*lacOp* rescue: Quality and quantity of DNA uptake in time dependence. Depicted is a picture of a Southern blot analysis. As target sequence the linearised plasmid NKrQ was used, as tracer for non-specific binding the linearised plasmid pHbaprl was used (see Material and Methods), which carries the *lac-operator* sequence. NK: linearised plasmid NKrQ. PH: linearised plasmid pHbaprl LMWF: In the sample of NKrQ low molecular weight DNA was present. This fraction was included into the evaluation. Standard (NK:pH): all samples derived from a mastermix, which was divided into samples of equal volume. The sample "standard" did not undergo the M450IZ-*lacOp* rescue. M450IZ-*lacOp* rescue: per sample in 100 μ l Volume (10mM Tris/HCl pH7,6; 10mM MgCl₂; 5% glycerol; 0,01%BSA) 500ng *Eco* RI/*Bam* HI digested chicken (*Gallus domesticus*) DNA (equivalent to 10⁵ diploid genomes), 500pg of the linearised plasmid NKrQ (ca. 10⁸ copies, carries *lac operator* sequence, see Material and Methods), 500pg of the linearised plasmid pHbaprl (reference plasmid for unspecific binding, ca. 10⁸ copies) and 600ng (7pmol) LacI were incubated under rotation for one hour at room temperature. This mixture was subsequently added to pelleted M450-Dynabeads® (10⁶ beads), which previously had been armed with α - β -Galactosidase monoclonal Antibodies (Promega), and incubated for different time spans (as indicated) on a rotator at room temperature. The supernatants were taken away, and the remaining beads rinsed two times (10mM Tris/HCl pH7,6; 10mM MgCl₂; 5% Glycerol; 0,01%BSA). Finally the beads were incubated with 20 μ l of high molar salt buffer (1M NaCl). All supernatants were taken away, transferred into equal buffer conditions, before they went into Southern blot analysis. As a radioactive probe a *Bam* HI/*Sal* I fragment of the neomycin resistance gene was used. The neomycin resistance gene is present both in NKrQ and pHbaprl (see Materials and Methods). Exposure time: 3h on a phosphorimager screen.

It was clear from this experiment that specificity of DNA uptake is not influenced to a notable degree by time of incubation. Therefore in later experiments consistently an incubation time of 90 minutes was chosen. It had also become obvious that the enrichment as

measured by uptake is disappointingly inefficient under the conditions chosen. That made further evaluation necessary with respect to buffer conditions during the different steps of the assay. One concern with respect to unspecific uptake of DNA is certainly the high amount of non-specific binding sites for DNA provided by LacIZ (ca. 10^{12} binding sites per assay for the target sequence; see Material and Methods). A possible way of trying to suppress unspecific uptake is to decrease the amount of binding sites provided by either decreasing the amount of LacIZ, or LacIZ and the magnetic beads. However for reasons of practicability we decided to attempt to block nonspecific binding by providing lac operator sequence in form of pBluescript.

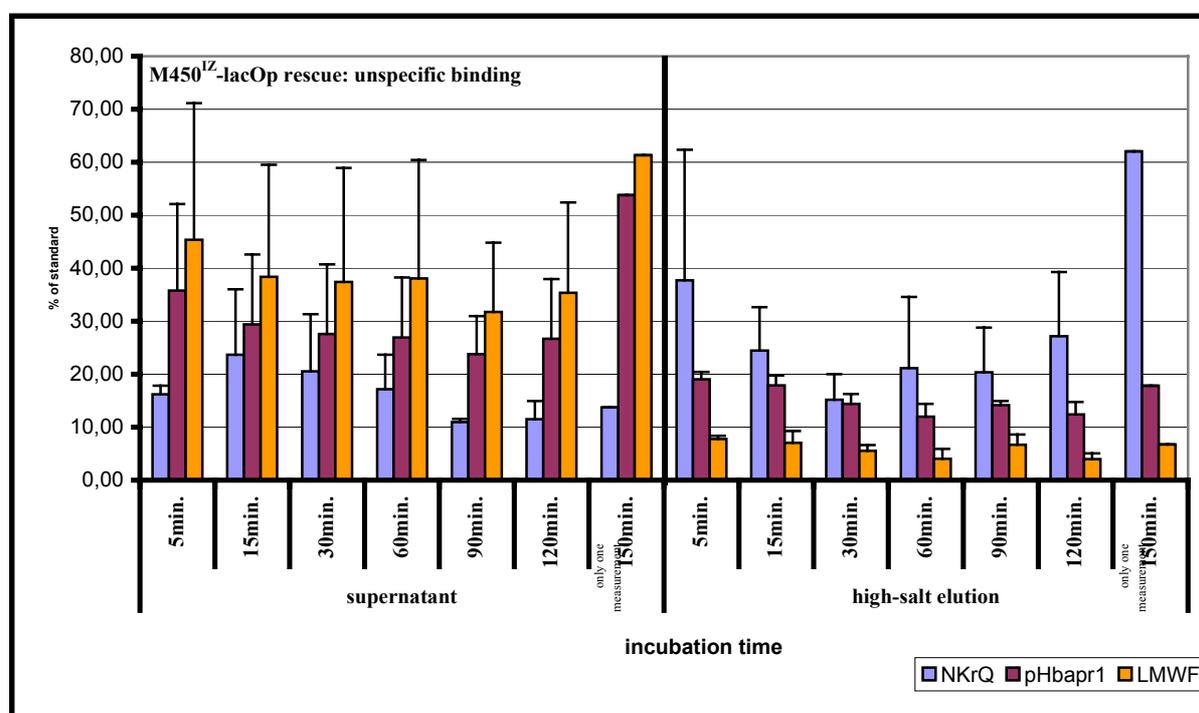


Figure 3-11. M450^{IZ}-lacOp rescue: unspecific binding of DNA. Graphical evaluation of three experiments as depicted in figure 3-10. Shown is the arithmetical mean of the fraction of DNA (per cent of Standard) which was detected and measured in the samples of the respective time points. The quantification was done with a phosphoimager (Software: TINAv2.08); evaluation with SPSS9.0 and EXCEL2000. Error bars: standard error.

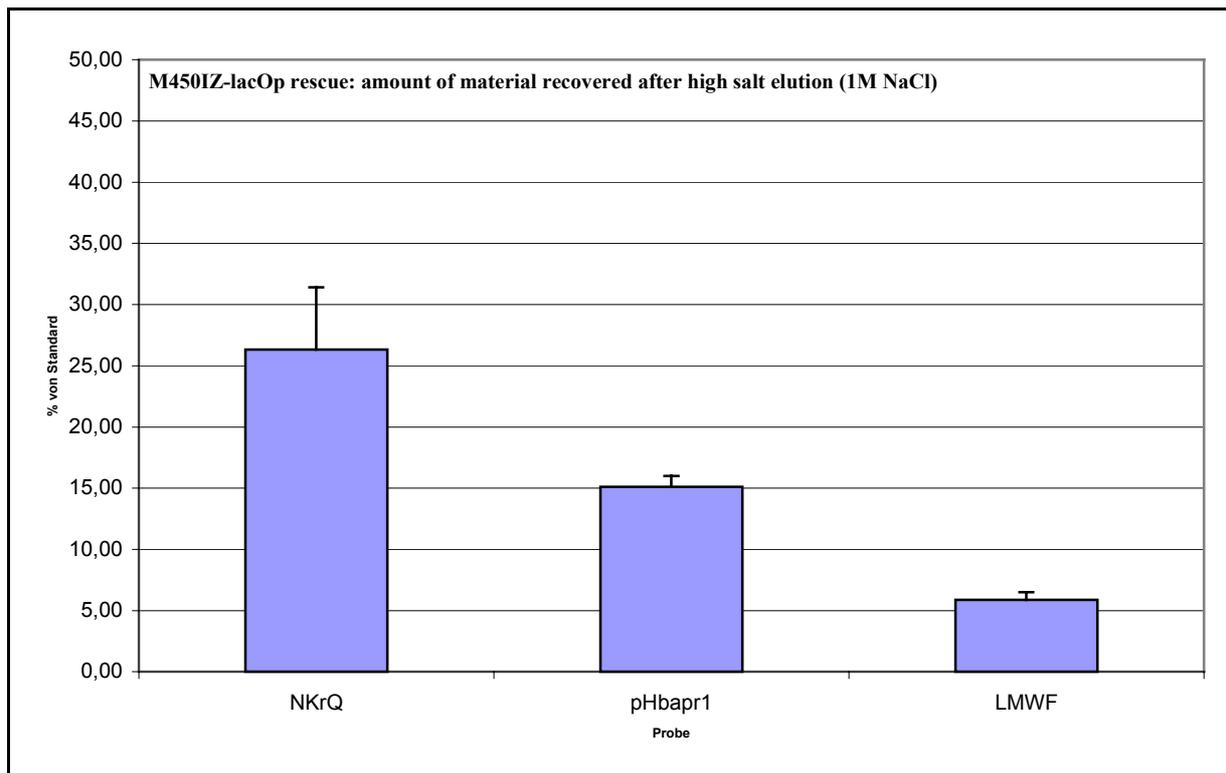


Figure 3-12. Recovery of DNA fragments after high salt elution with 1M NaCl. Depicted are the arithmetic means per sample of all time points combined. The quantification was done with a phosphoimager (Software: TINAv2.08); evaluation with SPSS9.0 and EXCEL2000. Error bars: standard error.

Blocking nonspecific binding of DNA: “spiking” of samples with pBluescript

Non-specific uptake of DNA during the M450^{IZ}-lacOp rescue is probably largely due to the nature of the Lac repressor protein. It has been shown that the Lac repressor has a certain general affinity to DNA (Lin *et al.* 1972; Hippel *et al.* 1974; Lin *et al.* 1975; Zingsheim *et al.* 1977; Kamashev *et al.* 1994). The affinity of the Lac repressor is mainly influenced by the similarity of DNA sequence to the *lac-operator* (Pfahl *et al.* 1978; Fried *et al.* 1981; Winter *et al.* 1981; Oehler *et al.* 1990). LacI also shows affinity to d(A-T) rich DNA regions (Riggs *et al.* 1972; Klug *et al.* 1979; Hogan *et al.* 1980), and in general the specificity of DNA-uptake is strongly affected by salt conditions (Lin *et al.* 1972; Lin *et al.* 1975). The effect of salt conditions will be looked at in later paragraphs. The LacIZ protein is available in high amounts during the rescue assay (ca. 10^{12} binding sites per assay for the target sequence; see Material and Methods). Even though studies of kinetics of LacI showed that the half life of the LacI/non-*lac-operator* DNA complexes is very short (in the range of milliseconds (Lin *et al.* 1975)), it is possible that non-specific uptake of DNA during the M450^{IZ}-lacOp rescue is reinforced by an avidity effect due to the density of LacI binding sites provided and the length of DNA.

We tested whether addition of *lac-operator* sequence in form of the plasmid pBluescript, leads to a successful competition for DNA binding sites thereby reducing

unspecific binding of DNA. For this purpose increasing amounts of the *EcoR* I linearised plasmid pBluescript were added to the samples which went through the M450^{Iz}-*lacOp* rescue. (Figure 3-13).

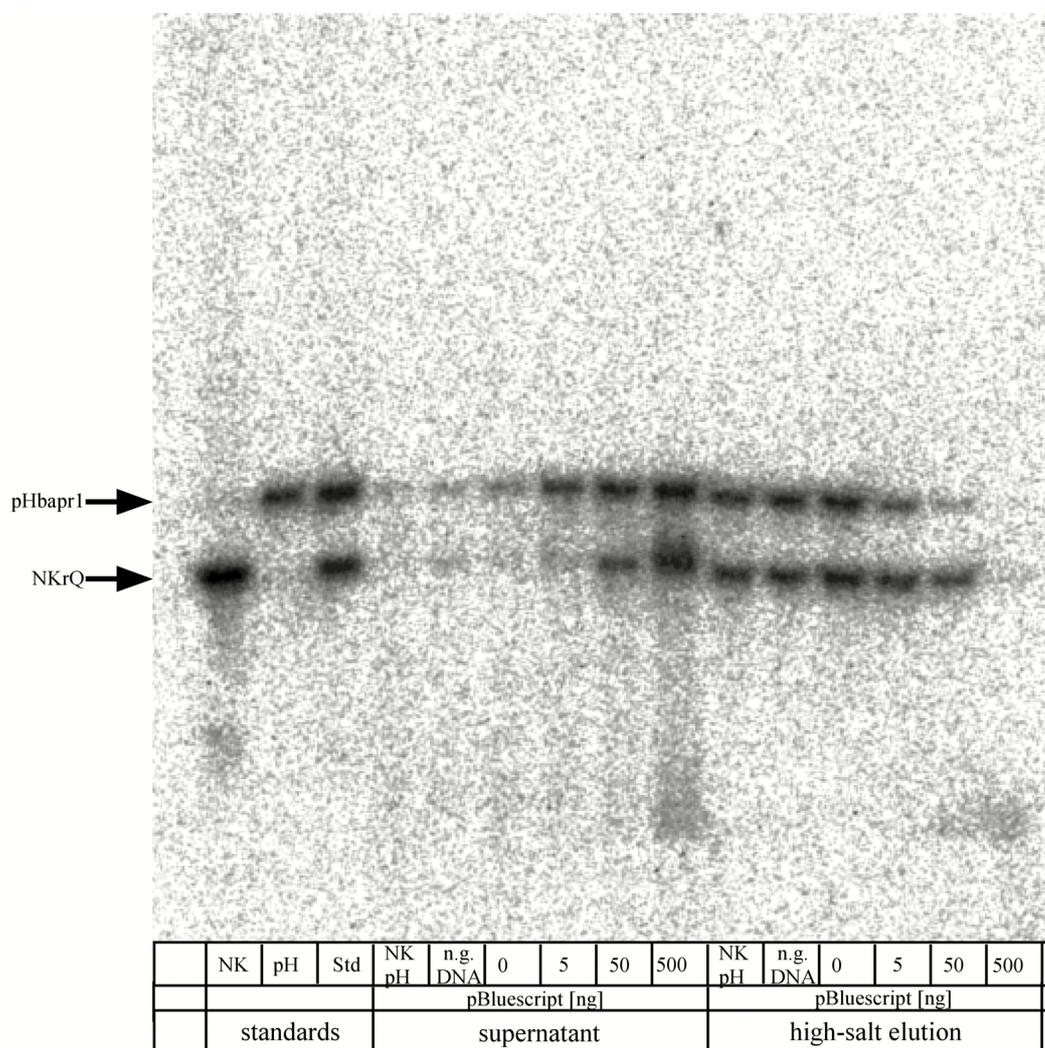


Figure 3-13. M450^{Iz}-*lacOp* rescue: Blocking of unspecific binding through addition of pBluescript. Depicted is a picture of a Southern blot analysis. As target sequence the linearised plasmid NKrQ (NK) was used, as tracer for non-specific binding the linearised plasmid pHbapr1 (pH) was used (see Material and Methods). NK: linearised plasmid NKrQ. PH: linearised plasmid pHbapr1. Std: standard (sample derived from mastermix, of which equal volume fractions went into the rescue assay). n.g. DNA: no genomic DNA present. M450^{Iz}-*lacOp* rescue: per sample of 20µl volume (10mM Tris/HCl pH7,6; 10mM MgCl₂; 5% Glycerol; 0,01%BSA) 500ng *EcoR* I/*Bam* HI digested genomic DNA of chicken erythrocytes (*Gallus domesticus*; equal to 10⁵ diploid genomes), 500pg of the linearised plasmid NKrQ, 500pg of the linearised plasmid pHbapr1 (each ca. 10⁸ copies) and 700ng of LacIz were mixed and incubated on a rotator at room temperature for 90 minutes. Subsequently this mixture was added to 10⁶ pelleted M450-Dynabeads[®], which had been armed under saturating conditions with anti-β-Galactosidase monoclonal antibodies, and incubated for further 90 minutes as before. The supernatants were taken off, and the beads rinsed two times (1volume; 10mM Tris/HCl pH7,6; 10mM MgCl₂; 5% Glycerol; 0,01%BSA). Finally the beads were incubated with 20µl high molar salt buffer (1M NaCl). All supernatants were taken away, transferred into equal salt conditions and analysed by southern blotting. As radioactively labelled probe served a *Bam* HI/*Sal* I fragment of the neomycin resistance gene (see Materials and Methods). As a radioactive probe a *Bam* HI/*Sal* I fragment of the neomycin resistance gene was used (see Materials and Methods). The neomycin resistance gene is present both in NKrQ and pHbapr1. Exposure time: 3h on a phosphorimager screen.

The specificity of binding of NKrQ could indeed be increased to a certain extent. The best ratio of target yield to purity was obtained under the conditions given, at 50ng pBluescript (figure 3-14 A). However the yield of target sequence was reduced compared to

the samples which did not contain any or less pBluescript (figure 3-14 B)). Theoretically 10^{12} binding sites for the *lac-operator* were provided per sample via LacIZ. 50ng of pBluescript is equivalent to ca. 10^{10} molecules. This means that upon addition of 50ng of pBluescript there is still a 100 fold excess of *lac-operator* binding sites. Due to the high degree of variation, and generally unsatisfactory performance both in enrichment and yield of target sequence, it was decided to use only 5ng of linearised pBluescript, and to test in further experiments the influence of buffer components at various steps of the protocol with respect to the efficiency of the M450^{IZ}-*lacOp* rescue.

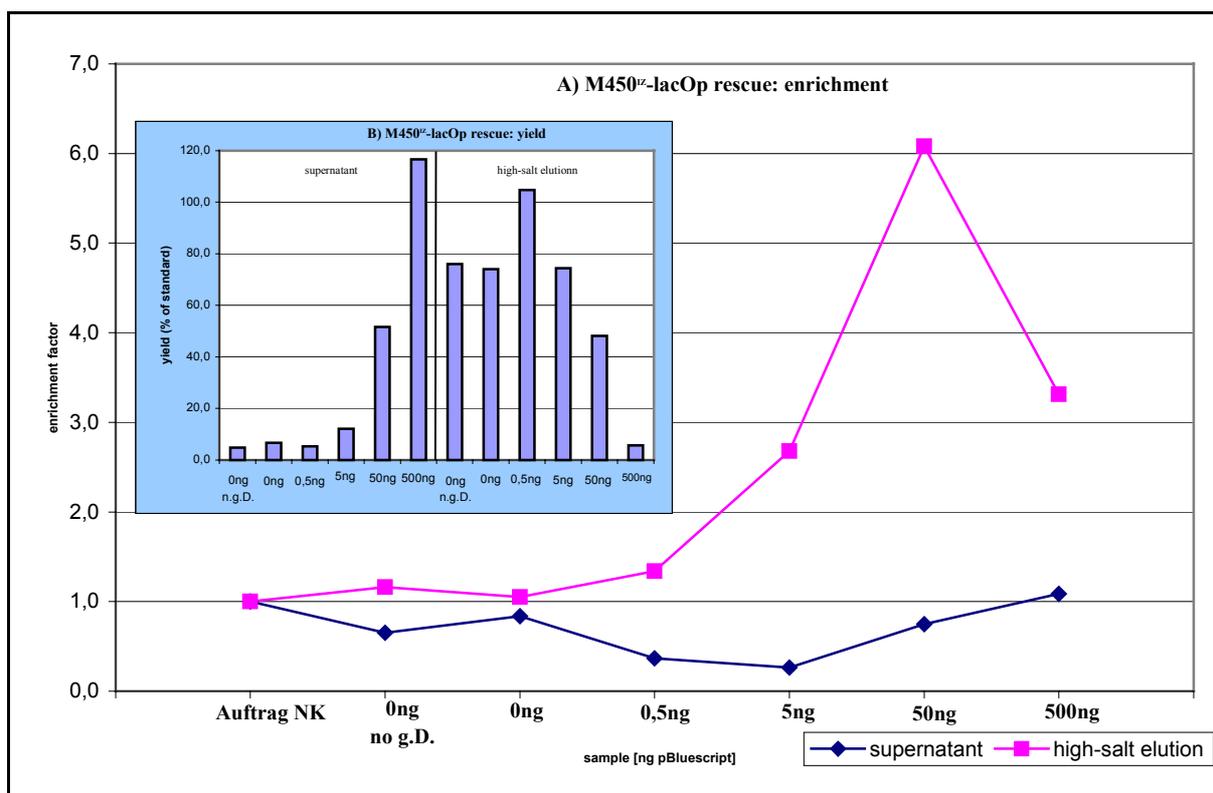


Figure 3-14. M450^{IZ}-*lacOp* rescue: Blocking of unspecific binding through addition of pBluescript. Graphical evaluation of experiments as depicted in Figure 3-13. Shown is in A) the enrichment of target sequence per sample. In B) the fraction of target-DNA (NKrQ; per cent of Standard) which was detected and measured in the samples of the respective time points. The quantification was done with a phosphoimager (Software: TINA v2.08); evaluation with EXCEL2000. n.g.D.: no genomic DNA present.

Specificity of uptake in relation to buffer salt concentration

Extensive studies have been performed on the binding of LacI to DNA in relation to salt concentration. High salt concentrations lead to a general decrease in affinity of the Lac repressor to DNA (Lin *et al.* 1972; Lin *et al.* 1975). Especially monovalent cations like K^+ and Na^+ seem to weaken electrostatic interactions between operator and repressor (Gilbert *et al.* 1967; Riggs *et al.* 1970; Riggs *et al.* 1970), while bivalent cations like Mg^{2+} and Mn^{2+} at low concentrations (mM range) have an enhancing influence on DNA binding of LacI. It was decided to evaluate for the M450^{IZ}-*lacOp* rescue assay whether the specificity of uptake of

DNA could be improved by changing the concentration of KCl and MgAc in the binding buffer.

a) KCl: As the specificity of uptake of NKrQ to M450^{Iz}-Dynabeads[®] was not satisfactory, we tested whether it could be improved upon increasing the salt concentration in the binding buffer, through addition of potassium chloride (figure 3-15). The addition of KCl had a strong effect on the specificity of target binding. The best enrichment (ca. 35-fold against pHβapr1) was achieved at 200mM KCl (figure 3-16).

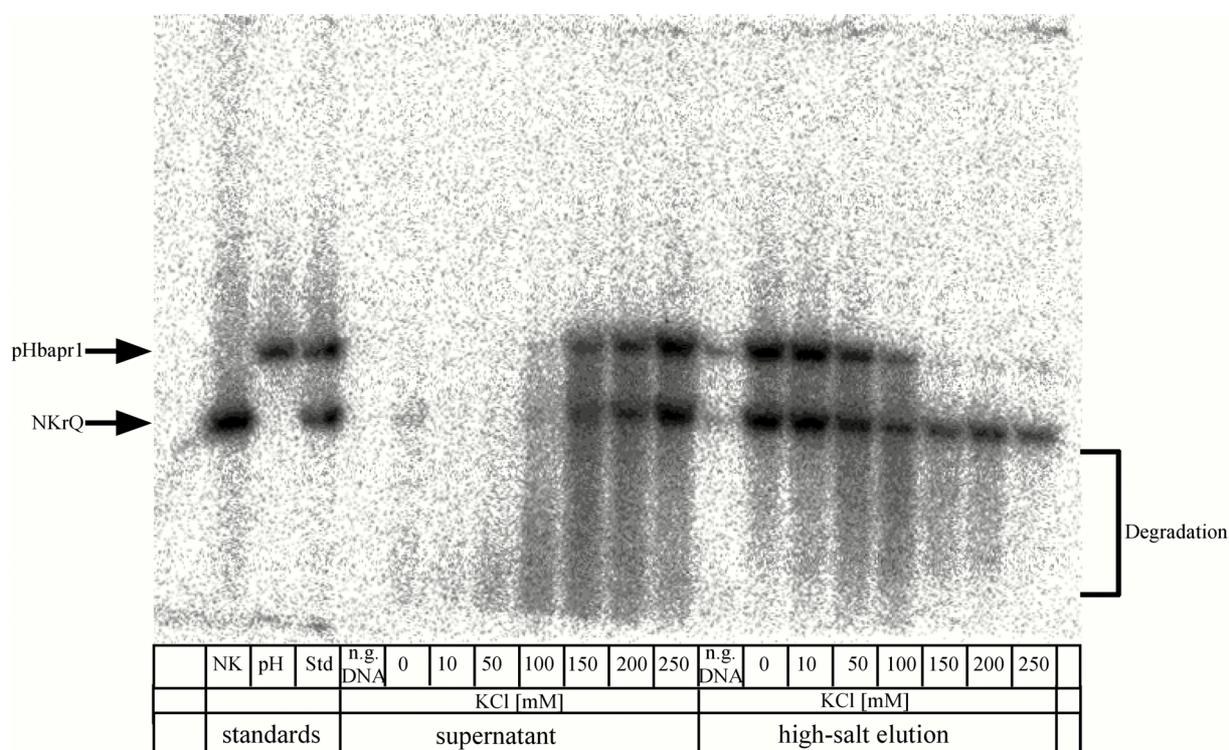


Figure 3.15. Specificity of binding in dependence of salt concentration (KCl). Depicted is a picture of a Southern blot analysis. As target sequence the linearised plasmid NKrQ (NK) was used which carries the *lac-operator* sequence, as tracer for non-specific binding the linearised plasmid pHβapr1 (pH) was used. (see Material and Methods). NK: linearised plasmid NKrQ. PH: linearised plasmid pHβapr1. Std: standard (sample derived from mastermix, of which equal volume fractions went into the rescue assay). n.g. DNA: no genomic DNA present. M450^{Iz}-*lacOP* rescue: per sample (20μl volume; 10mM Tris/HCl pH7; 10mM MgCl₂; 5% Glycerol; 0,01%BSA, KCl as indicated in legend), 500ng *Eco* RI/*Bam* HI digested chicken genomic DNA (*Gallus domesticus*; equivalent to 10⁵ diploid genomes), 500pg of the linearised Plasmid NKrQ, 500pg of the linearised plasmid pHβapr1 (each 10⁸ copies), and 700ng LacIz were incubated at room temperature for 90 minutes on a rotator. This mixture was added to 10⁶ pelleted M450-Dynabeads[®], which had been armed with anti-β-Galactosidase monoclonal antibodies, and incubated for another 90 minutes as before. The supernatants were taken off and the beads rinsed twice (1 volume; 10mM Tris/HCl pH7,6; 10mM MgCl₂; 5% Glycerol; 0,01%BSA). Finally the beads were incubated with 20μl high molar Salt buffer (1M NaCl). All supernatants were taken off, transferred into equal salt conditions and analysed by southern blotting. As a radioactive probe a *Bam* HI/*Sal* I fragment of the neomycin resistance gene was used. The neomycin resistance gene is present both in NKrQ and pHβapr1. Exposure: 15 minutes on a phosphoimager screen.

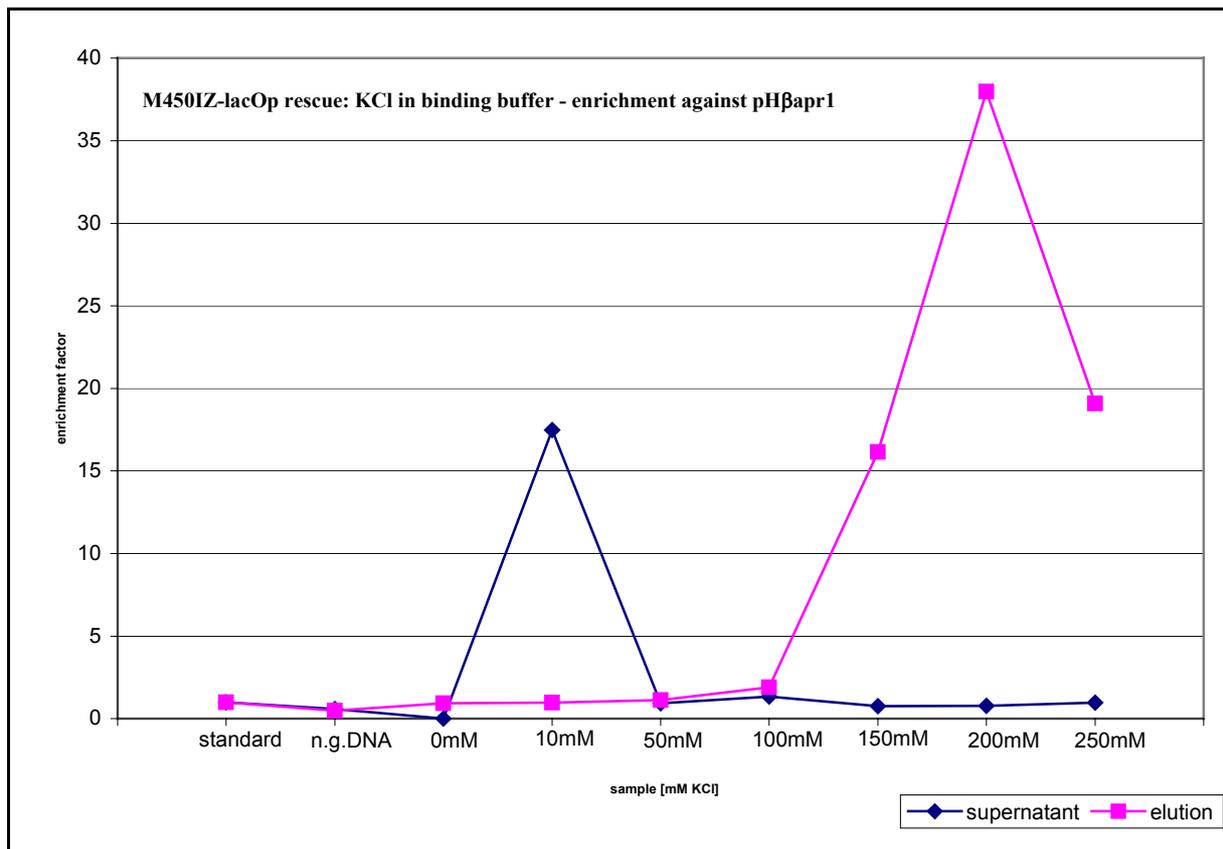


Figure 3-16. M450IZ-lacOp rescue: Influence of KCl concentration in the binding buffer on enrichment. Graphical evaluation of Figure 3-15. Plotted are the enrichment factors of the target sequence NKrQ, obtained with respect to the evaluation plasmid pH β apr1, at different concentrations of KCl, being present in the binding buffer. The values are plotted either for the supernatant after the binding step of the M450^{IZ}-lacOp rescue or for the elution step. Calculation of enrichment: The value of intensity (PSL-Photostimulated Luminescence), as calculated by the computerprogram TINA v2.08, per signal of NKrQ is divided by the intensity for pH β apr1 per sample. Each ratio obtained is subsequently divided by the ratio for the standard sample (see also Materials and Methods).

b) MgAc: The original protocol of Gossen et al. the binding buffer does not contain MgAc (Gossen *et al.* 1993). For reasons outlined above it was examined whether the addition of MgAc to the binding buffer influences the efficiency and specificity of the M450^{IZ}-lacOp rescue. Two experiments were done under exactly the same conditions. The result of one of the two experiments is shown in figure 3-17. As analysed in figure 3-18 a maximum of specificity was achieved at a MgAc concentration of 10mM. The enrichment factors vary considerably, depending on which reference DNA was used. It lies at a maximum of 100 relative to pH β apr1 (at 10mM MgAc); 50 relative to pSP72 (at 10mM MgAc) and is greater than 250 relative to pSP72u (at 10mM MgAc; pSP72u: unspecified DNA fragment in the sample of pSP72 (see figure 3-9)). For later experiments a final MgAc concentration of 10mM was taken. As can be seen from the high-salt elutions (figure 3-17 B)), the IPTG elution step only worked suboptimally in terms of yield. The maximal yield of target sequence lies at about 30% of the output at 3mM MgAc, and drops sharply at MgAc concentrations higher than 20mM (figure 3-19). It was attempted to improve it (see below). However it has to be stressed again that the major concern, with reference to the M450^{IZ}-lacOp rescue, in the

context of this project, is not yield but purity (enrichment) of the target sequence, for the reasons outlined before (see introduction).

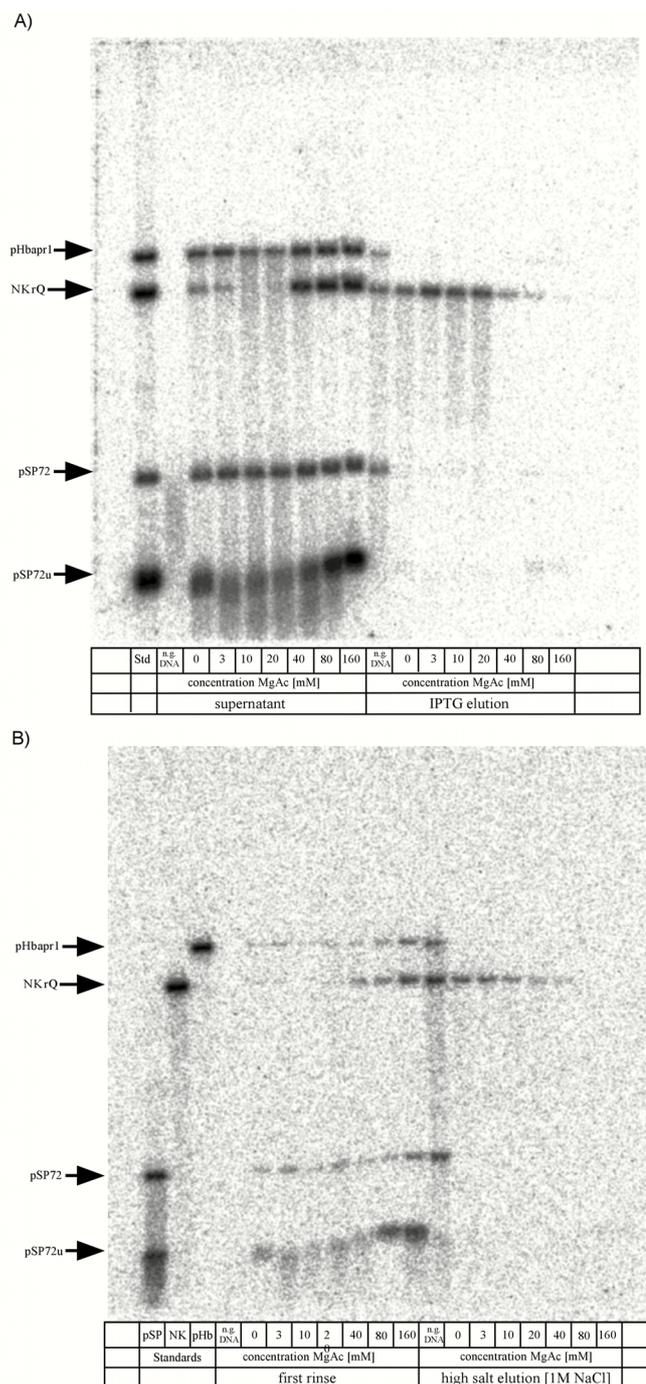


Figure 3-17. Binding specificity of target sequence in dependence of the MgAc concentration. Depicted is a picture of a Southern blot analysis. A) Supernatant and IPTG elution step. B) First rinsing step and high salt elution. As target sequence the linearised plasmid NKrQ was used which carries the *lac-operator* sequence, as tracer for non-specific binding the linearised plasmid pH β apr1, and pSP72 were used. (see Material and Methods). n.g.DNA: no genomic DNA. Std: standard (sample derived from mastermix, of which equal volume fractions went into the rescue assay). n.g. DNA: no genomic DNA present. M450^{IZ}-*lacOp* rescue: per sample (20 μ l volume; 10mM Tris/HCl pH7; 150mM KCl; 5% Glycerol; 0,01%BSA, MgAc as indicated in the legend), 500ng *EcoR* *V*/*Bam* HI digested chicken DNA (*Gallus domesticus*; equivalent to 10⁵ diploid genomes), 500pg of the linearised plasmid NKrQ, 500pg of the linearised plasmid pH β apr1, 500pg of the linearised plasmid pSP72 (each plasmid equals approximately 10⁸ copies), and 1 μ g LacZ were incubated for 90 minutes at room temperature on a rotator. This mixture was added to 10⁶ pelleted M450-Dynabeads[®], which had been armed with anti- β -Galactosidase monoclonal antibodies, and incubated for another 90 minutes as before. The supernatants were taken off and the beads rinsed twice (20 μ l Volume; 10mM Tris/HCl pH7; 150mM KCl; 5% Glycerol; 0,01%BSA, MgAc as indicated in the legend). Elution was performed with 1mM IPTG (20 μ l, 10mM Tris/HCl pH7; 50mM MgAc; 5% Glycerol; 0,01%BSA, 1mM IPTG). Finally the beads were incubated in 1 volume of high molar salt buffer (1M NaCl). 15 μ l of the supernatants of all steps went into the Southern blot analysis. As a radioactively labelled probe served a *Bgl* *I*/*Sca* I fragment of the β -*lactamase* gene (see Materials and Methods) which is present in all three plasmids. Exposure time: 5h on a phosphoimager screen.

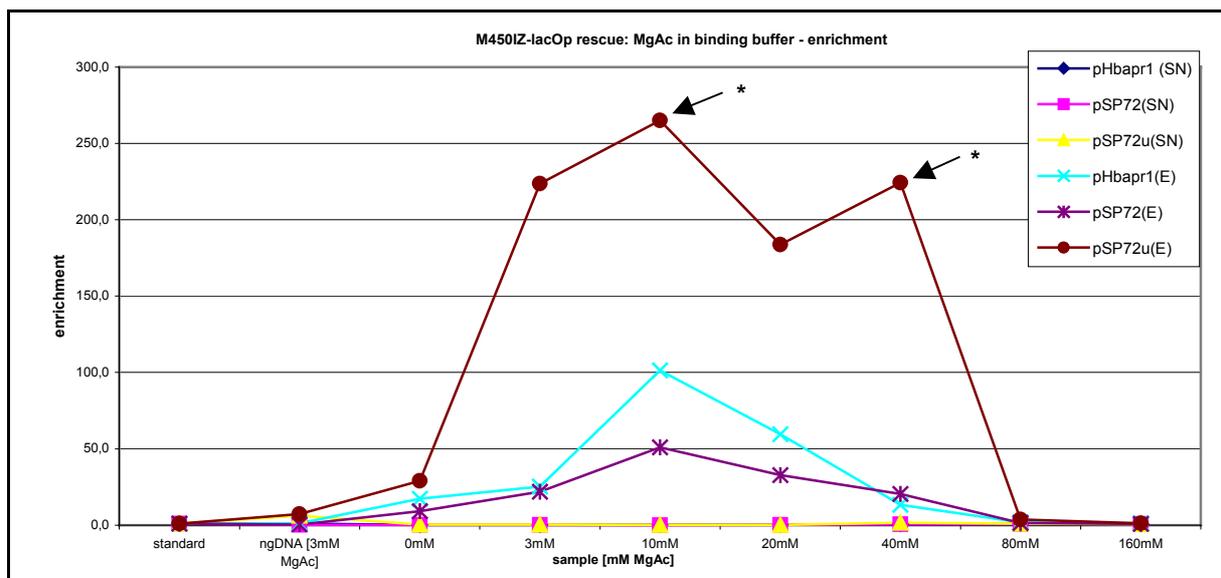


Figure 3-18. M450IZ-lacOp rescue: Influence of MgAc concentration in the binding buffer on enrichment. Graphical evaluation of two experiments (see figure 3-17) Plotted are the mean of enrichment factors of the target sequence NKrQ, obtained with respect to the evaluation plasmids pHbap1 and pSP72, at different concentrations of MgAc, being present in the binding buffer. The values are plotted either for the supernatant after the binding step of the M450IZ-lacOp rescue (SN) or for the elution step (E). Calculation of enrichment: The value of intensity (PSL-Photostimulated Luminescence), as calculated by the computerprogram TINAv2.08, per signal of NKrQ is divided by the intensity for one of the other plasmids of the same sample. Each ratio obtained is subsequently divided by the ratio for the standard sample (i.e. the sample that did not go through the rescue). pHb: pHbap1; pSP72u: unspecified DNA fragment in the sample of pSP72 (see figure 3-9); *: fictive minimal value: at 10mM and 40mM MgAc no pSP72u could be detected in the elution sample, for each concentration in one of the two experiments.

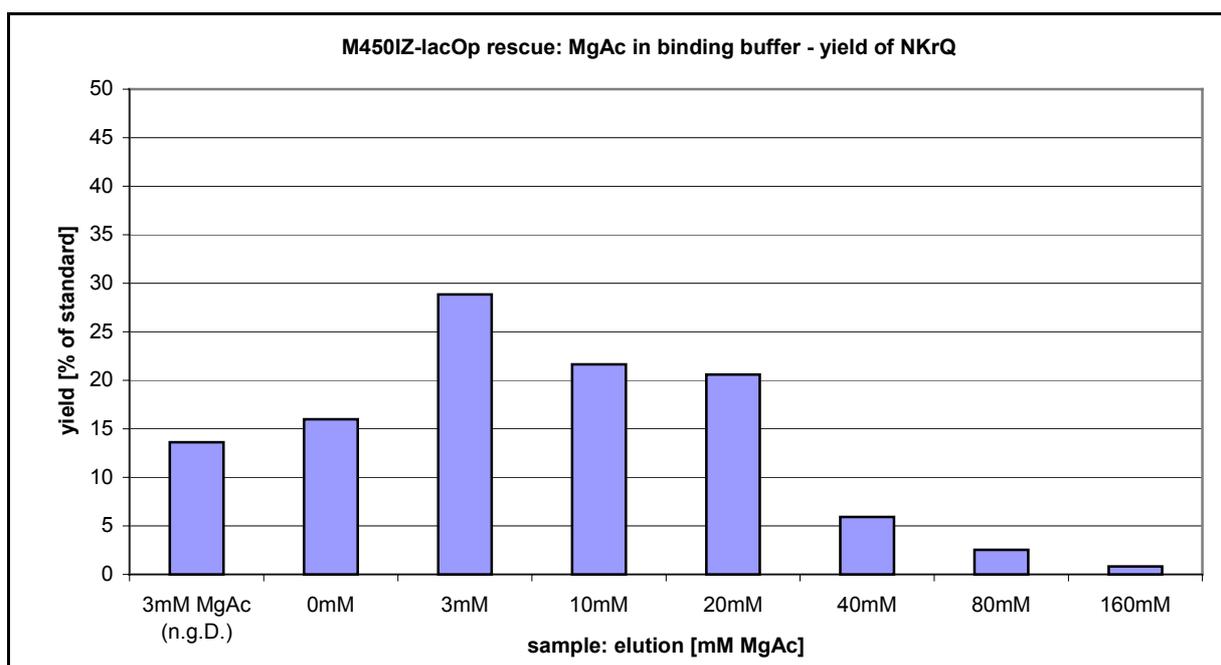


Figure 3-19. M450IZ-lacOp rescue - MgAc in binding buffer: yield Graphical evaluation of Figure 3-17 A) (elution). Plotted is the percentage of NKrQ recovered in the elution steps with respect to the output amount (standard). n.g.D.: no genomic DNA present.

Specific elution of *lac-operator* containing DNA fragments with IPTG

The final elutions for the experiments evaluating the DNA uptake during the M450^I-lacOp rescue were achieved non-specifically with 1M NaCl. Isopropyl-thio-galactoside (IPTG), a synthetic galactoside, and strong inducer of the *lac-operon* system, induces an allosteric conformational change in the Lac repressor which

leads to a 1000 fold reduction of affinity to the *lac-operator in vitro* (Barkley *et al.* 1975; Lewis *et al.* 1996). Experiments were carried out in order to assess whether upon final elution in the presence of IPTG further enrichment of the target sequence NKrQ can be achieved.

Elution in the presence of 1mM IPTG

Firstly the final elution was performed with a buffer containing 1mM IPTG. In parallel again the specificity of the M450^{IZ}-lacOp rescue was tested against KCl concentration in the binding buffer as above. As can be seen in figure 3-20 A) and B), and in the graphical evaluation (figure 3-21) the elution of NKrQ was highly specific. Enrichment factors greater than 300 could be obtained. It seems that the most efficient uptake of target sequence occurred at a KCl-concentration of 100 to 150mM (figure 3-20 A)), even though this is not reflected in the yield of target sequence (figure 3-22). This in turn could be due to the observed degradation of DNA which, for the elution samples is most obvious at 100 and 150mM KCl.

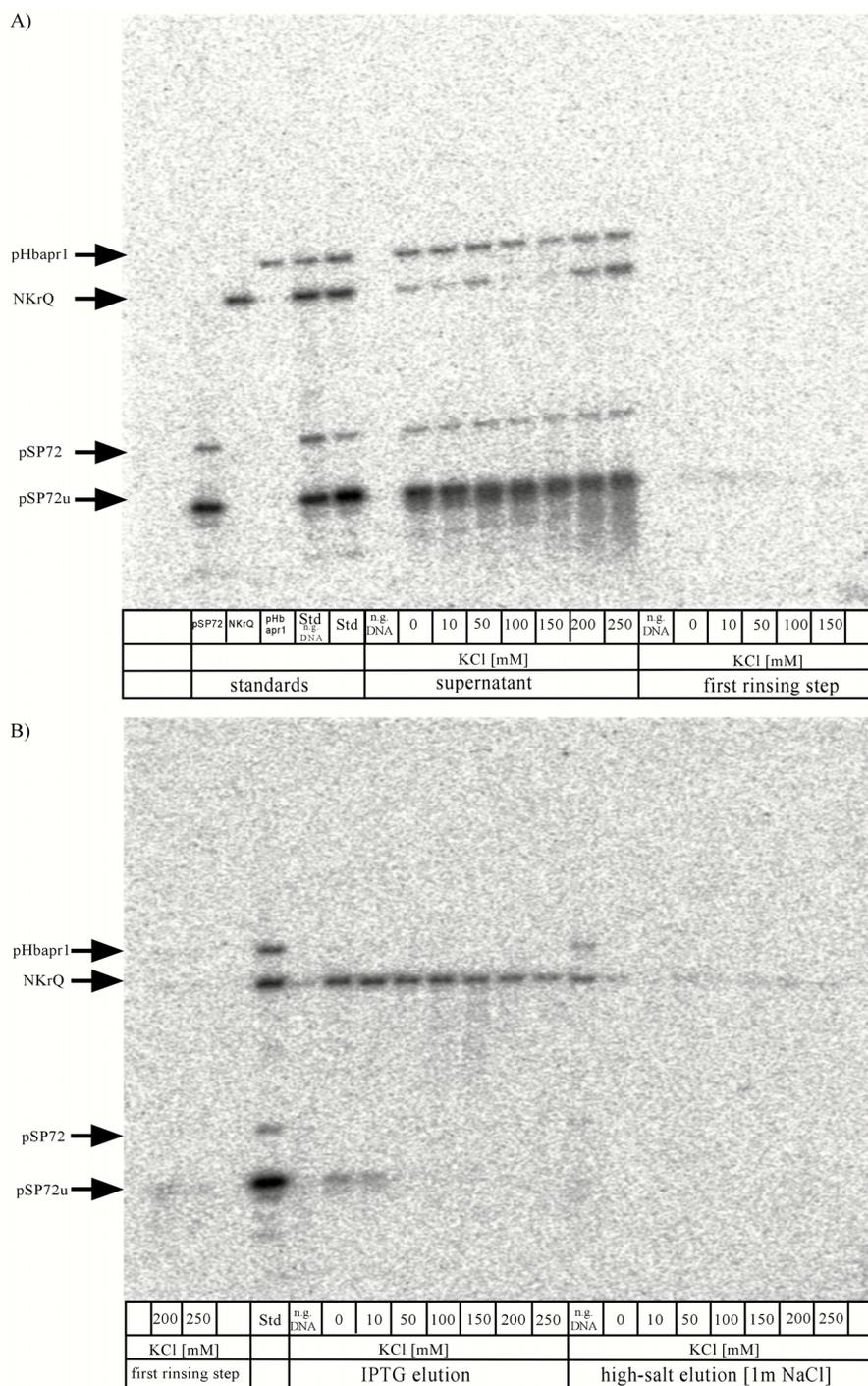


Figure 3-20. M450IZ-*lacOp* rescue - IPTG elution and influence of KCl concentration in the binding buffer on enrichment for NKrQ. A) Supernatant and first rinsing step. B) IPTG-elution and high salt elution. Depicted is a picture of a Southern blot analysis. As target sequence the linearised plasmid NKrQ was used which carries the *lac-operator* sequence, as tracer for non-specific binding the linearised plasmid pHbapr1; pSP72 and pSP72u were used (see Material and Methods). n.g. DNA: no genomic DNA. Std: standard (sample derived from mastermix, of which equal volume fractions went into the rescue assay). M450IZ-*lacOp* rescue: per sample (20 μ l volume; 10mM Tris/HCl pH7; 10mM MgCl₂; 5% Glycerol; 0,01%BSA, KCl, as indicated in legend), 500ng *EcoR* I/*Bam*HI digested chicken genomic DNA (*Gallus domesticus*; equivalent to 10⁵ diploid genomes), 500pg of the linearised Plasmid NKrQ, 500pg of the linearised plasmid pHbapr1 (see Material and Methods; each 10⁸ copies), 500pg of the linearised plasmid pSP72, and 1 μ g LacIZ, were incubated at room temperature for 90 minutes on a rotator. This mixture was added to 10⁶ pelleted M450-Dynabeads®, which had been armed with anti- β -Galactosidase monoclonal antibodies, and incubated for another 90 minutes as before. The supernatants were taken off and the beads rinsed twice (1 volume; 10mM Tris/HCl pH7,6; 10mM MgCl₂; 5% Glycerol; 0,01%BSA KCl as indicated in the legend). This step was followed by elution with 1mM IPTG (20 μ l; 10mM Tris/HCl pH7; 50mM MgAc; 5% Glycerol; 0,01%BSA, 1mM IPTG). Finally the beads were incubated with 20 μ l high molar salt buffer (1M NaCl). 15 μ l of each step of the assay went into Southern blot analysis. As a radioactively labelled probe served a *Bgl* I/*Sca* I fragment of the *β -lactamase* gene which is present in all three plasmids. Exposure: 8 hours on a phosphoimager screen.

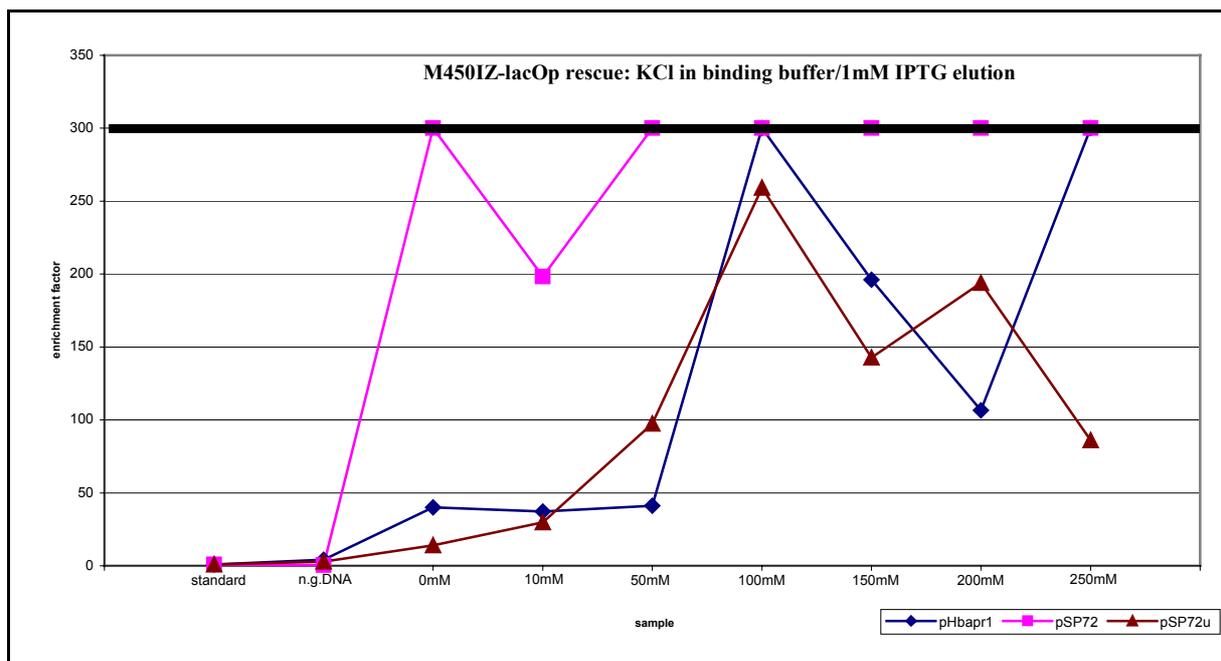


Figure 3-21. M450IZ-lacOp rescue: IPTG elution and influence of KCl concentration in the binding buffer on enrichment for NKrQ. Graphical evaluation of Figure 3-20 B) (elution). Plotted are the enrichment factors of the target sequence NKrQ, obtained with respect to the evaluation plasmid pHbapr1, pSP72, and pSP72u at different concentrations of KCl, being present in the binding buffer. The values are plotted for the supernatant of the elution step. Calculation of enrichment: The value of intensity (PSL-Photostimulated Luminescence), as determined by the computer program TINAv2.08, per signal of NKrQ is divided by the intensity for pHbapr1, pSP72, or pSP72u per sample. Each ratio obtained is subsequently divided by the ratio for the standard sample (see also Materials and Methods). Any value for the enrichment greater than 300 means that there was no trace of the plasmid in question detectable, and therefore the enrichment factor was not determinable.

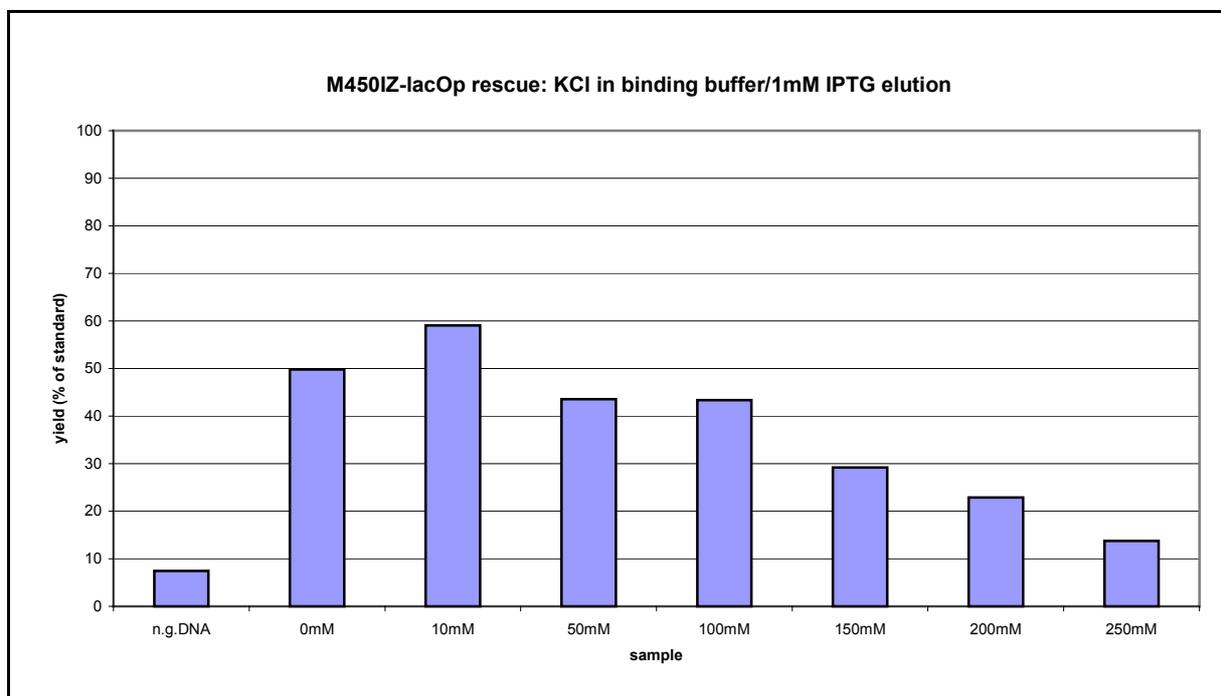


Figure 3-22. M450IZ-lacOp rescue: IPTG elution and influence of KCl concentration in the binding buffer on enrichment for NKrQ - yield. Graphical evaluation of Figure 3-20 A) (elution). Plotted is the percentage of NKrQ being recovered in the elution steps with respect to the output amount (standard).

Effect of Mg^{2+} ions on efficiency and specificity of elution

It was reported before that the presence of Mg^{2+} ions lowers the binding constant of the Lac-repressor complex to *lac-operator* sequence (Barkley *et al.* 1975). Therefore we explored for the M450^{IZ}-*lacOp* rescue assay whether and to what extent addition of MgAc to the elution buffer has a positive effect on the elution step. As above per assay the elution was performed in the presence of 1mM IPTG. Additionally the elution buffer contained increasing concentrations of MgAc. IPTG (MW=238) is present at approximately a 1000fold surplus over LacIZ (MW=150 000). The results are shown in figure 3-23. Maximal elution efficiency was obtained in the presence of 100mM to 150mM MgAc, which is nicely reflected in the lack of remaining target sequence NKrQ in the high salt washes (see figure 3-23 B) and figure 3-24 for graphical illustration). The enrichment in this experiment was very high, with no detectable traces of the evaluation plasmids pSP72 or pH β apr1 (see table 3-4). The maximal yield of the target sequence was obtained at 100 and 150mM MgAc with 20 to 30% of the output.

These promising results encouraged us to apply the optimised protocol for the M450^{IZ}-*lacOp* rescue, to conditions which are closer to our ultimate goal, that is to isolate single copy gene targets out of pools of genomic DNA. This approach is described in the next section.

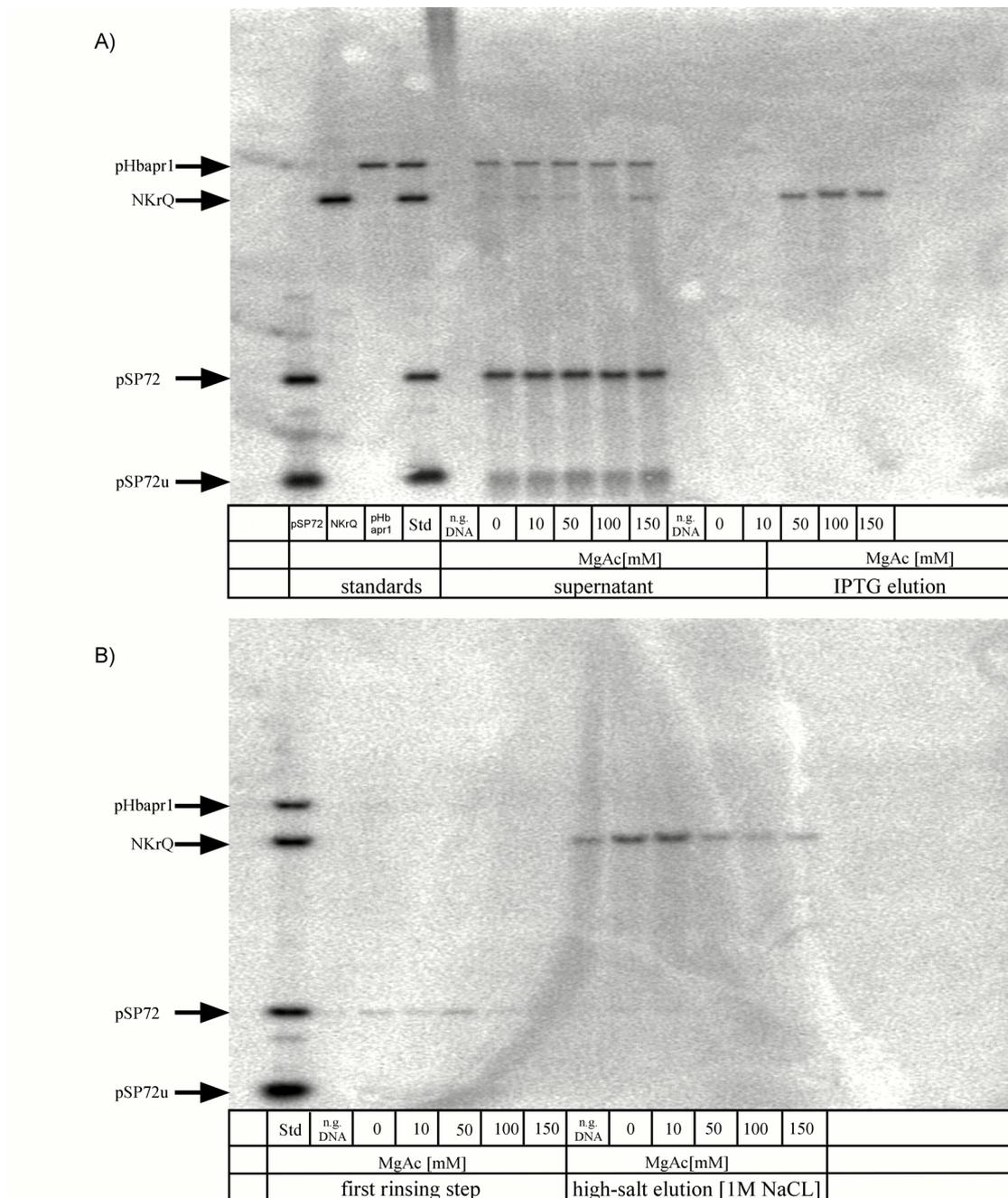


Figure 3-23. M450IZ-lacOp rescue - IPTG elution efficiency in relation to MgAc concentration. Depicted is a picture of a Southern blot analysis. A) Supernatant and IPTG elution; B) First rinsing step and high-salt [1M NaCl] elution. As target sequence the linearised plasmid NKrQ was used which carries the *lac-operator* sequence, as tracer for non-specific binding the linearised plasmid pH β aprl; pSP72 and pSP72u were used. n.g.DNA: no genomic DNA. Std: standard (sample derived from mastermix, of which equal volume fractions went into the rescue assay). M450IZ-lacOp rescue: per sample (20 μ l volume; 10mM Tris/HCl pH7; 150mM KCl; 10mM MgAc; 5% Glycerol; 0,01%BSA), 500ng *Eco* RI/*Bam* HI digested erythrocyte chicken DNA (*Gallus domesticus*; equivalent to 10⁵ diploid genomes), 500pg of the linearised plasmid NKrQ, 500pg of the linearised plasmid pH β aprl, 500pg of the linearised plasmid pSP72 (each plasmid equals approximately 10⁸ copies), 5ng of the *Eco* RI linearised plasmid pBluescript, and 1 μ g LacIZ were incubated for 90 minutes at room temperature on a rotator. This mixture was added to 10⁶ pelleted M450-Dynabeads[®], which had been armed with α - β -galactosidase monoclonal antibodies, and incubated for another 90 minutes as before. The supernatants were taken off and the beads rinsed twice (20 μ l Volume; 10mM Tris/HCl pH7; 150mM KCl; 10mM MgAc, 5% Glycerol; 0,01%BSA). Elution was performed with 1mM IPTG (20 μ l, 10mM Tris/HCl pH7; 5% Glycerol; 0,01%BSA, 1mM IPTG, MgAc as indicated in the legend). Finally the beads were incubated in 20 μ l of high molar salt buffer (1M NaCl). 15 μ l of the supernatants of all steps went into Southern blot analysis. As a radioactively labelled probe served a *Bgl* I/*Sca* I fragment of the β -lactamase gene which is present in all three plasmids.Exposure: 8 hours on a phosphoimager screen.

	NKrQ (PSL)	pHbaprl (PSL)	pSP72u (PSL)	pSP72 (PSL)
standard	1266,64	728,82	2717,39	1219,5
n.g. DNA /50mM MgAc	-5,76	0,12	-39,07	-6,45
0mM MgAc	11,23	-23,68	-32,15	-10,49
10mM MgAc	19,19	-34,56	-4,89	-6,65
50mM MgAc	244,15	-3,52	20,18	-0,02
100mM MgAc	365,59	11,68	-6,09	-5,52
150mM MgAc	284,95	-4,83	-78,12	-19,31

Table 3-4. Signal intensities (PSL-Photo Stimulated Luminescence) in elution samples (1mM IPTG; MgAc as indicated in respective lines). Listed are the signal intensities per plasmid and sample after background subtraction of figure 3-23.

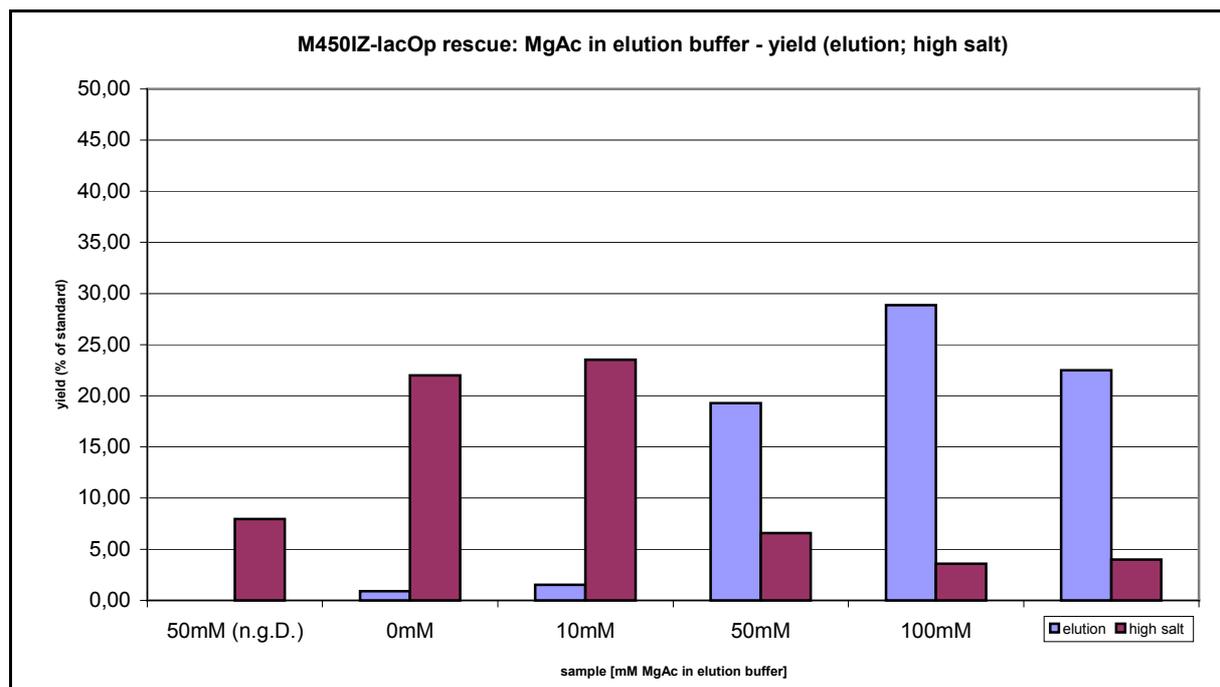


Figure 3-24. Comparison of yield of the target sequence NKrQ in the IPTG-elution (elution) step and the 1M NaCl (high salt) elution. Plotted is the percentage of the standard signal intensities measured for the respective samples.

3.1.4. Target extraction out of pools of genomic DNA: single copy gene extraction

With 10^3 copies per genome the ratio of target sequence copy number to genome number in the optimising experiments of section 3.1.3. was a thousand fold higher than the ratio we are aiming at. We explored therefore whether the extraction still produces satisfying results, if only one copy of target sequence is available per genome. The total amount of genomic DNA was kept at 500ng representing 10^5 diploid genomes, since we planned to perform an analysis on pools of 10^5 cells (see discussion). Due to the low starting amount of target sequence no direct assessment as in the optimising experiments presented before was possible. Therefore the evaluation was based on PCR of the samples that were analysed.

As target sequence 10^5 copies (700fg) of a 2kb long DNA-fragment were used, which contains the *E.coli lac-operator* sequence, as well as the first three exons of the *H2 D^b* gene ("ND" – see Material and Methods). As a reference, in order to evaluate the specificity of the target extraction 10^5 copies of a 2kb fragment were used which does not contain the

lac-operator sequence ("8-8PST" – see Material and Methods). Both fragments were mixed with 500ng *Eco* RI/*Bam* HI digested genomic erythrocyte chicken DNA (*Gallus domesticus*; equivalent to 10^5 diploid genomes). As has been shown in above, in one cycle of the M450^{Iz}-*lacOp* rescue assay for a target gene an enrichment factor in the order of 10^2 relative to a contaminating sequence can be obtained. However, despite this impressive magnitude of enrichment, this is not sufficient for our purposes, since we will be working in a multigene system, namely the *H2* (see introduction). In a pessimistic "Gedankenexperiment" we have to reckon with about 30 to 50 copies of (more or less) homologous *H2* genes per target gene (Trowsdale 1995), each of which could potentially interfere with our analysis by leading to in vitro recombinational artefacts resembling the mutations we are looking for (see introduction). If we enrich for the target gene by a factor of 100 relative to each of the remaining *H2* genes, the final ratio of target sequence to potentially "hazardous" sequences would be in the range of 1:1. We decided to apply for the following experiments two cycles of the M450^{Iz}-*lacOp* rescue. In order to evaluate the assay a PCR was performed on the eluted material, and in parallel on standard dilution rows (figure 3-25). The primers used were either specific for "ND" or for "8-8PST". As can be seen from comparison of the amplified standard dilution rows to the eluted material in figure 3-25 A) (first round of the M450^{Iz}-*lacOp* rescue), the signal strength of the tracer 8-8PST is undiluted maximally as strong as the signal arising from the 1:1000 dilution of the standard. On the other hand the signal strength of the target ND at the 1:100 dilution is at least as strong as the signal from the 1:1000 dilution of the standard. This implies an enrichment factor relative to 8-8PST after the first cycle of the assay of at least 10^2 .

Applying the same evaluation to the second cycle of the rescue assay a comparison of 8-8PST and ND to their respective standard dilution rows lead to an estimation of an enrichment factor of about 10^4 for the target sequence "ND", and a yield of 25% of the starting material (see figure 3-25 B)).

The enrichment achieved after two cycles of the M450^{Iz}-*lacOp* rescue in the model system presented is very promising. The result allowed us to pursue one further evaluating experiment, which finally mimics the ultimate goal of treating pools of germline derived cells with the M450^{Iz}-*lacOp* rescue assay in order to ultimately screen these for mutations. We have in our hands transgenic ES cells which harbour the *K^{blNC}* transgene (see introduction). We decided to assess whether starting from small pools of these cells the isolation of endogenous target sequence (the *K^{blNC}* gene) would be as effective as in the model experiments described above. This approach is described in the next section.

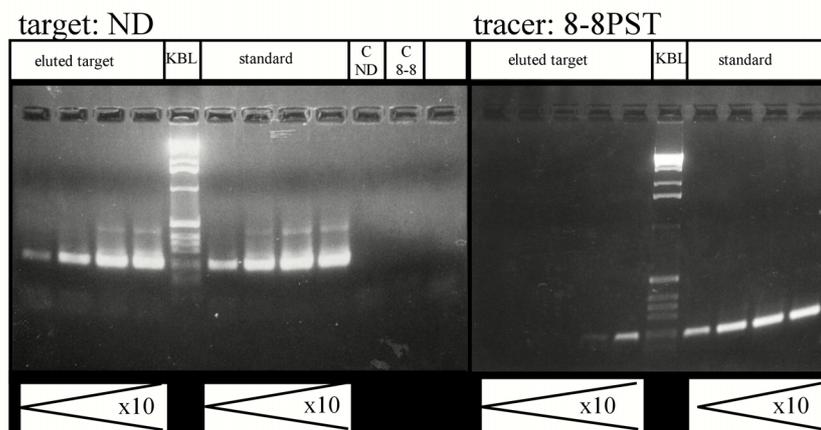
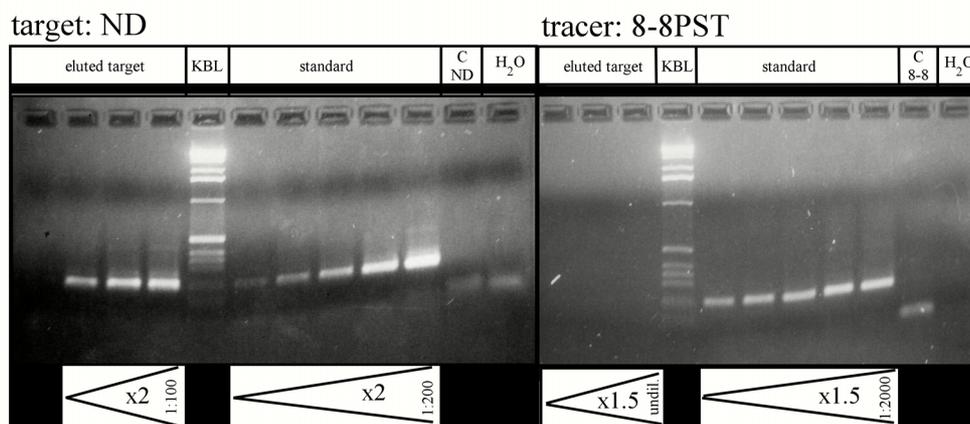
A) first round of M450^IZ-*lacOp* rescueB) second round of M450^IZ-*lacOp* rescue

Figure 3-25. M450^IZ-*lacOp* rescue in the context of genomic DNA: single copy gene extraction. Depicted are agarose gels (2%; 1XTBE; staining: 0.01% Ethidiumbromide). Loading volume: 30 μ l. A) first round of the M450^IZ-*lacOp* rescue. Eluted target: 10 fold dilution row of ND and 8-8PST after elution. Standard: tenfold dilution row of ND and 8-8PST starting with 10⁵ copies. C: negative control; chicken DNA that did neither contain ND nor 8-8PST, but which went through the enrichment protocol in parallel. The negative controls were subsequently amplified for ND (CND) or 8-8PST (C8-8). B) second round of the M450^IZ-*lacOp* rescue. Evaluation as in A), only that for ND a twofold dilution series starting from 1:100 was used, and for 8-8PST a 1.5 fold dilution row starting from 1:2000. Standard: Material of a sample mastermix which did not undergo the rescue protocol.

The rescue assay was performed as follows: 10⁵ copies of the lac operator carrying target fragment "ND" were mixed with 10⁵ copies of the *lacOp* negative fragment "8-8PST", and 500ng *EcoR* I/*Bam*H I digested chicken erythrocyte DNA (equal to about 10⁵ diploid genomes). Two rounds of the M450^IZ-*lacOp* rescue were employed. Binding to LacI Z: 20 μ l Volume; 1 μ g LacI Z; 10mM Tris/HCl pH7, 10mM MgAc, 150mM KCl; 5% Glycerol, 0.01% BSA, 5ng pBluescript; 90 minutes at room temperature on a rotator. Binding to anti- β -Galactosidase armed Dynabeads[®]: 10⁶ beads, 1 μ g LacI Z; 10mM Tris/HCl pH7, 10mM MgAc, 150mM KCl; 5% Glycerol, 0.01% BSA; 90 minutes on a rotator at room temperature. Rinsing: two times one volume; 10mM Tris/HCl pH7, 10mM MgAc, 150mM KCl; 5% Glycerol, 0.01% BSA. Elution: 20 μ l; 10mM Tris/HCl pH7, 150mM MgAc; 5% Glycerol, 0.01% BSA, 1mM IPTG; 30 minutes on a rotator at room temperature. The eluted material was transferred into 1 X binding buffer via Sephadex columns (Sephadex G-25; see Material and Methods), and a second cycle was performed as above. Evaluation was performed with PCR on dilution rows of eluted material and standards. A) Evaluation of first cycle of M450^IZ-*lacOp* rescue. PCR: 50 μ l; 2.5U Taq; 5 μ l template; 30 cycles; ND/GC40D for ND and AK/GC40K for 8-8PST (see Materials and Methods) B) Evaluation of second cycle of M450^IZ-*lacOp* rescue. PCR: 50 μ l; Taq; 2 μ l template; 10 cycles; ND/GC40D for ND and AK/GC40K for 8-8PST. Dilution rows as indicated.

3.1.5. Target extraction out of pools of cells: Isolation of the Kb-LNC transgene from LNC28.28 cells

The M450^IZ-*lacOp* rescue represents an element in the frame of a project which aims at the development of a streamlined protocol which would allow the isolation of *lac-operator* tagged MHC genes from the genome in order to facilitate the search for mutations in *H2*

genes of the mouse germline. Therefore primordial germcells or sperm would be used for the definitive analysis. In principle, however any cell type can be used for such an analysis.

In this laboratory Gaby Vopper and Dr. Lisbeth Guethlein generated embryonic stem cells which carry in the third intron of the *H2 K^b* gene the *E.coli lac-operator* sequence (K^{bLNC} , see also Introduction). These cells were generated by homologous replacement techniques (“knock-in”) and served ultimately to create transgenic mice (see Introduction and Materials and Methods). We wished to know whether, starting from these transgenic cells, the K^{bLNC} gene could be efficiently extracted from the mouse genome, using the M450^{IZ}-*lacOp* rescue protocol developed in the previous sections.

Genomic DNA was isolated from aliquots of 10^5 LNC28.28 cells, and digested with the restriction enzymes *Eco* RI and *Hind* III. This double digest creates a fragment of 3.8kb length which containing the first three exons, as well as the *lac-operator* sequence of K^{bLNC} (Lisbeth Guethlein and Gaby Vopper, unpublished data). The digested genomic DNA went through two cycles of the M450^{IZ}-*lacOp* rescue (see Materials and Methods). Afterwards PCR was employed to evaluate the assay. As a negative control, wildtype ES cells (ES14.1, (Kuhn *et al.* 1991)) were used. The K^{bLNC} -target sequence was recovered successfully as shown by amplification on the final IPTG-elution step for the target sequence (figure 3-26).

For the purpose of estimating the degree of suppression of contaminating DNA-sequences 18S rDNA was amplified. Of the 18S rDNA genes around 100 copies are present per haploid mouse genome (Alberts *et al.* 1994). In a first evaluation, equal amounts of supernatant derived from the first incubation of the genomic DNA with M450 Dynabeads and material from the final IPTG-elution step were amplified in parallel. Strong signals appeared for the supernatants, while no signals were detectable for the eluted material in agarose gels (see figure 3-27 A)). The evaluation was further extended through diluting the supernatant samples down to 1/25 600, and PCR was performed on the respective dilution steps in parallel to the undiluted material of the IPTG-elution (figure 3-27 B)). While in the supernatant samples a signal can be detected at least down to 1/12 800, no signal is detected in the elution samples. This shows clearly that the M450IZ-*lacOp* rescue is capable of significantly suppressing the presence of contaminating DNA-sequences.

The result presented here must certainly be regarded as the outcome of a pioneering experiment. It suggests strongly the feasibility of the approach for our purposes, but needs further exploration.

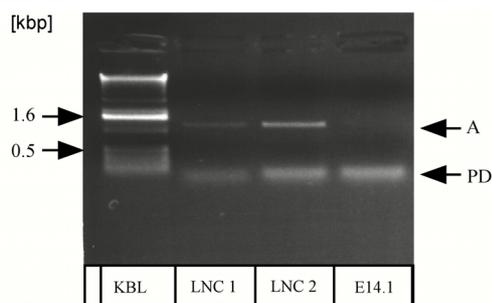
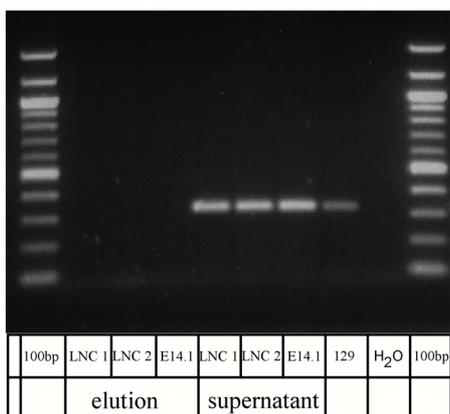


Figure 3-26. M450^{Iz}-lacOp rescue on of LNC28.28 ES cells: recovery of K^{bLNC} . Shown is the picture of an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide). 20 μ l of a 50 μ l PCR reaction were loaded: 1/10 of the final elution step were amplified with the primer pair LNCV/LNCK (50 μ l; 1.54U Pfu (Promega); 5pmol primer each; 35 cycles). A: Amplicon. PD: Primer Dimer.

A)



B)

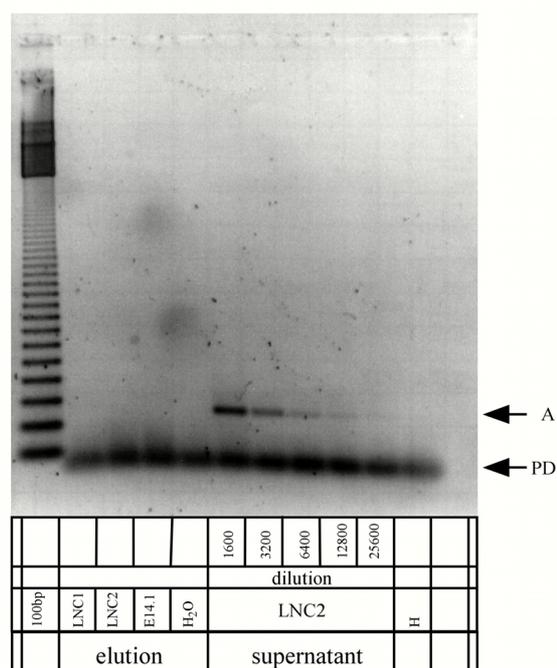


Figure 3-27. M450^{Iz}-lacOp rescue on of LNC28.28 ES cells: suppression of 18S rDNA. A) Depicted is the picture of an agarose gel (2%; 1XTAE; staining: 0.01% Ethidiumbromide). 20 μ l of a 50 μ l PCR reaction were loaded. 1/10 volume of the final IPTG-elution step (elution) as well as the material collected after the first incubation of the digested cell derived DNA with M450^{Iz} beads (supernatant), were amplified with the primer pair 18S-5'/18S-3' (50 μ l; 1.5U Taq; 2.5mM MgCl₂; 5pmol primer each; 35 cycles). LNC1, 2: two independent samples of 10⁵ LNC28.28 cells which went through two cycles of a M450^{Iz}-lacOp rescue. E14.1: wildtype ES-cells (negative control). 129: genomic DNA of 129sv mice (dilution 1/100; kindly provided by Iana Parnanova). B) Elution and supernatant (see A) after further amplification. The supernatant was diluted down to 1/25 600. Two rounds of PCR were employed. The first round as in A). 20 μ l of the samples derived from first amplification were purified (PCR-High Pure Purification Kit (Roche)), and eluted into 100 μ l 10mM Tris/HCl pH8. 1 μ l of the eluted material was amplified for additional 25 cycles as before. 20 μ l were loaded onto an agarose gel (1.5%; 1XTAE; staining: 0.01% Ethidiumbromide; post staining with Gel Star^o (1/10 000 in 1XTAE; FMC Bioproducts). 100bp: 100 basepair ladder (Gibco BRL). H: water control of first round of PCR. A: amplicon; PD: primer dimer.

3.2. Detection of *bg*-like mutations with K^{bLNC} -specific Polymerase Chain Reaction coupled to Sequence Specific Oligonucleotide Probing

As discussed in the introduction we set out to design and establish a protocol consisting of multiple components, which would allow us to detect and analyse mutations occurring in germline *H2* genes. We wanted to use a combination of different tools such as K^{bLNC} transgenic mice (see introduction), the M450^{LZ}-*lacOp* rescue as a target enrichment step, and enrichment of mutants by tube Denaturing Gradient Gel Electrophoresis (tube-DGGE; (Sousa 2000)) combined with cloning and sequencing, to unravel gene conversion like events and ultimately to address questions concerning the molecular machinery underlying such a process. Our ideal was not only to analyse mouse MHC *class I* genes, which is highly demanding, but also to obtain a maximal degree of freedom concerning the characterisation of such mutations (see introduction). This is in contrast to approaches which focus on documented ("historical") mutations (Hogstrand *et al.* 1994). However while working on the assembly of a "gene conversion detector" it became obvious that the characterisation of its components would take more effort than first anticipated, and also that the power of some components might only be suboptimal for our purposes (Sousa 2000). It was therefore decided to take advantage of the availability of pBluescript sequence in the third intron of the cloned K^{bLNC} gene, to develop a streamlined mutation detection protocol, consisting of a highly specific PCR coupled to Sequence Specific Oligonucleotide Probing (SSOP). The amplicon of the K^{bLNC} -PCR should surround a region which was repeatedly hit by mutations belonging to the *bg*-series (see introduction – table 1-1). To focus on the *bg*-series of mutations was decided as it is by far the most frequent one documented in the whole series of "classic" mouse mutations (discussed in (Martinsohn *et al.* 1999)). The SSOP protocol should be established and optimised with respect to specificity and sensitivity. The aim was to detect mutations which occur in a given pool of cells with a frequency of 1/100. A frequency of 1/100 was chosen as a desired detection limit, since at d13 *post coitum* (p.c.) a mouse gonad contains about 25 000 primordial germ cells (PGC; (Wylie 1993)). If 50% of the PGCs could be recovered, the analysis of one gonad could be covered by one 96-well plate. A lower detection limit was regarded to be too ambitious, since SSOP is normally rather used for the analysis of heterozygous states, but to detect rare mutations. A protocol should be set up, in which a given number of (primordial germ) cells would be aliquoted in pools such that in any given mutant pool the mutation occurs at a minimal frequency of 1%. The cells would

be lysed and the genomic DNA amplified by the LNC-PCR as to raise potential mutations to levels being detectable by SSOP. Subsequently SSOP would be employed to screen the amplified pools for the occurrence of mutations. In a positive pool the mutation could finally be brought to “visibility” by batch cloning techniques and sequencing.

In this section the approach of developing the combined K^{bLNC} -PCR/*bg*-SSOP assay will be described. The target region will be introduced as well as the attempt to optimise the components of the assay, that is the K^{bLNC} -PCR and the *bg*-specific SSOP. Finally an experiment will be presented which evaluates the power and feasibility of the complete approach in the context of the expected low frequency ($f=10^{-4}$) with which *bg*-like mutations are expected to occur.

3.2.1. The *bg*-target region

The *bg* mutation group includes a number of animals (altogether 9) that were detected in the historical *H2* mutation screening programs, based mainly on skin grafting (see introduction – table 1-1). In all cases the mutation, which hit the third exon of K^b , is characterised by either an A→T transversion at aminoacid position 116 (resulting in a Y→F change on the AA level) alone, or additionally by a T→C transition at aminoacid position 121 (resulting in a C→R change on the AA level). For the purpose of designing an effective PCR/SSOP assay it was decided to focus on the *bg* group as it appeared in roughly 50% of all detected K^b -mutations, and is therefore by far the most prevalent *H2* mutation known at present. In figure 3-28 the sequence region containing the characteristic *bg* base exchanges is depicted, aligned to Q4, the proposed donor gene for this event (Geliebter *et al.* 1986), as well as the anchor sites for the primer pairs LNCK/LNCV and NLNCK/NLNCV, which frame the amplicon covering the *bg* mutation. The original primer pair LNCK/LNCV frames an amplicon of 1.2kb length, spanning from the second exon to the pBluescript part of KbLNC, which is located in the third intron. The nested primer pair NLNCK/NLNCV was designed in anticipation for further need of specificity at some stage during the project. All primers were designed using the software program “Oligo” (vs6.0; Wojciech Rychlik). For the primer design of LNCK all other *H2 class I* genes of the *b*-haplotype were taken into account of which sequence information was available, to ensure maximal specificity. The length of the amplicon ($l=1232$ bp) was chosen in such a way that it covers the *bg*-region generously, in case detected mutational events do not just include the two *bg*-marker bases T and C, but also bases of the flanking regions of the *bg* area.



Figure 3-28 The bg -target region. Above: sketch of the exon-intron structure of K^{bLNC} . Below: Shown is an alignment of exon I to intron III of the $H2$ genes K^{bLNC} and $Q4$ (upper sequence), the proposed donor sequence for the bg -mutation (see table 1.1). Exon sequence of K^b is drawn in green, intron sequence in black. The sequence part of K^{bLNC} which derives from pBluescript is drawn in blue. Indicated are the positions of the primer pairs LNCK/LNCV (dark red); NLNCK/NLNCV (orange), and the positions of the SSOP oligonucleotides BLOT (light violet); BLOC (dark violet), as well as the Kb-oligo (turquoise), and KBO58 (blue). Furthermore indicated are the bg -specific bases derived from $Q4$ (in red): T (position 1527 in $Q4$), and C (position 1541) in $Q4$. The lac-operator sequence (lacOP; Gilbert and Maxam (1973)) is boxed in blue. The alignment was carried out with the program Bestfit (GCG-Sequence Analysis Software Packag; Version 6.0). Sequence of K^{bLNC} based on records provided by Gaby Vopper and Lisbeth Guethlein (personal communication).

3.2.2. LNC-PCR on cells

The first step in the combined K^{bLNC} -PCR/bg-SSOP assay consists of PCR on pools of cells. While the aliquoting of the cells to be examined can be regarded as an enrichment step for the mutation (as in any given mutant pool of cells the original mutant frequency of 10^{-4} is risen by a factor of 100 to 10^{-2}), the PCR serves as a "magnifier", bringing the mutation up to detectable levels for the SSOP step. Due to its low error rate of about 10^{-6} per position per cycle (Andre *et al.* 1997), we chose Pfu as the principal polymerase throughout the whole project. Since the anticipation was, that we would be in need for a high amount of Pfu, we decided to purify Pfu on our own behalf, following a protocol proposed by Lu and Erickson (Lu *et al.* 1997). The attempt to purify Pfu and to optimise it with respect to efficiency is described in the following section.

Isolation and performance of the *Pyrococcus furiosus* DNA Polymerase ("Pfu")

As outlined in the introduction, for our purposes *Pfu* was the Polymerase of choice to be used in the context of mutation detection due to its proof-reading activity, which is based on 3'-5' exonucleolytic activity (Cline *et al.* 1996). Since the frequency of "bg" mutations is presumably as low as 10^{-4} , it was obvious that in order to yield statistically relevant numbers of "bg-like" mutations, a great number of PCRs would have to be carried out. Even if the uptake of a mutated DNA fragment, being present at a frequency of 1% in a pool of DNA, to a detectable level, would be 100% efficient, one hundred PCRs would have to be carried out to get hold of one mutation. It is apparent that to come to meaningful conclusions about the nature and characteristics of mutations one relies on high numbers. This renders the analytical process cost intensive if one depends on polymerases which are only commercially available, motivating us to follow a published protocol of Lu and Erickson (Lu *et al.* 1997) in order to isolate Pfu ourselves.

We received from the American Type Culture Collection a 8kbp construct which is based on the expression vector pET11b (Novagen) and contains a 2.4kbp open reading frame encoding the wildtype *Pyrococcus furiosus* DNA polymerase (ATCC number 87496; sequence accession number: D12983; insert: nucleotides 237 to 2564 of D12983). This insert is cloned into the *NdeI* site of the pET11b polylinker (Lu *et al.* 1997). The construct was delivered in lyophilised *E.coli*, strain *DH5 α* .

The pET protein expression system is based on the T7 promoter-driven system originally developed by Studier *et al.* (Studier *et al.* 1986; Studier *et al.* 1990; Dubendorff *et*

al. 1991). The target gene on the pET plasmid is under control of the T7 promoter, which in turn is under control of the Lac repressor. The T7-polymerase gene of the host gene is under control of the *lac* promoter. Expression of the target gene is therefore strongly induced upon addition of IPTG. Additional control is achieved in the pLysS system: The host (*BL21DE3pLysS*) carries a plasmid encoding the T7 lysozyme, a natural inhibitor of the T7 polymerase, which is expressed at low basal levels. This ensures suppression of background expression of the target gene in uninduced cells (Studier 1991). *E.coli* bacteria of the strain *BL21(DE3)pLysS* (Promega) were transformed with the construct and an expression test was performed (figure 3-29). A 90kDa Protein appears in the samples that are induced with IPTG. An additional band visible upon induction at about 118kDa was also observed by Lu et al. (Lu *et al.* 1997). The authors claim that this 118kDa protein appears as well when expressing other proteins in the pLysS system.

Initially a simplified purification protocol was tried out, which included sonication of the samples and denaturation of background proteins by heat supply (figure 3-30). This purification worked surprisingly well, compared to the purification achieved by Lu and Erickson in a similar step (see figure 1 in Lu and Erickson, (1997)). The only other band being visible after purification in the Coomassie stained gel belongs to a very strongly expressed protein of about 25 to 30kDa.

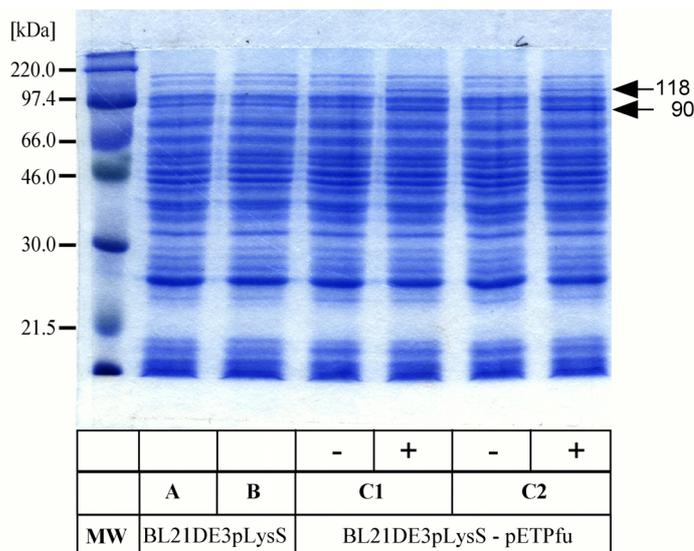


Figure 3-29. Expression of Pfu. Depicted is a 12.5% Coomassie blue stained SDS-PAGE gel. BL21(DE3)pLysS were transformed with the expression plasmid pET Pfu (two clones C1 and C2). Each clone was grown in 20ml LB in the presence of [25µg/ml] Chloramphenicol (Sigma) and [100µg/ml] Ampicillin (Sigma) in 3ml volume over night at 37°C. The over night culture was washed two times with 10ml LB and 1/50 volume was transferred into 20ml LB [25µg/ml] Chloramphenicol and [100µg/ml] Ampicillin for further growth. At an OD of about 0.5 the cultures were induced with 0.5mM IPTG and grown to a final OD of 2.8 to 3. 500µl of the culture was pelleted, taken up in 50µl PBS, mixed with 50µl 2Xprotein loading buffer and boiled for 10 minutes. 5µl were loaded onto the gel shown. A;B: BL21DE3pLysS untransformed; C1, C2: two clones of BL21DE3pLysS transformed with pETPfu; -: no induction with IPTG. +: induction with IPTG. 118: unspecified band, wat 118 kDa which was also reported by Lu and Erickson (1997). 90: induced band at the expected height of 90kDa. MW: Rainbow (Amersham).

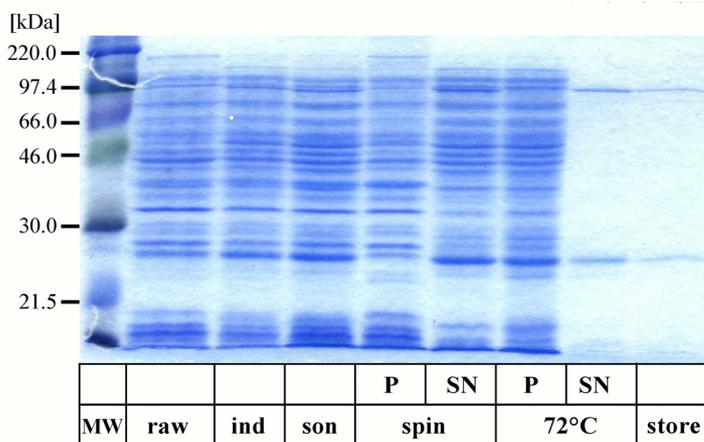


Figure 3-30. Purification of Pfu by denaturation. Depicted is a 12.5% Coomassie blue stained PAA gel. Loading volume per lane: 5µl. raw: culture before IPTG-induction (500µl of culture taken up in 50µl loading buffer). ind: culture after IPTG-induction (500µl of culture taken up in 50µl loading buffer). son: sonicated material (50µl in 100µl loading buffer). spin (P): pellet of 100 000g (50µl in 100µl loading buffer). spin (SN): supernatant of 100 000g (taken up in 500µl PBS; 1/10 in 100µl loading buffer). 72°C (P): pellet of 72°C, 100 000g (50µl in 100µl loading buffer). 72°C(SN): Pellet of 72°C, 100 000g (taken up in 500µl PBS; 1/10 in 100µl loading buffer). store: purified Pfu in storage buffer (5µl in 50µl loading buffer). Purification procedure: Starting from one colony a 20ml culture, was grown over night (o.n.) in LB ([25µg/ml] Chloramphenicol, [100µg/ml] Ampicillin). The o.n. culture was pelleted, washed twice in 50ml LB, and taken up in 20ml LB. 150ml LB ([25µg/ml] Chloramphenicol, [100µg/ml] Ampicillin) were inoculated with 7ml of the o.n. culture. At an OD of about 0.5 expression was induced by addition of IPTG (final c=0.5mM). After further growth at 37°C for 3h, the culture was pelleted and resuspended in 5ml cold resuspension buffer (50mM Tris/HCl pH8.0, 100mM NaCl, 1mM EDTA, 1mM PMSF (added freshly), 0.2mg/ml Lys-ozyme (freshly added)). Lysis occurred on ice for 2h. The suspension was shaken carefully from time to time and stored at -20°C until further processing. The frozen suspension was thawed at room temperature and MgCl₂ was added to a final concentration of 10mM, as well as DNaseI (final concentration 0.2µg/ml). After incubation for 30minutes at room temperature on a shaker sonication followed (Labsonic[®]; duty cycle 0.5; 3 X 30s; 30" pause). The sonicated material was spun down at 100 000g for 30 minutes. The supernatant was incubated at 72°C for 15 minutes to denature heat labile proteins, and spun down again at 100 000g for 30 minutes. The supernatant was concentrated and transferred into storage buffer (50mM Tris/HCl; pH8.2, 0.1mM EDTA, 1mM DTT, 0.1% NP-40, 0.1% Tween20, 50% w/v glycerol) by a Centriprep YM-30MW column (Amicon-Millipore). The purified Pfu was stored at -20°C. MW: Rainbow (Amersham).

The impressive purity obtained with this simple protocol encouraged us to use the purified Pfu at this stage in a PCR and compare its performance to commercially available Pfu (Promega). Only at high dilutions of Pfu, and high template numbers, amplification of target sequence by the home made Pfu could be detected (figure 3-31 A)) but as can be seen in figure 3-31 B)) this amplification was not very robust. At the height of primer dimer formation, a signal appears which strength correlates with the amount of Pfu put into the PCR reaction (figure 3-31 B))

It was taken into consideration that after the first purification step remaining nucleic acid of *E.coli* might interfere with the target amplification in the PCRs. OD measurements and loading of purified Protein samples on agarose gels revealed that indeed the samples contained high amounts of nucleic acids (see figure 3-32 and table 3-5).

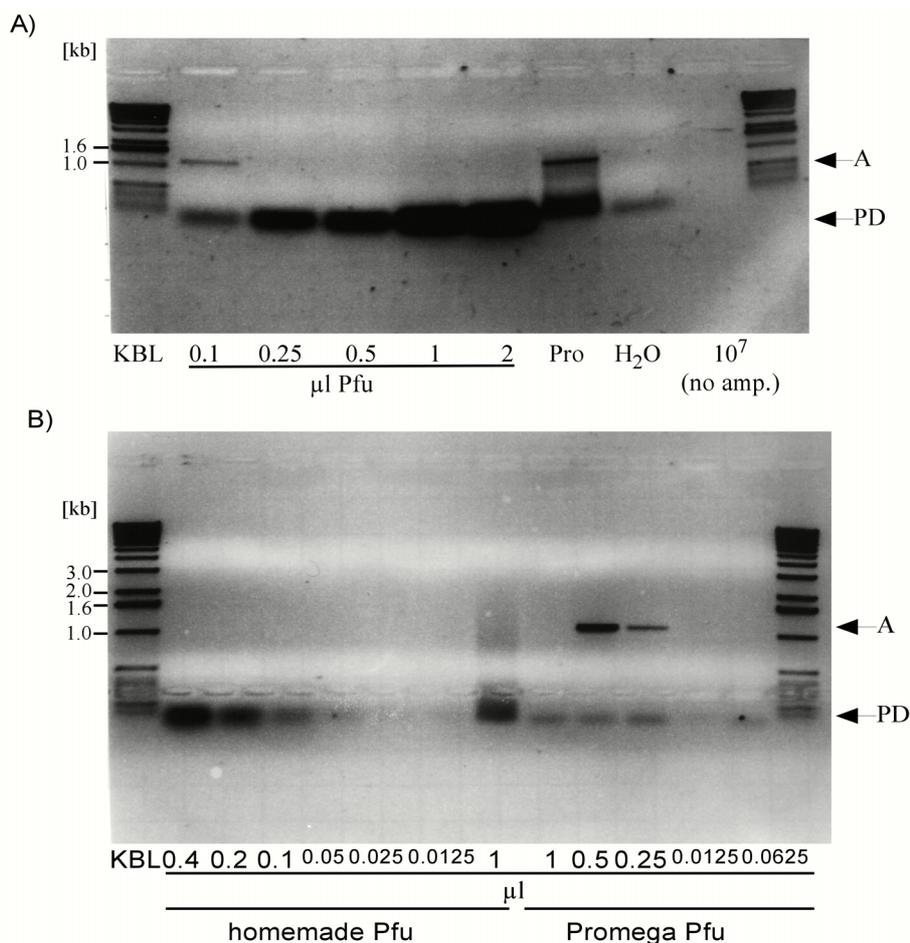


Figure 3-31. Amplification of target sequence with Pfu purified through denaturation. Depicted is an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide). Loading volume: 25µl. A) Dilution of homemade Pfu. Depicted is an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide). 10⁷ molecules of pBKb (see Materials and Methods) were amplified for 30 cycles (50µl; 0.2mM dNTP; 5pmol LNCV; 5pmol LNCK; 1 X Promega Pfu Buffer; Pfu as indicated; PCR: 3min., 95°C; 30 X [30s, 95°C; 30s, 64°C; 90s 75°C]; 10min. 75°C; ∞ 4°C). Loading volume: 25µl. B) Dilution of homemade and Promega Pfu. PCR conditions as above. A: amplicon of expected size. PD: height of primer dimer formation.

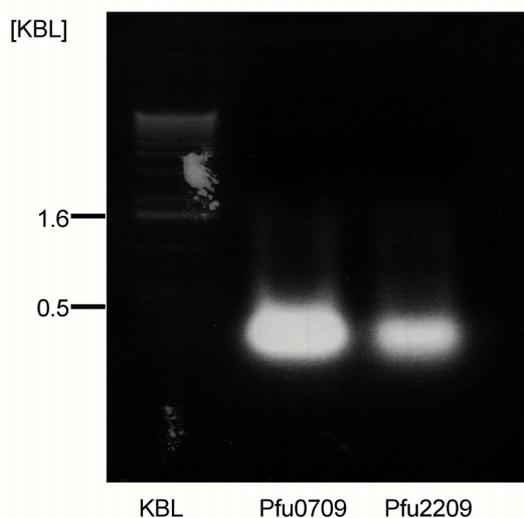


Figure 3-32. Nucleic acid content in purified home-made Pfu. Depicted is an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide). 1µl of two independently purified samples (Pfu0709 and Pfu2209) were loaded.

sample	dilution	A205	A280	A260	protein (A205) [mg/ml]	protein (A280) [mg/ml]	nucleic acid [mg/ml]
070999	1:100	1.7610	0.5382	0.9731	5.7	53	5
220999	1:10	n.m.	1.9024	2.377	n.m.	19	1.2
	1:100	1.5530	0.2289	0.2967	5	23	1.5

Table 3-5. OD measurements of two independent samples of purified home-made Pfu. A205 measures the peptide backbone ($x[\text{mg/ml}] \text{ Protein} = A_{205}/31$; (Coligan *et al.* 1997)).

We decided to add to the purification protocol an affinity based purification step by a Phosphocellulose column. In figure 3-33 a typical elution profile for a Phosphocellulose run is depicted.

The elution of Pfu occurred in all runs at about 210 to 220mM KCl, while Lu and Erickson (1997) claim that elution occurred at 480mM KCl. The yield (percent of the output) is not easy to estimate. Certainly for an estimation the FPLC running profile cannot be taken into account since the nucleic acids present interfere heavily with the absorption at 280nm. Estimates based on the Coomassie gel accompanying the elution profile (figure 3-33 B)) lead to approximately 90% yield (Output: 25 μ l of 5ml on gel; elution: 25 μ l of 3ml on gel). The purity is about 90%. The success of this purification step is clearly visible in the Coomassie stained gels since the low molecular weight band, at 25 to 30kDa disappeared. Also the accompanying agarose gels show that nucleic acids are no longer detectable (figure 3-33 C)).

The purified home-made Pfu was tested in PCR and Unit estimates were performed (figure 3-34). A comparison of signal strength in agarose gels of sample Pfu2509 to 1.5U Promega Pfu leads to approximately 15U/ μ l (the signal strength for sample Pfu2509 at 0.5 μ l 1:10 diluted is approximately half the signal strength for the Promega Pfu amplified target). For sample Pfu2709 this comparison leads to approximately 6U/ μ l. Altogether about 4000U home-made Pfu could be purified (final volume Pfu2509: 100 μ l; final volume Pfu2709: 400 μ l). Further dilutions of the home-made Pfu were tested (figure 3-34 B)). Disappointingly the peak activity of the home-made Pfu consistently was about ten times lower than that of Promega Pfu (as judged by comparison of signal strength in agarose gels).

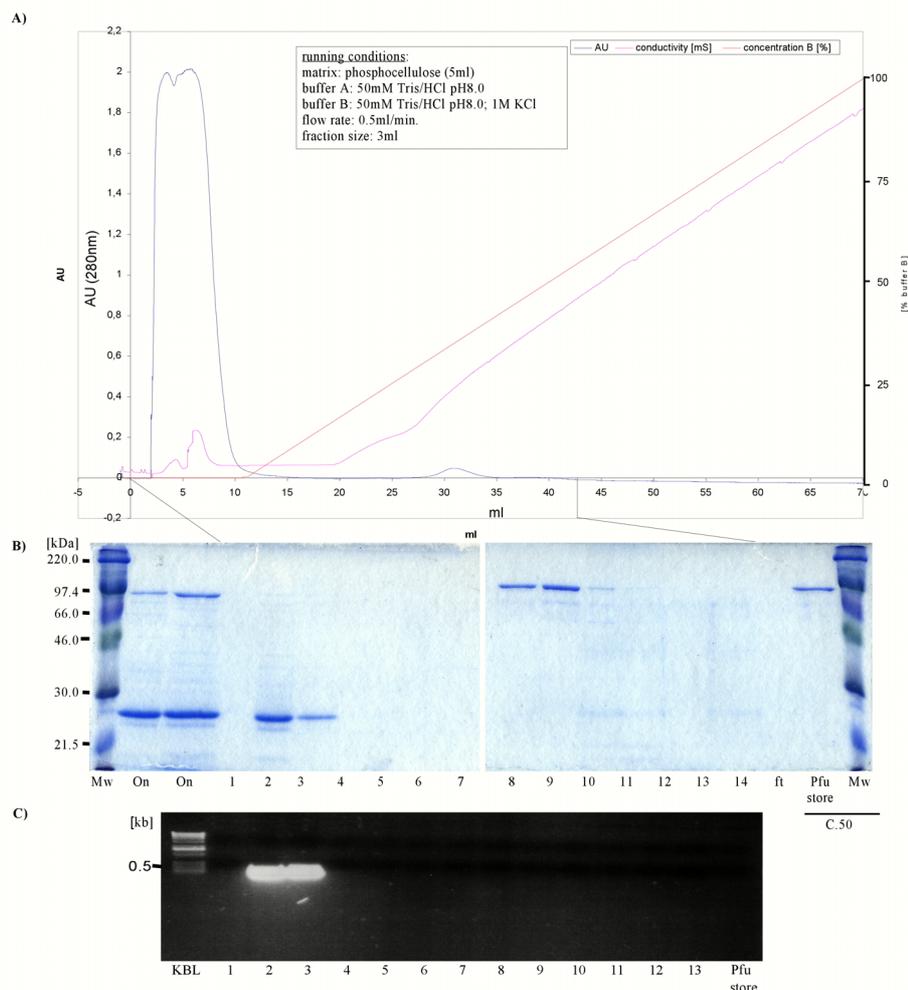


Figure 3-33. Phosphocellulose based purification of home-made Pfu. A) FPLC run: Starting from a 150ml culture which had been inoculated with 1/10th of an o.n. culture the purification was identical as in figure 3-30. The supernatant of the last 100 000g spin, after the heat denaturing step, was immediately loaded onto an equilibrated Phosphocellulose column (Whatman P11; 5ml volume; Pre-cycling according to manufacturers instructions). Loading conditions: 50mM Tris/HCl pH8.0, 1mM EDTA; Gradient: linear to 1M KCl over 60ml. Running conditions as indicated in figure. 3ml fractions were collected. Samples 8 and 9 were pooled and buffer exchange as well as concentration was accomplished in Centricon-50 spin columns (Amicon-Millipore): 3 times one volume buffer exchange (2X storage buffer: 100mM Tris/HCl pH8.0; 0.2mM EDTA; 2mM DTT; 0.2% NP-40; 0.2% Tween20). The final volume of samples 8 and 9 was 67 μ l. Glycerol was added to final concentration of 50%. Storage at -20°C. AU: Absorbance Units. B) Coomassie stained gel: 12.5% PAA; loading volume (except concentrated Pfu): 25 μ l of all samples were mixed with 1 Volume loading buffer. 10 μ l were loaded onto the gel. Concentrated Pfu: 5 μ l in 50 μ l loading buffer. 10 μ l were loaded onto the gel. ft: flow through of Centricon-50 spin column (see A)). Pfu store: stored Pfu (see A)). C.50: Samples (retained Pfu and flow through) which went through Amicon-50 spin column. MW: Rainbow (Amersham). C) Agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide): 100 μ l of all fractions were extracted by phenolization and treatment with Chloroform/Isoamylalcohol (24:1). After an Ethanol precipitation the pellets were taken up in 25 μ l H₂O of which 5 μ l were loaded onto the gel.

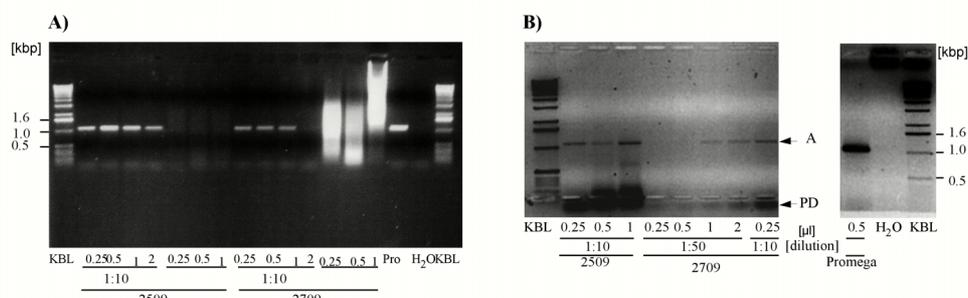


Figure 3-34. Phosphocellulose purified home-made Pfu : activity test. Depicted are agarose gels (1%; 1XTAE; staining: 0.01% Ethidiumbromide). Two independently purified samples (Pfu2509 and Pfu2709) were tested. 20 μ l of the 50 μ l PCR reactions were loaded. A) 10⁷ molecules of pBKb (see Materials and Methods) were amplified for 30 cycles as in Figure 3-31. Pfu amounts as indicated in Figure legend. B) 10⁷ molecules of pBKb (see Materials and Methods) were amplified for 30 cycles as in Figure 3-31. Pfu amounts [μ l] and dilutions as indicated in Figure legend. Sample Pfu2509 and Pfu2709 amplified with Stratagene 1X reaction buffer. Positive control: 0.5 μ l (ca. 1.5U) Promega Pfu. Note: the Promega Pfu control and samples Pfu2509 and Pfu2709 were originally loaded on one gel. For reasons of simplicity some samples of the gel were cut out graphically. A: Amplicon. PD: Primer Dimer.

It was suspected that the concentration and buffer exchange after the FPLC-purification, via the Amicon columns worked only suboptimal, and that this might lead to the unsatisfactory activity of the homemade Pfu. Therefore desalting was performed with a Sephadex G-50 column before transferring the purified samples into storage buffer. Typical purification and desalting profiles are depicted in figure 3-35. In this example fraction 23 and 24 were pooled and went without any further processing into the desalting procedure. Table 3-6 lists the output amount (amount after Phosphocellulose purification) and the yield. Approximately 18% of the output material were lost, mainly because fractions were chosen which correspond to the centre of the protein peak in figure 3-35, while the fractions covering the edges of the peaks were discarded.

Before buffer exchange					After buffer exchange				
Fraction 23		Fraction 24		Sum	Fraction 5		Fraction 6		Sum
OD ₂₇₈	[µg/ml]	OD ₂₇₈	[µg/ml]	[µg]	OD ₂₇₈	[µg/ml]	OD ₂₇₈	[µg/ml]	[µg]
0.176	238	0.135	182	420	0.118	160	0.138	186	346

Table 3-6. Buffer exchange of FPLC purified Pfu by Sephadex-G50: Protein amounts. Listed is the total Protein amount [mg/ml] of the fractions that were treated. OD₂₇₈[Pfu]: 1 ⇒ C_{Pfu} = 1.35mg/ml (Lu *et al.* 1997).

Testing the purified and Sephadex G-50 treated Pfu

The performance of the phosphocellulose purified Pfu was tested in some standard PCRs (figure 3-36). The activity reached peak levels, which were still about 10 times below to that of 0.5µl Promega Pfu (ca. 1.5 to 1.75U; figure 3-36 A)). The highest activity was reached at 4µl of a 1:10 dilution of the original material that had been purified by a phosphocellulose column (figure 3-36 B) and C)). In parallel a sample was tested which had, after the purification by a phosphocellulose column and desalting with sephadex G50, further been purified by a UNOQ-1 anion exchange column (BioRad; details see Materials and Methods). As can be see in figures 3-36 A) and C) no further increase in activity was observed. This is in agreement with Lu and Erickson, who used subsequently to the phosphocellulose step a MonoQ anion exchange column (HR10/10, Pharmacia (Lu *et al.* 1997).

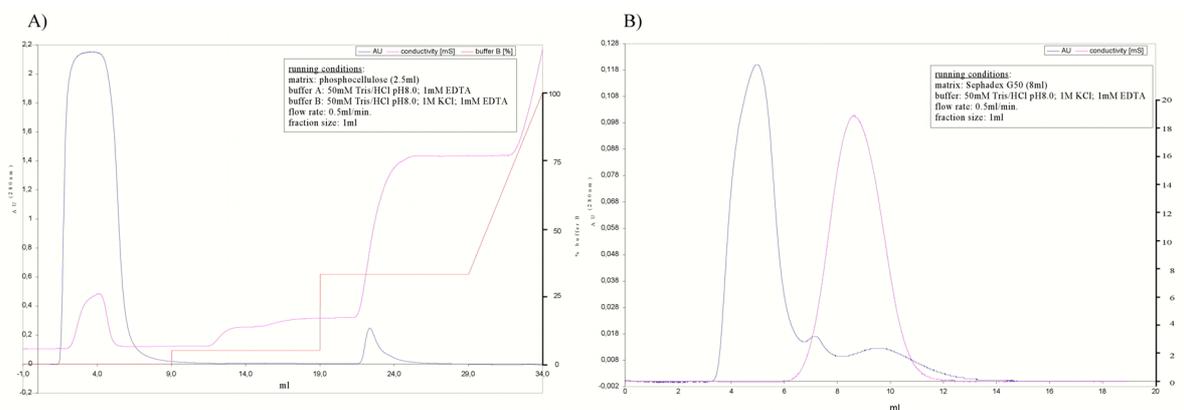


Figure 3-35. Purification of home-made Pfu and desalting with a Sephadex G-50 column.

A) typical FPLC-running profile of Pfu purification. Purification as in figure 3-33 except that instead of a linear gradient, a step gradient of the elution buffer was employed (1-10ml: 0%B; 10-20ml: 5%B; 20-30ml: 33%B; 30-35ml: 33% to 100% B). The fraction size was 1ml.

B) typical desalting profile upon usage of a Sephadex G-50 (medium; Pharmacia) column. The Sephadex was equilibrated in 50mM Tris/HCl pH 8.0; 1mM EDTA and a column of 1cm diameter packed. Final column volume: 8ml. Flow speed: 0.5ml/min. Fraction size: 1ml. The collected fractions 5 and 6 were stored over night at 4°C. 450µl of fraction 5 was directly mixed with one volume 2 X storage buffer (50mM Tris/HCl pH8.2; 2mM DTT; 0.2% v/v IGEAL; 0.2% v/v Tween20; 90%w/v Glycerol) and stored at -20°C. This purified Pfu sample was called "Pfu1210S". AU: Absorbance Units.

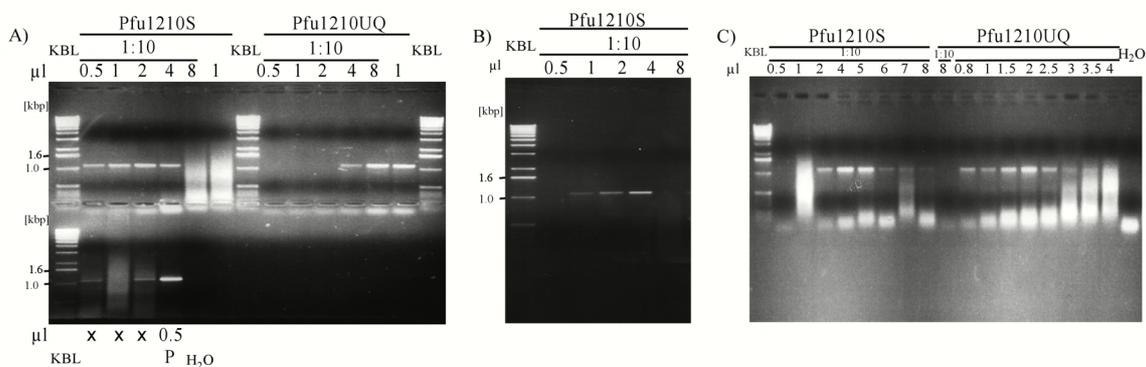


Figure 3-36. Activity of Pfu1210S and Pfu1210UQ. Depicted are agarose gels (1%; 1XTAE; staining: 0.01% Ethidiumbromide). 35µl of the 50µl PCR reactions were loaded. 10^7 molecules of pBKb (see Materials and Methods) were amplified for 30 cycles as in Figure 4. Pfu amounts and dilutions as indicated in Figure legend.

A) Sample Pfu1210S and Pfu1210UQ in comparison to Promega Pfu.

B) repetition of **A**).

C) extended dilution rows of the Pfu samples tested. Pfu1210S: Pfu purified by phosphocellulose column, and sample desalted with Sephadex G50 column. Pfu1210UnoQ: sample derived from Pfu1210S batch but further purification by UNOQ-1 (BioRad) column. x: samples of an unrelated PCR evaluation, not discussed here.

Large scale purification of Pfu

Even though the peak activity of the home-made Pfu remained below that of Promega Pfu, it was decided to carry on with the purification. The idea was to scale up the purification protocol as to produce a large batch of home-made Pfu which could be optimised for performance and used throughout the whole project as such.

The purification protocol was in principle the same as above, but the culture being induced for expression had a volume of 1000ml (2X500ml). In figure 3-37 the elution profiles of the Phosphocellulose step is illustrated. The final yield was about 7mg purified Pfu. This batch of Pfu was called “LSPfu”.

Performance of Large Scale purified Pfu (LSPfu)

The performance and optimal concentration of LSPfu was tested by dilution rows and comparison to Promega Pfu (figure 3-38). The amplification was done on the plasmid pSP72 (Promega) with the primer pair 72F and 72R (see Material and Methods). The highest activity was obtained at a dilution of 1:20; and 0.8 to 1.6 μ l volume. About 10% of the activity of Promega Pfu (0.5 μ l; 1.5U) was obtained. Therefore the activity of the original stock (before 1:20 dilution) was about 5U/ μ l. this means that altogether around 25 000U of LSPfu could be purified (total Volume of batch: 5ml).

The activity for LSPfu was also compared to that of Promega Pfu for the LNC-PCR system, using the primer pairs LNCV/LNCK and NLNCV/NLNCK (see Materials and Methods and figure 3-28). As can be seen in figure 3-38 still the activity was about a factor 10 lower than for the Promega Pfu.

It was tested whether the storage buffer, which was used for the Pfu's purified in this lab (1 X Pfu-storage buffer: 25mM Tris/HCl pH8.2; 1mM DTT; 0.1% v/v IGEPAL; 0.1% v/v Tween20; 45%w/v Glycerol) has an inhibitory effect on the PCR activity. Therefore PCRs with Turbo Pfu (Stratagene), to which increasing amounts of Pfu storage buffer were added (figure 3-39). Only at high concentrations of Storage buffer, the PCR is inhibited (from 2 μ l on, which is equivalent to 4% v/v end concentration, compared to 0.1% v/v if 1 μ l of 1:20 dilution of LSPfu is used for the PCR). In fact apparently at low concentrations the addition of storage buffer has a rather “stimulatory” effect.

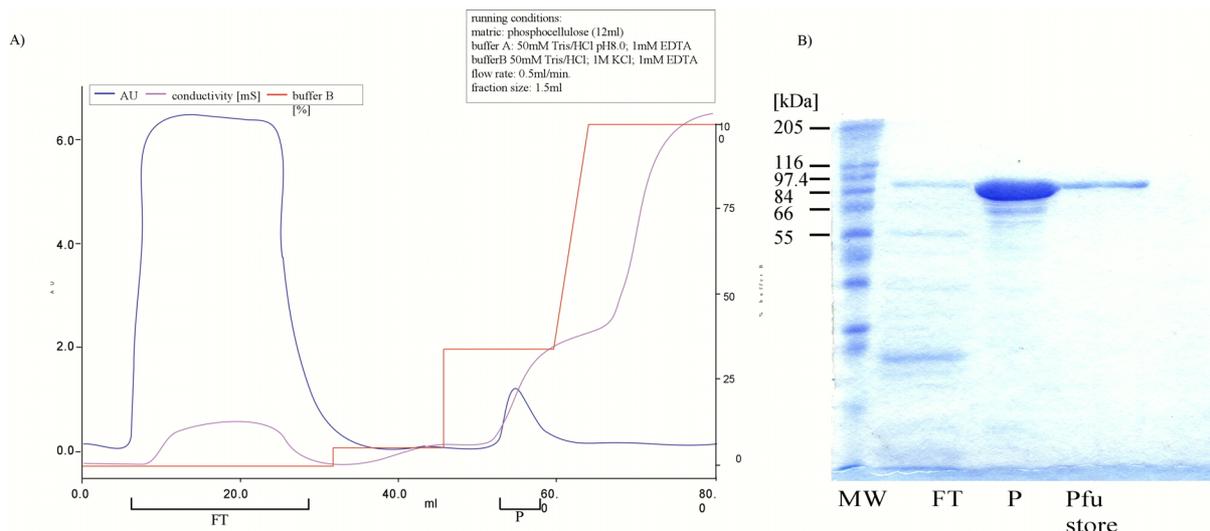


Figure 3-37. (BiocFPLC) Large scale purification of Pfu. A) Shown is the FPLC elution profiles of the phosphocellulose purification step. In B) a 12.5% Coomassie stained gel is depicted on which the purified samples have been loaded (8µl of the pooled samples FT; 8µl of the pooled samples P; 1µl of Pfu store). In short, after treatment with *DnaseI* (Roche) and sonication the raw extract (20ml Volume) was centrifuged at 100 000g for 30 minutes and immediately loaded onto a preequilibrated Phosphocellulose column (Volume: 12ml). The FPLC run was performed as indicated in A). Four fractions (each 1.5ml) were pooled (indicated in A) as P), and the volume decreased to 2ml (Amicon Centriplus 30 spin column). After centrifugation at 100 000g for 30 minutes the supernatant was loaded onto a preequilibrated Sephadex G-50 column (10ml volume; 50mM Tris/HCl pH8.2; 0.1mM EDTA). The condition were as in figure 3-35B. Four fractions were collected (each 1.5ml volume) and the volume decreased to 2.3ml (Amicon Centriplus 30 spin column). The sample was transferred into storage buffer and stored at -20°C. The purity of the home-made Pfu was at least 90% as estimated from the Coomassie stained gel depicted in B). FT: flow through. P: fractions belonging to peak area of Pfu. Pfu store: Pfu after transfer into storage buffer. MW: Wide Range (Sigma)

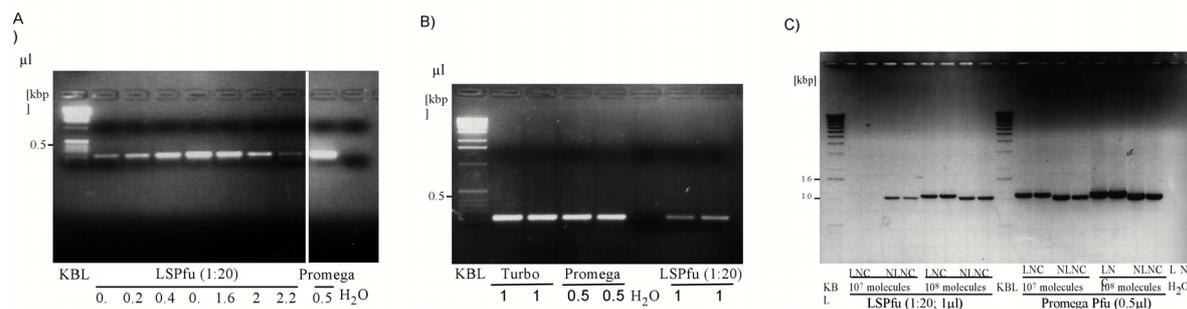


Figure 3-38. Performance of LSPfu. Depicted are agarose gels (1%; 1XTAE; staining: 0.01% Ethidiumbromide; 20µl loading volume). A) 10⁷ molecules of pSP72 (Promega), amplified with LSPfu (diluted 1:20 in H₂O) and Promega Pfu. Amounts of polymerase as indicated. In all cases the PCR was performed as follows: 50µl; number of target molecule pSP72 (Promega) as indicated; the expected amplicon length is 234bp; 0.2mM dNTP; 5pmol psp72F; 5pmol psp72R; 1XPromega Pfu Buffer; Pfu polymerase as indicated; PCR: 2min., 95°C; 25X[30s, 95°C; 30s, 60°C; 30s 75°C]; 10min. 75°C; ∞ 4°C. B) comparison of 1µl LS Pfu (1:20) to 1µl (2.5U) Turbo Pfu (Stratagene) and 0.5µl Promega Pfu (1.5U). C) Depicted is a positive image of an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide; 20µl loading volume). Either 10⁷ or 10⁸ molecules of pBKb were amplified for 25 cycles with 1µl LSPfu (1:20 in H₂O) or 0.5µl Promega Pfu. The primer pairs used were LNCV/LNCK and NLNCV/NLNC (see Materials and Methods). LNC-PCR in 50µl as before.

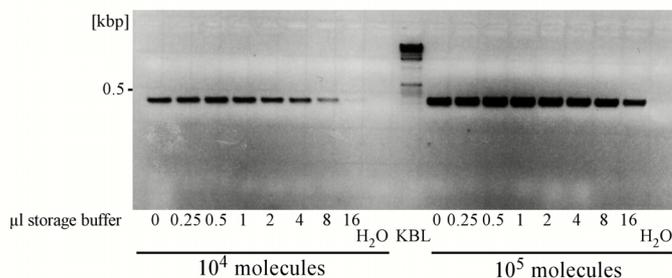


Figure 3-39. Influence of Pfu storage buffer on PCR performance. Depicted is an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide; 20µl loading volume). 10⁴, 10⁵ molecules were amplified with 2.5 Units of Turbo Pfu (Stratagene) for 25cycles in the presence of increasing amounts of storage buffer (25mM Tris/HCl pH8.2; 1mM DTT; 0.1% v/v IGEPAL; 0.1% v/v Tween20; 45%w/v Glycerol), amounts as indicated in gel). Target was the plasmid pSP72, which was amplified as in Figure 3-38 A).

PCR performance of LSPfu on cells

The original purpose was to use the home-made Pfu in PCRs on cells. For this reason, even though the overall performance of the purified Pfu was unsatisfactory in PCRs performed on plasmids compared to commercially available Pfu-Polymerases, PCR tests were done on aliquoted cells. This was done also because it was observed that PCRs performed on cells in general work more efficiently than PCRs on pure plasmid DNA. In first experiments pools of 10^2 to 10^4 cells were lysed and then a PCR, consisting of two rounds directly performed on these pools. The first round implied 28 cycles and was executed either with Turbo Pfu, or with home-made Pfu. For the second round $1/10^{\text{th}}$ volume of round I was taken and amplified further for 14 cycles with 2.5U of Taq. Figure 3-40 depicts a typical result of a series of three of such PCRs which showed very similar outcomes. Under these conditions the LSPfu did not show any activity, while clearly visible signals appeared for Turbo Pfu already in the first round for 10^3 cells. Stratagene claims that Turbo Pfu shows a higher efficiency, compared to "normal" Pfu's (Bornes *et al.* 1998). In my experience this is only true to a very limited degree (see for example figure 3-38 B)). In the LNC-Cell PCR system this is not the case as shown below in a comparison of Turbo Pfu to the Promega Pfu (figure 3-44). Interestingly the PCR efficiency on cells peaks at 10^3 cells and shows a sharp efficiency decline at 10^4 cells. This was observed in every PCR performed on cells under these conditions. It indicates that the system is balanced in an optimal way (Pfu amount; cell amount; volume) at 10^3 cells under these conditions (see also below - LNC-PCR on cells: final protocol). In parallel it was tested whether hot start, that is the addition of the Polymerase only after the sample reached a temperature of 95°C , is advantageous to a 95°C start (addition of PCR tube to pre-heated PCR-machine block). As can be seen in figure 3-40 this is not the case. 95°C starts were as good in terms of efficiency and specificity.

In later experiments the cycle number was increased to 35 cycles and a second round of PCR with Taq omitted. In figure 3-41 a comparison of Pfu1210S to Promega Pfu is shown under such conditions. Even though at 10^3 cells the yield of PCR-product generated with Pfu1210S is quite satisfactory, at 10^2 cells it lies clearly below that of the samples being processed with Promega Pfu. Also the primer dimer formation and general background smear was considerably increased in the samples amplified with Pfu1210S.

In the same way Pfu1210S was compared to LSPfu (figure 3-42). Surprisingly while for 10^2 cells no amplification occurred with LSPfu, very strong signals appeared for 10^4 cells, which was not observed with any of the used Pfus before under these conditions. This might

indicate that optimal PCR-amplification on cells is very sensitive to a balanced environment concerning cell amount, Pfu amount, volume, buffer composition. Again excessive Primer dimer formation was observed.

To this moment it is not obvious why the Pfu isolated in this laboratory does not perform as well, and not as reliably as the commercially available counterparts. Evidently the activity of the polymerase produced in this laboratory remains consistently at a 10% level of the Pfu sold by Promega and Stratagene. It might be that the growth conditions during the expression of Pfu could be further optimised. However Lu and Erickson (1997) report in their publication that the activity and performance of the Pfu expressed in *E.coli* was similar to that of other available Pfu's. In later experiments the LNC-PCR on cells was done with Promega Pfu, since the results were impressively consistent (see below). No further analysis has been done on the Pfu purified in this laboratory, due to time limitations. Further optimisation should be carried out at later stages.

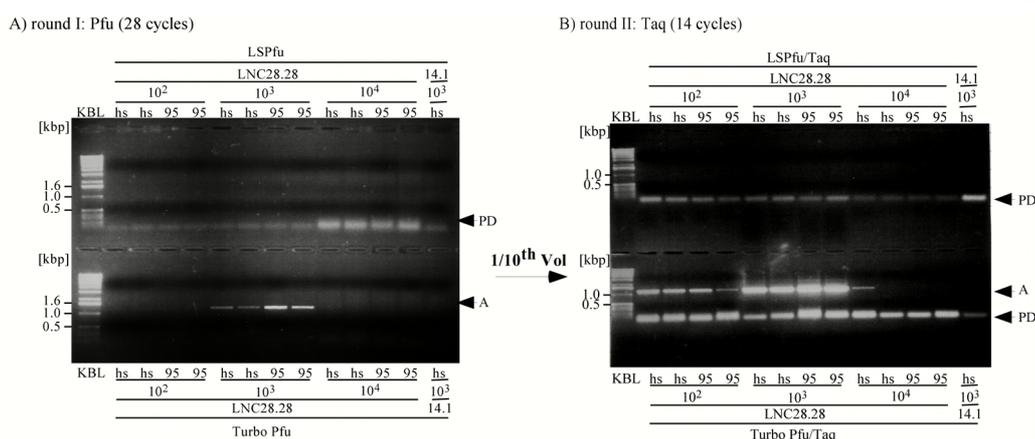


Figure 3-40. Pfu based PCR on cells: comparison of LSPfu to Turbo Pfu (Stratagene). Depicted are agarose gels (1%; 1XTAE; staining: 0.01% Ethidiumbromide; 20 μ l loading volume). **A)** 10^2 ; 10^3 ; or 10^4 of aliquoted cells were lysed in 1 X Promega reaction buffer in the presence of 5 μ g Proteinase K (Merck). After inactivation (95°C; 10 minutes), a PCR with the primer pair LNCK/LNCV was performed under the standard conditions, with the exception that either "traditional" hot starts (hs) or 95°C starts (95; see text) were applied. PCR: 28 cycles; 0.5 μ l of LSPfu (1:20); 2.5U Turbo Pfu (Stratagene). **B)** 5 μ l of A) went into another 14 cycles of PCR with 2.5U Taq (expressed and purified in this laboratory). 14.1: ES cells, negative control; hs: hot start: the samples were heated to 95°C before Pfu was added. 95: the heating block was heated to 95°C before placing the PCR on the heating block. PD: signal at height of primer dimer formation; A: signal of amplicon.

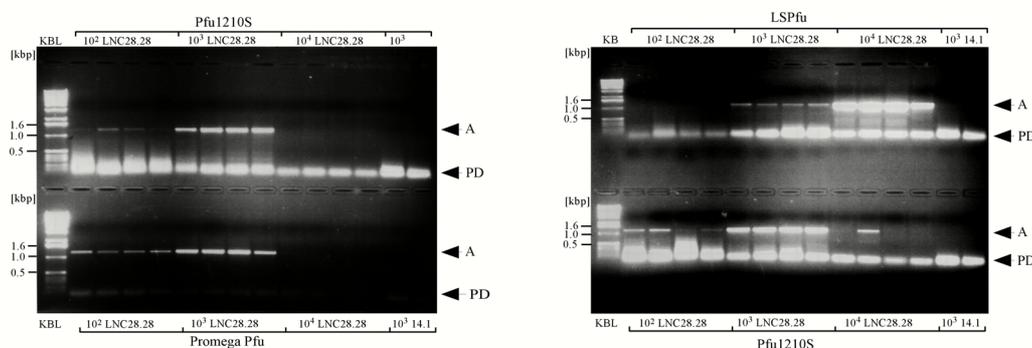


Figure 3-41. Pfu based PCR on cells: comparison of Pfu1210S to Promega Pfu. Depicted is a picture of an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide; 20 μ l loading volume). A 35 cycle LNC-PCR on the indicated amount of lysed cells was performed with either Pfu1210S (0.55 μ l/sample) or Promega Pfu (0.5 μ l/sample). PD: signal at height of primer dimer formation; A: signal of amplicon.

Figure 3-42. Pfu based PCR on cells: comparison of Pfu1210S to LSPfu. Depicted is a picture of an agarose gel (1%; 1XTAE; staining: 0.01% Ethidiumbromide; 20 μ l loading volume). A 35 cycle LNC-PCR on the indicated amount of lysed cells was performed with either Pfu1210S (0.55 μ l/sample) or LSPfu (1 μ l of a 1:20 dilution/sample).

Test of HEPES buffer for Pfu based PCR amplification

The standard buffer used in most applications in molecular biology is Tris (Tris(hydroxymethyl)-aminomethane). This is due to its buffering range (pKa=8.06), which covers well the optimum for most enzymes (BRENDA - Enzyme and Metabolic database (www.brenda.uni-koeln.de)), as well as the fact that it is uncomplicated to handle (unlike Phosphate buffers (Beynon *et al.* 1996), and comparably inexpensive. However, some neglected, still rather adverse characteristics are connected to this buffer. R.J. Beynon and J.S. Easteby (1996) comment on Tris in their book "Buffer solutions": "*A ubiquitous buffer, that has several problems, including a high temperature sensitivity, reactivity as a primary amine, the need for Tris-competent pH electrodes and some undesirable effects on some biological systems. Its continued use may be more to do with familiarity and to published recipes than to scientific justification*".

In particular the temperature sensitivity (i.e. lowering pH with increasing temperature) obviously sets a dilemma in the context of PCR, as a low pH leads to increased hydrolysis of DNA (Friedberg *et al.* 1995). Also DNA-directed DNA Polymerases often show a distinct activity optimum in dependence of pH (BRENDA - Enzyme and Metabolic database (www.brenda.uni-koeln.de)) On the other hand there exist buffers which are much less temperature sensitive. One of those is HEPES (4-(2-Hydroxyethyl)piperazine-I-ethanesulphonic acid). In table 3-7 some features of Tris and Hepes are contrasted. Dr. Ana Sousa (2000) showed that DNA is much more protected from heat (as detectable by degradation) in HEPES buffered solutions than in Tris buffered solutions. In table 3-8 pH values which would be expected during typical PCR reactions are listed. It is obvious that HEPES represents a more suitable buffer for PCR reactions, when it comes to temperature dependence.

	Tris (Tris(hydroxymethyl)-aminomethane)	HEPES (4-(2-Hydroxyethyl)piperazine-I-ethanesulphonic acid)
pKa	8.06	7.66
Temperature sensitivity [dpKa/dT]	-0.028/°C	-0.014/°C
MW	121.14 (free base)	238.3 (free acid)

Table 3-7. Comparison of Tris and HEPES: general features.

Temperature	Tris	HEPES	HEPES
4°C	9.4	9.1	8.6
25°C	8.8	8.8	8.3
72°C	7.5	8.1	7.6
75°C	7.4	8.1	7.6
95°C	6.8	7.8	7.3

Table 3-8. Temperature dependence of Tris and Hepes. The pH of Pfu 1Xreaction buffer is 8.8 at 25°C (Promega and Stratagene).

We tested whether in principle HEPES is compatible with PCR carried out with the Pfu polymerase. Either 10^6 or 10^7 molecules template were amplified in increasing concentrations of HEPES. The other components of the PCR-reaction buffer were the same as in the Promega Pfu buffer. It could be shown that the PCR works as well as in a HEPES- as in a Tris-buffered system (figure 3-43). The optimal concentration for HEPES is between 15 and 30mM. This shows clearly that HEPES represents an attractive alternative to Tris as the buffer of choice in PCR reactions. Further, more refined analysis on the influence of HEPES on the performance of DNA polymerases should be done.

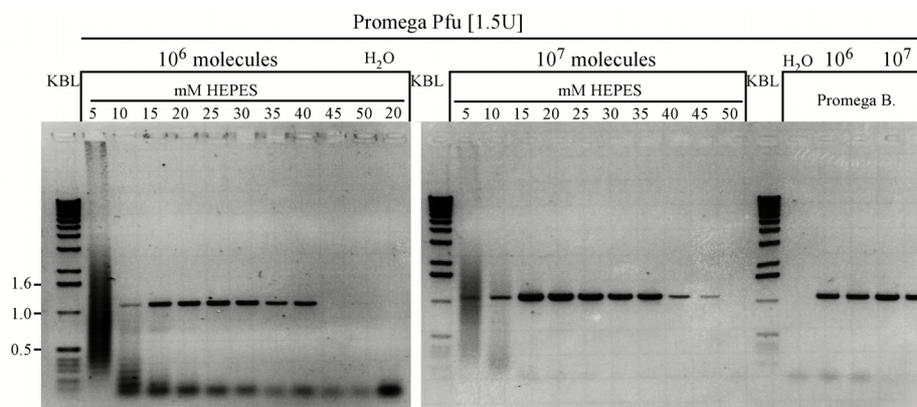


Figure 3-43. PCR performance in HEPES buffered systems. Depicted are pictures of agarose gels (1%; 1XTAE; staining: 0.01% Ethidiumbromide). Loading volume: 25µl. 10^6 or 10^7 molecules of pBKb (see Materials and Methods) were amplified for 30 cycles (50µl; 0.2mM dNTP; 5µmol LNCV; 5µmol LNCK; 0.5µl Promega Pfu; PCR: 3min., 95°C; 30 X [30s, 95°C; 30s, 64°C; 90s 75°C]; 10min. 75°C; ∞ 4°C). 1 X Reaction buffer: 10mM KCl; 2mM MgSO₄; 10mM (NH₄)₂SO₄; 0.1% Triton-100; 0.1mg/ml BSA. HEPES: pH 8.3 at 25°C. Concentration as indicated in figure legend. Promega 1 X reaction buffer: 20mM Tris/HCl pH 8.8 at 25°C.

LNC-PCR on cells: final protocol

Since the final mutation detection protocol should be starting on pooled cells (see introductory section to this chapter), a PCR step had to be set up which efficiently amplifies target sequences from aliquoted cells. The basic protocol consists of a lysis step with ProteinaseK in the presence of the PCR buffer, which precedes the actual PCR. After inactivation of Proteinase K by heat supply the samples are immediately further processed by the PCR step. For reasons of simplicity and processing speed lysis and PCR were done on each sample in the same well of a 96-well plate (see also Materials and Methods). The aim was to produce clearly detectable signals of the amplicon in agarose gels after amplification on pools of 50 and 100 cells. The pool size of 50 and 100 cells was chosen as finally the detection limit of the SSOP assay should be at least at 1 mutant molecule in 99 wildtype molecules. If in one cell of a given pool one allele of K^{bLNC} is mutant than the frequency (with respect to amplicon) is 1:100 for K^{bLNC}/K^{bLNC} homozygous cells with 50 cells and, with 100 cells for haploid K^{bLNC}/K^b cells. Based on the results obtained in tests of the SSOP dot blot system (see 3.2.3.), it was assumed that this would allow for a reliable detection of a mutation occurring in a frequency of 10^{-2} in a given pools of cells. The PCR should be performed in 96-well plates as to ensure a high throughput during the mutation screen. As has been shown in figure 3-40 in first tests two rounds of PCR were employed. The first round consisted of 28 cycles with either Promega or Turbo Pfu, the second one consisted of 14 cycles on 1/10th of the material of the first round, and was carried out with Taq. At the beginning test PCRs had been done with Pfu that has been isolated in this laboratory (see also figure 3-41). However the feeling was that using this PCR was neither reliable nor efficient enough at this stage (see above), and therefore it was decided to proceed with either Turbo Pfu from Stratagene or Pfu obtained from Promega. As can be seen in figure 3-44 A) and B) both polymerases work equally well on the cells. Finally it was decided to use Promega Pfu. At later stages the second PCR round was omitted and only one round of 35 cycles was used. This led to extremely consistent and efficient PCRs. In figure 3-45 typical outcomes of such PCR's are depicted.

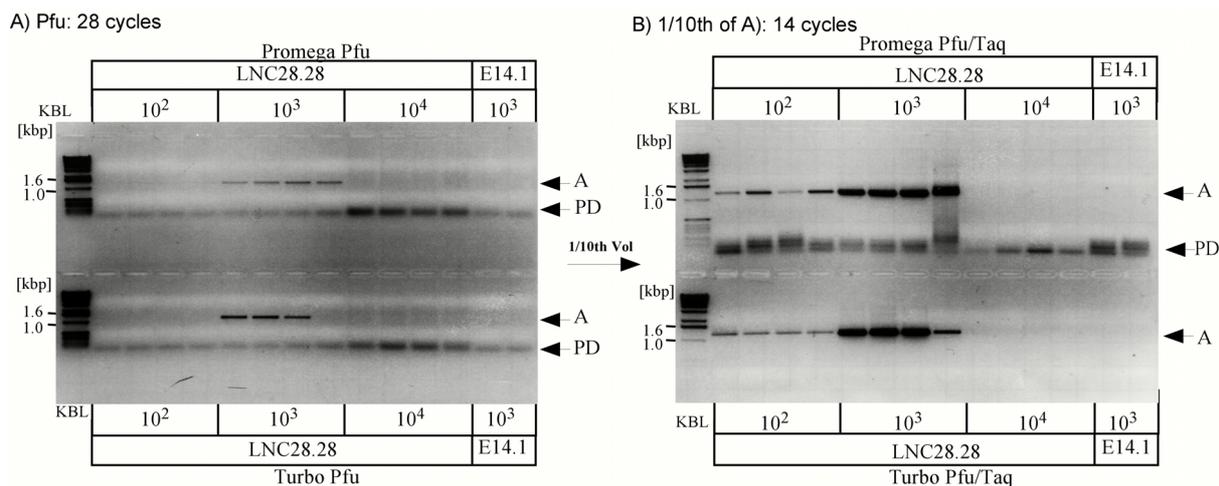


Figure 3-44. Comparison of Promega and Turbo Pfu. Depicted are pictures of agarose gels (1%; 1XTAE; staining: 0.01% Ethidiumbromide). Loading volume: 20µl.

A) First round of PCR: 28 cycles; Pfu

B) Second round of PCR: 1/10th volume of first PCR; 14 cycles; Taq. Lysis of cells: 30µl; 0.5µl Proteinase K (Merck; 10mg/ml) in respective 1XPCR buffer; 50°C, 1h; 95°C, 10min.; PCR conditions of first cycle: 50µl; 0.5µl Promega Pfu (ca. 2U); Promega buffer; 1µl Turbo Pfu (2.5U); Stratagene Pfu buffer; 5pmol LNCV/LNCK. PCR conditions of second cycle: 50µl; 1µl Taq (ca. 5U); 2.5mM MgCl₂; 5pmol of LNCV/LNCK. PD: signal at height of pimer dimer formation: A: signal of amplicon.

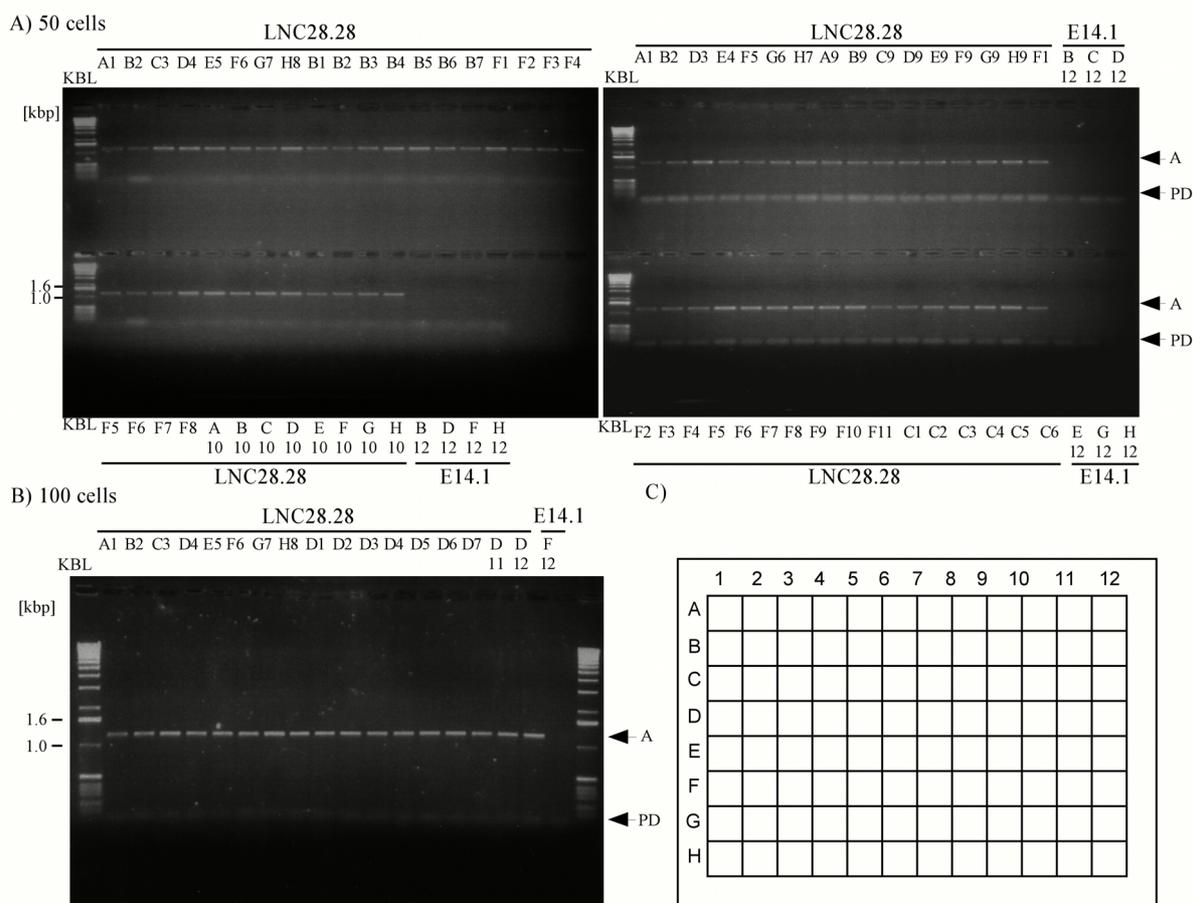


Figure 3-45. LNC-PCR on cells. Depicted are pictures of agarose gels (1%; 1XTAE; staining: 0.01% Ethidiumbromide). Loading volume: 5µl.

A) LNC-PCR performed on 50 cells. Shown are two exemplary results .

B) LNC-PCR performed on 100 cells. Shown is an exemplary result. All PCR's were performed in 96-well plates. Procedure: Lysis of cells. 30µl; 0.2µl Proteinase K (Merck; 10mg/ml); 50°C, 1h; 95°C, 10min. in 1 X PCR buffer (20mM Tris/HCl pH8.8 (25°C); 10mM KCl; 10mM (NH₄)₂SO₄; 2mM MgSO₄; 0.025% Tx-100; 0.1mg/ml BSA). PCR: 50µl; 0.5µl Promega Pfu (ca. 1U); 1 X PCR buffer (see above); 5pmol LNCV/LNCK; 35 cycles. PD: signal at height of pimer dimer formation: A: signal of amplicon.

C) scheme of a 96 well plate for identification of samples loaded onto the gels.

3.2.3. Development of a *bg* specific SSOP-dot blot assay

In the initial proposal to detect and analyse gene conversion like events in the germline of mice, a protocol should be developed and established in which a target enrichment (M450^{LZ}-*lacOp* rescue) precedes a mutation enrichment (tube-DGGE; (Sousa 2000)). As already pointed out (see above and introduction) it became apparent during this project, that the development of such a protocol was highly ambitious and imposed by many obstacles. Therefore it was decided to branch off into a second avenue, which would allow to effectively enrich for mutations, and which could be coupled to the target rescue and the tube-DGGE based mutation enrichment.

It was decided to develop a Sequence Specific Oligonucleotide Probing assay (SSOP) focussing on the detection of the most prevalent *H2* mutation *bg* (see introduction – table 1-1). This step should be preceded by a PCR (discussed above in 3.2.2.), covering the area which contains the *bg* characteristic base exchanges. Advantage was taken of the availability of *K^{bLNC}* transgenic mice and a *K^{bLNC}* transgenic ES-cell line (LNC28.28; see introduction). This PCR should bring mutations to detectable levels.

The SSOP assay should be carried out in a 96-well dot blot system. This would allow for high throughput since the products from the PCR preceding the SSOP, which is done in 96-well plates, can directly be transferred onto filters in this dot blot system. Two oligonucleotides, "BGOT" and "BGOB", each of which is specific for one of the two base exchanges characterising the *bg*-series of *H2* mutations (see also figure 3-28), are employed in the *bg*-specific SSOP assay. It was decided to use a non-radioactive oligonucleotide labelling system, the 3'-oligolabelling and CDP-Star detection system from Amersham (see Material and Methods). We anticipated that the handling of these probes would be easier than that of radioactively labelled probes (for example long-term storage of labelled oligonucleotides is possible, which allows the preparation of large batches of "standardized" labelled probes), and also the sensitivity was, according to the manufacturers information about as good as radioactive labels (http://www.apbiotech.com/technical/technical_index.html). First optimisation tests were done using plasmid DNA, which either contained *K^b* sequence (pBKb) and served as the wildtype species, or Q4 sequence(pBQ4). The latter was used as to imitate the "*bg*" mutant (see Introduction and Materials and Methods and figure 3-28). At later stages the adaptability of the SSOP to PCR products was tested, where more background was anticipated. As already mentioned a mutation detection limit of 10^{-2} was aimed at for the SSOP assay, the reason

being that this would in theory allow to cover 10 000 cells (about one gonad to be analysed) with one to two 96-well plates. This section will introduce the oligonucleotides BGOT and BGOC and describe the power of the *bg*-specific SSOP under the respective conditions employed.

Optimisation of the oligonucleotides BGOT and BGOC

Both the oligonucleotides BGOT and BGOC are 18-mers. BGOT covers the A to T transversion at aminoacid position 116 of *Kb* (leading to Y→F; see table 1-1), while BGOC covers the T to C transition at aminoacid position 121 (leading to C→R; see table 1-1). Their position within the third exon of *K^b* is depicted in figure 3-28. The calculated melting temperatures ($4X[GC]+2X[AT]$) are 56°C for BGOT and 58°C for BGOC (see figure 3-46).

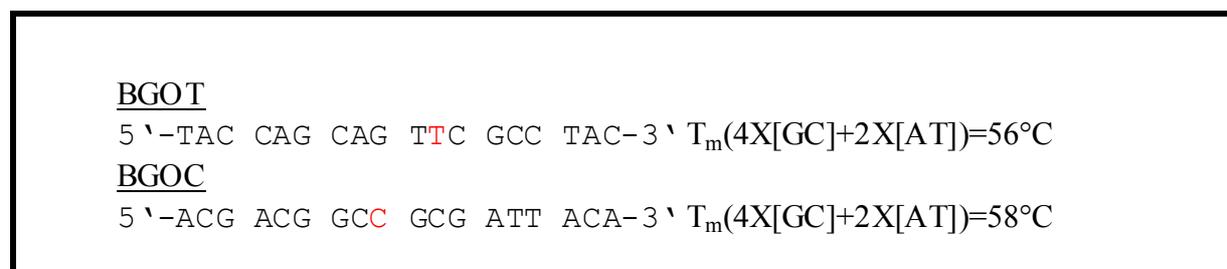


Figure 3-46. Base composition and theoretical melting temperatures of BGOT and BGOC. In red: bases exchanged in the *bg*-mutation.

We tested, whether for the high stringency washes during the blotting procedure, the same temperature could be used. This would facilitate the handling during the assay. As can be seen in figure 3-47 only slight differences in the sensitivity can be observed in a 56°C/58°C interval of the washing steps. As 56°C seems to be slightly advantageous for BGOT, and 58°C for BGOC it was decided to carry out the high stringency washes at these respective temperatures. The specificity is very high (down to 1:125) even at an input of such a high number of background molecules. Interestingly once pBQ4 was diluted into a constant amount of pBKb the detection limit lowered from 1/125 to 1/625.

It was suspected that this “background effect” promoting the sensitivity of the assay is due to the general DNA content present in the system, which is here provided by the pBluescript. A test series was performed in which pBQ4 was either blotted alone, or diluted into pBKb or/and pBluescript (figure 3-48). This blot demonstrates that the background created in the system is directly related to the amount of DNA present.

It could be shown that under ideal conditions, the detection limit for pBQ4, diluted into pBKb can go down to 1/800, compared to pBkB alone (figure 3-49). In the blot presented

this is only visible for BGOC at a molecule amount of 10^{12} copies, and 25 minutes exposure. This stresses that the detection limit in this dot blot system depends on multiple factors, such as total DNA amount and exposure time. However detection limits greater than 1:100, mutant sequence to wildtype were reached repeatedly for both BGOC and BGOT.

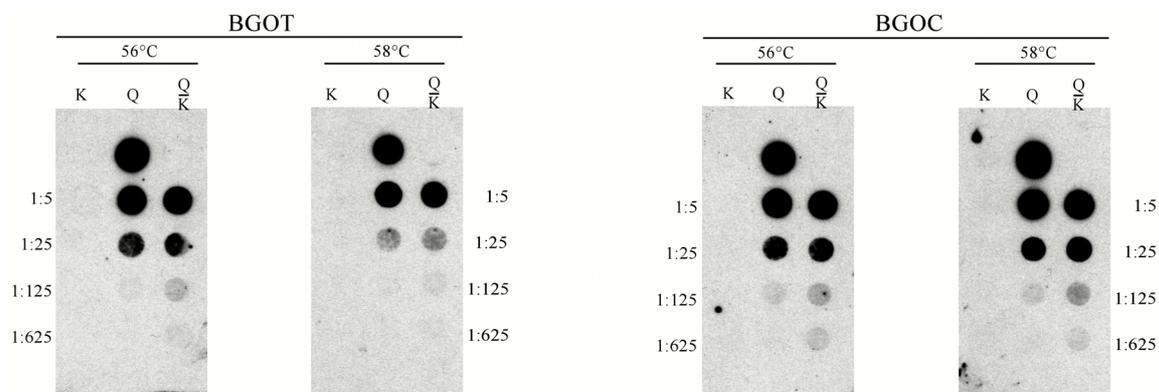


Figure 3-47. High stringency washes during the *bg*-specific SSOP. Depicted is the image of a DNA-dot blot. Starting from 5×10^{13} molecules either the plasmid pBKb (K; background) or pBQ4 (Q, target) or pBQ4 diluted into pBKb (Q/K; here pBQ4 was diluted into pBKb, and the total plasmid copy number was held constant) was blotted in a fivefold dilution series, and probed with BGOT or BGOC. High stringency washes were performed at 56°C and 58°C. The temperature was controlled in boxes identical to those used for the filters, containing the washing buffers. The dot blot assay was carried out under standard conditions as described in Material and Methods. The plasmids are linearised with *Sca* I. Exposure: 20 minutes.

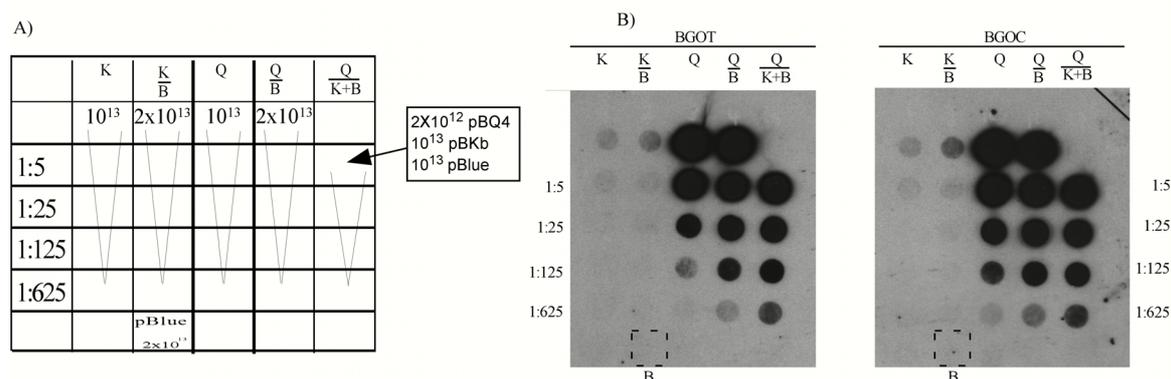


Figure 3-48. Effect of background DNA on sensitivity and specificity of the *bg* specific SSOP. Depicted is the image of a DNA-dot blot. Starting from 10^{13} molecules plasmids (target or background) were diluted in a fivefold series into each other, and probed with either BGOT or BGOC. **A)** Pipetting scheme of the blot. K: pBKb alone, starting from 10^{13} molecules. K/B: pBKb diluted into pBluescript (constant 10^{13} molecules), starting from 10^{13} molecules pBKb. Q: pBQ4 alone starting from 10^{13} molecules. Q/B: pBQ4 diluted into pBluescript (constant 10^{13} molecules) starting from 10^{13} molecules pBQ4. Q/K+B: pBQ4 diluted into pBKb (constant amount 10^{13} molecules), and pBluescript (constant amount of 10^{13} molecules), starting from 2×10^{12} molecules pBQ4. **B)** pBluescript (10^{13} molecules) alone. The dot blot assay was carried out under standard conditions as described in Material and Methods. The plasmids are linearised with *Sca* I (pBQ4, pBKb) or *Eco* RI (pBluescript). Exposure: 30 minutes.

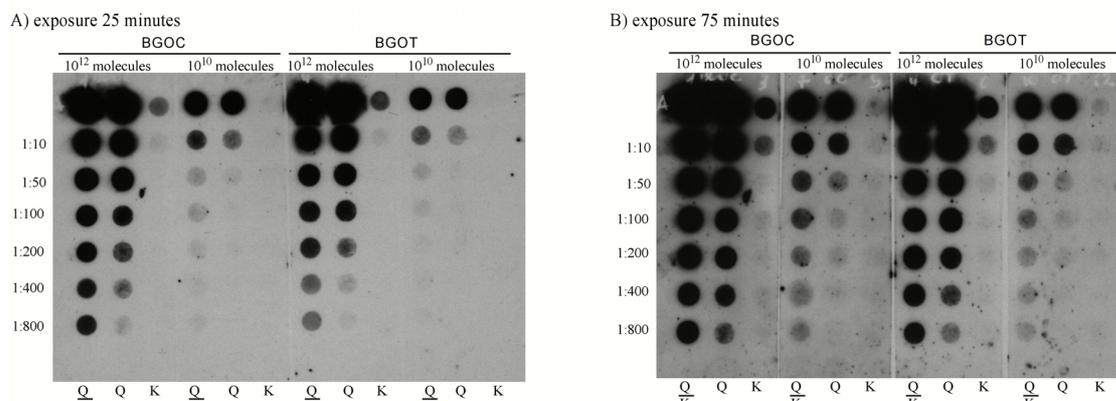


Figure 3-49. Detection limit for pBQ4 in the *bg*-specific SSOP. Either starting from 10^{12} or 10^{10} molecules the plasmid pBQ4 (Q) was diluted into pBKb (K). The dot blot assay was carried out under standard conditions as described in Material and Methods. The plasmids are linearised with *Sca* I (pBQ4, pBKb). Probe (BGOC;BGOT) as indicated. **A)** exposure time 25 minutes. **B)** exposure time 75 minutes.

SSOP on PCR-products: Triton X-100 effect

The *bg*-specific SSOP was supposed to be performed with PCR products. In first attempts it became obvious that the efficiency of blotting PCR products which derived from Pfu-amplified material was by far below that of blotting plasmid DNA or Taq derived PCR products. In figure 3-50 some examples are depicted. In figure 3-50 A) and B) Pfu amplified PCR products are depicted in parallel to plasmid DNA (A)) or Taq amplified material (B)). As soon as PCR products which have been amplified with Pfu are blotted, the signals get dispersed and show a "corona-like" outer ring. It turned out that this "dispersion effect" of the signal, is caused by the detergent Triton X-100. As can be seen in figure 3-50, as soon as Plasmid DNA is blotted in the presence of Triton X-100, the Pfu-PCR effect appears to an impressive extent. The signals basically disappear. Clearly under these conditions the SSOP assay is dramatically handicapped. At the same time figure 3-50 C) illustrates how it is possible to reinstate the vanishing signals. After purification of the PCR product (in this case with the High-Pure PCR-Product Purification Kit (Roche)), the signal "reappears" without any corona effect. As we wanted to avoid any further step in the LNC-PCR/bgSSOP assay, we decided to test whether we can simply lower the concentration of Triton X-100 in the Pfu-PCR buffer, instead of introducing a post-PCR purification step. The result is shown in figure 3-51. A standard LNC-PCR on cells was performed with Pfu-PCR reaction buffer containing decreasing amounts of the detergent. The original 10 X Pfu reaction buffers, supplied by Promega and Stratagene (see respective catalogues) contain 1% Triton X-100. Already by lowering the detergent content by a factor of 2 to 0.5% brings back the normal signal strength, apparently without lowering the PCR-efficiency. Based on this result we decided to prepare and use for all subsequent PCR reactions employing Pfu, a 10X reaction buffer containing 0.25% Triton X-100.

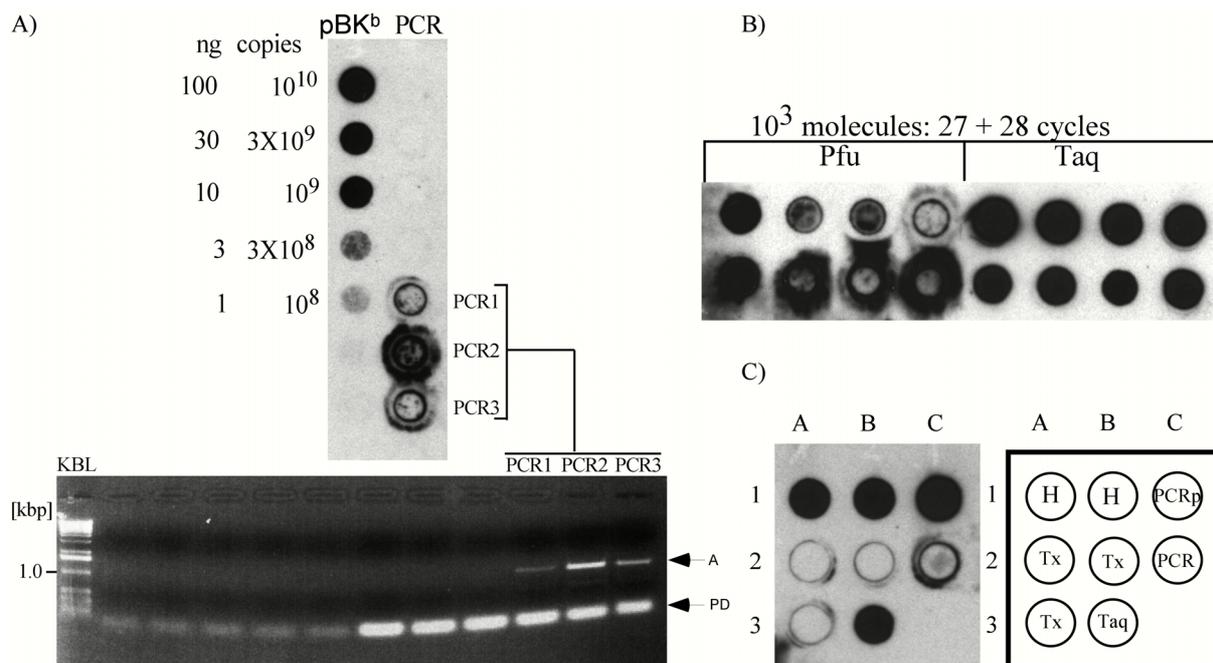


Figure 3-50. Triton X-100 effect on blotted PCR products in the *bg*-specific SSOP.

A) shown is a dot blot SSOP onto which a dilution series, starting with 10^{10} molecules, of the plasmid pBKb has been blotted. In parallel three LNC-PCR products (template pBKb), which are also depicted on the lower picture of the agarose gel have been blotted. The SSOP assay was done under standard conditions (see Materials and Methods). As a probe the 3'-fluorescein dUTP labelled oligonucleotide KB-oligo was used. PCR: Two rounds (25 and 28 cycles on 1/10th volume of first round) of LNC-PCR on 10^2 molecules of pBKb. 20 μ l of the PCR product went into the SSOP assay, 20 μ l were loaded onto the agarose gel depicted (1%; 1x TAE, 0.01% Ethidiumbromide).

B) dot blot SSOP of PCR products, either amplified with Pfu or Taq. 20 μ l of a LNC-PCR reaction, consisting of two rounds (27 and 28 cycles on 1/10th volume of first round; template 10^3 molecules pBKb were blotted and probed as in A).

C) Comparison of plasmid DNA blotted in presence of 0.1% Triton X-100 (A2; A3; B2), or in H $_2$ O (A1; B1), and of PCR product of a Taq amplification (B3) to a Pfu amplification (C2), as well as to a purified Pfu-PCR product (C1; High-Pure PCR-Product Purification Kit; Hoffmann La Roche). Blotted were either 10^{10} molecules of the plasmid pBKb, or 10 μ l of a LNC-PCR reaction performed as in B). Blotting and probing as in A).

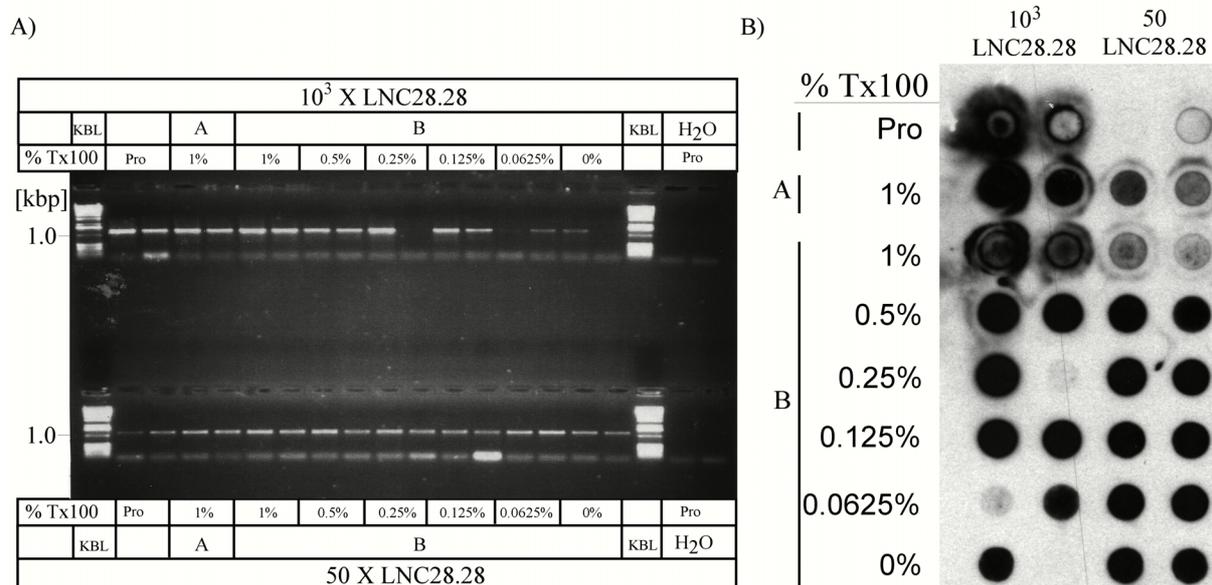


Figure 3-51. Blotting efficiency in dependence of Triton X-100 concentration in the *bg*-specific SSOP.

A) Depicted is an agarose gel (1%; 1XTAE; 0.01% Ethidiumbromide), onto which 5 μ l of a cell LNC-PCR product performed in a twofold dilution series of Triton X-100 was loaded. Indicated are the concentrations of Triton X-100 of the respective 10XPCR reaction buffers used.

B) SSOP dot blot of the PCR products presented in A). PCR: Standard LNC-PCR performed on either 10^3 or 50 LNC28.28 cells (negative control: 10^3 E14.1 cells; see Material and Methods), in the presence of the indicated concentration of Triton X-100 (10XPfu PCR reaction buffer). The Lysis of the cells was performed in the same buffer conditions. SSOP: 25 μ l of the PCR product was blotted, and probed under standard conditions with the 3'-fluorescein dUTP labelled oligonucleotide KB-oligo (see Materials and Methods). Each test was performed and is depicted in duplicates. Pro: Promega PCR-reaction buffer. A; B: two Pfu buffers, prepared independently.

SSOP: signal intensity dependence on exposure time in the CDP-Star™ detection system

From Amersham no information about the signal saturation behaviour of 3'-fluorescein-dUTP labelled oligonucleotides and CDP-star detection (see Materials and Methods) in dependence of exposure time was available (personal communication). Therefore in one experiment the saturation behaviour was tested by exposing a filter which had been probed with 3'-fluorescein-dUTP labelled Kb-oligo and developed with the CDP-star detection kit, for different time intervals (see figure 3-52). In this case between 30 seconds and 20 minutes exposure the signal strength increased exponentially. At some stage beyond 20 minutes exposure the background of the film is too high as to detect any signal. It has to be said that in my experience the "efficiency" of the flourolabelled oligonucleotides varies considerably. As will be seen in section 3.2.5. different batches of oligonucleotides need very different exposure times to get to the same signal intensities. Still it is probably true that the time interval were one can expect an unsaturated signal strength upon probing with a fluoresceinlabelled oligonucleotide stay more or less the same.

Both the K^{bLNC} -PCR and the SSOP part of the mutation detection assay presented here were promising with respect to identify mutations in the germline of mice. Still, facing the low expected frequency of these mutations we felt that an experiment had to be designed which would come closer to the ultimate conditions that we had to expect. Such an "evaluation-experiment" should enable us to conclude how powerful the detection assay finally will. This attempt will be described in the next chapters.

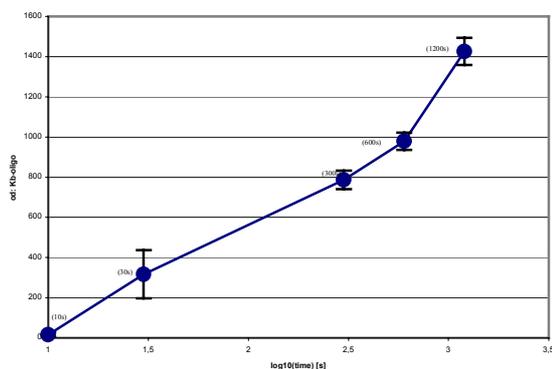


Figure 3-52. SSOP using fluorescein-dUTP labelled probes and CDP-star detection (Amersham): Signal intensities against exposure time. Shown is a semi logarithmically scaled graph, into which the signal intensities obtained for a fluorescein-labelled Kb-oligonucleotide probed filter are plotted against exposure times. In short 50 cells went into a standard LNC-PCR/SSOP assay (see Material and Methods). The filter was first probed with BGOC and after stripping with Kb-oligo. The oligonucleotides had been labelled freshly. Evaluation: The measured od of each sample per time point (sum:88 samples) was collected and the arithmetic mean plus standard deviation (error bars) calculated. Note: measurements were taken until 2400s exposure time. At later time points than 1200s the general background increases to such an extent that the signal are no longer detectable.

3.2.4. Design and cloning of the evaluation plasmid pBbgMCB

The SSOP experiments using plasmid DNA were very promising in terms of specificity. As shown in figure 3-49 a detection limit greater than 10^{-2} for mutant diluted into wildtype sequence could be envisioned. However the PCR step preceding the dot blot assay turned out to be a weaker element of the mutation detection procedure, mainly due to statistical fluctuation. The PCR serves to elevate mutations to a level at which they can be detected with SSOP. At a theoretical 100% efficiency of the PCR reactions the distribution of mutations should follow a Poisson distribution due to the rareness with which they occur (Sokal *et al.* 1998). Additionally to the introduction of statistical fluctuation it is also expected that PCR introduces a higher degree of background into the SSOP system, due to impurities of the PCR products. In order to evaluate the power of the LNC-PCR/SSOP mutation detection system it was decided to clone a plasmid which would enable us to closely imitate the ultimate *bg*-detection assay. This plasmid should carry the two base exchanges characterising the *bg*-mutation embedded in K^b sequence. Also it should be amplifiable with the primer pair LNCV/LNCK. Still the amplification product resulting from this plasmid should be unambiguously distinguishable from any real mutation occurring in the K^b gene. This plasmid was called pBbgMCB, and its design and cloning will be described here.

The design of the plasmid pBbgMCB is outlined in figure 3-53. It is based on the plasmid pBKb (see Material and Methods). A synthesized insert should be created by annealing and ligation of 12 oligonucleotides. This insert covers the region of the third exon of K^b , and harbour the *bg*-characteristic base exchanges. Additionally it is framed by two sequence stretches of 20bp length, which derive from the *chlorophyll a/b-binding* gene (*CAB2*) of the common ice plant *Mesembryanthemum crystallinum* (Michalowsky and Bohnert; EMBL Genbank accession number AF003128). These “buffer sequences” should ensure that the artificial mutant cannot be mistaken for a real mutation in later mutation detection experiments carried out in this laboratory. The insert should replace a *PpuM I/Bcl I* fragment cut out of pBKb. In the fragment of *Kb* being integrated into pBluescript three *PpuM I*-sites are present (position 1206; 1307 and 1419). The *PpuM I*-site at position 1206 should be transformed into a *Nla V* site by site directed mutagenesis. Subsequently the created *bg*-carrying insert should be integrated into the *PpuM I/Bcl I* restricted plasmid. The resulting LNC/LNCV amplicon is 112bp shorter than the endogenous amplicon, which additionally facilitates the identification of this artificial mutant. The plasmid pBKb is based on the plasmid pBluescript KS (Stratagene), which is no longer commercially available, while

the pBluescript part of the replacement vector K^{bLNC} is based on the plasmid pBluescript II KS (Guethlein and Vopper, unpublished data). The anchor site of the primer LNCV covers a *Bss*H II site (see figure 3-28) which is available in pBluescript II but not in pBluescript (see Stratagene catalogue). Therefore this site had to be inserted as to ensure an efficient PCR. This was also done by site directed mutagenesis.

Assembly of the insert bgMCB

Twelve oligonucleotides as depicted in figure 3-54 were ordered (ABI) with the 5' ends already phosphorylated. They were annealed, and all six resulting double stranded fragments ligated (either first fragments [1+2+3] and [4+5+6], or immediately all six fragments at once [1-6]). As can be seen in figure 3-55 this ligation approach not one distinct 200bp fragment was created but apparently a mixture of fragments of different sizes. The ligation sample in which all six fragments were put together ([1-6]), was taken and amplified with the oligonucleotides 1s and 6a. A distinct fragment of the expected size of 200bp could be recovered successfully by this means (figure 3-56). In order to create *PpuM I/Bcl I* overhangs, PCRs with the primers O1sPpu and O6aBcl2, were performed on a fraction of the ligated DNA fragments (figure 3-57). The oligo O6aBcl2 additionally reintroduced a *Bcl I* site which, due to the original oligonucleotide-design, had been transformed into a *Mbo I* site (see figure 3-53). The resulting material was digested with *PpuM I* and *Bcl I* in succession, purified, and used in the following cloning steps (figure 3-58).

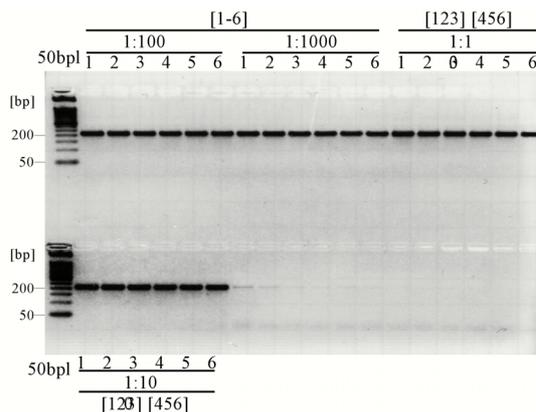


Figure 3-56. PCR rescue of "bgMCB" ligation products. Depicted is an Agarosegel (2%; 1XTAE; 0.01% Ethidiumbromide). Loading volume: 20µl. The ligation samples were diluted as indicated and 1µl was amplified with oli1s and oli6a (see figure3-54) for 30 cycles (Biometra Tgradient™; 50µl; 1µl template; 2.5U Taq; 2.5mM MgCl₂; 5pmol of each primer; T_{annealing}: 1=72.3°C; 2=73.1°C; 3=73.8°C; 4=74.4°C; 5=74.8°C; 6=75°C. PCR cycles: 95°C, 15s;29X[95°C, 15s; T_{annealing}, 15s; 75°C; 30s]; 75°C, 5minutes; 4°C, ∞. [1-6]: ligation [1-6] of figure 3-55; [123] [456]: ligations [123] and [456] (figure 3-55) ligated to each others.

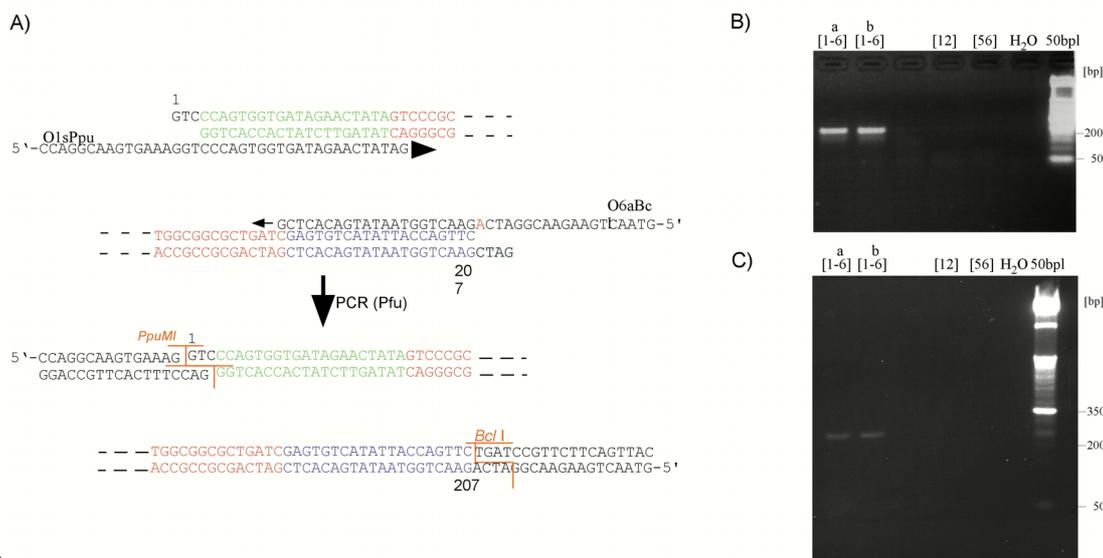


Figure 3-57. PCR rescue of "bgMCB" ligation products and recovery of the PpuMI and BclI site. A) sketch of annealing of O1sPpu and O6aBcl to the opposite ends of the insert bgMCB, and outcome of the PCR. The mismatched nucleotide A in O6aBcl creating the Bcl I site is depicted in red. B) Depicted is an Agarosegel (3%; 1XTAE; 0.01% Ethidiumbromide). Loading volume: 10µl. Short run of products of the PCR performed on ligation samples. Shown are the PCRs on two independent ligations of fragments 1 to 6 ([1-6] a and b; see also figure 3-55 [12] and [56] are samples of fragment 1 ligated to 2 and 5 to 6, and serve in addition to the water control (H₂O) as negative controls. C) long run of B). PCR: 50µl; Template: 1µl of a 1:100 dilution of the ligation samples; 1.5U Pfu (Promega); 5pmol O1sPpu/O6aBcl; 0.2mM dNTP; (95°C, 15"; 30X[95°C, 15"; 60°C, 15"; 75°C, 30"]; 75°C, 5'; 4°C, ∞). Marker: 50bp ladder (Gibco BRL).

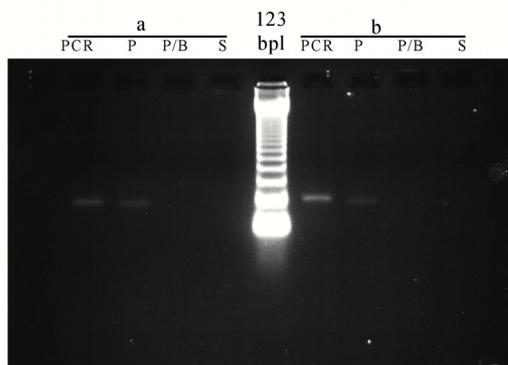


Figure 3-58. Recovery of "bgMCB" ligation product after PpuMI/Bcl I digest. Depicted is an Agarosegel (3%; 1XTAE; 0.01% Ethidiumbromide). Loading volume: 6µl. The PCR rescued "bgMMCB" ligation products (samples a and b of Figure 30) were successively digested with PpuMI and Bcl I and purified by running through a Sephadex G-50 column (Pharmacia; DNA-grade). PCR: sample after PCR of Figure 3-57. P: sample after PpuMI digest. P/B: sample after digest with Bcl I. S: sample after running through Sephadex column.

Cloning of the plasmid pBbgMCB

Starting material was the plasmid pBKb (see Material and Methods). A *PpuM* I site at position 1206 (figure 3-59) was transformed into a *NlaI* V site, following the Stratagene “Quick Change™ Site directed Mutagenesis Kit” instructions (protocol see Material and Methods and www.stratagene.com/mutagenesis/quikchngxl.htm). The primers used are depicted in (figure 3-59). The resulting material (figure 3-60) was used to transform a *dam*-methylation negative *E.coli* strain (*SCS110*, Stratagene). One of the resulting clones was used to produce plasmid DNA and digested successively with *PpuM* I and *Bcl* I. The gel purified plasmid was used for the ligation step with the insert bgMMCB. The ligated material was transformed into *E.coli* strain *DH5a* and minipreps with resulting clones tested in SSOP dot blots with the oligonucleotides BGOc and BGOT (figure 3-61). Clone 4, was used to introduce a *BssH* II site like before to create the LNCV primer anchor region via site directed Mutagenesis (figure 3-62 A) to C)). The success of this operation was confirmed with an *BssH* II digest (figure 3-62 D)) on several clones and sequencing (see Appendix). Clone 1 was finally chosen for all later experiments as plasmid pBbgMCB.

Test of pBbgMCB in the LNC-PCR system

The effect of converting pBKb into pBbgMCB on the LNC-PCR was tested in comparison to the plasmid pBKb, both with or without the presence of cells, and in comparison to the PCR carried out directly on LNC28.28 cells (figure 3-63). The LNC-PCR performed on pBbgMCB is about 100 times as efficient as being performed on pBKb. The successful cloning of pBbgMCB resulted in the availability of a valuable tool to evaluate the combined KblNC/bgSSOP assay. Its characterisation with respect to the power with which it is able to unravel *bg* mutations will be described in the following section.

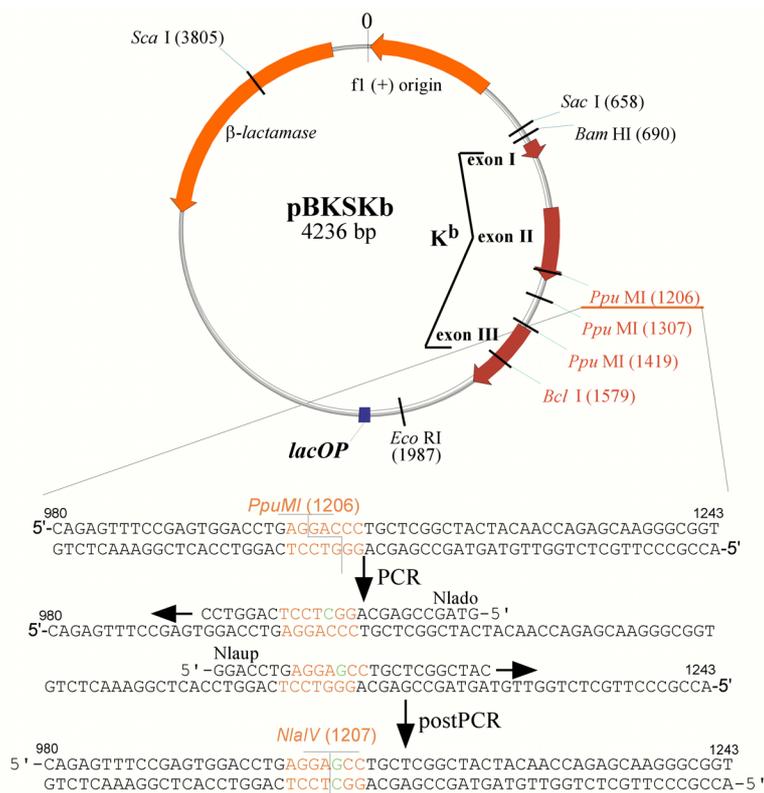


Figure 3.61 Elimination of the Ppu MI site at position 1206 of pBKb. Depicted is a drawing of the plasmid pBKb (see Material and Methods) containing the first three exons of the mouse H-2 gene Kb. The Ppu MI site (position 1206) was eliminated by using a PCR-based site directed mutagenesis protocol (see text) employing the primers Nlado and Nlaup.

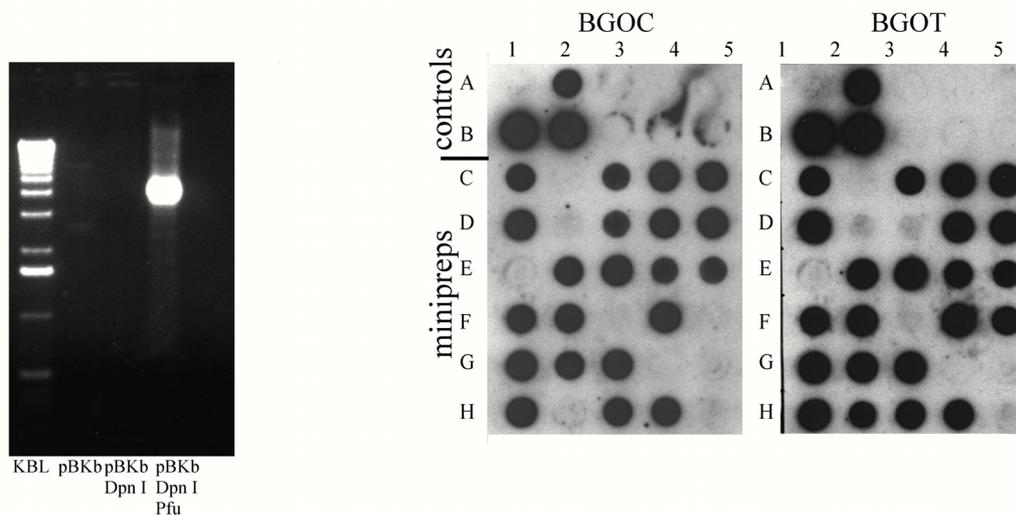


Figure 3.62 Elimination of the Ppu MI site at position 1206 of pBKb. Depicted is an Agarosegel (1%; 1XTAE; 0.01% Ethidiumbromide). Loading volume: 15 μ l. A site directed mutagenesis protocol was carried out as suggested by Stratagene. 15 μ l of the treated product (total volume: 50 μ l) were loaded onto an agarose gel for control purposes. pBKb: untreated sample. Amount of plasmid pBKSkb that went into the site directed mutagenesis protocol. pBKb DpnI: As pBKb, but treatment with Dpn I. pBKb Dpn I Pfu: sample after site directed mutagenesis protocol. Site directed mutagenesis (in short, see also Quick ChangeTM XL Site-directed Mutagenesis Kit at www.stratagene.com/mutagenesis/quickchngxl.htm) PCR: 50 μ l; 25ng pBKSkb; 2.5U Turbo Pfu (Stratagene); 16pmol Nlaup; 16pmol Nlado; 0.2mM dNTP; (95 $^{\circ}$ C, 30"; 12X[95 $^{\circ}$ C, 30"; 55 $^{\circ}$ C, 1'; 68 $^{\circ}$ C, 8'40"]; 4 $^{\circ}$ C, ∞)

Figure 3.63 Test of Minipreps for clones of pBbgMCB. 2 μ l of each miniprep were blotted onto a Nylon Membrane (Amersham; Hybond N+), in 50 μ l total volume. The blots were probed with the oligonucleotides BGOC and BGOT. The clone C4 was chosen for further steps. C1 to H4: samples of Minipreps. A1, A2, B1, B2: controls. A1: pBKSkb (100ng); A2: pBQ4 (100ng); B1, B2: PCR product of Figure 30 samples [1-6] a (B1) and b (B2) (10 μ l). The SSOP dot-blot was done under standard conditions as described in Material and Methods.

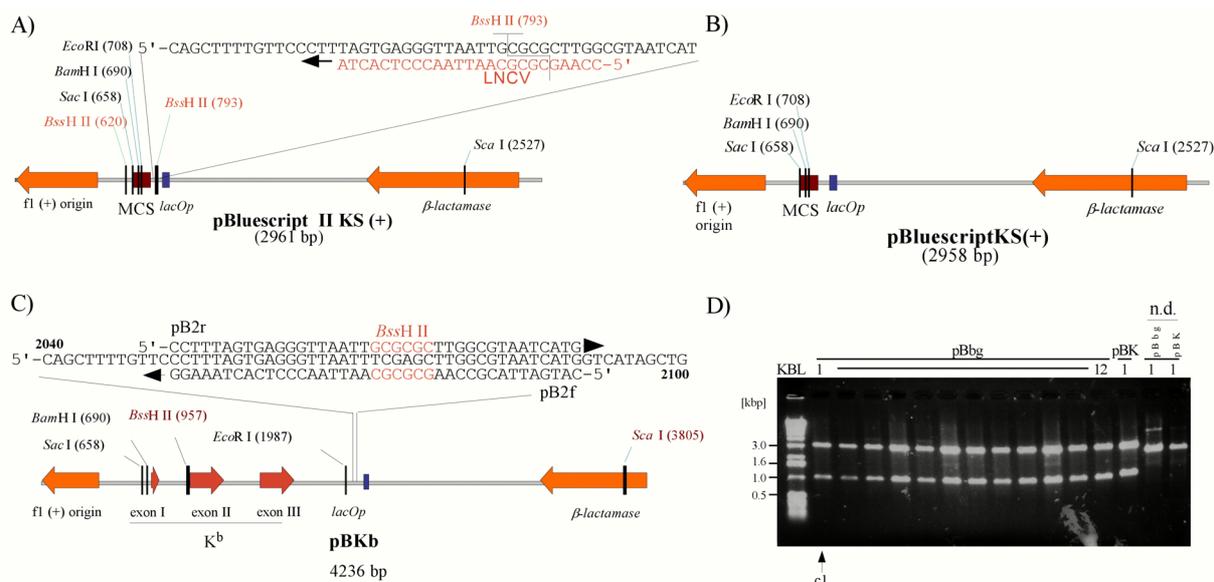


Figure 3-62. Introduction of the BssH II site into pBKSkb. Depicted is a drawing of the plasmids pBluescript KS II (+) (in **A**) and pBluescript KS (+) (in **B**). In pBluescript II a BssH I site is present (Position 793), which is part of the LNCV-anchor site, and which is absent in pBluescript (see Stratagene Catalogue). **C**) drawing of pBKb which is based on pBluescript KS (see Materials and Methods). Therefore the BSS HII site at position 2072 which is part of the LNCV anchor site is lacking. It was introduced by PCR-based site directed mutagenesis (as described in figure 3-60) with the primer pair pB2r and pB2f. **D**) Control of the successful integration of a BssH I site into pBbgMCB. Depicted is an agarose gel of BssH II control digests (1%: 1XTAE; 20µl loading volume). Expected are fragments of 1.1kb and 3.1kb length (see C). pBbg: clones that derived from site directed mutagenis to introduce BssH II site. pBK: In parallel to the plasmid that should be transformed into pBbgMCB, the into the plasmids pBKb, and pBQ4 the BssH II site was also introduced. Only one clone of pBKb is shown here, which derives from this attempt. n.d.: samples (pBbg, clone 1; pBKb, clone1) which were not digested with BssH I. Clone 1 of pBbg (indicated by the arrow) was chosen for all later experiments.

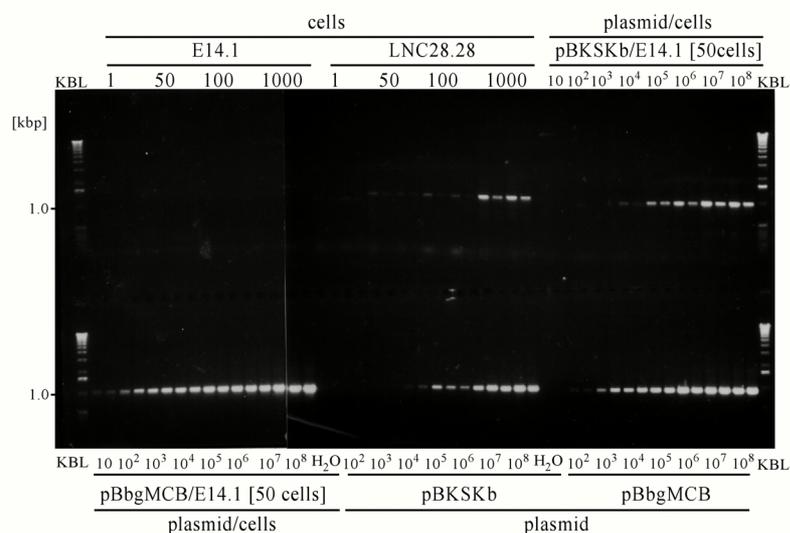


Figure 3-63. PCR on pBbgMCB, pBKb, and LNC28.28 cells. Depicted is an agarose gel (1%; 1XTAE; loading volume: 10µl). All samples were amplified on a 96-well plate. Samples containing cells underwent Protein lysis before PCR. A standard LNC-PCR was performed as described in Materials and Methods. E14.1: E14.1 stem cells, negative controls, umbers indicate cell number that went into PCR. LNC28.28: numbers indicate cell numbers. pBKb/E14.1 [50 cells]: plasmid pBKSkb amplified in presence of 50 E14.1 cells. numbers indicate plasmid copie number. pBbgMCB/E14.1 [50 cells]: plasmid pBbgMCB amplified in presence of 50 E14.1 cells. numbers indicate plasmid copie number. pBKb: plasmid pBKb amplified. numbers indicate plasmid copie number. pBKb: plasmid pBbgMCB amplified. numbers indicate plasmid copie number.

3.2.5. Evaluation of the power of the K^{bLNC} -PCR/SSOP based *bg* mutation detection assay

With the successful cloning of pBbgMCB, and an established protocol for a K^{bLNC} -specific PCR on LNC28.28 cells, it became possible to design an experiment which would allow us to estimate the mutation detection power of the K^{bLNC} -PCR/bgSSOP assay. To design and execute such an experiment was necessary, as it had become apparent that despite the fact that the components of the detection assay were described fairly well, it was not clear how effectively the assay would finally be able to discover *bg*-like mutations occurring in mouse cells. The major handicap concerning this enterprise is the low expected frequency with which such mutations are suspected to occur, combined with the high potential risk of creating *in vitro* artefacts as discussed in the introduction. The low frequency implies a considerable degree of statistical fluctuation, rendering the analysis complicated. Furthermore our mutation frequency estimate is based on published data on *in vivo* mutations mainly derived from C57BL/6 mice (Martinson *et al.* 1999). In fact the sample size is not big enough to allow for reliable predictions, weakening the working hypothesis of an expected frequency in the range of 10^{-4} . Additionally the assay is made up of elements which are themselves vulnerable to variation. For the SSOP this concerns background, even though it proved to be a generally reliable and robust method. The PCR-step serves to elevate mutations to detectable levels. If one assumes that the occurrence of mutations follows a binomial distribution, and that the overall frequency is indeed 10^{-4} , then in average 1% of all pools containing 100 cells should harbour at least one mutation (Sokal *et al.* 1998). In other words in average per 96-well plate to analyse one mutation should be detectable. This is of course only true if the PCR is 100% efficient, i.e. each mutation is indeed elevated to a detectable level by the PCR. It has been shown that the PCR conditions used here on cells lead to a very reliable and robust amplification of the target (see figure 3-47). However since the mutation itself presents a rare event in the cell pool to be analysed, again statistical fluctuation is to be expected. It is for example of importance in which cycle during the amplification process the mutation is picked up by the PCR.

For this reason it was our feeling that to leap directly into the analysis of germline derived cells of mice, for analytical purposes would not be advisable. Beforehand it would be helpful to get an estimate of the effort that would have to be put forward to find at least one mutation at a starting frequency of 10^{-4} . So starting from a known mutation frequency, being above 10^{-4} , a deduction should be drawn about what effort has to be made to detect mutations. Also only this evaluation would enable us later to estimate mutation frequencies from real

observed events, as we are able to include the observed efficiency of the assay into our estimate.

For this purpose the plasmid pBbgMCB was utilized as a representative for a "bg-mutant" K^{bLNC} sequence, and mixed at different dilutions steps with LNC28.28 cells. These mixtures went through the LNC-PCR/SSOP based mutation detection procedure. The number of cell pools carrying pBbgMCB (potentially mutant pools) was estimated upon transferring the PCR products of the LNC-PCR into the SSOP-assay (Materials and Methods) and probing with either the oligonucleotide BGOc or BGOT. Two series of this kind of experiment were carried out. In the first case pools of 50 cells were analysed harbouring pBbgMCB at dilutions of 1/50; 1/150, and 1/450 (note: the dilutions refer to the number of cells and therefore to the number of LNC-PCR amplifiable target sequences as the LNC28.28 cells are haploid for K^{bLNC}). Secondly pools of 100 cells were examined, harbouring pBbgMCB at dilutions of 1/100, 1/200, 1/400, and 1/800.

The evaluation is based on the assumption that the distribution of mutants follows the binomial or Poisson distribution (Sokal and Rohlf 1995). This means that if the PCR step during the assay works with 100% efficiency, at a dilution of 1/100 about 60% of all examined wells should be positive (i.e. containing AT LEAST one mutant molecule). The number of detected potential mutants along the dilution series should be compared to the expected frequencies, and a trend for higher dilutions deduced, which finally should lead to an estimation, about how many plates would have to be analysed to detect at least one mutation at a given frequency of 1/10 000.

We also examined the general characteristics of the combined LNC-PCR/bg-specific SSOP assay. This included a comparison of the performance of the two oligonucleotides BGOT and BGOc. We furthermore used the assay to estimate the LNC-PCR efficiency, an important point with respect to detection of rare mutation events. This efficiency test stimulated us to follow the fate of rare specimen in PCR assays in computer simulations. That is to say we were interested to know whether, starting a PCR on two specimen of molecules at given ratio, this ratio is maintained throughout the amplification or shifted towards or away from the rare specimen.

Detection of pBbgMCB diluted into pools of 50 cells

Starting from mastermixes, 96-well plates were prepared which contained 50 cells per well, into which pBbgMCB had been diluted at steps of 1/50, 1/150, and 1/450. As

negative controls each plate contained two pools of E14.1 ES cells, which do not provide the LNC-amplicon. (see figure 3-64). The cells were lysed and the LNC-PCR was performed as described in Material and Methods. Two aliquots of each PCR-sample were transferred onto separate nylon membranes in a dot blot system. At this stage a positive control for the blotting and detection efficiency was introduced by transferring 10^{10} molecules pBbgMCB onto each filter (sample 96; see figure 3-64). The filters were first probed for the “mutant molecule” pBbgMCB with either BGOc or BGOT, and (after stripping) subsequently with KBO58 which recognises as well the K^b -wildtype sequence, as pBbgMCB. Examples of developed blots are depicted in figure 3-65.

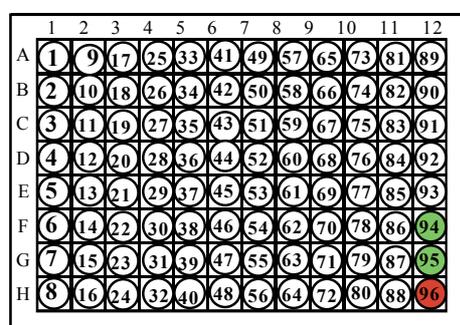


Figure 3-64 Numbering of samples in dot-blot system. Samples 94, 95: negative controls (E.14.1 cells). Sample 96: positive control (10^{10} molecules pBbgMCB)

The intensity of each dot was measured by densitometry. In figure 3-66 examples of obtained values at different exposure times for the oligonucleotides BGOT and KBO58 are plotted against each others in a two-dimensional X-Y coordinate system. Based on these plots certain dots were considered to represent possible candidates of cell pools containing the plasmid pBbgMCB as a mutant. The choice of such candidates was based on visual inspection judging the degree with which points in the coordinate system stuck out of the bulk population. In parallel, cluster analysis was applied as an independent heuristic measure (program used: SPSS9.0). This also served for the purpose of determining if, and to what degree the whole procedure could be automated. As can be seen in the overview in the scatter plots (figure 3-67), the resolution of potential mutant cell pools seems to be favoured at exposures of 5 minutes in contrast to 90 seconds for the mutant specific oligonucleotide BGOT and 20 minutes exposures in contrast to 5 minutes exposures for KBO58 (see also section Comparison of BGOT to BGOc). However it is also evident that finding the optimal exposure time is a “*va banque*” game. At 5 minutes exposure time for BGOT the od-values of the bulk population start to “stretch out” along the y-axis (od: BGOT), and ultimately the resolution for the potential mutant candidates would get lost. This suggests that it is advisable

to work with some "explorative mock filters", to test out a certain batch of fluorolabelled oligonucleotides with respect to optimal exposure time before treating the experimental filters (see also 3.2.3. - SSOP: signal intensities in dependence of exposure time in the CDP-Star™ detection system).

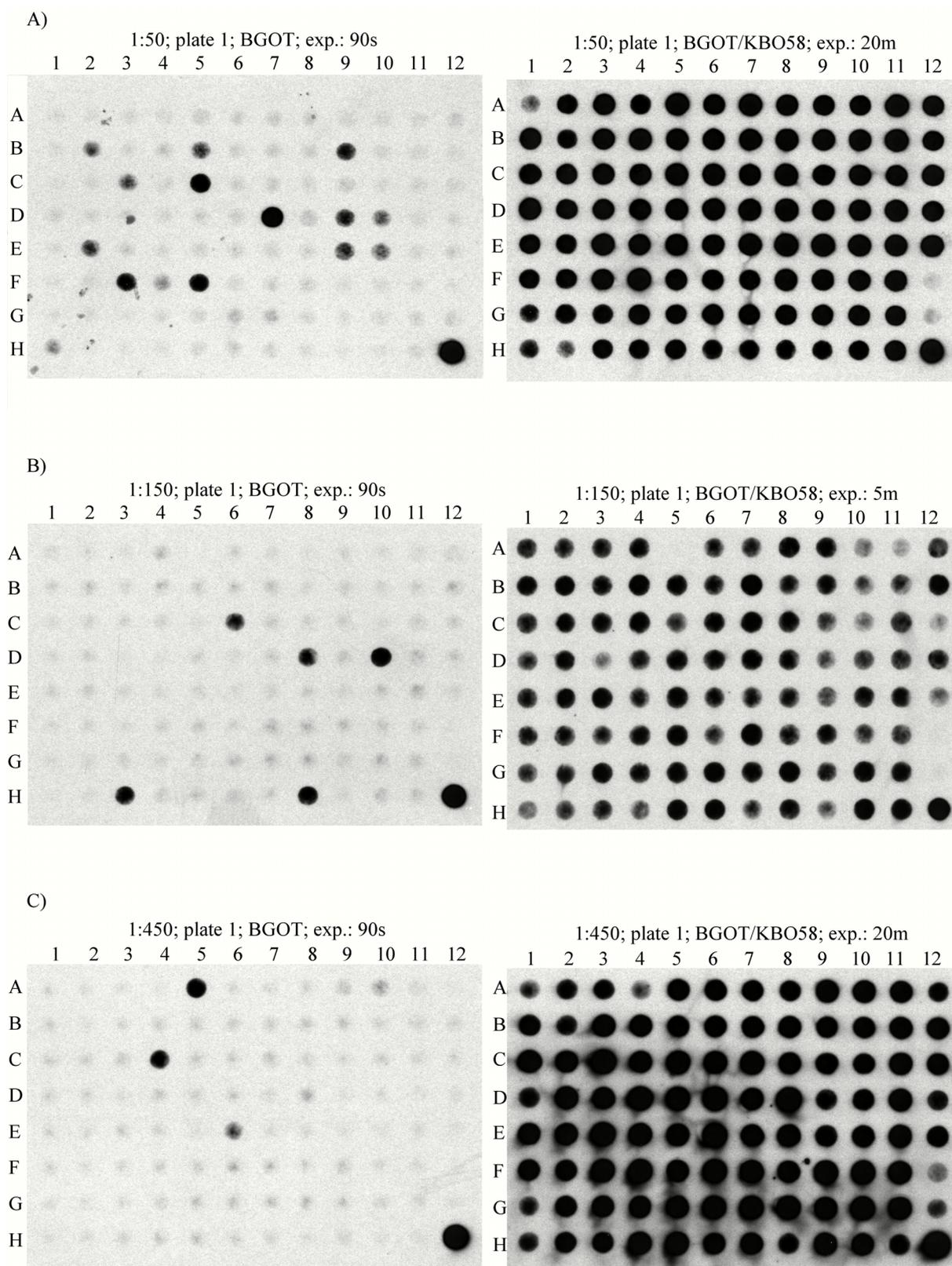
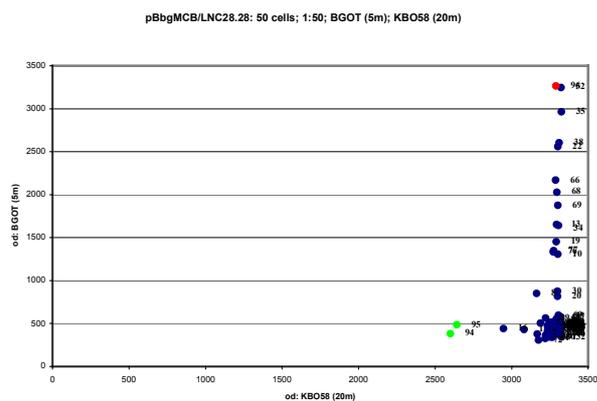
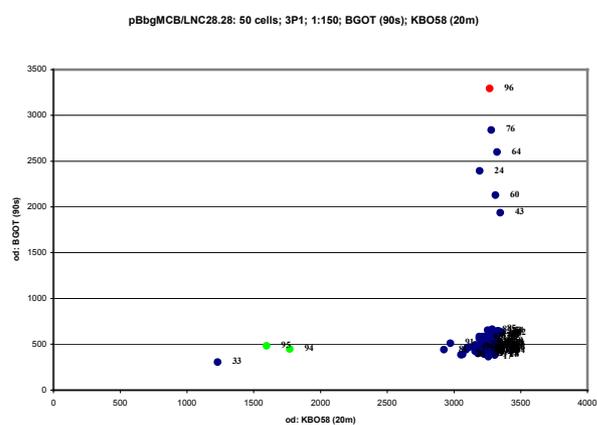


Figure 3-65. pBbgMCB/LNC28.28 (50 cells) - examples of SSOP dot blots. Depicted are examples of filters derived from the 50 cell experiment which were probed with BGOT (left panel) and subsequently to stripping with the oligonucleotide KBO58 (right panel). Each well represents 20 μ l of a LNC-PCR, being transferred onto the filter. **A)** 1:50 dilution; BGOT (exposure 90 seconds); KBO58 (exposure 20 minutes). **B)** 1:150 dilution; plate 1; BGOT (exposure 90 seconds); KBO58 (exposure 5 minutes). **C)** 1:450 dilution; plate 1; BGOT (exposure 90 seconds); KBO58 (exposure 20 minutes).

A)



B)



C)

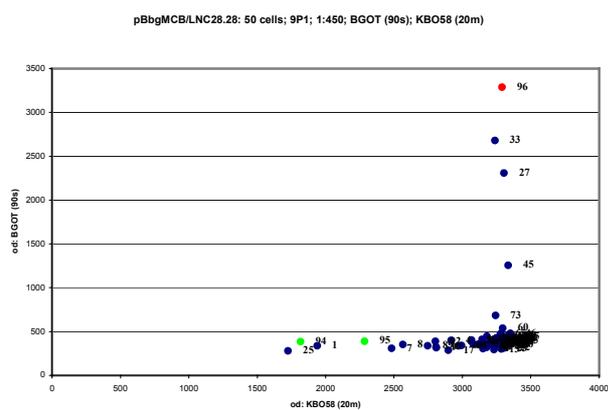


Figure 3-66. pBbgMCB/LNC28.28 (50 cells) - examples of od-data plotted into X-Y coordinate systems. Depicted are examples of the signal intensity data obtained from the Program TINA4.0, after scanning filters which had been probed with BGOT and subsequently with KBO58, plotted into X-Y coordinate systems. A) 1:50 dilution; BGOT (exposure 5 minutes); KBO58 (exposure 20 minutes). B) 1:150 dilution; plate 1; BGOT (exposure 90 seconds); KBO58 (exposure 20 minutes). C) 1:450 dilution; plate 1; BGOT (exposure 90 seconds); KBO58 (exposure 20 minutes).

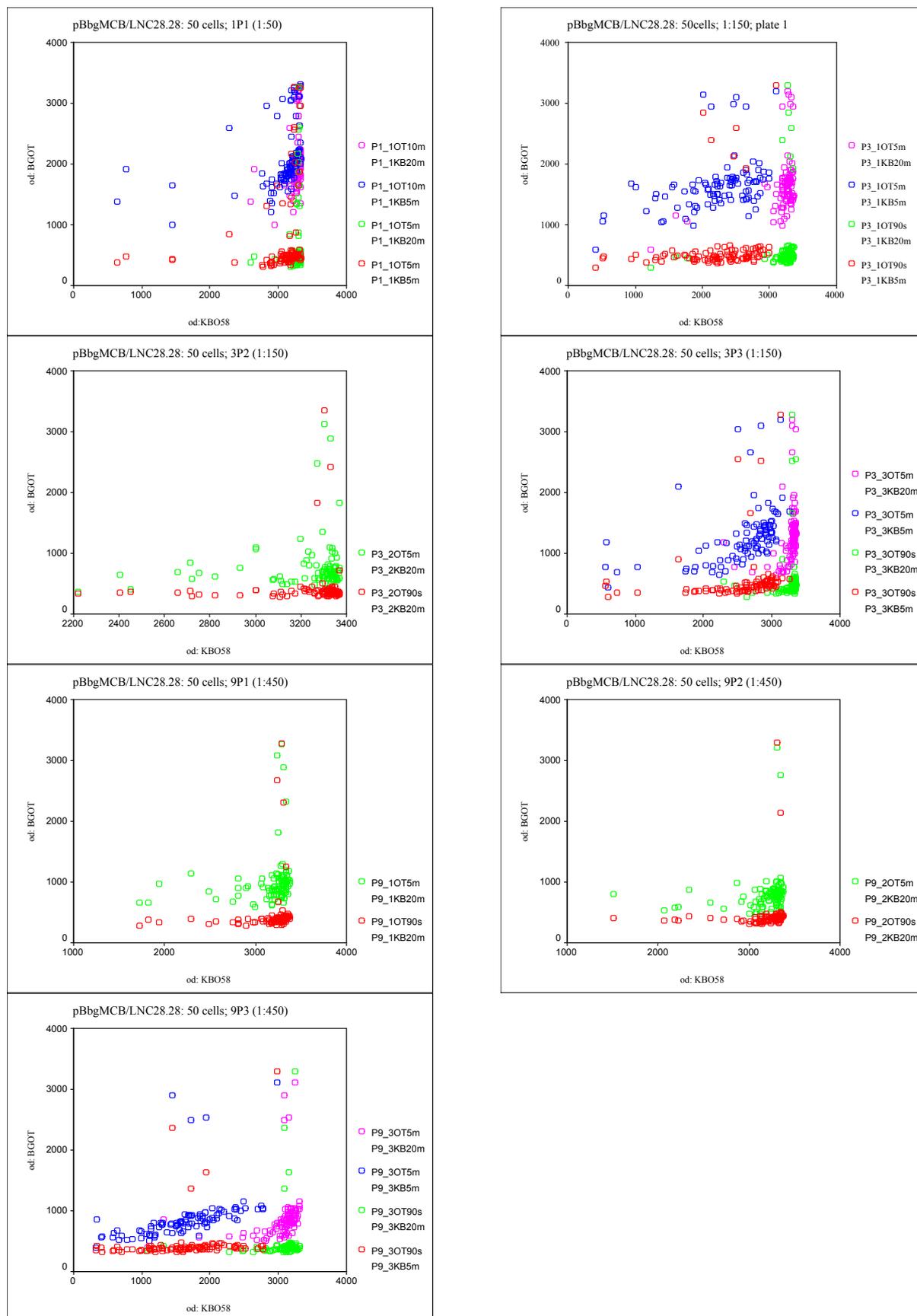


Figure 3-67 Overlaid scatter diagrams of SSOP-filters probed with either BGOT/KBO58, or BGOC/KBO58. Depicted are combinations of od-values for BGOT with KBO58 obtained for different exposure times of the films. The respective exposure times are indicated in the legend (e.g.: P1_1OT10m: plate number 1, probe: BGOT, exposure: 10 minutes; P1_1KB20m: plate number 1, probe: KBO58, exposure: 20 minutes). The perspective dilutions are given in the headings in brackets.

Initially a two dimensional cluster analysis was performed, taking the od-values for BGOT and KBO58 into account. However it became obvious that the results were very similar to identical to a one dimensional cluster analysis, only taking the od-values for BGOT into account. A typical result of a cluster analysis protocol is shown in figure 3-68 for plate 1 of the 1/150 dilution. In table 3-9 and figure 3-69 the number of selected candidates per plate are listed. Since overall the differences between the evaluation by visual inspection (i.e. choosing potential candidates from the X-Y coordinate system plots), and per cluster analysis are not very large, all values were taken into the subsequent evaluation.

exposure (BGOT/KBO58)	plate 1 (1/150)						plate 1 (1/150)						plate 2 (1/150)						plate 3 (1/150)						
	visual inspection				cluster analysis		visual inspection				cluster analysis		visual inspection				cluster analysis		visual inspection				cluster analysis		
	5m/5m	10m/5m	5m/20	10m/20	BGOT (5m)	BGOT (10m)	90s/5m	5m/5	90s/20	5m/20	BGOT (90s)	BGOT (5m)	90s/5	5m/5	90s/20	5m/20	BGOT (90s)	BGOT (5m)	90s/5m	5m/5	90s/20	5m/20	BGOT (90s)	BGOT (5m)	
8	8	8	8	8	13	8	24	24	24	24	24	24	32	32	32	32	32	4	4	4	4	4	4	4	
10	10	10	10			10	43	43	43	43	43	43	43	43	43	43	43	42	42	42	42	42	42	42	
13	13	13	13			13	60	60	60	60	60	60	60	60	60	60	60	71	71	71	71	71	71	71	
19	19	19	19			19	64	64	64	64	64	64	64	64	64	64	64	50	50	50	50	50	50	50	
20	20	20	20			20	76	76	76	76	76	76	76	76	76	76	76								
22	22	22	22			22													79	79	79	79	79	79	79
30	30	30	30			30																			82
				33																					
34	34	34	34			34																			
35	35	35	35			35																			
38	38	38	38			38																			
52	52	52	52			52																			
60	60	60	60			60																			
66	66	66	66			66																			
67	67	67	67			67																			
68	68	68	68			68																			
69	69	69	69			69																			
76	76	76	76			76																			
77	77	77	77			77																			
sum of candidates	16	16	15	18	9	17	5	5	5	5	5	5	n.m.	n.m.	3	3	2	3	5	5	5	4	3	7	

exposure (BGOT/KBO58)	plate 1 (1/450)				plate 2 (1/450)				plate 3 P1(450)							
	visual inspection		cluster analysis		visual inspection		cluster analysis		visual inspection		cluster analysis					
	90s/5m	5m/5	90s/20	5m/20	90s/5m	5m/5	90s/20	5m/20	90s/5	5m/5	90s/20	5m/20				
			27	27			37	37			17	17	17	17	34	17
			33	33			33	33			34	34	34	34	34	34
			45	45			45	45			74	74	74	74	74	74
			73	73			73	73								
sum of candidates	n.m.	n.m.	4	4	2	3	n.m.	n.m.	1	1	1	1	3	3	3	3

Table 3-9. Number and identity of potential mutant candidate pools found through analysis of X-Y coordinate system plotting or cluster analysis. Indicated is the Plate analysed per dilution, as well as the exposure times examined either by visual inspection of the X-Y coordinate systems (visual inspection) or cluster analysis (CA). The numbers indicate the well numbers of the examined 96-well plate (see figure 3-64)

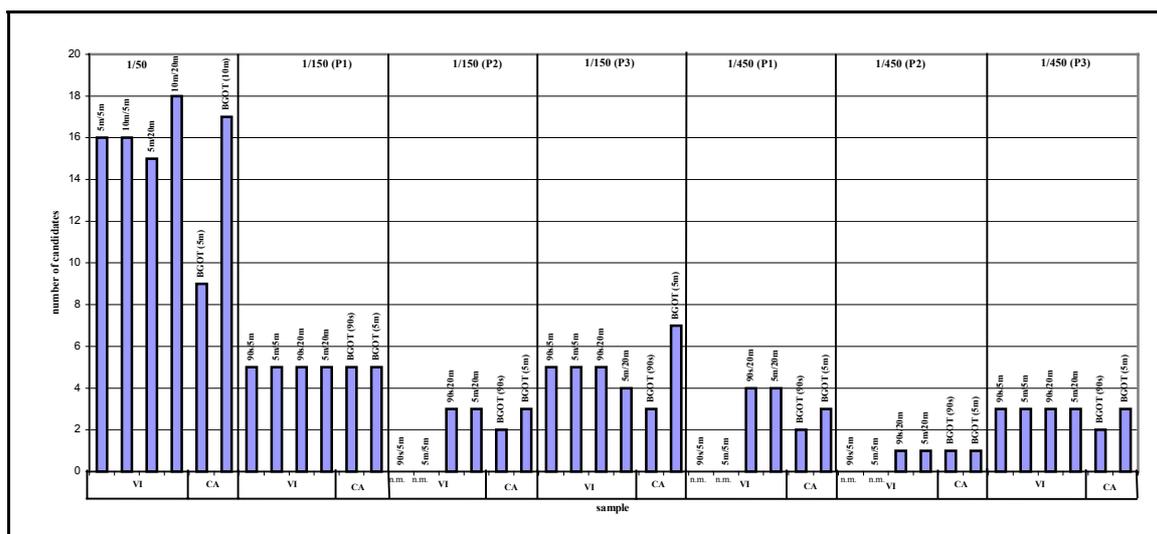


Figure 3-69. Number of potential mutant candidate pools found through visual inspection of X-Y coordinate system plotting or cluster analysis. Depicted are the number of potential mutant candidates which were found either through visual inspection (VI) of plotting of the through densitometry obtained values for the dot blots, or through cluster analysis (CA). 1/x (Px): dilution, (number of plate); x/y: on top of the bars: exposure time BGOT/KBO58 (only BGOT for cluster analysis); VI: analysis through visual inspection of X-Y coordinate system plotting; CA: analysis through cluster analysis (SPSS9.0).

Taking the numbers of potential candidates carrying pBbgMCB as a mutant, a nonlinear regression analysis per plate and evaluation method (visual inspection; cluster analysis) was performed (an example of a resulting file is given for the 100 cell experiment – see below). As the model formula served: $f(x) = C * 93 * (1 - (1 - 1/\text{dilution})^{50})$. The term $(1 - 1/\text{dilution})^{50}$ describes the probability, of having zero mutant candidates in a pool of 50 cells at a given dilution. Therefore $(1 - (1 - 1/\text{dilution})^{50})$ describes the probability of finding at least 1 mutant candidate in a pool of 50 cells. The factor 93 derives from the fact that per plate 93 wells are examined (in each plate three wells (94; 95; 96) served as controls (figure 3-64). The term $93 * (1 - (1 - 1/\text{dilution})^{50})$ builds the *expected value* for the number of mutant cell pools. The constant C was introduced to be able to calculate deviations from the *expected value*. At 100% efficiency (i.e. wildtype (K^{BLNC}) and mutant (pBbgMCB) molecules have an equal chance of being lifted to a detectable level), C is equal to 1. Nonlinear regression analysis based on the observed values leads to an estimate of C, which can subsequently be used to calculate expected numbers of mutant candidates at any dilution (table 3-10 and figure 3-70).

		visual inspection								cluster analysis				
		BGOT(90s) KBO58(5m)		BGOT(5m) KBO58(5m)		BGOT(90s) KBO58(20m)		BGOT(5m) KBO58(20m)		BGOT(90s)		BGOT(5m)		
plate	dilution	observed	predicted: (C=0,248; R ² =0,934)	observed	predicted: (C=0,203; R ² =0,681)	observed	predicted: (C=0,222; R ² =0,874)	observed	predicted: (C=0,248; R ² =0,831)	observed	predicted: (C=0,144; R ² =0,879)	observed	predicted: (C=0,25; R ² =0,884)	mean of predicted (C=0,219)
1	50	16	14,69	16	12,00	15	13,16	18	14,67	9	8,5	17	14,78	12,95
1	150	5	6,57	5	5,36	5	5,88	5	6,56	5	3,8	5	6,61	5,79
2	150	n.m.	6,57	n.m.	5,36	3	5,88	3	6,56	2	3,8	3	6,61	5,79
3	150	5	6,57	5	5,36	5	5,88	4	6,56	3	3,8	7	6,61	5,79
1	450	n.m.	2,43	n.m.	1,99	4	2,18	4	2,43	2	1,41	3	2,45	2,14
2	450	n.m.	2,43	n.m.	1,99	1	2,18	1	2,43	1	1,41	1	2,45	2,14
3	450	3	2,43	3	1,99	3	2,18	3	2,43	2	1,41	3	2,45	2,14
	10000	n.m.	0,12	n.m.	0,09	n.m.	0,1	n.m.	0,12	n.m.	0,07	n.m.	0,12	0,10

Table 3-10. Nonlinear regression : pBbgMCB in LNC28.28 (50 cells). Listed are the number of potential mutant candidates observed, and the number of potential mutant candidates predicted after nonlinear regression analysis. Used were the values obtained for the oligonucleotides indicated at distinct exposure times (in brackets). Visual inspection: candidates chosen from X-Y coordinate system plots (see figure 3-66). Cluster analysis: candidates chosen by cluster analysis (SPSS9.0). Plate: plate number of respective dilution step. N.m.: not measured. Mean of predicted: constant C derived from the mean of all C's obtained by nonlinear regression. Nonlinear regression analysis (SPSS9.0): Used was the equation $f(x) = C * 93 * (1 - (1 - 1/\text{dilution})^{50})$. For the nonlinear regression analysis the startig value C for the iterations during the calculation process was 1 (Sokal and Rohlf 1995).

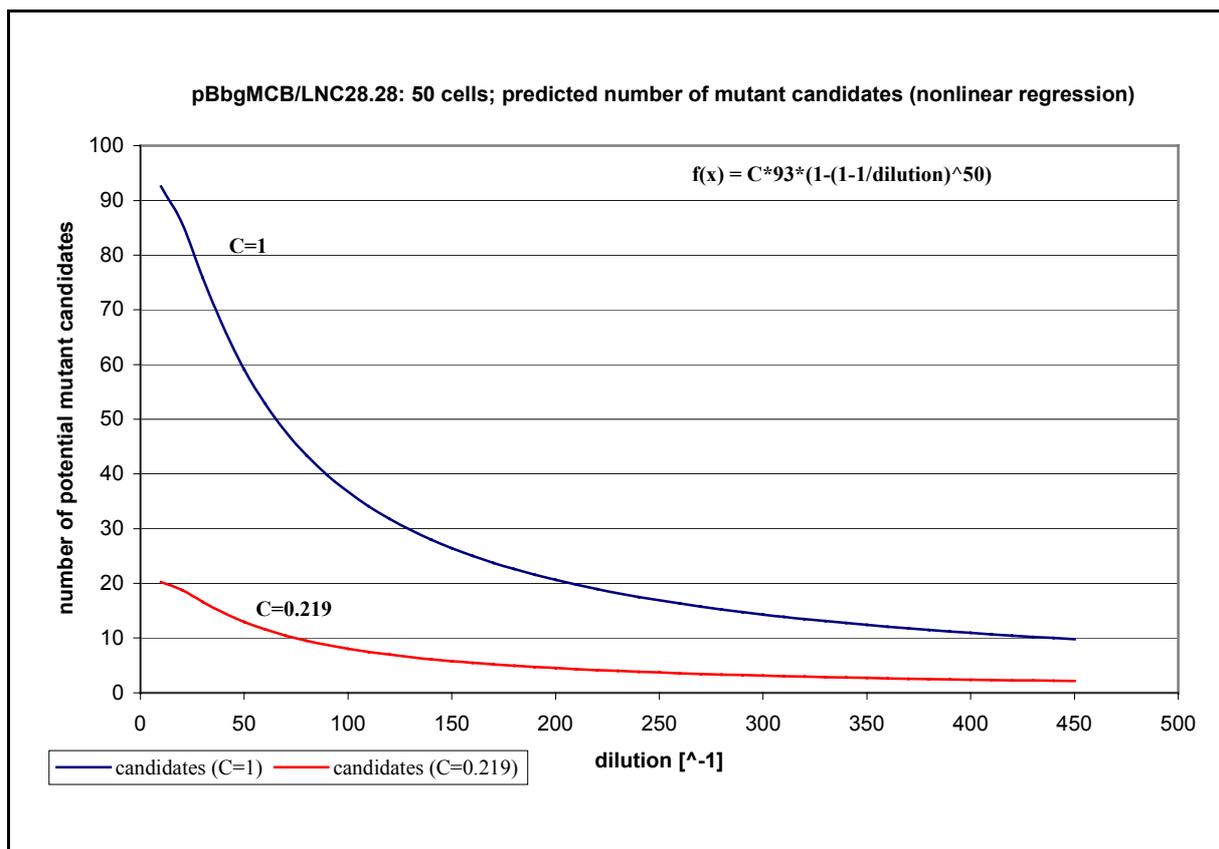


Figure 3-70. Nonlinear regression : pBbgMCB in LNC28.28 (50 cells). Depicted are the curves which result from the equation $f(x) = C \cdot 93 \cdot (1 - (1 - 1/\text{dilution})^{50})$, for $C=1$ and $C=0.219$.

The value of C was calculated to be 0.219. This means that the detection of mutant candidates is 5 times less efficient than predicted for $C=1$. In other words: For $C=1$ in average the analysis of about two 96-well plates ($x = 93 \cdot (1 - (1 - 1/10\,000)^{50}) = 0,46$) would lead to the detection of one mutant candidate pool of 50 cells at an overall mutant frequency of 10^{-4} . For $C=0.219$ in average about 10 plates would be needed ($x = 0.219 \cdot 93 \cdot (1 - (1 - 1/10\,000)^{50}) = 0,10$).

Detection of pBbgMCB diluted into pools of 100 cells

In this experiment pools of 100 LNC28.28 cells were analysed, with the LNC-PCR/SSOP approach. The plasmid pBbgMCB was diluted at steps of 1/100 (one 96-well plate), 1/200 (three 96-well plates), 1/400 (four 96-well plates), and 1/800 (four 96-well plates) into LNC28.28 cells. Examples of resulting filters of dot blots, probed with BGOT, which derived from this experiment are depicted in figure 3-71. The resulting densitometry data was transferred into X-Y coordinate systems (for examples see figure 3-72). A summary of the identity and number of potential mutant candidates obtained per plate and method (visual evaluation by plotting the data into X-Y coordinate systems and cluster analysis) is given in table 311 and figure 3-73.

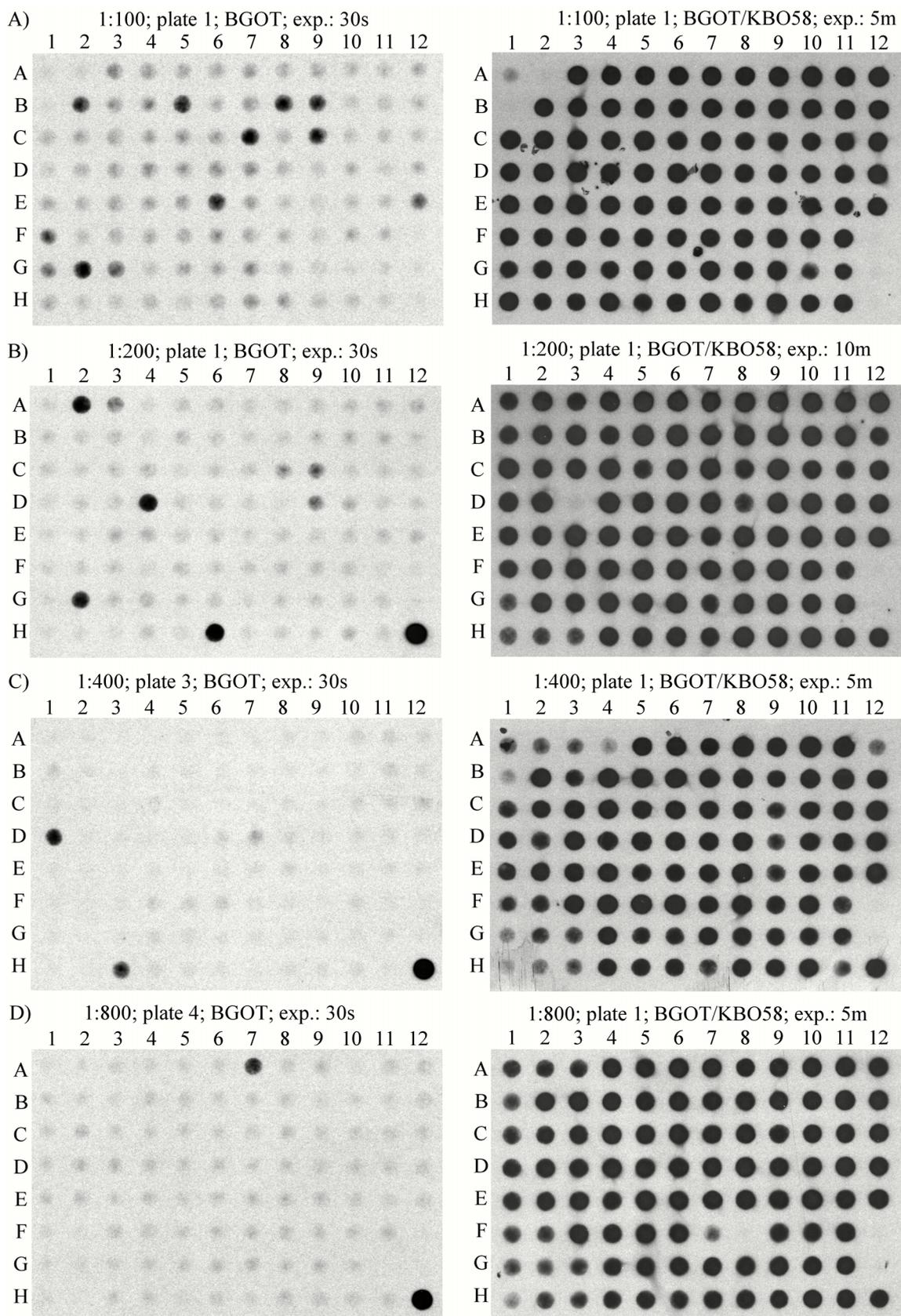
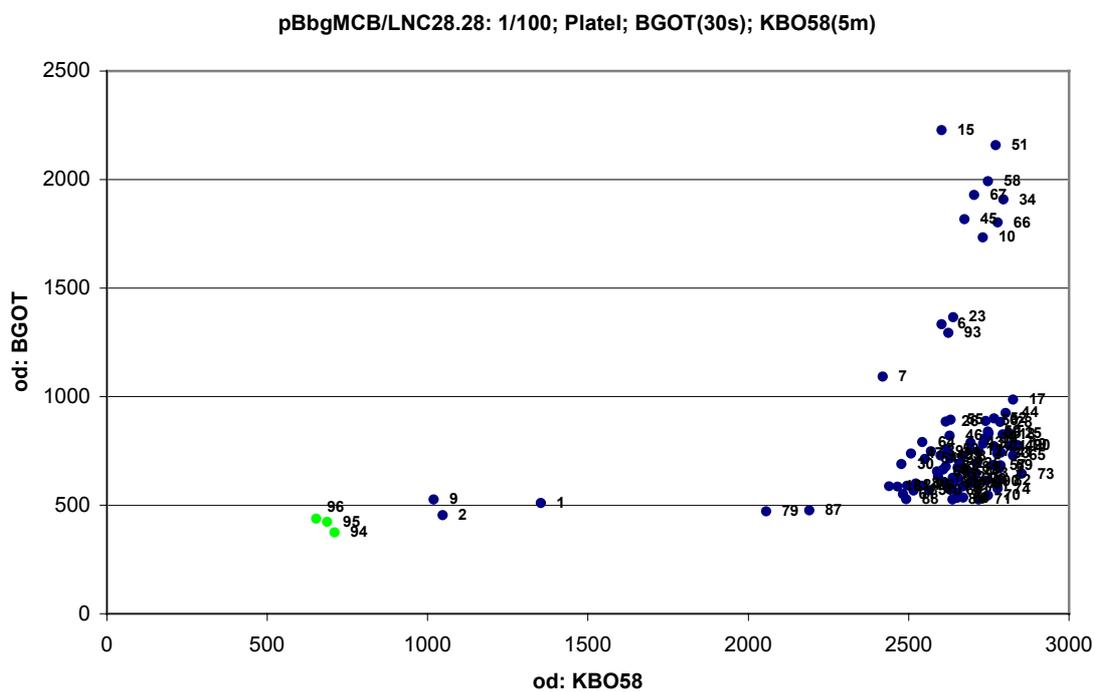
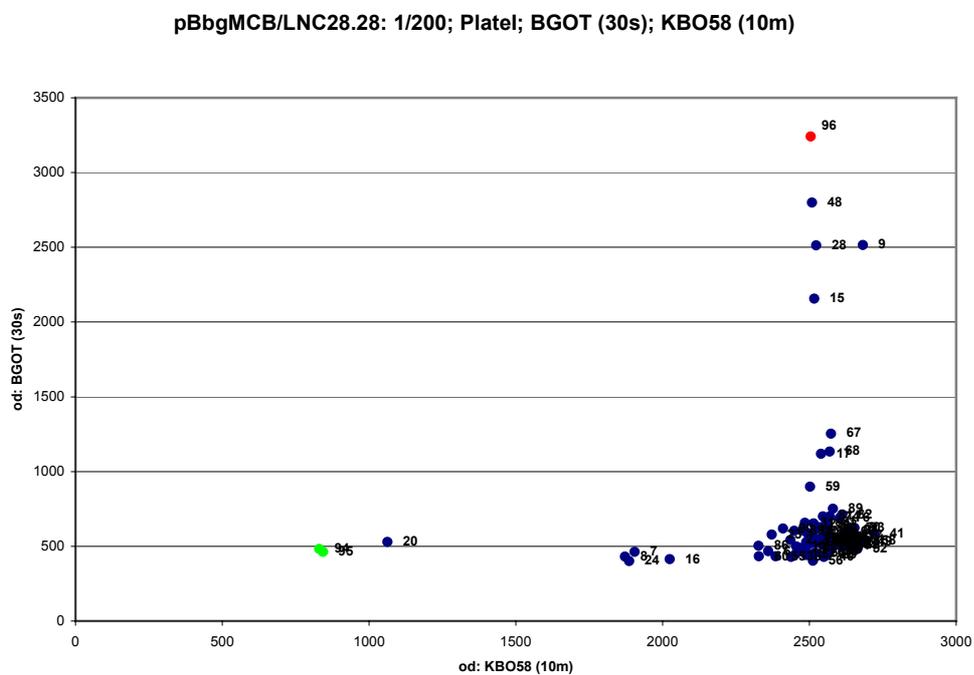


Figure 3-71. pBbgMCB/LNC28.28 (100 cells) - examples of SSOP dot blots. Depicted are examples of filters derived from the 100 cell experiment which were probed with BGOT (left panel) and subsequently to stripping with the oligonucleotide KBO58. Each well represents 20 μ l of a LNC-PCR, being transferred onto the filter. **A)** 1:100 dilution; BGOT (exposure 30 seconds); KBO58 (exposure 5 minutes). **B)** 1:200 dilution; plate 1; BGOT (exposure 30 seconds); KBO58 (exposure 10 minutes). **C)** 1:400 dilution; plate 1; BGOT (exposure 30 seconds); KBO58 (exposure 5 minutes). **D)** 1:800 dilution; plate 4; BGOT (exposure 30 seconds); KBO58 (exposure 5 minutes).

A)

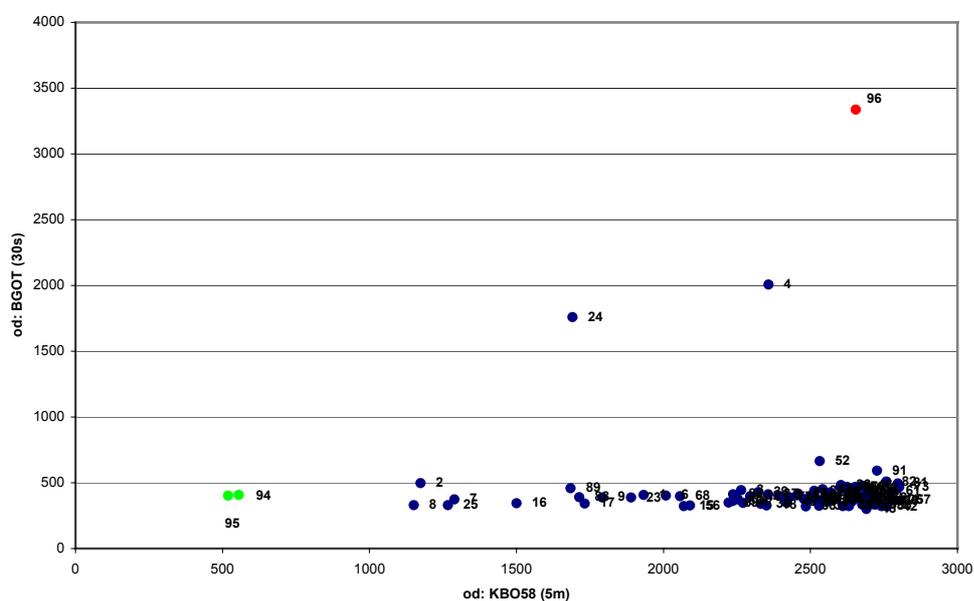


B)



C)

pBbgMCB/LNC28.28: 1/400; Plate 3; BGOT (30s); KBO58 (5m)



D)

pBbgMCB/LNC28.28: 1/800; Plate 4; BGOT (30s); KBO58 (5m)

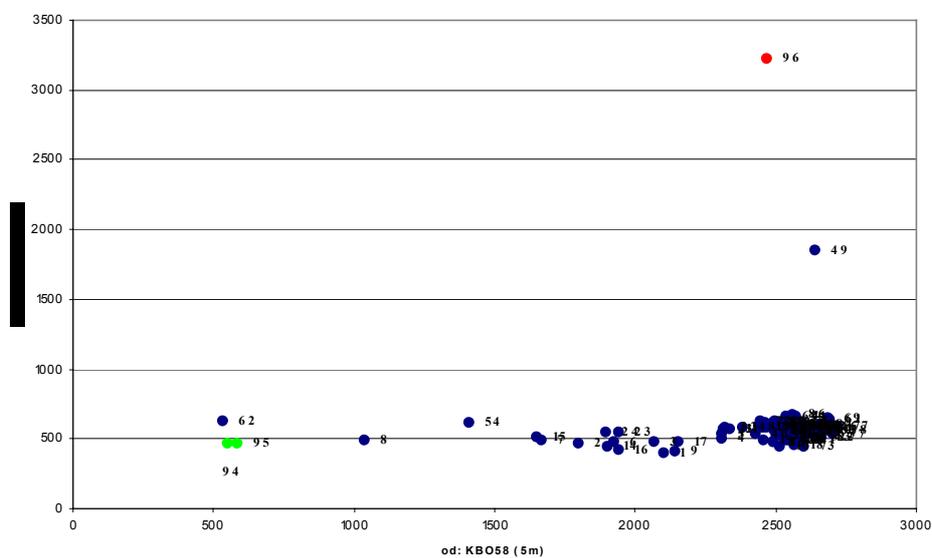


Figure 3-72. pBbgMCB/LNC28.28 (100 cells) - examples of od-data plotted into X-Y coordinate systems. Depicted are examples of the signal intensity data obtained from the Program TINA4.0, after scanning filters which had been probed with BGOT and subsequently with KBO58, plotted into X-Y coordinate systems. A) 1:100 dilution; BGOT (exposure 30 seconds); KBO58 (exposure 5 minutes), this plate did not contain pBbgMCB in well 96. . B) 1:200 dilution; plate 1; BGOT (exposure 30 seconds); KBO58 (exposure 10 minutes). C) 1:400 dilution; plate 3; BGOT (exposure 30 seconds); KBO58 (exposure 5 minutes). D) 1:800 dilution; plate 4; BGOT (exposure 30 seconds); KBO58 (exposure 5 minutes).

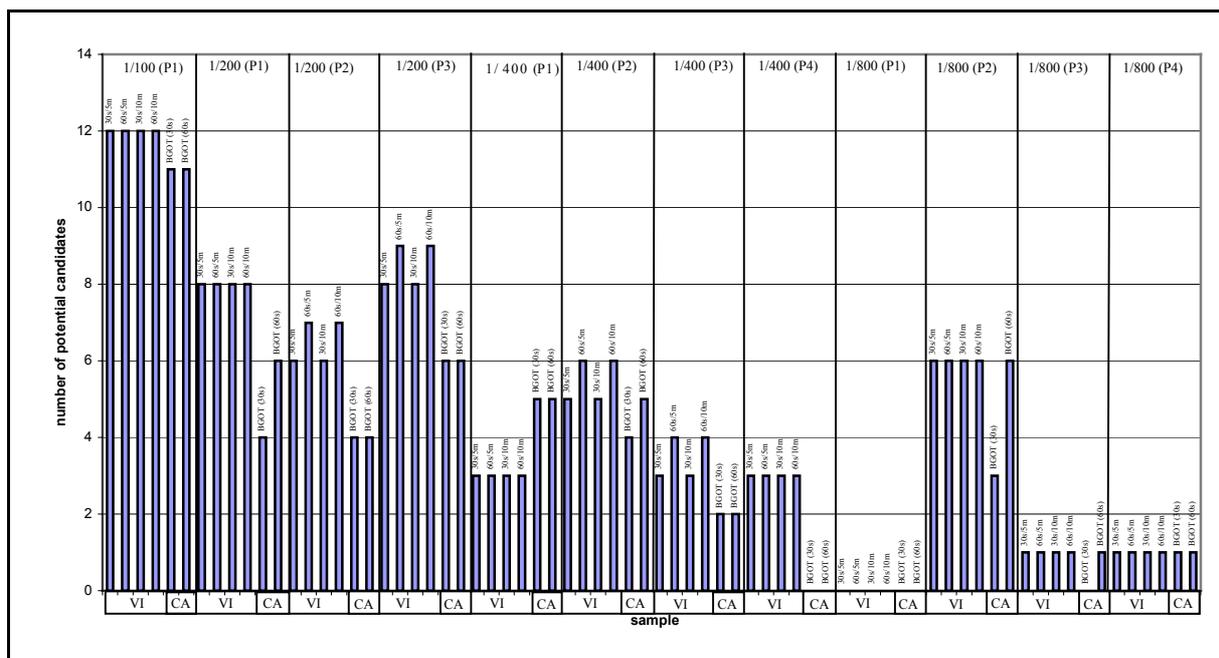


Figure3-73. Number of mutant candidate pools per plate and evaluation method. Depicted is the total number of potential mutant candidate pools per dilution step and plate. The analysis was either performed by visual inspection of the od-values for each pool, being plotted into an X-Y coordinate system, or by cluster analysis. For the visual inspection of the X-Y coordinate systems the probing with the two oligonucleotides BGOT (detection of mutant) and KBO58 (detection of wildtype) were taken into account. For BGOT the od-values obtained from the exposure times 30 seconds (30s) and 60 seconds(60s) were used, for KBO58 5 minutes (5m) and 10 minutes (10m). For the cluster analysis only the od-values obtained with the oligonucleotide BGOT at 30 seconds and 60 seconds were used.

As in the 50 cell experiment above the numbers of potential mutant candidate pools was taken to perform a nonlinear regression analysis with SPSS9.0. An example for a resulting viewer file is given in figure 3-74. This time the equation employed is as follows:
 $f(x) = C * 93 * (1 - (1 - 1/\text{dilution})^{100})$.

```
All the derivatives will be calculated numerically.
Iteration  Residual SS          C
1           6301,689283    1,00000000
1.1         29,15530380    ,195166874
2           29,15530380    ,195166874
2.1         29,15530380    ,195166873
Run stopped after 4 model evaluations and 2 derivative evaluations.
Iterations have been stopped because the relative reduction between
successive residual sums of squares is at most SSCON = 1,000E-08
Nonlinear Regression Summary Statistics      Dependent Variable ET3K10
Source           DF  Sum of Squares  Mean Square
Regression        1    368,84470      368,84470
Residual          11    29,15530      2,65048
Uncorrected Total 12    398,00000
(Corrected Total) 11    136,66667
R squared = 1 - Residual SS / Corrected SS = ,78667
Asymptotic 95 %
Confidence Interval
Parameter  Estimate  Std. Error  Lower  Upper
C           ,195166873  ,016544238  ,158753250  ,231580496
```

Figure 3.74. Result (viewer file) of nonlinear regression analysis performed by SPSS9.0. Depicted is an example of a nonlinear regression analysis performed with the values for numbers of potential mutant candidate pools obtained by visual inspection at 30 seconds exposure for BGOT and 10 minutes exposure for KBO58. Used was the equation $f(x) = C * 93 * (1 - (1 - 1/\text{dilution})^{100})$. The starting value for C was 1. The resulting value for C is 0.195166873; R^2 is 0.78667 (compare to table 3-12; column [observed; VI BGOT(30s) KBOT(10m)]).

Plate	dilution	observed VI BGOT(30s) KBOT(5m)	predicted regression (C=0.195; R ² =0.787)	observed VI BGOT(60s) KBOT(5m)	predicted regression (C=0.207; R ² =0.787)	observed VI BGOT(30s) KBOT(10m)	predicted regression (C=0.195; R ² =0.787)	observed VI BGOT(60s) KBOT(10m)	predicted regression (C=0.207; R ² =0.787)	observed CA BGOT (30s)	predicted regression (C=0.187; R ² =0.191)	observed CA BGOT (60s)	predicted regression (C=0.205; R ² =0.214)
1P1	1/100	12	11.51	12	12.2	12	11.51	12	12.2	11	11.02	11	12.11
2P1	1/200	8	7.16	8	7.59	8	7.16	8	7.59	4	6.85	7	7.53
2P2	1/200	6	7.16	7	7.59	6	7.16	7	7.59	5	6.85	4	7.53
2P3	1/200	8	7.16	9	7.59	8	7.16	9	7.59	6	6.85	8	7.53
4P1	1/400	3	4.02	3	4.26	3	4.02	3	4.26	17	3.85	17	4.23
4P2	1/400	5	4.02	6	4.26	5	4.02	6	4.26	4	3.85	5	4.23
4P3	1/400	3	4.02	4	4.26	3	4.02	4	4.26	2	3.85	2	4.23
4P4	1/400	3	4.02	3	4.26	3	4.02	3	4.26	3	3.85	3	4.23
8P1	1/800	0	2.13	0	2.26	0	2.13	0	2.26	0	2.04	0	2.25
8P2	1/800	6	2.13	6	2.26	6	2.13	6	2.26	5	2.04	6	2.25
8P3	1/800	1	2.13	1	2.26	1	2.13	1	2.26	1	2.04	1	2.25
8P4	1/800	1	2.13	1	2.26	1	2.13	1	2.26	1	2.04	1	2.25
	1/10000	n.m.	0.18	n.m.	0.19	n.m.	0.18	n.m.	0.19	n.m.	0.17	n.m.	0.19

Table 3-12. Nonlinear regression : pBbgMCB in LNC28.28 (100 cells). Listed are the number of potential mutant candidates observed , and the number of potential mutant candidates predicted after nonlinear regression analysis. Used were the values obtained for the oligonucleotides indicated at distinct exposure times (in brackets). Visual inspection: candidates chosen from X-Y coordinate system plots (see figure 3-72). Cluster analysis (CA): candidates chosen by cluster analysis (SPSS9.0). Plate: plate number of respective dilution step (e.g., 1P1:1/100 plate number 1). N.m.: not measured. Used was the equation $f(x)=C*93*(1-(1-1/dilution)^{100})$. The starting value C for the iterations during the calculation process was 1. n.m.: not measured.

sample	VI OT(30s)KB(5m)	VI OT(60s)KB(5m)	VI OT(30s)KB(10m)	VI OT(60s)KB(10m)	CA BGOT (30s)	CA BGOT (60s)	mean of C
C	0.195	0.207	0.195	0.207	0.187	0.205	0.201

Table 3-13. Mean value of calculated C per sample. Listed are the individual constants C of table 3-12, obtained through nonlinear regression analysis, and their arithmetic mean.

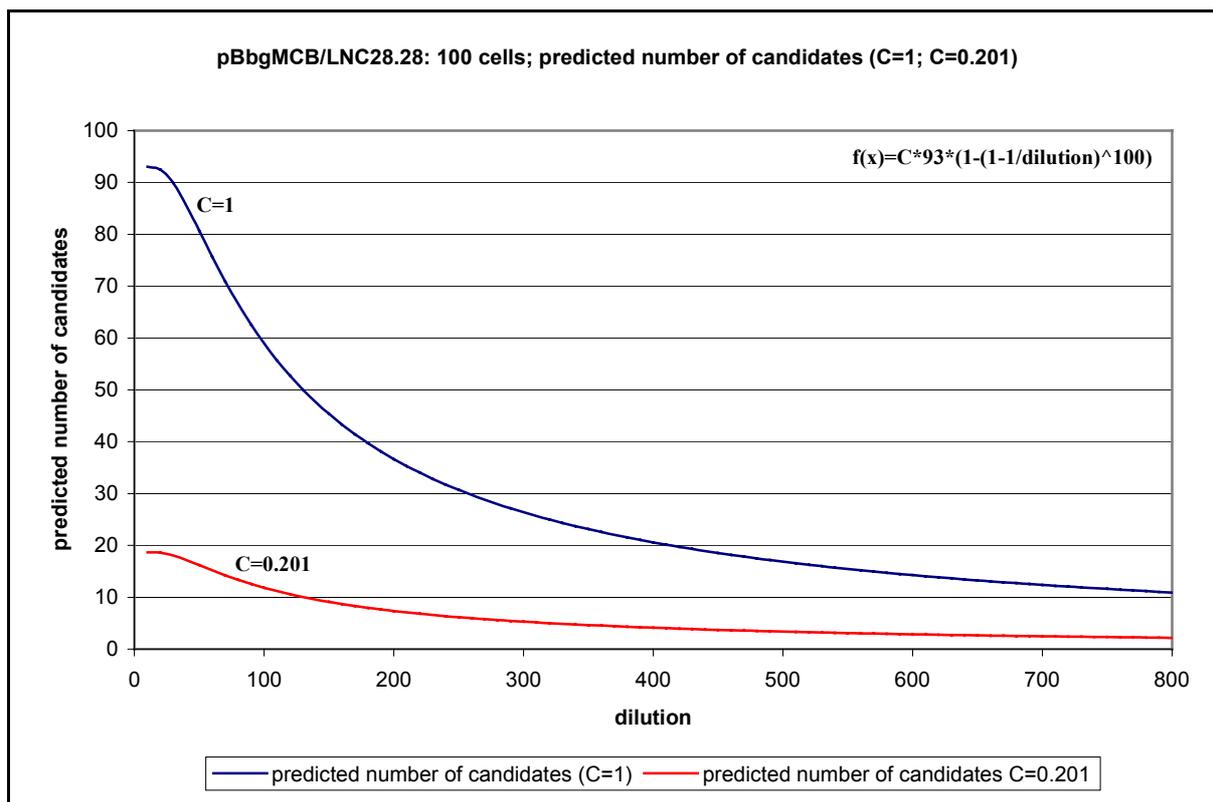


Figure 3-75. Nonlinear regression : pBbgMCB in LNC28.28 (100 cells). Depicted are the curves which result from the equation $f(x)=C*93*(1-(1-1/dilution)^{100})$, for C=1 and C=0.201.

The value of C for this experiment, using 100 cells, was calculated to be 0.201 (see table 3-12 and for mean of C table 3-13). The resulting curves for the predicted number of mutant candidate pools in dependence of the dilution (from 1:10 down to 1:800) for C=1 and C=0.201 are depicted in figure 3-75. The result means that the detection of mutant candidates

is 5 times less efficient than predicted for $C=1$. In other words: At a mutant frequency of 10^{-4} , for $C=1$ in average the analysis of about one 96-well plate with pools of 100 cells would lead to the detection of one mutant candidate pool ($x=93*(1-(1-1/10\ 000)^{100})=0,93$). For $C=0.201$ in average about 5 plates would be needed ($x=0.201*93*(1-(1-1/10\ 000)^{100})=0,19$).

The $k \geq 2$ problem

As mentioned in the introduction to this chapter, we assume that the distribution of mutants *in vivo*, as well as in the test experiments presented here, follow a binomial distribution, or to a good approximation the Poisson distribution (Sokal *et al.* 1998). This would mean that if we look at a 1/100 dilution of mutant in wildtype sequence, for a 96-well plate onto which pools of 100 cells have been aliquoted, 37% of all pools (36 wells) should not contain a mutation and appear in the SSOP-assay negative, while 63% (60 wells) should AT LEAST contain one mutation and therefore appear positive. The data presented here show the actual detection efficiency lies a factor 5 below expectation both in the 50 cell and 100 cell experiment. *A priori* it is not possible to know whether the detected positive pools contain one or more mutant molecules. Following the binomial distribution, of the 63% of all positive cell pools; about half contain EXACTLY one mutant molecule, and half two or more mutant molecules. Therefore the observed numbers for positive pools could as such be explained by the presence of two or more molecules, i.e. detection is only possible if at the starting point of the assay, before the LNC-PCR, two or more mutant molecules are available. For the result of the evaluation of the LNC-PCR/SSOP assay presented above this is of no significance. Independently of the actual number of mutant molecules present in the cell pools the result is that the efficiency of mutation detection lies 5 fold underneath expectation. Possible reasons will be discussed (see Discussion). However in the context of the low expected mutation frequency of *in vivo* mutations, which lies in the range of 10^{-4} , it has to be certain that a starting copy number of one mutant molecule is sufficient to reliably detect the mutant with the mutation detection assay. If one assumes strict stochastic behavior of the mutants, the occurrence of two or more mutants in the same cell pool is basically *nil* (following the binomial distribution the probability of finding exactly one mutant at a dilution of 10^{-4} is 1%, and of finding two or more mutants is about 0.005%). The data on PCR and SSOP presented above and in section - An estimate of the LNC-PCR efficiency - does strongly indicate that one mutant molecule can be detected in the KbLNC/SSOP assay. However we were still interested in assessing which of the models a) detection of one or more molecules, or b) detection of two or more molecules does better fit to the data collected. A nonlinear

regression analysis was performed on the number of positive pools detected in the 100 cell experiment, upon probing with BGOT, whereby the equations used either described the presence of at least one mutant molecule, or the presence of two or more mutant molecules. The term "R squared", which describes the "goodness of fit" (Sokal *et al.* 1998) gives an indication, which of the two models fit better to the data observed. For the calculations two exposure time combinations for BGOT and KBO58 were taken into account: (1) 30 seconds for BGOT, 10 minutes for KBO58 and (2) 60 seconds for BGOT, 10 minutes for KBO58. The results are presented in table 3-14 to figure 3-76. The values for R^2 indicate that the goodness of fit after the nonlinear regression is better for $k \geq 1$ ($R^2 = 0.79$ for (1) and (2)) than for $k \geq 2$ ($R^2 = 0.43$ (1) and 0.32 (2)). This is also reflected by the degree of correspondence of the respective curves deduced from the observed values (visual inspection; BGOT (60s), KBO58 (10m)) and the predicted values based on $C = 0.207$ ($k \geq 1$; $R^2 = 0.787$) and $C = 0.634$ ($k \geq 2$; $R^2 = 0.324$). This suggests that the model $k \geq 1$ is more likely to be true than the model $k \geq 2$.

Plate	dilution	VI BGOT(30s) KBOT(10m)			VI BGOT(60s) KBOT(10m)		
		observed	predicted		observed	predicted	
			$k \geq 1$ regression (C=0.195; $R^2=0.787$)	$k \geq 2$ regression (C=0.618; $R^2=0.430$)		$k \geq 1$ regression (C=0.207; $R^2=0.787$)	$k \geq 2$ regression (C=0.634; $R^2=0.324$)
1P1	1/100	12	11,51	15,18	12	12,2	15,81
2P1	1/200	8	7,16	5,16	8	7,59	5,38
2P2	1/200	6	7,16	5,16	7	7,59	5,38
2P3	1/200	8	7,16	5,16	9	7,59	5,38
mean 2P(1-3)	1/200	7,3	7,16	5,16	8	7,59	5,38
4P1	1/400	3	4,02	1,51	3	4,26	1,57
4P2	1/400	5	4,02	1,51	6	4,26	1,57
4P3	1/400	3	4,02	1,51	4	4,26	1,57
4P4	1/400	3	4,02	1,51	3	4,26	1,57
mean 4P(1-4)	1/400	3,5	4,02	1,51	4	4,26	1,57
8P1	1/800	0	2,13	0,41	0	2,26	0,43
8P2	1/800	6	2,13	0,41	6	2,26	0,43
8P3	1/800	1	2,13	0,41	1	2,26	0,43
8P4	1/800	1	2,13	0,41	1	2,26	0,43
mean 8P(1-4)	1/800	2	2,13	0,41	2	2,26	0,43
	1/10000	n.m.	0,18	0,00	n.m.	0,19	0,00

Table 3-14. Observed and predicted values (nonlinear regression) for potential mutant candidate pools: 100 cells. Listed are the observed numbers of potential mutant pools (visual inspection; see also table 3-11) at an exposure time of 30 seconds for BGOT and 10 minutes for KBO58 (VI BGOT(30s) KBOT(10m)) as well as 60 seconds for BGOT and 10 minutes for KBO58 (VI BGOT(60s) KBOT(10m)). Included are for each exposure time combination the predicted number of mutant candidate pools for $k \geq 1$ and $k \geq 2$. The nonlinear regression analysis was performed as in table 3-12. For $k \geq 1$ the equation $f(x) = C * 93 * (1 - (1 - 1/\text{dilution})^{100})$ was employed. For $k \geq 2$ the equation $f(x) = C * 93 * (1 - ((1 - 1/\text{dil.}]^{100} + ([100/\text{dil.}] * (1 - 1/\text{dil.}]^{99})))$ was used. Starting value for C in the nonlinear regression process (SPSS9.0) was in both cases 1. the calculated C was used to calculate the predicted values per dilution step.

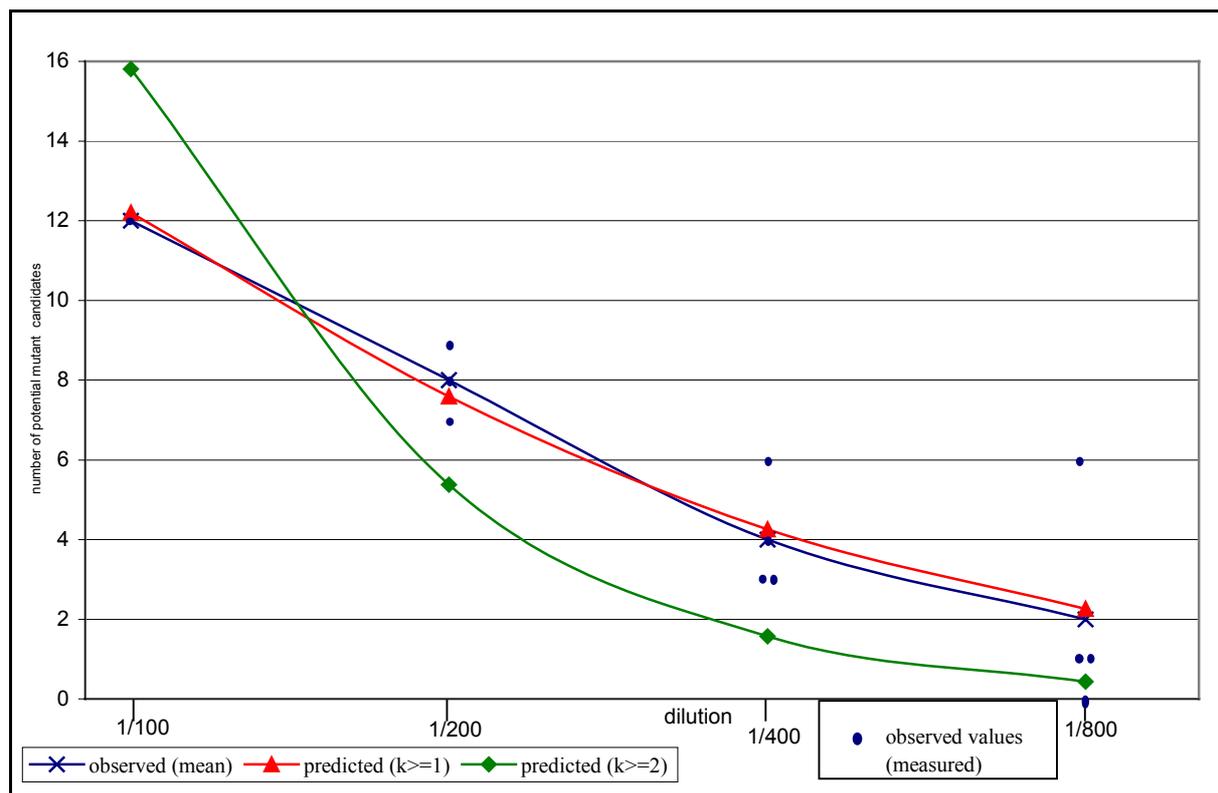


Figure 3-76. Observed and predicted values (nonlinear regression) for potential mutant candidate pools: 100 cells. (Semi logarithmic scale.) The graph depicts the mean (see table 3-14) along with the actual number of observed potential mutant pools as well as the predicted number for $k \geq 1$ and $k \geq 2$ per dilution step. The values for C were taken from table 3-14. X-axis (dilution): $-\log_2(\text{dilution})$. The values were taken for VI BGOT(60s) KBOT(10m) (see table 3-14).

Comparison of the *bg*-mutation detection oligonucleotides BGOT and BGOC

The above evaluations have been performed with the data obtained for the oligonucleotide BGOT. As to access whether the oligonucleotide BGOC is equally efficient in the LNC-PCR/SSOP assay, the number and identity of potential mutant pools was compared for the two oligonucleotides.

The overlaid scatter diagrams (see figure 3-77); here for the exposure times 90 seconds for BGOT and BGOC, and 20 minutes for KBO58), show a very good fit of distribution of signals for the two oligonucleotides. In table 3-15 the results for 50 cells are summarized in further detail. In principle the number and identity of oligonucleotides is the about the same. Especially for the cluster analysis, longer exposure times (here 5 minutes as to compared to 90 seconds) seem to result in higher numbers of identified potential mutant pools, and more consistency as well between the two oligonucleotides as to the results obtained by visual inspection of the X-Y coordinate systems plots. This is explained by better resolution of potential positive pools at longer exposure times as already seen in figure 3-67. However generally the fit of data between BGOT and BGOT is sufficiently good to confirm the results obtained with BGOT before.

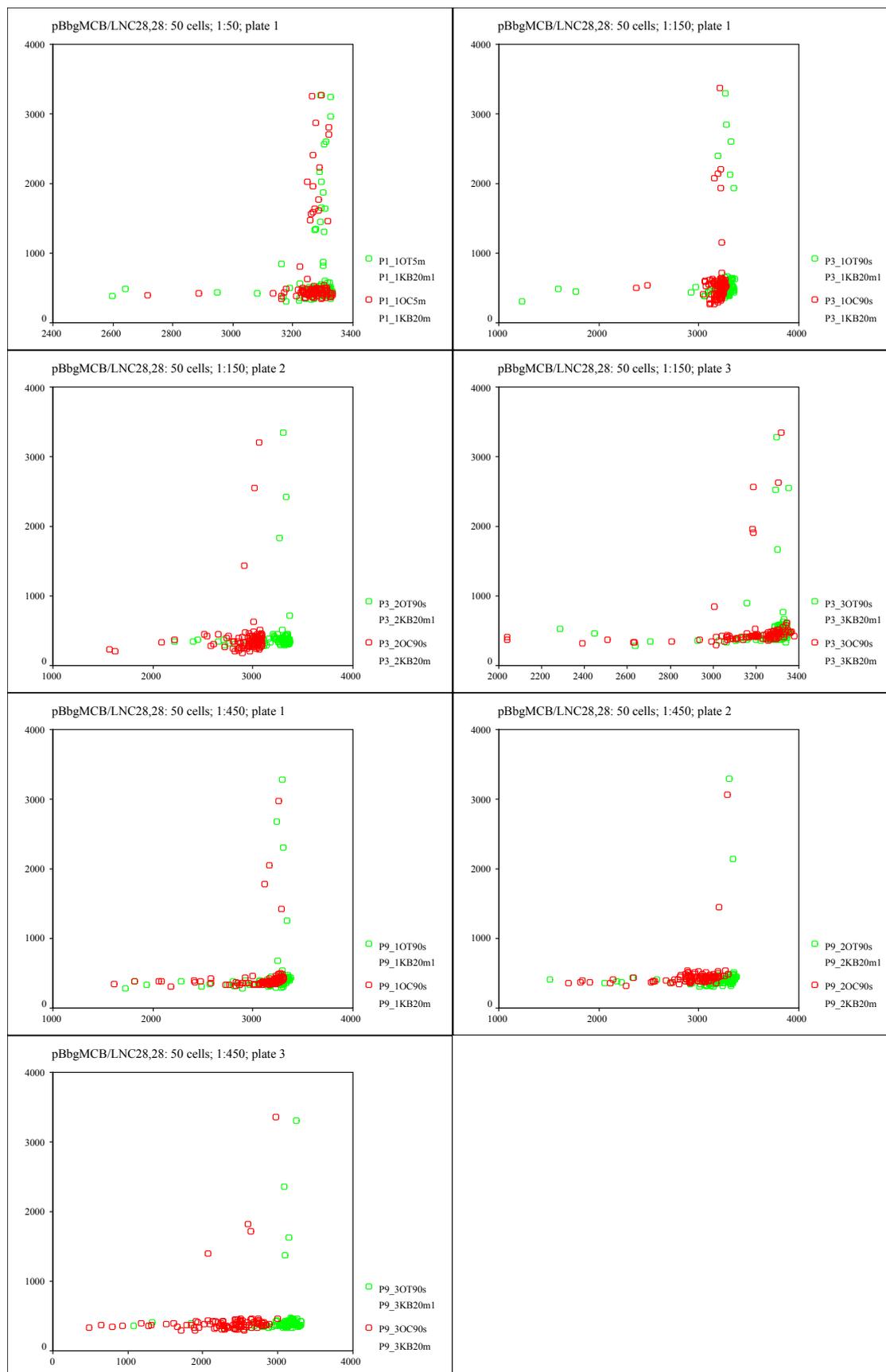


Figure 3-77. pBbgMCB/LNC28.28 (50 cells): overlaid scatter diagrams. Depicted are the od-values obtained per dilution and plate for the blots being probed with BGOT (90 seconds exposure) and KBO58 (20 minutes exposure), green dots, and with BGOC (90 seconds exposure) and KBO58 (20 minutes exposure), red dots).

For the 100 cell dilution series the obtained od-values for BGOT and BGOC are depicted in figure 3-78. The resolution of the BGOC signals is not as good as that of the BGOT signals. This is reflected in the summary in table 3-16. The results obtained by visual inspection of the od-values plotted into X-Y coordinate systems differ more than for the 50 cell dilution series. For BGOC the cluster analysis is in some cases not able to identify unambiguously candidate pools (1:200, plate 3; 1:400, plate 2; 1:800 plate 3). A good example illustrating the loss of resolution with BGOC in figure 3-78 is plate 3 of the 1:200 dilution. For both oligonucleotides BGOT and BGOC the exposure time was 60 seconds, and the range of od-values is the same. Yet the bulk population of the samples derived from the filter probed with BGOC shows a scattered distribution, while the bulk population of the BGOT-samples is quite distinct, making it easy to identify outliers. As can also be seen from the original blots figure 3-79 for BGOC the general background is very high. This leads to the identification of potential mutant candidate pools which are clearly negative for the filters probed with BGOT (1:200, plate3, sample 72 and 84; table 3-16 and figure 3-79). The reason for the unspecific behaviour of BGOC is unclear. As has been shown for the 50 cell dilution experiment, as well as in model experiments before (see e.g. figure 3-49) BGOC behaves normally as specific as BGOT. Possibly during the incubation of the filters with BGOC, or during the high stringency washing step, the temperature control was not exact. However this is unlikely because the temperature of the water baths was controlled by independent measurements regularly and proved to be very exact in every case. As for every 96-well plate the PCR samples were transferred onto the filters, which later should be probed with either BGOT or BGOC, in parallel, and as the lack of resolution effect for BGOC is visible for the whole dilution series, the filters as such cannot be the source of this effect. It is a possibility that the flourolabelled oligonucleotide BGOC itself leads to this effect. For every dilution series being examined, a new batch of flourolabelled oligonucleotide was used. It is not clear what the origin of such unspecific behaviour of a batch of a flourolabelled oligonucleotide could be (technical service – Amersham) more so even because purposely the labeling conditions were held constant. Since a hybridization solution, containing a flourolabelled oligonucleotide can be used several times, in future it is advisable to first test a reference filter with a newly probe as to ensure a good resolution. This would also be useful with respect to calibrating exposure times as discussed in 3.2.3. - SSOP: signal intensities in dependence of exposure time in the CDP-Star™ detection system.

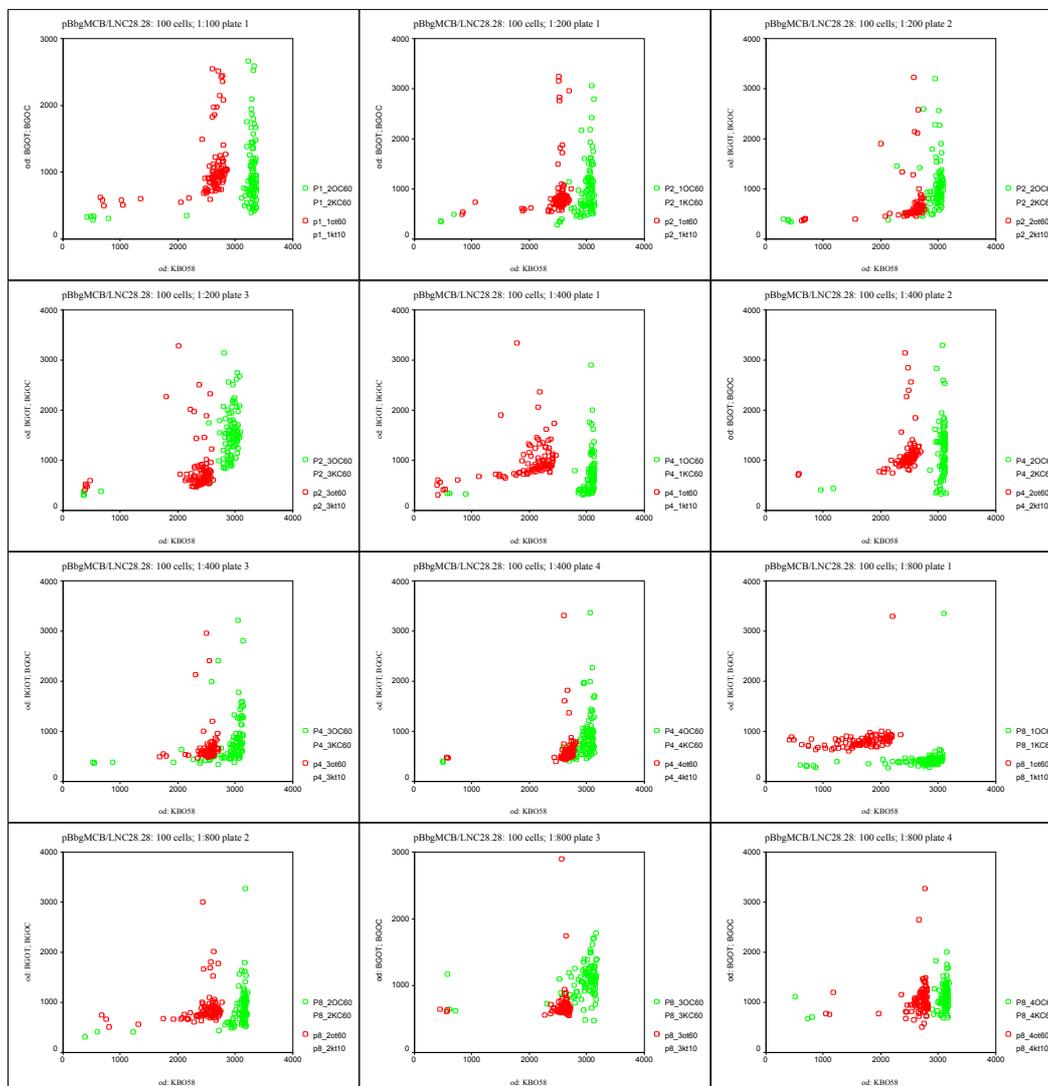


Figure 3-78. pBhgMCB/LNC28.28 (50 cells): overlaid scatter diagrams. Depicted are overlaid scatter plots of the od-values obtained for the filters probed with either BGOT (exposure 60 seconds) and KBO58 (exposure 10 minutes), red dots, or BGOC (exposure 60 seconds) and KBO58 (exposure 10 minutes), green dots.

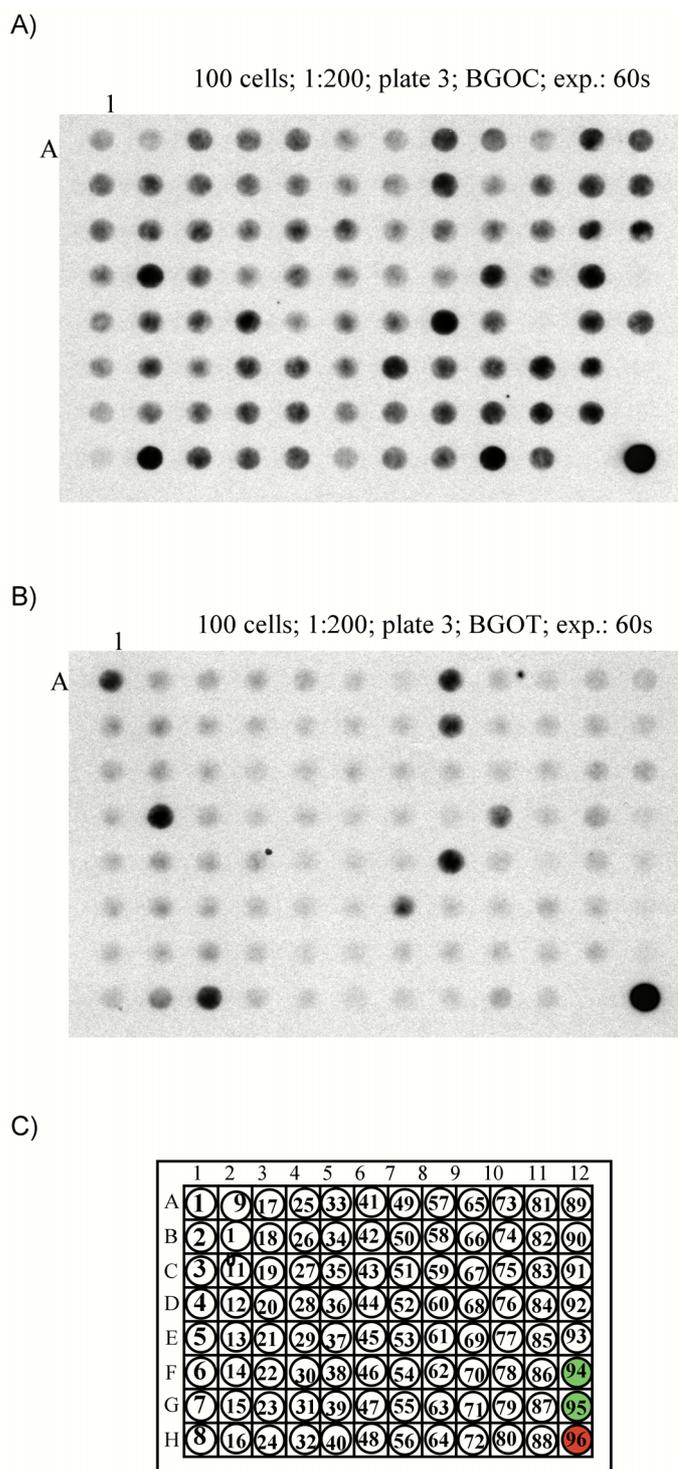


Figure 3-79. pBbgMCB/LNC28.28 (100 cells) - loss of resolution for BGOC (example). Depicted are the filters derived from the 100 cell experiment (dilution 1:200; plate 3) which were probed with BGOC (upper picture) and BGOT (lower picture). Exposure time was in both cases 60 seconds.

wildtype molecule which is set before the PCR, is on average maintained during the PCR-cycles. or whether biases in favour or to the disadvantage of the rare specimen are introduced.

Both questions about the general amplification power of the LNC-PCR as well as the expected fate of rare mutants in such a PCR are addressed in the following two sections.

Efficiency of amplification of its target region by LNC-PCR

To gain insight into the power of the LNC-PCR, the filters of the 50 and 100 cell experiments which have been probed with KBO58 were used. The oligonucleotide KBO58 detects both the amplicon generated by amplification on pBbgMCB as on LNC28.28 cells (see figure 3-28), and can therefore be used as a tool to obtain a readout of the PCR efficiency. On each filter of the dot-blot, dot number 96 represents 10^{10} molecules of pBbgMCB as a positive control for blotting efficiency (see figure 3-64). The measured signal intensity derived upon probing with KBO58 of sample number 96 was compared to the mean of signal intensities of the other samples (except 94 and 95, which are negative controls). The evaluation was performed with the data obtained from the 100 cell experiment. As a control for consistency also part of the data obtained for the 50 cell experiment was evaluated. For each evaluation exposure times for KBO58 were taken for which the obtained signal intensities were not yet in the saturated range (as judged by eye and see also figure 3-80). The results are summarized in table 3-17 to 3-19. Based on this evaluation, about 10^{10} molecules are produced after 35 cycles of the LNC-PCR, compared to an expected amount of 10^{12} molecules at 100% efficiency of the PCR (table 3-19). As shown in table 3-20 this result is very consistent with the 50 cell experiment. This estimated number of produced copy numbers can be used to calculate the PCR efficiency per cycle. Therefore the equation $N_f = N_0(1+Y)^n$ can be used, whereby N_f is the final copy number; N_0 the initial copy number, Y the efficiency of primer extension per cycle, and n the number of cycles (Dieffenbach *et al.* 1995). Transformed for Y the equation reads:

$$Y = \frac{\sqrt[n]{N_f}}{\sqrt[n]{N_0}} - 1$$

Y is for the LNC-PCR on cells presented here 0.7. This value fits very well to other published values on PCR efficiencies (Hermann *et al.* (1997); and (Dieffenbach *et al.* 1995)), and is actually above a reported value of 0.6 for Pfu (P. Andre, unpublished results, cited in

Dieffenbach (1995)). Therefore it can be assumed that the conditions for the LNC-PCR on cells are well chosen.

For the mutant, if present as a single copy at the beginning of the assay, the calculated efficiency means that the SSOP should be able to detect 10^8 molecules (20pg of a 1200bp fragment). This sensitivity is to be expected from data on plasmid DNA being blotted, and also due to the manufacturers claim (see http://www.apbiotech.com/technical/technical_index.html). Additionally it has been shown that starting from 10 molecules of target sequence a signal is easily detectable in the SSOP after 35 cycles of LNC PCR (figure 3-81). Finally it has to be said the this estimate presented here is rather conservative, since the 10^{10} molecules pBbgMCB of sample 96 blotted onto the filters are compared to less than half of the PCR product blotted (20 μ l of 50 μ l total PCR volume). Also the detection limit of DNA in agarose gels stained with ethidium bromide lies about 10ng (Maniatis *et al.* 1989). As can be seen in figure 3-45, 5 μ l of the product of the 35 cycle LNC-PCR on LNC28.28 cells is easily detectable in ethidium bromide stained agarose gels, and the yield is remarkably constant. If one assumes that the bands represent in average between 10to 100ng DNA, than about 40ng to 400ng (20 μ l were loaded pere sample per filter) would in average go into the *bg*-SSOP assay. This is equivalent to 3×10^{10} to 3×10^{11} molecules, which fits very well to the above estimate.

100 cells: BGOT/KBO58 (10m)			100 cells: BGOC/KBO58 (60s)		
Fälle	908		Fälle	1100	
Mittelwert		2566,3204	Mittelwert		3014,0432
Standardfehler		4,5320	Standardfehler		5,9126
95% Konfidenzintervall des Mittelwerts	Untergrenze	2557,4261	95% Konfidenzintervall des Mittelwerts	Untergrenze	3002,4420
	Obergrenze	2575,2148		Obergrenze	3025,6444
Median		2589,0900	Median		3057,8910
Varianz		18649,210	Varianz		38454,172
Standardabweichung		136,5621	Standardabweichung		196,0974
Minimum		2000,80	Minimum		1234,34
Maximum		2819,43	Maximum		3368,61
Spannweite		818,63	Spannweite		2134,27

Table 3-17. Mean of od obtained upon probing with KBO58. Evaluated were the od measurements of the 100 cell experiment. The filters had been probed either first with BGOT and afterwards with KBO58 (exposure time: 10 minutes; BGOT/KBO58 (10m)), or first with BGOC and afterwards with KBO58 (exposure time 60 seconds; BGOC/KBO58 (60s)). Evaluation procedure: The od values per filter were collected and compared to the negative controls (samples 94 and 95). The lowest values were regarded as “failed PCR’s” and taken out of the evaluation (2% of all cases for BGOT/KBO58 and BGOC/KBO58). The remaining values were pooled and the mean was calculated. Note: For BGOT/KBO58 (10m) the 1/400 dilution, plate 1 and 1/800 dilution, plate 1 were taken out entirely as the od--values of the samples lied consistently significantly beneath those of the rest of the series. Computing done with SPSS9.0.

S96 (BGOT/KBO58)			S96 (BGOC/KBO58)		
Fälle		9	Fälle		11
Mittelwert		2484,1022	Mittelwert		3037,0397
Standardfehler		68,0658	Standardfehler		30,4032
95% Konfidenzintervall des Mittelwerts	Untergrenze	2327,1422	95% Konfidenzintervall des Mittelwerts	Untergrenze	2969,2971
	Obergrenze	2641,0623		Obergrenze	3104,7823
Median		2504,6500	Median		3071,5290
Varianz		41696,598	Varianz		10167,919
Standardabweichung		204,1974	Standardabweichung		100,8361
Minimum		2010,93	Minimum		2803,42
Maximum		2759,28	Maximum		3173,46
Spannweite		748,35	Spannweite		370,05

Table 3-18. Mean od-values of samples 96 obtained upon probing with KBO58. Sample 96 serves as a positive control and represents 10^{10} molecules. The filters had been probed either first with BGOT and afterwards with KBO58 (exposure time: 10 minutes; BGOT/KBO58), or first with BGOC and afterwards with KBO58 (exposure time 60 seconds; BGOC/KBO58). Evaluation procedure: The od-values obtained for sample 96 per filter were pooled and the mean value calculated. This mean value was subsequently compared to the mean value obtained for the LNC-PCR products (see table 3-18). Computing done with SPSS9.0.

sample	mean od	molecules
BGOT/KBO58	2566.3204	1,03E+10
S96 (BGOT/KBO58)	2484.1022	1,00E+10
BGOC/KBO58	3014.0432	9,92E+09
S96 (BGOC/KBO58)	3037.0397	1,00E+10

Table 3-19. Calculated number of molecules produced after 35 cycles of LNC-PCR. The mean values of sample 96 (table 3-19) representing 10^{10} molecules were used to calculate the molecule number of the means obtained in table 3-18 for the filter series BGOT/KBO58 and BGOC/KBO58 (depicted in bold).

50 cells: BGOT/KBO58 (5m)			S96 (BGOT/KBO58)		
Fälle		356	Fälle		4
Mittelwert		2474,5208	Mittelwert		3109,8550
		32,9048	Standardfehler		51,3210
95% Konfidenzintervall des Mittelwerts	Untergrenze	2409,8079	95% Konfidenzintervall des Mittelwerts	Untergrenze	2946,5287
	Obergrenze	2539,2337		Obergrenze	3273,1813
Median		2620,0850	Median		3110,6950
Varianz		385451,058	Varianz		10535,378
Standardabweichung		620,8470	Standardabweichung		102,6420
Minimum		1006,94	Minimum		2983,95
Maximum		3332,23	Maximum		3234,08
Spannweite		2325,29	Spannweite		250,13

Table 3-20. Mean of od obtained upon probing with KBO58. Evaluated were the od measurements of the 50 cell experiment. The filters had been probed first with BGOT and afterwards with KBO58 (exposure time: 5 minutes; BGOT/KBO58 (5m)). Evaluation procedure: The od values per filter were collected and compared to the negative controls (samples 94 and 95). The lowest values were regarded as “failed PCR’s” and taken out of the evaluation (4% of all cases). The remaining values were pooled and the mean was calculated. Right panels: mean of od-values obtained for sample 96 (10^{10} molecules pBbgMCB). Computing done with SPSS9.0.

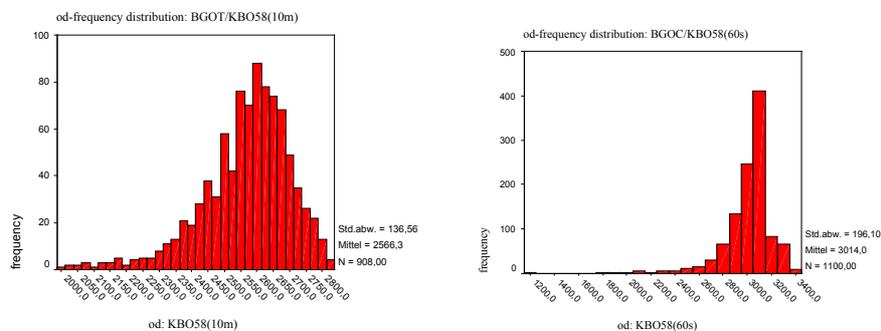


Figure 3-80. Distribution of the od-values for BGOT/KBO58(10m) and BGOC/KBO58(60s). The Histograms depict the distribution of the od-values that went into the calculation of the means for the 100-cell experiment (see table 3-18).

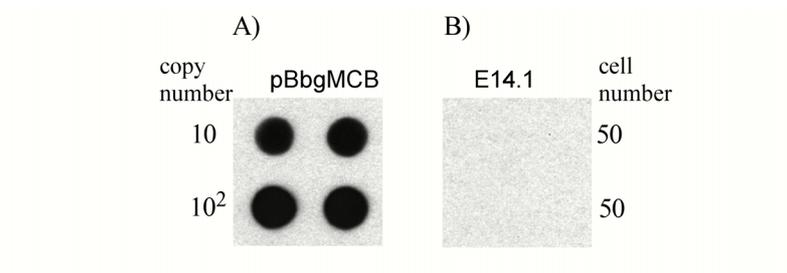


Figure 3-81. LNC-PCR on pBbgMCB. Evaluation by SSOP. Depicted are the images of filters, onto which 20µl of the PCR products of a 35cycle LNC-PCR on pBbgMCB have been transferred. **A)** the plasmid pBbgMCB was amplified in the presence of 50 E14.1 cells. Template copy number as indicated. **B)** Amplification on indicated number of E14.1 cells, as negative control. The LNC-PCR and SSOP procedure was done under standard conditions as described in Material and Methods. Probe: BGOC (exposure time: 60 seconds). All PCR's have been done in doublets for A) and four times for B).

Following the fate of rare specimen in PCRs with computer simulations

(This work was done in collaboration with Gert Osterburg)

Ultimately we would like to amplify on pools of 100 copies the K^{bLNC} target region. In a fraction of these pools, which size depends on the ultimate frequency of the mutants, we expect the presence of a mutant molecule at a ratio of 1/100. We wanted to know to what extent the starting ratio of the rare mutant molecule will be maintained during the PCR. Therefore we initiated computer simulations of the PCR-amplification process (16 cycles), assuming different efficiencies Y of the PCR ((Dieffenbach *et al.* 1995); note: for the calculations presented here the value q representing the “PCR-inefficiency” ($q=1-Y$) was used!). The PCR simulations show that with decreasing PCR-efficiencies the probability to shift the ratio of mutant to wildtype molecule towards the wildtype molecule increases (see table 3-21). In other words, the less efficient a PCR is, the more likely it becomes to lose a rare molecule (mutant). At our assumed efficiency of $Y=0.7$ for the LNC-PCR (see above: PCR efficiency) in about 50% of all sampled cases the final frequency will be lower than the starting frequency of 10^{-2} . This means that if the detection limit of the SSOP lies exactly at 1:100 (mutant to wildtype molecule) 50% of all mutant molecules would get lost due to failure of detection. Examples for distributions of observed values for the end ratio after 16 cycles of PCR are shown in figure 3-82. It is noteworthy that with decreasing PCR efficiency the distributions get skewed to the right. It seems that lower PCR efficiencies lead more often to end-ratios favouring the wildtype, but that this effect is accompanied by the (rare) occurrence of end ratios in favour of the mutant molecule. This is also visible in table 3-21 (see column “max”). Since we could not exceed 16 PCR-cycles (computational limits), we finally tested whether the number of cycles has an influence on this effect (table 3-22). In simulations we assumed a constant PCR-efficiency of $f=0.9$, and employed increasing numbers of cycles (10 to 16). No effect on the distribution of end ratios could be detected.

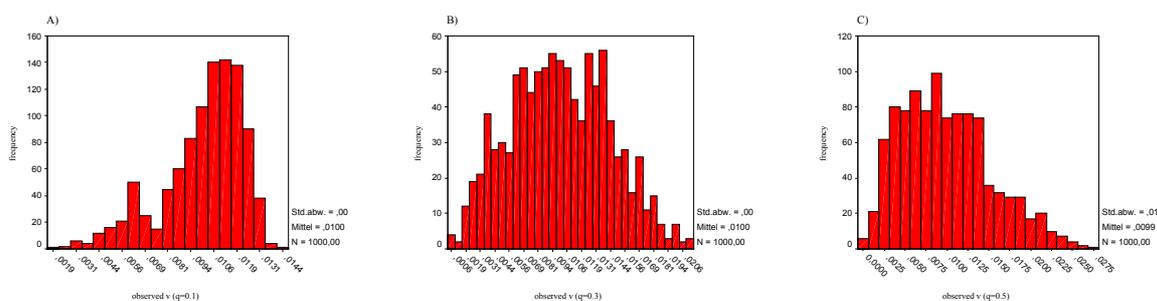


Figure 3-82. Histograms of distributions of end ratios rare: common molecule in dependence of PCR-efficiency. Depicted are the frequency distributions of end ratios rare:common molecule after 16 cycles of PCR (1000 PCR -simulations). A) $q=0.1$; $f(\text{PCR efficiency})=0.9$. B) $q=0.3$; $f=0.7$. C) $q=0.5$; $f=0.5$.

q	V	mean	stddev	min	Q25	median	Q75	max	max[f_L]
0.025	0.01	0.0101	0.0010	0.0041	0.0099	0.0104	0.0107	0.0112	0.7240
0.050	0.01	0.0101	0.0017	0.0027	0.0096	0.0105	0.0111	0.0123	0.6570
0.075	0.01	0.0101	0.0020	0.0016	0.0093	0.0107	0.0115	0.0132	0.6520
0.100	0.01	0.0100	0.0024	0.0015	0.0086	0.0106	0.0118	0.0143	0.5910
0.125	0.01	0.0101	0.0026	0.0014	0.0085	0.0106	0.0120	0.0151	0.5810
0.150	0.01	0.0101	0.0029	0.0014	0.0080	0.0105	0.0123	0.0161	0.5650
0.175	0.01	0.0102	0.0031	0.0005	0.0079	0.0106	0.0126	0.0177	0.5630
0.200	0.01	0.0100	0.0034	0.0017	0.0075	0.0103	0.0127	0.0189	0.5330
0.225	0.01	0.0102	0.0036	0.0007	0.0076	0.0106	0.0129	0.0190	0.5580
0.250	0.01	0.0101	0.0038	0.0006	0.0071	0.0101	0.0130	0.0202	0.5140
0.275	0.01	0.0102	0.0040	0.0006	0.0071	0.0104	0.0131	0.0223	0.5340
0.300	0.01	0.0103	0.0043	0.0007	0.0070	0.0101	0.0135	0.0227	0.5160
0.325	0.01	0.0100	0.0045	0.0004	0.0065	0.0097	0.0132	0.0236	0.4670
0.350	0.01	0.0101	0.0048	0.0004	0.0064	0.0099	0.0132	0.0271	0.4870
0.375	0.01	0.0102	0.0049	0.0001	0.0064	0.0097	0.0137	0.0247	0.4780
0.400	0.01	0.0102	0.0052	0.0002	0.0060	0.0098	0.0135	0.0274	0.4870
0.425	0.01	0.0100	0.0053	0.0003	0.0060	0.0093	0.0133	0.0269	0.4470
0.450	0.01	0.0102	0.0056	0.0001	0.0059	0.0094	0.0138	0.0274	0.4630
0.475	0.01	0.0102	0.0055	0.0003	0.0061	0.0095	0.0135	0.0302	0.4610
0.500	0.01	0.0102	0.0061	0.0001	0.0052	0.0097	0.0139	0.0345	0.4850

Table 3-21. End ratio rare:common molecule in dependence of PCR efficiency (16 cycles; 1000 PCR-simulations per experiment). Listed is the value for the frequency with which the starting ratio of $V=0.01$ or greater ($\max[f_L]$) appears for a given q ("PCR-inefficiency"; $q=1-Y$ (Y : PCR efficiency, Dieffenbach (1995))). Computation: To start with N_0 mutant and M_0 wildtype molecules are given. The term $N_0/(N_0+M_0)$ describes the ratio of mutant molecules N to the total amount of molecules. Iterations: steps of $k+1$; each step k is equivalent to one amplification cycle in a PCR. At each step k of the $[kN+kM]$ molecules a portion of the size $F*(kN+kM)$ is chosen randomly and doubled. Since the choice is random, it is not known how many of the mutant and how many of the wildtype molecules are taken up at each k . F describes the fraction size and is equivalent to the PCR efficiency constant Y . The remaining "nonchosen" molecules go into the next amplification cycle without duplication. Present are therefore $(k+1)N$ mutant and $(k+1)M$ wildtype molecules. The total molecules calculates as follows: $\Sigma = (k+1)N+(k+1)M = (kN+kM)*(1+F)$. After n cycles as described above (in the simulation presented here 16) the total number of molecules calculates as follows: $(N_0+M_0)*(1+F)^n$. At the same time a distinct end-ratio of mutant molecules to total amount of molecules $N_n/(N_n+M_n)$ is built up by chance. q : $(1-[\text{PCR-efficiency}])$. V : starting ratio (mutant : wildtype). Mean: arithmetic mean of computed end ratios. stddev: standard deviation. Min: minimum of computed end ratios. Q25: first quartile. Median: Median of computed end ratios. Q75: third quartile. Max: maximum of computed end ratios. $\max[f_L]$: fraction of the starting ratio of $V=0.01$ or greater. In bold: values for $q=0.3$ which is equivalent to our calculated efficiency of $f=0.7$ for the LNC-PCR (see section 3.2.5. - An estimate of the LNC-PCR efficiency), and for $q=0.4$ which is equivalent to an estimated PCR efficiency for Pfu (Dieffenbach, 1995).

cycles	q	V	mean	stddev	min	25%	median	75%	max	Max[F(L)]
10	0.100	0.01	0.0100	0.00241	0.00288	0.00888	0.01037	0.01187	0.01386	0.56333
11	0.100	0.01	0.0101	0.00233	0.00301	0.00869	0.01069	0.01175	0.01389	0.62000
12	0.100	0.01	0.0102	0.00220	0.00310	0.00859	0.01075	0.01184	0.01389	0.60333
13	0.100	0.01	0.0101	0.00236	0.00250	0.00907	0.01050	0.01182	0.01392	0.61333
14	0.100	0.01	0.0102	0.00227	0.00307	0.00907	0.01084	0.01186	0.01449	0.63667
15	0.100	0.01	0.0103	0.00222	0.00229	0.00924	0.01081	0.01180	0.01390	0.64000
16	0.100	0.01	0.0101	0.00241	0.00240	0.00877	0.01083	0.01192	0.01355	0.60000

Table 3.22 End ratio rare:common molecule in dependence of PCR cycle number ($q=0.1$; 300 PCR simulations per experiment). Listed is the value for the frequency with which the starting ratio of $V=0.01$ or greater ($\max[f_L]$) appears for a given number of PCR cycles (cycles). Computation as in table 3.23. Cycles: number of PCR cycles. q : $(1-[\text{PCR-efficiency}])$. V : starting ratio. Mean: arithmetic mean of computed end ratios. stddev: standard deviation. Min: minimum of computed end ratios. Q25: first quartile. Median: Median of computed end ratios. Q75: third quartile. Max: maximum of computed end ratios. $\max[f_L]$: fraction of the starting ratio of $V=0.01$ or greater.

4. DISCUSSION

We set out to examine gene conversion as a possible molecular mechanism underlying the polymorphism observed for MHC genes. It is tempting to think of gene conversion as an efficient process to generate optimised sequence variability in the MHC. It would represent a mutational process contributing to the resistance against fast evolving pathogens, in accordance with the hypothesis that natural selection in host-pathogen systems favours the protrusion of elevated mutation rates in resistance genes as put forward by Haldane ((Haldane 1949), and see also introduction). Gene conversion could be of high significance for the adaptation of the vertebrate immune system to a threatening and constantly changing environment, and it could itself be the target of adaptive regulatory mechanisms (it would than follow the “strong form” of the gene conversion hypothesis of MHC evolution as described in the introduction). It was our intention to develop an assay to detect germline *H2 K^b* mutations to light, at a frequency allowing an answer to questions concerning the molecular mechanism behind it as well as the adaptive significance of it

The major driving force, motivating us to pursue this path, was certainly the 40 *in vivo H2* mutants that had been recorded (see table 1-1). They were a stimulus for the exposure of many theories, ideas and vivid discussions around the evolution of the MHC (e.g. (Loh *et al.* 1984; Geliebter *et al.* 1986; Klein 1986; Pease 1986; Howard 1995)). The gene conversion hypothesis of MHC evolution found much support from indirect evidence such as the phylogenetic analysis of sequence patterns in MHC genes (e.g. (Parham *et al.* 1989; Parham *et al.* 1991; Yuhki *et al.* 1994)), as well as direct experimental approaches (e.g. (Högstrand and Böhme 1994; Huang *et al.* 1995)). However it was our feeling that due to a lack of tools which would make MHC genes more “accessible”, a systematic analysis with respect to mutations was still in need. Most of the questions which surround the hypothesis of gene conversion in MHC genes, such as the targeting of such events within the genes, a potential sex bias, onset during development, and whether it is a meiotic or mitotic event remain unanswered (see also introduction). We wanted to clear away the major obstacles (the low frequency combined with the presence of an array of highly homologous genes (see introduction) believing that this would finally open the portal to a more systematic view on gene conversion in the MHC. Obviously we did not yet manage to open the portal. The question remains, whether our approach did get as anywhere near our goal, and which steps could be taken next. This will be discussed in this final section.

4.1. Views on a multiple component gene conversion detector

In its original form the protocol which we planned to assemble consisted of a target gene enrichment step, the M450^{IZ}-*lacOp* rescue (section 3.1.), and as a mutation enrichment step, Denaturing Gradient Gel Electrophoresis (DGGE, (Sousa 2000)). Starting with the transgenic K^{bLNC} mice (see Materials and Methods) pools of primordial germ cells should have been isolated and the M450^{IZ}-*lacOp* rescue lead to enrichment of the K^{bLNC} gene. After a PCR step, the material should be further enriched for mutations by the tube DGGE. Finally analysis of the mutants should be done by cloning steps and sequencing. As explained in section 3.2. we altered the course slightly during the continuation of this project, by introducing a combined PCR/Sequence Specific Oligonucleotide Probing (SSOP) assay. This was supposed to add flexibility into the overall approach. The components of the gene conversion detection assay (with the exception of the DGGE) will be discussed in the following paragraphs.

4.1.1. Enrichment of target sequences through the M450^{IZ}-*lacOp* rescue

Since the *H2 K^b* gene represents an element of a multigene family, with several highly homologous members, we had to find a way to suppress the risk of creating *in vitro* artefacts arising during PCR steps of the mutation detection assay. We showed that the enrichment of target sequences via extraction of *E.coli lac-operator* tagged DNA fragments has a very high potential. The system is based on a protocol published by Gossen et al., in which transgenic mice have been generated carrying about 40 *lac-operator* containing plasmid copies integrated into their genome (Gossen *et al.* 1993). In their case the rescue was evaluated by recirculation of the plasmids coupled with estimates of the transformation efficiency of *E.coli*. Due to the design of the rescue in their case enrichment was not only achieved through the exploitation of the high affinity of *lac-operator* to the Lac repressor and specific elution with IPTG, but also through the transformation step of *E.coli*. Since we wanted to work with minimally changed endogenous gene sequences we had to adapt the assay for our purposes, relying solely on the *lac-operator*/LacI binding, and the specific elution with IPTG. Therefore direct ways to assess the enrichment power such as capillary electrophoresis, southern blotting and comparative PCR, were employed to monitor the efficiency of the assay during optimisation. We wanted to show that the presence of contaminating *lacOp*-negative sequences can be reduced so that the occurrence of *in vitro* artefacts resembling gene conversion events could be largely excluded. An important step is

the successful affinity based purification of LacIZ, using *lac-operator* charged NHS-columns. This procedure enriches for active protein, since it is based on the proper function of specific binding as well as effective IPTG elution of the LacIZ protein. Activity losses of LacI and LacI/LacZ chimeric proteins during purification have been observed by me and others ((Kania 1976; Fried *et al.* 1981), Dynal cooperation Oslo (personal communication) and discussed in (Martinson 1996)). Also, which in the context of our project is more significant, during the characterisation of LacI/LacZ chimeras higher molecular weight aggregates were found which showed unspecific binding to DNA, but no Lac repressor activity (Kania *et al.* 1974). The affinity based purification of LacIZ proved to be an efficient and economic method which can yield high amounts of active LacIZ.

The purification of active LacIZ allowed us to further characterise further the M450^{IZ}-*lacOp* rescue with regard to efficient enrichment of *lacOp*-tagged genes or fragments out of the genome of mice. While the original protocol led to impressive enrichment factors in the plasmid system based on pBluescript (section 3.1.2.), it had to be optimised for the ultimate goal. This was achieved through a test series on samples containing genomic DNA which had been experimentally contaminated with the plasmid NKrQ as the target sequence, and appropriate *lac-operator* negative plasmids. The optimisation of the rescue assay ultimately allowed us to assess its value working with cells.

The LNC28.28 ES-cells that served to create the K^{bLNC} transgenic mice (see Material and Methods), were ideally suited to assess the power of the M450^{IZ}-*lacOp* rescue assay in the context of the isolation of single copy *lacOp* tagged genes out of whole genomes from tissues of transgenic mice. In the mouse MHC about 30 class I genes are present which could potentially represent exactly the class of homologous sequences which lead to the danger of introducing in vitro artefacts (Klein *et al.* 1986; Trowsdale 1995). By employing the 18S rDNA genes as standards for contaminating DNA-sequences we were able to show that we can get rid of such sequences to an impressive degree. To what extent the actual risk of introducing in vitro artefacts during the mutation detection assay is suppressed can not be answered precisely yet. Future experiments will attempt to assess this risk and its suppression by the M450^{IZ}-*lacOp* rescue more directly. In principle one could also attempt to assess the reduction of introduction of in vitro artefacts in a plasmid based system, by using two plasmids containing homologous sequences. One of the plasmids does, the other one does not contain the *lac-operator* sequence. The two plasmids could be mixed at different ratios and the M450^{IZ}-*lacOp* rescue performed on a fraction of the mixtures. After PCR the degree of

introduced mutated molecules per assay could be assessed. However it would be impossible to transfer any obtained result directly to the situation we encounter by analysis of members of the mouse MHC and naturally the more one attempts to imitate this system the more one risks to contaminate the working environment with exactly the class of artefacts one attempts to suppress. Also future experiments aiming at the detection of mutations in MHC genes will be accompanied by appropriate controls, such as the comparative detection of mutations in the same tissue with and without the M450^{I^Z}-*lacOp* rescue assay. A further control is the comparison of germline derived cells to somatic cells where we do not expect gene conversion to occur (see also (Hogstrand *et al.* 1994)).

Finally we will also explore whether we can reduce the sample size of the M450^{I^Z}-*lacOp* rescue assay. We decided to work on samples representing 10⁵ genomes, since we were confident that pools of 10⁵ cells are a suitable size for the rescue assay. However by day 13 p.c. approximately 25 000 primordial germ cells (PGC) have colonized each gonad primordium of the mouse (Hogan *et al.* 1994). Depending on the nature of question we ask in future experiments it can be of more value to work on a gonad by gonad basis, instead of having to pool a number of gonads. In this respect, concerning the isolation and purity of PGC's it is noteworthy that recently a monoclonal antibody (TG-1) has become available which allows the specific purification of PGC's with the help of paramagnetic beads (Pesce *et al.* 1995).

In summary it is reasonable to say that the M450^{I^Z}-*lacOp* rescue as such has proven to possess a high potential to enrich for *lac-operator* tagged DNA sequences with an impressive power. It can in principle be combined with any other component of a gene conversion detection protocol. Further exploration will show how robust the assay is, and also of what value it is when it comes to the detection of gene conversion events in the K^{bLNC} gene.

4.1.2. Enrichment and detection of mutations through the K^{bLNC} -PCR coupled to a *bg* specific Sequence Specific Oligonucleotide Probing Assay

The decision to introduce an assay consisting of a PCR step which is directly performed on pools of cells and targeted to the K^{bLNC} transgene, combined with a sequence specific oligonucleotide probing assay (SSOP) for the *bg* mutations meant at the same time to economise on effort to bring MHC mutations to light and abandoning the original intention to screen for unknown sequence variants. This needs to be put in context. In our gene conversion detection project, we started out attempting to combine the M450^{I^Z}-*lacOp* rescue with

Denaturant Gradient Gel Electrophoresis. Our intention was to employ DGGE exactly because we wanted to detect unknown sequence variants. However it turned out that even though DGGE is definitely a valuable tool in the mutation detection field, for our purposes (high throughput) it was considerably laborious and also that intrinsic characteristics to the method lower the mutation detection power noticeably (discussed in Sousa (2000) and see below). The *bg*-specific SSOP on the other hand is restricted to the detection of a particular documented mutation event, the *bg*-series, but allows to be carried out in a streamlined way in multiarray systems. The assay is still dependent on the availability of the K^{bLNC} transgene, but can be employed independently of the $M450^{IZ}$ -*lacOp* rescue and of the mutation enriching DGGE step originally foreseen to be part of the whole gene conversion detection assay (Sousa 2000). However it could be combined with one or all of the other components if desired. A major advantage of this assay lies in its potential for high throughput and automation. A major drawback is the focus on one historical mutation. This eliminates immediately the question of whether gene conversion events in the MHC are targeted preferentially to certain regions within the MHC genes (but see below). Furthermore statements about frequencies would be restricted to the mutation in question. However other issues such as a sex bias, or the onset of the gene conversion process during development, meiosis versus mitotic process remain unaffected.

When the tools of the combined PCR/SSOP assay, the K^{bLNC} -mice; the LNC-PCR on cells; and the *bg* specific oligonucleotide probing were available or else established in model systems, we decided not to attempt directly to bring gene conversion events to light in our laboratory, but rather to cautiously approach this intention. We had good reasons to follow this reserved strategy. We were confident about the general reliability of the LNC-PCR on cells, but not so much about the stochastic characteristics inherent to it. Sequence specific oligonucleotide probing assays are routinely employed to assess heterozygote states in clinical research, but it was not clear to us whether a detection limit of 1/100 mutant to wildtype sequence was within a reasonable range with the *bg*-specific oligonucleotides, most of all since we were looking for single base substitutions, and not for mutations leaving more prominent marks. We aimed at this detection limit of 10^{-2} due to the expected rareness of the *bg*-like mutations. As explained in the results (section 3.2.2) the reason was that strictly speaking the SSOP serves as mutation enrichment step. Any pool of cells showing up positive in the SSOP assay would have to be further analysed by batch cloning techniques and sequencing to finally uncover the detected mutation. A detection limit of 10^{-2} means to a first

approximation that 10 000 cells could be analysed with one 96-well plate, which seemed at that time to be a reasonable number relative to our expected mutation frequency of 10^{-4} . This detection limit was reached. So the components of the PCR/SSOP based mutation detection were at our disposition, and we could have searched for mutations in germline derived cells from 129^{KbLNC} mice. Our hesitation to do so was again founded on the low frequency we expected to encounter and all uncertainties which go along with it. Without further characterisation of the assay under conditions very similar to the ultimate experiment involving germline derived cells from mice, strong statements about the frequency with which such mutations occur would not have been possible. This motivated us to set up the experiment using pBbgMCB as a mutant molecule which we diluted at different steps into LNC28.28 cells. The results enabled us to extrapolate an expectation of what effort we would have to make as to detect mutations which occur at a frequency in the range of 10^{-4} . It will also allow us at later time points to deduce based on an observed mutation frequency in experiments with germline derived mouse cells, the real frequency which probably underlies the phenomenon. We had to be very careful not to contaminate the working environment with molecules which possess sequence characteristics resembling the events we are looking for, and on the other hand we wanted to come as close as possible to the actual analysis. This led to the cloning the plasmid pBbgMCB, which represents a valuable tool in the context of this project.

The assay as such turned out to be remarkably robust. In the 50 cell experiment 651 pools of cells were screened for pBbgMCB, in the 100 cell experiment altogether 1116 pools. The PCR proved to be extraordinarily consistent and also efficient (see in 3.2.5. - An estimate of the LNC-PCR efficiency). Also in none of the examined plates the negative controls were contaminated which stresses the feasibility of the set-up using 96-well plates. The SSOP part of the assay is in principle a very practicable approach and it worked well. There is a slight reservation to this statement, which concerns the consistency of signal strength in relation to exposure time of filters. In most cases the data obtained in the SSOP assay were quite consistent. However it occurs that in some cases the exposure time to signal strength ratios for the same oligonucleotide (but different batches of labelled oligonucleotide), were quite different. Also, as discussed in 3.2.5. (Comparison of the bg-mutation detection oligonucleotides BGOT and BGOC) in the 100 cell experiment the probing with BGOC led to unsatisfactory resolution of the filter images. I suspect that these inconsistencies have their origin in the labelling efficiency of the oligonucleotides, since within one probed filter series

the obtained values are quite similar (one filter series represents all 96-well plates of either the 100 cell or 50 cell experiment and was either probed with BGOc or BGOT and subsequently with KBO58 (see also Materials and Methods). For each filter series, a new batch of labelled oligonucleotide was used. We initially thought that this would ensure a stable sensitivity of the assay. It is uncertain what the actual reason for the suspected variable labelling efficiency is, since the labelling conditions were held purposely very constant. However this potential drawback of the SSOP assay can in future be dealt with by testing a batch of labelled oligonucleotides with a testfilter. This would also allow to adjust for the most appropriate exposure time.

4.1.3. The mutation detection efficiency of the LNC-PCR/bgSSOP assay

The two experimental test series presented in section 3.2.5. were mainly accomplished to assess the power of the mutation detection assay which is based on a LNC-PCR combined with the *bg* specific SSOP. Both series involving the 50 or 100 cells, represent two independent experiments. The detection of potential mutant cell pools lied consistently a factor five below the prediction for a 100% efficiency of the assay. This consistency concerning the power of the assay indicates that inherent characteristics of the assay, or the way the experiment was set up underlie the factor five. A remarkable result in this respect is the observation in computer simulations that inefficient PCRs on mixtures of abundant and rare molecules tend to lead to the loss of the rare specimen (see in 3.2.5. - Following the fate of rare specimen in PCRs with Computer simulations). Interestingly this phenomenon is distinct enough to contribute to the lower than expected efficiency by a factor of two (due to the way the computer simulations were initiated this of course is only true, if any final mutation/wildtype ratio lower than 1/100 is not detectable any more). Other factors lowering the power of the proposed assay have to be speculative at present. One feature of the way the experiment is set up and which might be suboptimal is the fact that the rare mutant was added as “naked” plasmid DNA to the abundant wildtype molecule which was put into the assay as LNC28.28 cells. The protocol starts with the lysis of cells which takes place for one hour at an elevated temperature of 55°C in a Tris buffered system (1X Pfu Polymerase buffer). As discussed in 3.2.2 (Test of HEPES buffer for Pfu based amplification) heat supply in aqueous solution leads to degradation of DNA. Furthermore the amounts of material that went into the assay were estimated by different measures (pBbgMCB: photospectrometric measurement. cells: counting with a hemocytometer). While the “loss of rare specimen” effect is intrinsic to the assay, it is possible that other factors which diminish the power of the assay

could be removed. However in my opinion the remaining factor of 2.5 is not of major importance. Concerning the frequency estimate of mutations that could be derived from the detection of mutants, a factor of three is negligible in the context of magnitudes which are discussed here. If it is true that this factor is only inherent to the way the explorative experiments described here were set up, this would mean that the effort that goes along with detection of real mutants at a frequency of 10^{-4} would be three times less than expected. This has to be put into the context of the general feasibility of the assay, and the question whether there are (meanwhile) reasonable alternatives for our aim available.

4.1.4. Hunting for gene conversion events in germline MHC genes of Mice: The perspectives in the year 2001

Where do we stand with the mutation detection assay discussed here with respect to the uncovering of germline mutations in the *H2* of mice? – This has to remain a rather hypothetical question since no attempt was undertaken yet to use it for the discovery of *bg* mutations in the germline of mice. Still it is possible to appraise its value with regard to our aim, and to compare it to other methods to detect mutations. It is probably true that the *M450^{IZ}-lacOp* rescue assay is of great value when it comes to the enrichment of target sequences derived from the MHC. Yet it must be considered that we look for very rare events, but ultimately ask for high numbers of those since we would like to deduce conclusive answers to the questions raised in the introduction of this thesis. Every step in this assay is carried out manually. The volumes that are employed during the assay are small and every sample has to be treated with care and in a concentrated way to avoid contamination and great losses of the magnetic beads and therefore the target sequence. I estimate that for a person familiar with the assay it is possible to treat maximally ten samples in parallel, and at maximum to perform one assay in one day, starting from the lysis of cells. Since the protocol suggests to use pools containing 10^5 cells, this would be equivalent to the treatment of 10^6 cells at once. This is in the light of an expected frequency of mutants of 10^{-4} an acceptable number even considering the estimated yield of 10%.

The LNC-PCR/*bg*-specific SSOP assay is performed in 96-well plates. Per day two of these plates can be processed starting from the cell-lysis step until the transfer onto filters. The probing could be done over night and the development of the filters the next day. Nevertheless it would probably be more advisable to split up the analysis into a “period” of amplification and blotting and a period of filter development, each time involving a series of 96-well plates or filters respectively. This would economise on effort and minimise variation

of the conditions under which the assay is performed (in fact the 50 cell and 100 cell experiment presented here were executed like this). I estimate that this way per month about 30 plates could be processed.

Up to now three attempts to unravel gene conversion in germline genes of mammals using modern analytical methods stand out. These are the PCR based mouse class II gene *Ab* approach of Högstrand and Böhme (Hogstrand *et al.* 1994), the PCR/sequencing approach of Huang *et al.* (Huang *et al.* 1995) and the allele specific PCR/differential oligonucleotide typing approach of Zangenberg *et al.* (Zangenberg *et al.* 1995). All three approaches were presented in the introduction. These experiments were of great value for the discussion around gene conversion taking place in the MHC of mammals. Since the “classical” skin grafting experiments no new data had been produced which looked directly at mammalian genes. However this data stands out like a beacon until today, confirming that gene conversion takes place in the germline of mammals, but not allowing any strong claims about its significance or the molecular mechanisms being involved. This stresses that it is still justified and necessary to design a gene conversion detection assay which could be employed on a routine basis, with high throughput and in a reliable way. The approach discussed here represents a step into this direction. This is because, due to the availability of the K^{bLNC} transgenic mice, for the first time we are able to look at a mouse MHC I gene in a systematic way. We are able to suppress the risk of introducing in vitro artefacts by employing either the M450^{I_Z}-*lacOp* rescue, or the LNC-specific PCR. This is an important point since up to today it is the data of MHC class I genes obtained in the mutation screen of mice which has the highest significance in terms of numbers and stimulus around gene conversion in MHC genes (see Introduction and table 1-1). Also due to the region that the LNC-PCR amplicon covers (see figure 3-28) we are not solely restricted to the two base pair exchanges characterising the *bg*-series of mutations when it comes to the outcome of gene conversion events. Due to the oligonucleotides we employ we do rely on at least one of the two bases being exchanged. However the amplicon spans upstream and downstream of the *bg* marker region an area which shows a high number of non-homologous base positions leaving room for the detection of gene conversion events which show different features as the *bg* series. In fact since the amplicon covers the second intron as well as parts of the third intron even the question about the targeting of gene conversions (see introduction) could be addressed.

One important concern remains the throughput of the assay. It is worthwhile to look at other mutation detection methods with respect to their feasibility in the context of our project and the potential to combine them with our approach.

4.1.5. Features of mutation detection methods presently available

Generally one can divide mutation screening assays into physical methods, chemical methods, methods based on enzymatic recognition of mismatches, PCR based methods, and methods which are based on differential hybridisation techniques (Cotton *et al.* 1998). Often these methods overlap or are combined.

Physical methods of mutation detection

Denaturing Gradient Gel Electrophoresis (DGGE) separates DNA fragments on the basis of their melting properties (Fischer *et al.* 1983). DNA fragments run through a gradient of denaturant agents at elevated temperatures. Upon melting of the DNA strands these build forklike structures and are retarded. This can be exploited for analytical purposes (confirmation of the presence of mutant sequences) as well as for preparative purposes (enrichment of mutant sequences). DGGE is a very widely used method for the detection of mutant sequences. This is due to its wide range of applications, since native melting properties of DNA molecules can be manipulated by attachment of artificial clamps by PCR (Skopek 1991), and also because the detection rates of mutants are close to 100% (Lerman *et al.* 1998). In principle DGGE is ideally suited in the context of our project. Roberto Lorenzi attempted to employ DGGE to enrich for mutations in a particular fragment of the *H2 K^b* gene (Lorenzi 1994). He came to the conclusion that under his conditions the enrichment power would not be sufficiently high to detect gene conversion events which occur at a frequency of 10^{-4} on a reliable basis. Lorenzi attributed this mainly to a lack of stringency in temperature control. Later Ana Sousa in our laboratory established a tube-DGGE and characterised thoroughly its properties (Sousa 2000). Despite very stringent temperature control she concluded that the mutation enriching of the DGGE system is not great enough to enrich sufficiently for mutations in the context of our project. This was mainly due to background factors inherent to the approach, which could not be suppressed. A related method which was also tested in our laboratory is constant denaturing capillary electrophoresis (CDCE; Kirstin Hebenbrock – unpublished data). At that time *H2* class I test sequences were used in order to optimise the separation of mutant from wildtype DNA sequences, but the resolution remained suboptimal. This was mainly attributed to disadvantageous melting properties, due to the high

GC content of the sequences. A major advantage of CE is that due to the beneficial surface to volume ratio of the capillaries, very high excitation voltages can be applied, ensuring short running times in the minute range. Also the detection limit of DNA is much lower than in traditional flat bed gel systems (Landers 1993). Furthermore the collection of samples to be further processed (in our case the enriched mutant fraction) is easy and straight forward (the samples of interest can just be collected by letting them run into a PCR tube which serves at the same time as the anode). This is of very great value for us. However at that time we were working with a self assembled CE system, which only allowed to run one sample at a time taking away the advantage of short running times. This was one of the reasons why later the tube-DGGE system was introduced by Ana Sousa (see above), since it allowed the parallel processing of up to twelve samples. It is worth having another look at a CE system in the context of our project. Meanwhile multiarray systems (96 samples!) are available and using CE in mutation research is established (Gao *et al.* 2000; Larsen *et al.* 2000; Ren 2000). CE has been linked to modern highly sophisticated methods in order to improve the sensitivity (Mitchelson 2001). It should also be looked at again in the context of DNA damage due to heating, since Ana Sousa could show that this effect contributes significantly to the background which lowers the enrichment power of her tube-DGGE system (Sousa 2000). While in DGGE the samples are exposed several hours to elevated temperatures, this is only the case for several minutes in CE-systems. This should have a major effect on the integrity of the DNA and therefore lower the background considerably.

Both DGGE and CDGE belong to physical methods of mutation detection since physical properties of DNA lead to different running behaviour in gel systems, which serve to separate mutant from wildtype molecules. Another mutation detection method belonging to this category is single strand conformation analysis (SSCA; (Nataraj *et al.* 1999)). In SSCA electrophoretic mobility differences between wildtype and mutant DNA-strands are due to different folding of single strands in non-denaturing gels. A major disadvantage is that the detection rate is low compared to DGGE and that the running behaviour of unknown mutant strands is unpredictable (Hayashi *et al.* 1993).

Chemical cleavage of mismatch

Chemical Cleavage of Mismatch (CCM) is based on the reaction of certain chemicals with mismatched bases in heteroduplexes. This reaction makes these sites vulnerable to cleavage by piperidine, which can subsequently be detected by electrophoresis. Despite a

detection rate of 100% and high sensitivity the assay as such is very demanding and therefore not widely used (Cotton 1999).

Enzymatic detection of mutations

MutS based mutation detection

In recent years there has been growing interest in DNA repair mechanisms since the enzymes involved in it take part in a variety of biological phenomenon (Cleaver 1994; Koshland 1994; Koshland *et al.* 1994; Rosenberg 1997; Cummings *et al.* 1998; Diaz *et al.* 1998). This boost of interest in repair mechanisms resulted also in the attempt to exploit the capability of some elements of the repair machinery to recognise and bind to DNA base mismatches. Especially the MutHLS system of *E.coli* has been evaluated in this respect (Smith *et al.* 1996). An assay was proposed using MutS in an immobilised form to bind to mismatched bases, and to enrich thereby for mutations (Wagner *et al.* 1995). A major disadvantage of this assay lies in the fact that MutS shows a strong preference for certain mismatches. However work is in progress looking for MutS mutants which show much higher affinity for mismatched bases, combined with a lack of the disadvantageous preference for bases observed for the wildtype MutS (Miroslav Radman – personal communication). Another interesting approach combining the mismatch recognition characteristics of MutS with PCR will be described below.

T4 endonuclease VII based mutation detection

The characterisation of T4 endonuclease VII has a long history (Kemper *et al.* 1973). It is the product of gene 49 (gp49) of bacteriophage T4. This endonuclease is a Holliday-structure resolvase (X-solvase) responsible for clearing branched replicative DNA prior to packaging and therefore an integral part of the packaging machine of phage T4 (Golz *et al.* 1999). Its properties and involvement in repair mechanisms (Kleff *et al.* 1988; Solaro *et al.* 1993) prompted the attempt to use it as a mutation screening tool (Youil *et al.* 1995). Now this resolvase is used in enzyme mismatch cleavage (EMC), a technique capable of rapidly scanning DNA fragments for mutations. It relies on the ability of the bacteriophage T4 endonuclease VII, to cleave DNA at single base pair mismatches and small heteroduplex loops. Originally the process was performed using radioactively labeled DNA and the results analysed after denaturing polyacrylamide gel electrophoresis and autoradiography. Access to systems capable of detecting fluorescent species migrating through a gel and the widespread

availability of fluorescently tagged primers have improved upon the original technique. Fluorescent EMC detected the presence, position and number of mutations in DNA fragments as large as 1 kbp. It is claimed that the use of fluorescent tagged primers improves the original method with respect ease of use, increase in signal-to-noise ratio and the ability to multiplex samples by labelling DNA fragments with different fluorophores. It appears that this technique provides a sensitive, robust way for mutation detection (Babon, 1999). Recently the use of resolvases T4 endonuclease VII was combined with T7 endonuclease I in mutation detection to further improve on the sensitivity (Babon *et al.* 2000).

DNA amplification based detection of mutants

PCR

The polymerase chain reaction is extensively used to enrich for particular DNA sequences relative to homologous sequences by taking advantage of the fact that the PCR efficiency can be greatly inhibited due to one or a few mismatches in one of the primers (Parsons and Heflich refer to enrichment based of a particular DNA sequence relative to another closely-related DNA sequence based only on a change of one or a few bases as “genotypic selection methods“ (Parsons *et al.* 1997)). Also, in the case of the restriction fragment length polymorphism polymerase chain reaction (RFLP/PCR) the usage of restriction sites which are available in the amplicon of the wildtype sequence but absent in the mutant sequence can lead to impressive detection levels (down to 10^{-8}), once combined with other techniques such as cloning and oligonucleotide probing (Felley-Bosco *et al.* 1991; Parsons *et al.* 1997). A very effective Method with respect to sensitivity is the mismatch Amplification Mutation Assay (MAMA; (Cha *et al.* 1992)). In this assay a mutant specific primer is employed which contains one mismatch to the mutant allele but two mismatches to the wildtype allele. It was claimed that detection limits can go down to 10^{-5} . Another method combines the MutS approach with a mutation enriching PCR step. In a series of pioneering experiments Parsons and Heflich achieve detection limits down to 10^{-7} for a particular mutation by combining MutEx (see below) enrichment and allele-specific competitive blocker PCR (ACB-PCR) (Parsons *et al.* 1998; Parsons *et al.* 1998b). In the MutEx enrichment heteroduplex formation between the abundant wildtype and rare mutant sequences is initiated. MutS recognises and binds to the mismatches, thereby protecting those from subsequent exonucleolytic treatment. Afterwards the ACB-PCR is employed. So called blocker primers are designed which hybridise within the wildtype sequence to the region of base exchange of

the mutant sequence. They prevent elongation of the wildtype sequence throughout the PCR. At the same time normal primers are used which frame the region of interest in the mutant sequence, so that normal amplification occurs.

Mutation detection based on differential hybridisation combined with PCR

In the recent years techniques were developed which allow the detection of mutant molecules due to the usage of mutant specific fluorescent labelled oligonucleotide probes which emit signals only upon successful amplification of the targets. For the time being two major systems are available. They are of interest to us because they are quite sensitive, allow for online detection of mutants and are suitable for very high throughput. The first system is based on the TaqMan[®] probes (ABI; www.appliedbiosystems.com/molecularbiology/about/pcr/sds/). These are oligonucleotides specific for the mutant sequence. In their “native” state a fluorescent label (reporter dye) is attached to the 5' end of the probe, while a quencher at the 3' end. Due to the spatial proximity of quencher to reporter, no signal is emitted. During the PCR the probe binds to the complementary sequence. During the elongation cycle the 5'-3' exonucleolytic activity (Holland *et al.* 1991) of Taq destroys the TaqMan[®] probe, upon which a signal is emitted. With each cycle the signal intensity increases proportional to the amount of amplicon produced. In certain PCR machines (e.g. ABI Prism[™] 7700 and the Lightcycler[®] (Roche)) the signal emission can be followed by online monitoring. Another similar system is using so called “molecular beacons” (Tyagi *et al.* 1996). Again oligonucleotide probes are labelled with a fluorescent dye at one end and a quencher at the other end. Additionally the oligonucleotide is designed in such a way that it does not only specifically hybridise to the sequence looked at (the mutant) but also that at both ends there is a short region of complementarity so that a stem-like structure is formed (they resemble in this respect a little bit the tRNA molecules (Alberts *et al.* 1994), which brings the fluorescent label and the quencher into close proximity. Unless the probe binds to its target sequence no signal is emitted. Upon binding to the target sequence the “molecular beacon” gets spread out and the fluorescent label starts to emit its signal. Both the TaqMan[®] probes and the molecular beacons are used on a routine basis in clinical research, but, as for SSOP in almost all the cases for the examination of heterozygote states. The use of one of the two methods presented in our context would ask for at least as much optimisation as the *bg* specific SSOP assay described here (Häseler (ABI-technical service) - personal communication).

Rolling circle amplification

Rolling circle amplification (RCA) is another very promising approach for the detection of mutations (Baner 1998; Lizardi 1998). It is also a further example (like the usage of certain enzymes, see above) for the introduction of strategies provided by nature, into the mutation detection field. The rolling-circle mechanism of DNA replication is used by small prokaryotic genomes, such as single-stranded phages and plasmids. However, phages and plasmids have adapted the rolling-circle mechanism differently to suit their contrasting biological needs. Over the past several years, there has been a revival in studies on RC replication as a result of the discovery that many plasmids replicate by this mechanism and major advances have recently been made in the understanding of plasmid RC replication, including the characterisation of the biochemical activities of the plasmid initiator proteins and their interaction with the double-strand origin, the domain structure of the initiator proteins and the molecular basis for the function of single-strand origins in plasmid lagging strand synthesis (Novick 1998; Khan 2000). Interestingly in a recent report a link between DNA-repair and RCA was found in *E.coli* (Bruand *et al.* 2000).

RCA driven by DNA polymerase can replicate circularized oligonucleotide probes with either linear or geometric kinetics under isothermal conditions. In the presence of two primers, one hybridizing to the (+)-strand, and the other to the (-)-strand of DNA, a complex pattern of DNA strand displacement ensues that generates 10^9 or more copies of each circle in 90 minutes, enabling detection of point mutations in human genomic DNA. Using a single primer, RCA generates hundreds of tandemly linked copies of a covalently closed circle in a few minutes. If matrix-associated, the DNA product remains bound at the site of synthesis, where it may be tagged, condensed and imaged as a point light source. Linear oligonucleotide probes bound covalently on a glass surface can generate RCA signals, the colour of which indicates the allele status of the target, depending on the outcome of specific, target-directed ligation events. As RCA permits millions of individual probe molecules to be counted and sorted using colour codes, it is particularly amenable for the analysis of rare somatic mutations. RCA also shows promise for the detection of padlock probes bound to single-copy genes in cytological preparations (Baner *et al.* 1998; Lizardi *et al.* 1998; Hatch *et al.* 1999).

Lately major advances of RCA in the application range and in the context of microarrays have been made (Schweitzer *et al.* 2001). In a new report RCA was combined in

a very sophisticated way with allele-discriminating oligonucleotide probes. It was shown that signal amplification by RCA can be coupled to nucleic acid hybridization and multicolor fluorescence imaging to detect single nucleotide changes in DNA within a cytological context or in single DNA molecules. The authors claim that their approach provides a means for direct physical haplotyping and the analysis of somatic mutations on a cell-by-cell basis (Zhong *et al.* 2001), which is of course of great interest to us!

DNA chip technology

The microarray began as a way of looking at gene expression patterns in a given organism. The first arrays were made by spotting the cDNA or complementary mRNA of a number of genes onto a membrane in fixed locations and then hybridising it with the radiolabelled mRNA of the target organism. By exposing the membrane to x-ray film, a pattern of dots can be seen which correspond to the locations of the genes which are being expressed. From this it is possible to see which genes are being expressed. By quantifying the signal strengths it is possible to determine the level of mRNA expression of each gene relative to the other genes. Another method of making microarrays was later developed. This method involved spotting cDNA onto a glass slide and labelling samples with fluorescent dyes and has several advantages over the previous method. It uses non toxic fluorescent dyes instead of radioactivity therefore is much safer to use, glass slides are also easier to handle than membranes. DNA chips were the next step in microarrays. Like the glass slide microarrays, DNA chips have the advantage of being able to be used with two different samples at once. In addition to this, DNA chips can contain > 60 000 genes. The concept behind DNA chip or microarray technology is similar to that underlying northern and Southern blots. A machine known as an "arrayer" can robotically spot cDNAs or ESTs onto a glass slide so that each sequence has a known well arrayed onto a thumb-nail-sized glass chip, at a rate of 800 spots a minute. Alternative methods for creating a DNA array include photolithography, fragment-based DNA printing, and an ink-jet-based method, which are currently being refined (Cheung *et al.* 1999; Kurian *et al.* 1999). Meanwhile there are instructions available which help to build and set up a microarray facility (see e.g.: Stanford Medical School; www.cmgm.stanford.edu/pbrown/index.html). DNA chip techniques are also used in mutational research (Hacia 1999; Hacia *et al.* 1999; Wilgenbus *et al.* 1999) and a very high throughput of samples can be reached (Buetow *et al.* 2001). In future the DNA chip technology will certainly be of great value for projects as ours. The principle applied is the same as for the SSOP (specific hybridisation to target sequences) but naturally the big

advantage lies in the high sensitivity of the approach, the great amount of samples that can be screened at the same time and the fact that at the same time mutant and wildtype signal can be monitored making a standardization easier.

The royal road to mutation detection: Sequencing

In the context of the human genome project, sequencing has become a routine method. Meanwhile machines are available which allow the sequencing of samples with an impressive efficiency and processing speed. An example are systems based on capillary electrophoresis (Kheterpal *et al.* 1999). I would like to draw the attention on a very creative approach which was invented a few years ago by a young Swedish group, and which they presented at the Fourth International Workshop on Mutation Detection in Brno (1997). They called their approach “Pyrosequencing” (Ronaghi *et al.* 1996). This sequencing technique is based on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate (PPi) detection assay (ELIDA; (Nyren 1987)). It allows for real-time sequencing, without the need for electrophoresis. The assay consists of five steps:

- Step 1

A sequencing primer is hybridized to a single stranded DNA template, and incubated with the enzymes, DNA polymerase, ATP sulfurylase, luciferase and apyrase, and the substrates, adenosine 5' phosphosulfate (APS) and luciferin.

- Step 2

The first of four deoxynucleotide triphosphates (dNTP) is added to the reaction. DNA polymerase catalyzes the incorporation of the deoxynucleotide triphosphate into the DNA strand, if it is complementary to the base in the template strand. Each incorporation event is accompanied by release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide.

- Step 3

ATP sulfurylase quantitatively converts PPi to ATP in the presence of adenosine 5' phosphosulfate. This ATP drives the luciferase mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera. The height of each peak (light signal) is proportional to the number of nucleotides incorporated.

- Step 4

Apyrase, a nucleotide degrading enzyme, continuously degrades ATP and unincorporated dNTPs. This switches off the light and regenerates the reaction solution. The next dNTP is then added.

- Step 5

Addition of dNTPs is performed one at a time. Deoxyadenosine alfa-thio triphosphate (dATP α S) is used as a substitute for the natural deoxyadenosine triphosphate (dATP) since it is efficiently used by the DNA polymerase, but not recognized by the

luciferase. As the process continues, the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks. (Ronaghi *et al.* 1998)

Meanwhile pyrosequencing is an established method. While in 1997 the assay only allowed to read up to 15 bases ((Ronaghi *et al.* 1996) and Fourth International Workshop on Mutation Detection in Brno (1997)), now readouts up to 200bp are possible, and also a high throughput system has been developed ((Ronaghi 2001) and; - the producers claim that 96 samples can be analysed in less than an hour, and due to the set-up of the assay the sequencing can be monitored online (see www.pyrosequencing.com/pages/products_sa.html). The assay is constantly improved (Nordstrom *et al.* 2000; Ronaghi 2000). Pyrosequencing is widely used for SNP analysis and found its way into mutation analysis (Ahmadian *et al.* 2000; Garcia *et al.* 2000). It is to my opinion a method of great promise also for project like ours. I am confident that soon it will become feasible even for laboratories in the academic field of research (in contrast to industrial laboratories) to perform sequencing on a large scale. In the context of our gene conversion project this means that we might at the same stage be able to do without a mutation enrichment step. Theoretically it would be possible to sequence directly starting from PCR`s with single (haploid) cells such as sperm cells. We would probably still have to rely on the M450^{IZ}-*lacOp* rescue as a target enrichment step. It could be envisioned that pools of cells go through the target enrichment protocol, and after a subsequent batch cloning step directly into sequence analysis.

Some of the techniques for mutation detection presented here would be worth considering for our project. In my opinion this is especially true for the TaqMan[®]/molecular beacon approach, as well as the usage of DNA chip technology, since they are amenable for the very high throughput of samples, and also since it was shown that the PCR used here is remarkably stable. Of course a major disadvantage is the fact that these techniques are not suitable for the detection of unknown mutations. To search for these probably CDCE should be worth to be (re)considered, even though to achieve high throughput with this technique one would probably have to rely on a commercially available multiarray machine because otherwise the assay gets too tedious.

4.2. And still not all converted yet!

We look back on more than 60 years of research on the major histocompatibility complex (reviewed in (Klein 1986)), and about 20 years passed by since the gene conversion hypothesis of MHC evolution loiters the scientific world (see introduction).

It is no longer necessary to convince anyone about the occurrence of gene conversion in the mammalian genome (see Schimenti 1994), and it seems to me that this holds true for the MHC, even in the community of people being interested in MHC evolutionⁱ. This in combination with the absence of sufficient new and exciting data about the matter is probably responsible for a certain lack of emotion and productivity in the field, which is in sharp contrast to the lively debates of the early days (Klein 1986; Pease 1986), a “Brouhaha” as Lawlor and his colleagues put it once nicely (Lawlor *et al.* 1990).

This is astonishing, since none of the questions raised in the introduction to this thesis is answered conclusively yet:

- With what frequency occur gene conversion events finally in the mouse MHC germline genes?
- Are there differences between the inbred mouse strains (and outbred wildtype mice!)?
- At which time point of the development is the onset of gene conversion?
- Is there a sex bias of gene conversion?
- Are gene conversion events preferentially targeted to certain regions within MHC genes?
- Is gene conversion directional with regard to usage of donor and acceptor sequences?

And ultimately:

- Does the “strong hypothesis of the gene conversion hypothesis of MHC evolution” apply? - is gene conversion an adaptive mutator under selection, or is it just a process which happens due to the organization of the MHC, with no special adaptive significance?

On the other hand this calm might just reflect that many scientists felt overwhelmed by the obstacles that the exploration of this phenomenon poses. Perhaps now, at times when high throughput screening methods and automation have become very fashionable, there is another chance to push the MHC-gene conversion question back, closer into the sphere of interest of the scientific community. Maybe these new techniques will initiate a *renaissance* of research on gene conversion in the MHC of mammals. It would be a pity to miss this chance, since

ⁱ An exception to this might be found in some of the pioneers of the “dawning days” of MHC research who, probably also for nostalgic reasons still can get quite excited about the sloppy usage of the term “gene conversion” in this context... (Walter Bodmer – personal communication at the VIIIth CEPH Annual Conference (2000) in Paris)

clearly the question of gene conversion taking place in the MHC of mammals potentially harbours answers (and questions!) which imply a wide radius of scientific fields of biological and philosophical nature. Hopefully soon the gene conversion hypothesis of MHC evolution will step out of its undeserved wall-flower existence. - And who knows may be one day one could say: It was a long and winding road, but it was worth it....

5. REFERENCES

- Ahmadian, A., B. Gharizadeh, et al. (2000). "Single-nucleotide polymorphism analysis by pyrosequencing." *Anal Biochem* **280**(1): 103-10.
- Alberts, B., D. Bray, et al. (1994). *Molecular Biology of the Cell*. New York, Garland Publishing.
- Andre, P., A. Kim, et al. (1997). "Fidelity and mutational spectrum of Pfu DNA polymerase on a human mitochondrial DNA sequence." *Genome Res* **7**(8): 843-52.
- Ausubel, F. M., R. Brent, et al., Eds. (1998). *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.
- Babon, J. J., M. McKenzie, et al. (1999). "Mutation detection using fluorescent enzyme mismatch cleavage with T4 endonuclease VII." *Electrophoresis* **20**(6): 1162-70.
- Babon, J. J., M. McKenzie, et al. (2000). "The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection." *Methods Mol Biol* **152**: 187-99.
- Bailey, D. W. (1966). "Heritable histocompatibility changes: lysogeny in mice?" *Transplantation* **4**(4): 482-8.
- Bailey, D. W. and H. I. Kohn (1965). "Inherited histocompatibility changes in progeny of irradiated and unirradiated inbred mice." *Genet Res* **6**(3): 330-40.
- Bailey, D. W. and B. Usama (1960). "A rapid method of grafting skin on tails of mice." *Transplantation Bulletin* **7**: 424-425.
- Baner, J., M. Nilsson, et al. (1998). "Signal amplification of padlock probes by rolling circle replication." *Nucleic Acids Res* **26**(22): 5073-8.
- Barkley, M. D., A. D. Riggs, et al. (1975). "Interaction of Effecting Ligands with Lac Repressor and Repressor-Operator Complex." *Biochemistry* **14**(8): 1700-1710.
- Becker, R. S. and K. L. Knight (1990). "Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits." *Cell* **63**(5): 987-97.
- Bergström, T. F., A. Josefsson, et al. (1998). "Recent origin of HLA-DRB1 alleles and implications for human evolution." *Nature Genetics* **18**: 237-242.
- Beynon, R. J. and J. S. Easterby (1996). *Buffer Solutions -the basics-*. Oxford, IRL Press.
- Bezzubova, O. Y. and J. M. Buerstedde (1994). "Gene conversion in the chicken immunoglobulin locus: a paradigm of homologous recombination in higher eukaryotes." *Experientia* **50**(3): 270-6.
- Biotechnology, P. L. (1995). "Micropurification of DNA-binding proteins by DNA-affinity chromatography (NHS-columns)." *Technical Note SMART™ System*.
- Bjorkman, P. J. and P. Parham (1990). "Structure, function, and diversity of class I major histocompatibility complex molecules." *Annu Rev Biochem* **59**: 253-88.
- Boerigter, M. E. T. I., M. E. T. Dolle, et al. (1995). "Plasmid-based transgenic mouse model for studying in vivo mutations." *Nature* **377**: 657-659.
- Bollag, R. J. and R. M. Liskay (1988). "Conservative intrachromosomal recombination between inverted repeats in mouse cells: association between reciprocal exchange and gene conversion." *Genetics* **119**(1): 161-9.
- Bollag, R. J., A. S. Waldman, et al. (1989). "Homologous recombination in mammalian cells." *Annu Rev Genet* **23**: 199-225.
- Bornes, M., J. Cline, et al. (1998). "Comparing PCR Performance and Fidelity of Commercial Pfu DNA Polymerases." *Strategies* **12**(1): 40-42.
- Brake, A. J., A. V. Fowler, et al. (1978). "β-Galactosidase chimeras: Primary structure of a lac repressor-β-galactosidase protein." *Proceedings of the National Academy of Sciences* **75**(10): 4824-4827.
- Brenneman, M., F. S. Gimble, et al. (1996). "Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonucleases." *Proc Natl Acad Sci U S A* **93**(8): 3608-12.
- Briles, W. E., R. W. Briles, et al. (1983). "Resistance to a malignant lymphoma in chickens is mapped to subregion of major histocompatibility (B) complex." *Science* **219**(4587): 977-9.
- Briles, W. E., H. A. Stone, et al. (1977). "Marek's disease: effects of B histocompatibility alloalleles in resistant and susceptible chicken lines." *Science* **195**(4274): 193-5.
- Bruand, C. and S. D. Ehrlich (2000). "UvrD-dependent replication of rolling-circle plasmids in Escherichia coli." *Mol Microbiol* **35**(1): 204-10.
- Buetow, K. H., M. Edmonson, et al. (2001). "High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip- based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry." *Proc Natl Acad Sci U S A* **98**(2): 581-4.
- Busch, R., R. C. Doebele, et al. (2000). "Accessory molecules for MHC class II peptide loading." *Curr Opin Immunol* **12**(1): 99-106.
- Butler, J. E., J. Sun, et al. (1996). "The VH and CH immunoglobulin genes of swine: implications for repertoire development." *Vet Immunol Immunopathol* **54**(1-4): 7-17.
- Buus, S. (1999). "Description and prediction of peptide-MHC binding: the 'human MHC project'." *Curr Opin Immunol* **11**(2): 209-13.
- Cadavid, L. F. and D. I. Watkins (1997). "Heirs of the jaguar and the anaconda: HLA, conquest and disease in the indigenous populations of the Americas [corrected and republished article originally printed in Tissue Antigens 1997 Sep;50(3):209-18]." *Tissue Antigens* **50**(6): 702-11.
- Carrington, M., G. W. Nelson, et al. (1999). "HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage." *Science* **283**: 1748-1752.
- Cha, R. S., H. Zarbl, et al. (1992). "Mismatch amplification mutation assay (MAMA): application to the c-H- ras gene." *PCR Methods Appl* **2**(1): 14-20.
- Cheung, V. G., M. Morley, et al. (1999). "Making and reading microarrays." *Nat Genet* **21**(1 Suppl): 15-9.
- Cleaver, J. E. (1994). "It Was a Very Good Year for DNA Repair." *Cell* **76**: 1-4.
- Cline, J., J. C. Braman, et al. (1996). "PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases." *Nucleic Acids Res* **15**(24): 3546-51.

- Coligan, J. E., B. M. Dunn, et al., Eds. (1997). Current Protocols in Protein Science, John Wiley Sons, Inc.
- consortium, T. M. s. (1999). "Complete sequence and gene map of a human major histocompatibility complex." Nature **401**(6756): 921-3.
- Cotton, R. G. (1999). "Mutation detection by chemical cleavage." Genet Anal **14**(5-6): 165-8.
- Cotton, R. G. H., E. Edkins, et al. (1998). Mutation Detection. Oxford, Oxford University Press.
- Cummings, W. J. and M. E. Zolan (1998). "Functions of DNA repair genes during meiosis." Curr Top Dev Biol **37**: 117-40.
- de Wind, N., M. Dekker, et al. (1995). "Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer." Cell **82**(2): 321-30.
- Diaz, M. and M. F. Flajnik (1998). "Evolution of somatic hypermutation and gene conversion in adaptive immunity." Immunol Rev **162**: 13-24.
- Dieffenbach, C. W. and G. S. Dveksler, Eds. (1995). PCR Primer, Cold Spring Harbor Laboratory Press.
- Dolle, M. E. T., H. J. Martus, et al. (1996). "Evaluation of a Plasmid-Based Transgenic Mouse Model For Detecting In-Vivo Mutations." Mutagenesis **11**(1): 111-118.
- Donoho, G., M. Jasin, et al. (1998). "Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells." Mol Cell Biol **18**(7): 4070-8.
- Dubendorff, J. W. and F. W. Studier (1991). "Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor." J Mol Biol **5**(1): 45-59.
- Egorov, O. S. and I. K. Egorov (1984). "H-2^{bm23}, a new K^b mutant similar to, but not identical with H-2^{bm3}." Immunogenetics **20**: 83-87.
- Eismann, E. K. (1989). Untersuchungen zur Wirkungsweise des Lacrepressors bei der Operatorerkennung und über große Distanzen. Institut für Genetik. Köln, Universität zu Köln.
- Elliott, B., C. Richardson, et al. (1998). "Gene conversion tracts from double-strand break repair in mammalian cells." Mol Cell Biol **18**(1): 93-101.
- Evans, G. A., D. H. Margulies, et al. (1982). "Structure and expression of a mouse major histocompatibility antigen gene, H-2Ld." Proc Natl Acad Sci U S A **79**(6): 1994-8.
- Felley-Bosco, E., C. Pourzand, et al. (1991). "A genotypic mutation system measuring mutations in restriction recognition sequences." Nucleic Acids Res **19**(11): 2913-9.
- Fickert, R. and B. Müller-Hill (1992). "How Lac Repressor Finds lac Operator in vitro." Journal of Molecular Biology **226**: 59-68.
- Fischer Lindahl, K. (1997). "On naming H2 haplotypes: functional significance of MHC class Ib alleles." Immunogenetics **46**(1): 53-62.
- Fischer, S. G. and L. S. Lerman (1983). "DNA Fragments differing by single base-pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory." Proceedings of the National academy of Sciences USA **80**(6): 1579-1583.
- Flaherty, L. (1988). "Major histocompatibility complex polymorphism: a nonimmune theory for selection." Hum Immunol **21**(1): 3-13.
- Fried, M. and D. M. Crothers (1981). "Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis." Nucleic Acids research **9**(23): 6505-6525.
- Friedberg, E. C., G. C. Walker, et al. (1995). DNA Repair and Mutagenesis. Washington D.C., ASM Press.
- Gachelin, G., B. Dumas, et al. (1982). "Mouse genes coding for the major class I transplantation antigens: a mosaic structure might be related to the antigenic polymorphism." Ann Immunol (Paris) **133C**(1): 3-20.
- Gao, Q. and E. S. Yeung (2000). "High-throughput detection of unknown mutations by using multiplexed capillary electrophoresis with poly(vinylpyrrolidone) solution." Anal Chem **72**(11): 2499-506.
- Garcia, C. A., A. Ahmadian, et al. (2000). "Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene." Gene **253**(2): 249-57.
- Gaur, L. K. and G. T. Nepom (1996). "Ancestral major histocompatibility complex DRB genes beget conserved patterns of localized polymorphisms." Proc Natl Acad Sci U S A **93**(11): 5380-3.
- Geliebter, J. and S. G. Nathenson (1987). "Recombination and the concerted evolution of the murine MHC." TIG **3**(4): 107-112.
- Geliebter, J., R. A. Zeff, et al. (1986). "Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: Kbm9 and Kbm6." Proc Natl Acad Sci U S A **83**(10): 3371-5.
- Geliebter, J., R. A. Zeff, et al. (1986). "Interaction between Kb and Q4 gene sequences generates the Kbm6 mutation." Mol Cell Biol **6**(2): 645-52.
- Gilbert, W. and A. Maxam (1973). "The Nucleotide Sequence of the lac Operator." Proceedings of the National Academy of Sciences **70**(12): 3581-3584.
- Gilbert, W. and B. Müller-Hill (1967). "The lac Operator is DNA." Proceedings of the National Academy of Sciences **58**: 2415-2421.
- Gilbertson, L. A. and F. W. Stahl (1996). "A test of the double-strand break repair model for meiotic recombination in *Saccharomyces cerevisiae*." Genetics **144**(1): 27-41.
- Ginsburg, M., M. H. Snow, et al. (1990). "Primordial germ cells in the mouse embryo during gastrulation." Development **110**(2): 521-8.
- Gobin, S. J. and P. J. van den Elsen (2000). "Transcriptional regulation of the MHC class Ib genes HLA-E, HLA-F, and HLA-G." Hum Immunol **61**(11): 1102-7.
- Golz, S. and B. Kemper (1999). "Association of holliday-structure resolving endonuclease VII with gp20 from the packaging machine of phage T4." J Mol Biol **285**(3): 1131-44.
- Gossen, J. A., W. J. F. d. Leeuw, et al. (1993). "Plasmid Rescue from Transgenic Mouse DNA Using LacI Repressor Protein conjugated to magnetic beads." BioTechniques **14**(4): 624-629.
- Gossen, J. A., W. J. F. d. Leeuw, et al. (1994). "LacZ transgenic mouse models: their application in genetic toxicology." Mutation Research **307**: 451-459.
- Gossen, J. A., H.-J. Martus, et al. (1995). "Spontaneous and X-ray-induced deletion mutations in a LacZ plasmid-based transgenic mouse model." Mutation Research **331**: 89-97.
- Gruen, J. and W. SM (1997). "Evolving views of the major histocompatibility complex." Blood **90**(11): 4252-65.
- Gunning, P., J. Leavitt, et al. (1987). "A human beta-actin expression vector system directs high-level accumulation of antisense transcripts." Proc Natl Acad Sci U S A **84**(14): 4831-4835.

- Hacia, J. G. (1999). "Resequencing and mutational analysis using oligonucleotide microarrays." *Nat Genet* **21**(1 Suppl): 42-7.
- Hacia, J. G. and F. S. Collins (1999). "Mutational analysis using oligonucleotide microarrays." *J Med Genet* **36**(10): 730-6.
- Haldane, J. B. S. (1949). "Disease and Evolution." *La Ricerca Scientifica* **19 Supplemento Anno 19**: 68-75.
- Hamilton, M. S. and I. Hellstrom (1978). "Selection for histoincompatible progeny in mice." *Biol Reprod* **19**(2): 267-70.
- Hanneman, W. H., K. J. Schimenti, et al. (1997). "Molecular analysis of gene conversion in spermatids from transgenic mice." *Gene* **200**(1-2): 185-92.
- Haring, V., J. E. Gray, et al. (1990). "Self-incompatibility: a self-recognition system in plants." *Science* **250**(4983): 937-41.
- Hatch, A., T. Sano, et al. (1999). "Rolling circle amplification of DNA immobilized on solid surfaces and its application to multiplex mutation detection." *Genet Anal* **15**(2): 35-40.
- Hayashi, K. and D. W. Yandell (1993). "How sensitive is PCR-SSCP?" *Hum Mutat* **2**(5): 338-46.
- Hervieu, G. (1997). A quick and safe method for destaining Coomassie-Blue-stained protein gels, Elsevier Trends Journals Technical Tips Online. **1997**.
- Hickson, R. E. and R. L. Cann (1997). "Mhc allelic diversity and modern human origins." *J Mol Evol* **45**(6): 589-98.
- Hill, A. V. S., C. E. M. Allsopp, et al. (1991). "Common West African HLA antigens are associated with protection from severe malaria." *Nature* **352**: 595-600.
- Hippel, P. H. v., A. Revzin, et al. (1974). "Non-specific Binding of genome Regulating Proteins as a Biological Control Mechanism: 1. The lac Operon: Equilibrium Aspects." *Proceedings of the National Academy of Sciences* **71**(12): 4808-4812.
- Ho, H. N., Y. S. Yang, et al. (1994). "Sharing of human leukocyte antigens in couples with unexplained infertility affects the success of in vitro fertilization and tubal embryo transfer." *Am J Obstet Gynecol* **170**(1 Pt 1): 63-71.
- Hogan, B., R. beddington, et al. (1994). *Manipulating the Mouse Embryo*. New York, Cold Spring Harbour Laboratory Press.
- Hogan, M., D. Wemmer, et al. (1980). "On the Interaction of the lac Repressor Headpiece with Nucleic Acids." *FEBS LETTERS* **124**(2): 202-203.
- Hogstrand, K. and J. Bohme (1994). "A determination of the frequency of gene conversion in unmanipulated mouse sperm." *Proc Natl Acad Sci U S A* **91**(21): 9921-5.
- Hogstrand, K. and J. Bohme (1997). "Gene Conversion of Major Histocompatibility Complex Genes in the Mouse Spermatogenesis is a Premeiotic Event." *Mol Biol Cell* **8**(12): 2511-7.
- Hogstrand, K. and J. Bohme (1999). "DNA damage caused by etoposide and gamma-irradiation induces gene conversion of the MHC in a mouse non-germline testis cell line." *Mutat Res* **423**(1-2): 155-69.
- Holland, J. H. (1992). "Genetic Algorithms." *Scientific American* **267**: 44-50.
- Holland, P. M., R. D. Abramson, et al. (1991). "Detection of specific polymerase chain reaction products by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase." *PNAS* **88**: 7276-7280.
- Holliday, R. (1964). "A mechanism for gene conversion in fungi." *Genetic Research* **5**: 282-304.
- Horton, R. M., B. E. Loveland, et al. (1991). "Characterization of the spontaneous mutant H-2K^{bm29} indicates that gene conversion in H-2 occurs at a higher frequency than detected by skin grafting." *J. Immunol.* **147**: 3180-3184.
- Howard, J. (1995). "Not all converted yet [news; comment]." *Nat Genet* **10**(4): 371-3.
- Huang, M. M., H. A. Erlich, et al. (1995). "Analysis of mutational changes at the HLA locus in single human sperm." *Hum Mutat* **6**(4): 303-10.
- Hughes, A. L. and M. Nei (1989). "Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection." *Proceedings of the National Academy of Sciences USA* **86**: 958-962.
- Hughes, A. L. and M. Yeager (1998). "Natural selection at major histocompatibility complex loci of vertebrates." *Annu Rev Genet* **32**: 415-35.
- Jasin, M. (1996). "Genetic manipulation of genomes with rare-cutting endonucleases." *Trends Genet* **12**(6): 224-8.
- Kamashev, D. E., K. K. Ebralidze, et al. (1994). "Electrostatic Interactions and Differences in the Specific and Unspecific Interactions of the Lac Repressor with DNA." *Biophysics* **39**(6): 1169-1174.
- Kania, J. (1976). Proteinchemische Untersuchungen an lac-Repressor- β -Galactosidase-Chimären aus *E.coli*. *Institut für Genetik*. Köln, Universität zu Köln: 110.
- Kania, J. and D. T. Brown (1976). "The functional repressor parts of a tetrameric lac repressor- β -galactosidase chimaera are organized as dimers." *Proceedings of the National Academy of Sciences of the United States of America* **73**(10): 3529-3533.
- Kania, J. and B. Müller-Hill (1974). "Lac repressor can be fused to β -galactosidase." *Nature* **249**: 561-563.
- Kemper, B. and J. Hurwitz (1973). "Studies on T4-induced nucleases. Isolation and characterization of a manganese-activated T4-induced endonuclease." *J Biol Chem* **248**(1): 91-9.
- Khan, S. A. (2000). "Plasmid rolling-circle replication: recent developments." *Mol Microbiol* **37**(3): 477-84.
- Kheterpal, I. and R. A. Mathies (1999). "Capillary array electrophoresis DNA sequencing." *Anal Chem* **71**(1): 31A-37A.
- Kleff, S. and B. Kemper (1988). "Initiation of heteroduplex-loop repair by T4-encoded endonuclease VII in vitro." *Embo J* **7**(5): 1527-35.
- Klein, J. (1978). "H-2 Mutations: Their Genetics and Effect on Immune Functions." *Advances in Immunology* **26**(4380): 55-146.
- Klein, J. (1986). "The continuation of the debate on gene conversion." *Transplantation* **42**(1): 102-5.
- Klein, J. (1986). *Natural History of the Major Histocompatibility Complex*. New York, John Wiley & Sons.
- Klein, J. (1987). "Origin of major histocompatibility complex polymorphism: the trans- species hypothesis." *Hum Immunol* **19**(3): 155-62.
- Klein, J. and F. Figuera (1986). "Evolution of the major histocompatibility complex." *Critical Reviews in Immunology* **6**(4): 295-375.
- Klein, J., Y. Satta, et al. (1993). "Trans-specific Mhc polymorphism and the origin of species in primates." *J Med Primatol* **22**(1): 57-64.
- Klug, A., A. Jack, et al. (1979). "A Hypothesis on a Specific Sequence-dependent Conformation of DNA and its relation to the Binding of the lac-repressor Protein." *Journal of Molecular Biology* **131**(669-680).
- Koshland, D. E. (1994). "Molecule of the Year: The DNA Repair Enzyme." *Science* **266**: 1925.
- Koshland, D. E. and E. Culotta (1994). "DNA Repair Works Its Way to the Top." *Science* **266**: 1926-1929.

- Kuhn, R., K. Rajewsky, et al. (1991). "Generation and analysis of interleukin-4 deficient mice." *Science* **254**(5032): 707-10.
- Kunieda, T. and Y. Toyoda (1992). "Nucleotide sequence of mouse Sry gene is different between Y chromosome originating from *Mus musculus musculus* and *Mus musculus domesticus*." *Genomics* **13**(1): 236-7.
- Kurian, K. M., C. J. Watson, et al. (1999). "DNA chip technology." *J Pathol* **187**(3): 267-71.
- Landers, J. P. (1993). *Handbook of Capillary Electrophoresis*. London, CRC Press.
- Larsen, L. A., M. Christiansen, et al. (2000). "High throughput mutation screening by automated capillary electrophoresis." *Comb Chem High Throughput Screen* **3**(5): 393-409.
- Larson, C. J. and G. L. Verdin (1992). "A high-capacity column for affinity purification of sequence-specific DNA-binding proteins." *Nucleic Acids Research* **20**(13): 3525.
- Lawlor, D. A., J. Zemmour, et al. (1990). "Evolution of class-I MHC genes and proteins: from natural selection to thymic selection." *Annu Rev Immunol* **8**(9): 23-63.
- Lehming, N., J. Sartorius, et al. (1987). "The interaction of the recognition helix of lac repressor with lac operator." *The EMBO Journal* **6**(10): 3145-3153.
- Lerman, L. S. and C. Beldjord (1998). *Comprehensive Mutation Detection with Denaturing Gradient Gel Electrophoresis*. R. G. H. Cotton, E. Edkins and S. Forrest. Oxford, Oxford University Press: 35-61.
- Lewis, M., G. Chang, et al. (1996). "Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer." *Science* **271**: 1247-1254.
- Liang, F., M. Han, et al. (1998). "Homology-directed repair is a major double-strand break repair pathway in mammalian cells." *Proc Natl Acad Sci U S A* **95**(9): 5172-7.
- Liang, F., P. J. Romanienko, et al. (1996). "Chromosomal double-strand break repair in Ku80-deficient cells." *Proc Natl Acad Sci U S A* **93**(17): 8929-33.
- Lienert, K. and P. Parham (1996). "Evolution of MHC class I genes in higher primates." *Immunol Cell Biol* **74**(4): 349-56.
- Lin, S. and A. D. Riggs (1975). "The General Affinity of lac Repressor for E.coli DNA: Implications for Gene Regulations in Prokaryotes and Eucaryotes." *Cell* **4**: 107.
- Lin, S.-Y. and A. D. Riggs (1972). "lac Repressor binding to Non-operator DNA: Detailed Studies and a Comparison of Equilibrium and Rate Competition Methods." *Journal of Molecular Biology* **72**: 671-690.
- Liskay, R. M. and J. L. Stachelek (1986). "Information transfer between duplicated chromosomal sequences in mammalian cells involves contiguous regions of DNA." *Proc Natl Acad Sci U S A* **83**(6): 1802-6.
- Liskay, R. M., J. L. Stachelek, et al. (1984). "Homologous recombination between repeated chromosomal sequences in mouse cells." *Cold Spring Harb Symp Quant Biol* **49**(1): 183-9.
- Lizardi, P. M., X. Huang, et al. (1998). "Mutation detection and single-molecule counting using isothermal rolling-circle amplification." *Nat Genet* **19**(3): 225-32.
- Loh, D. Y. and D. Baltimore (1984). "Sexual preference of apparent gene conversion events in MHC genes of mice." *Nature* **309**: 639-640.
- Lorenzi, R. (1994). *Studies on gene conversion as a mutational mechanism in the evolution of Major Histocompatibility Complex genes*. Department of Immunology, The Babraham Institute. Cambridge, St. John's College: 165.
- Lu, C. and H. P. Erickson (1997). "Expression in *Escherichia coli* of the thermostable DNA polymerase from *Pyrococcus furiosus*." *Protein Expr Purif* **11**(2): 179-84.
- Lundberg, A. S. and H. O. McDevitt (1992). "Evolution of major histocompatibility complex class II allelic diversity: direct descent in mice and humans." *Proc Natl Acad Sci U S A* **89**(14): 6545-9.
- Maniatis, T., E. F. Fritsch, et al. (1989). *Molecular Cloning*. Cold Spring Harbor, Cold spring Harbor laboratory press.
- Martinsohn, J. T. (1996). *Analyse von Genkonversionsereignissen im MHC I der Maus -Entwicklung einer Methode zur spezifischen Isolierung von Genen-*. Institut für Genetik. Köln, Universität zu Köln: 162.
- Martinsohn, J. T., A. B. Sousa, et al. (1999). "The gene conversion hypothesis of MHC evolution: a review." *Immunogenetics* **50**(3-4): 168-200.
- McConnell, T. J., W. S. Talbot, et al. (1988). "The origin of MHC class II gene polymorphism within the genus *Mus*." *Nature* **332**(6165): 651-4.
- McCormack, W. T., L. W. Tjoelker, et al. (1991). "Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion." *Annu Rev Immunol* **9**: 219-41.
- McKenzie, I. F., G. M. Morgan, et al. (1979). "B6.C-H-2bm12. A new H-2 mutation in the I region in the mouse." *J Exp Med* **150**(6): 1323-38.
- Mellor, A. L., E. H. Weiss, et al. (1983). "A potential donor gene for the bm1 gene conversion event in the C57BL mouse." *Nature* **306**(5945): 792-5.
- Melvold, R. W., H. I. Kohn, et al. (1982). "History and genealogy of the H-2Kb mutants from the C57BL/6Kh colony." *Immunogenetics* **15**(2): 177-85.
- Mitchelson, K. R. (2001). "The application of capillary electrophoresis for DNA polymorphism analysis." *Methods Mol Biol* **162-163**: 3-26.
- Moonsamy, P. V., V. C. Suraj, et al. (1992). "Genetic diversity within the HLA class II region: ten new DPB1 alleles and their population distribution." *Tissue Antigens* **40**: 153-157.
- Moynahan, M. E. and M. Jasin (1997). "Loss of heterozygosity induced by a chromosomal double-strand break." *Proc Natl Acad Sci U S A* **94**(17): 8988-93.
- Murti, J. R., M. Bumbulis, et al. (1992). "High-frequency germ line gene conversion in transgenic mice." *Mol Cell Biol* **12**(6): 2545-52.
- Nataraj, A. J., I. Olivos-Glander, et al. (1999). "Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection." *Electrophoresis* **20**(6): 1177-85.
- Natarajan, K., H. Li, et al. (1999). "MHC class I molecules, structure and function." *Rev Immunogenet* **1**(1): 32-46.
- Nelson, C. A. and D. H. Fremont (1999). "Structural principles of MHC class II antigen presentation." *Rev Immunogenet* **1**(1): 47-59.
- Nordstrom, T., K. Nourizad, et al. (2000). "Method enabling pyrosequencing on double-stranded DNA." *Anal Biochem* **282**(2): 186-93.
- Novick, R. P. (1998). "Contrasting lifestyles of rolling-circle phages and plasmids." *Trends Biochem Sci* **23**(11): 434-8.
- Nyren, P. (1987). "Enzymatic Method for Continuous Monitoring of DNA Polymerase Activity." *Analytical Biochemistry* **167**: 235-238.

- Ober, C., L. R. Weitkamp, et al. (1997). "HLA and mate choice in humans." *Am J Hum Genet* **61**(3): 497-504.
- Oehler, S., E. R. Eismann, et al. (1990). "The three operators of the lac operon cooperate in repression." *The EMBO Journal* **9**(4): 973-979.
- Orr-Weaver, T. L. and J. W. Szostak (1985). "Fungal Recombination." *Microbiological Reviews* **49**(1): 33-58.
- Parham, P., E. J. Adams, et al. (1995). "The origins of HLA-A,B,C polymorphism." *Immunological Reviews* **143**: 141-180.
- Parham, P., R. J. Benjamin, et al. (1989). "Diversity of class I HLA molecules: functional and evolutionary interactions with T cells." *Cold Spring Harbor Symp. Quant. Biol.* **54**: 529-543.
- Parham, P. and D. A. Lawlor (1991). "Evolution of class I major histocompatibility complex genes and molecules in humans and apes." *Hum Immunol* **30**(2): 119-28.
- Parham, P., D. A. Lawlor, et al. (1989). "Diversity and diversification of HLA-A,B,C alleles." *J Immunol* **142**(11): 3937-50.
- Parham, P. and T. Ohta (1996). "Population biology of antigen presentation by MHC class I molecules [see comments]." *Science* **272**(5258): 67-74.
- Parng, C. L., S. Hansal, et al. (1996). "Gene conversion contributes to Ig light chain diversity in cattle." *J Immunol* **157**(12): 5478-86.
- Parsons, B. L. and R. H. Heflich (1997). "Genotypic selection methods for the direct analysis of point mutations." *Mutat Res* **387**(2): 97-121.
- Parsons, B. L. and R. H. Heflich (1998). "Detection of a mouse H-ras codon 61 mutation using a modified allele-specific competitive blocker PCR genotypic selection method." *Mutagenesis* **13**(6): 581-8.
- Parsons, B. L. and R. H. Heflich (1998). "Detection of basepair substitution mutation at a frequency of 1×10^{-7} by combining two genotypic selection methods, MutEx enrichment and allele-specific competitive blocker PCR." *Environmental and Molecular Mutagenesis* **32**(3): 200-211.
- Paterson, S., K. Wilson, et al. (1998). "Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population." *Proc Natl Acad Sci U S A* **95**(7): 3714-9.
- Pease, L. (1986). "Answer to 'The Continuation of the Debate on Gene Conversion' from J. Klein." *Transplantation* **42**(1): 104-105.
- Pease, L. (1986). "The continuation of the debate on gene conversion (answer to J. Klein)." *Transplantation* **42**(1): 102-5.
- Pease, L. R., R. M. Horton, et al. (1991). "Structure and diversity of class I antigen presenting molecules in the mouse." *Crit Rev Immunol* **11**(1): 1-32.
- Pease, L. R., R. M. Horton, et al. (1993). "Unusual mutation clusters provide insight into class I gene conversion mechanisms." *Mol Cell Biol* **13**(7): 4374-81.
- Penn, D. and W. Potts (1998). "The Evolution of Mating Preferences and Major Histocompatibility Complex Genes." *The American Naturalist* **153**: 145-164.
- Pepling, M. E., M. de Cuevas, et al. (1999). "Germline cysts: a conserved phase of germ cell development? [In Process Citation]." *Trends Cell Biol* **9**(7): 257-62.
- Pepling, M. E. and A. C. Spradling (1998). "Female mouse germ cells form synchronously dividing cysts." *Development* **125**(17): 3323-8.
- Pesce, M. and M. De Felici (1995). "Purification of mouse primordial germ cells by MiniMACS magnetic separation system." *Dev Biol* **170**(2): 722-5.
- Pfahl, M., V. Gulde, et al. (1978). "'Second' and 'Third Operator' of the lac Operon: An Investigation of their Role in the Regulatory Mechanism." *Journal of Molecular Biology* **127**: 339-344.
- Pittman, D. L. and J. C. Schimenti (1998). "Recombination in the mammalian germ line." *Curr Top Dev Biol* **37**: 1-35.
- Potter, D. A., C. J. Larson, et al. (1993). "Purification of the major histocompatibility complex class I transcription factor H2TF1." *J Biol Chem* **268**(25): 18882-90.
- Potts, W. K., C. J. Manning, et al. (1991). "Mating patterns in seminatural populations of mice influenced by MHC genotype." *Nature* **352**: 619-621.
- Potts, W. K. and P. R. Slev (1995). "Pathogen-Based Models Favoring MHC Genetic Diversity." *Immunological Reviews* **143**: 181-192.
- Potts, W. K. and E. K. Wakeland (1993). "Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference." *Trends Genet* **9**(12): 408-12.
- Radman, M., I. Matic, et al. (1999). "Evolution of evolvability." *Ann N Y Acad Sci* **870**: 146-55.
- Ratcliffe, M. J. and K. A. Jacobsen (1994). "Rearrangement of immunoglobulin genes in chicken B cell development." *Semin Immunol* **6**(3): 175-84.
- Ren, J. (2000). "High-throughput screening of genetic mutations/polymorphisms by capillary electrophoresis." *Comb Chem High Throughput Screen* **3**(1): 11-25.
- Reynaud, C. A., V. Anquez, et al. (1985). "A single rearrangement event generates most of the chicken immunoglobulin light chain diversity." *Cell* **40**(2): 283-91.
- Reynaud, C. A., V. Anquez, et al. (1987). "A hyperconversion mechanism generates the chicken light chain preimmune repertoire." *Cell* **48**(3): 379-88.
- Reynaud, C. A., A. Dahan, et al. (1989). "Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region." *Cell* **59**(1): 171-83.
- Richardson, C., M. E. Moynahan, et al. (1998). "Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations." *Genes Dev* **12**(24): 3831-42.
- Riggs, A. D., S. Bourgeois, et al. (1970). "The lac Repressor-Operator Interaction III. Kinetic Studies." *Journal of Molecular Biology* **53**: 401-417.
- Riggs, A. D., S. Lin, et al. (1972). "Lac Repressor Binding to Synthetic DNAs of Defined Nucleotide Sequence." *Proceedings of the national Academy of Sciences USA* **69**(3): 761-764.
- Riggs, A. D., R. F. Newby, et al. (1970). "lac Repressor-Operator Interaction II. Effect of Galactosides and other Ligands." *Journal of Molecular Biology* **51**: 303-314.
- Ronaghi, M. (2000). "Improved performance of pyrosequencing using single-stranded DNA-binding protein." *Anal Biochem* **286**(2): 282-8.

- Ronaghi, M. (2001). "Pyrosequencing sheds light on DNA sequencing." *Genome Res* **11**(1): 3-11.
- Ronaghi, M., S. Karamohamed, et al. (1996). "Real-time DNA sequencing using detection of pyrophosphate release." *Anal Biochem* **242**(1): 84-9.
- Rosenberg, S. M. (1997). "Mutation for survival." *Curr Opin Genet Dev* **7**(6): 829-34.
- Russel, L. B. and W. L. Russel (1992). "Frequency and nature of specific-locus mutations induced in female mice by radiations and chemicals: a review." *Mutation Research* **296**: 107-127.
- Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." *Proc Natl Acad Sci U S A* **74**: 5463-7.
- Sargent, R. G., M. A. Brennehan, et al. (1997). "Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination." *Mol Cell Biol* **17**(1): 267-77.
- Satta, Y., C. O'HUigin, et al. (1993). "The synonymous substitution rate of the major histocompatibility complex loci in primates." *Proc Natl Acad Sci U S A* **90**(16): 7480-4.
- Schimenti, J. C. (1994). "Gene conversion and the evolution of gene families in mammals." *Soc Gen Physiol Ser* **49**(2): 85-91.
- Schimenti, J. C. (1999). "Mice and the role of unequal recombination in gene-family evolution." *Am J Hum Genet* **64**(1): 40-5.
- Schlager, G. (1972). "Natural mutation rates in the house mouse: estimates for recessive visible mutations." *Mutat Res* **14**(2): 254-8.
- Schlager, G. and M. M. Dickie (1971). "Natural mutation rates in the house mouse. Estimates for five specific loci and dominant mutations." *Mutat Res* **11**(1): 89-96.
- Schwacha, A. and N. Kleckner (1995). "Identification of double Holliday junctions as intermediates in meiotic recombination." *Cell* **83**(5): 783-91.
- Schweitzer, B. and S. Kingsmore (2001). "Combining nucleic acid amplification and detection." *Curr Opin Biotechnol* **12**(1): 21-7.
- Scofield, V. L., J. M. Schlumpberger, et al. (1982). "Protochordate allorecognition is controlled by a MHC-like gene system." *Nature* **295**(5849): 499-502.
- Sette, A. and J. Sidney (1999). "Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism." *Immunogenetics* **50**(3-4): 201-12.
- Skopek, T. R. (1991). "Analysis of Mutations Using PCR and Denaturing Gradient Gel Electrophoresis." *Environmental and Molecular Mutagenesis* **18**: 249-254.
- Smith, J. and P. Modrich (1996). "Mutation detection with MthH, MutL, and MutS mismatch repair proteins." *Proc Natl Acad Sci U S A* **93**(9): 4374-9.
- Snell, G. D. (1986). "Some recollections of Peter Gorer and his work on this fiftieth anniversary of his discovery of H-2." *Immunogenetics* **24**(6): 339-40.
- Sokal, R. R. and F. J. Rohlf (1998). *Biometry*. New York, W.H. Freeman and Company.
- Solaro, P. C., K. Birkenkamp, et al. (1993). "Endonuclease VII of phage T4 triggers mismatch correction in vitro." *J Mol Biol* **230**(3): 868-77.
- Soloski, M. J., M. E. Szperka, et al. (2000). "Host immune response to intracellular bacteria: A role for MHC-linked class-Ib antigen-presenting molecules." *Proc Soc Exp Biol Med* **224**(4): 231-9.
- Sousa, A. B. (2000). *The Gene Conversion Hypothesis of MHC Evolution: Studies on the Detection of Rare Segmental Exchange Variants in the Laboratory Mouse*. Institute of Genetics. Cologne, Universität zu Köln: 94.
- Spassky, A., S. Busby, et al. (1984). "On the action of the cyclic AMP-cyclic AMP receptor protein complex at the Escherichia coli lactose and galactose promoter regions." *EMBO J* **3**(1): 43-50.
- Stahl, F. W. (1994). "The Holliday Junction on its Thirtieth Anniversary." *Genetics* **138**: 241-246.
- Stemmer, W. P. C. (1995). "Searching sequence space." *Biotechnology* **13**: 549-553.
- Studier, F. W. (1991). "Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system." *J Mol Biol* **219**(1): 37-44.
- Studier, F. W. and B. A. Moffatt (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." *J Mol Biol* **189**(1): 113-30.
- Studier, F. W., A. H. Rosenberg, et al. (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." *Methods Enzymol* **185**(1): 60-89.
- Sun, H., D. Treco, et al. (1991). "Extensive 3'-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site." *Cell* **64**(6): 1155-61.
- Szostack, J. W., T. L. Orr-Weaver, et al. (1983). "The Double Strand break Repair Model for Recombination." *Cell* **33**: 25-35.
- Taddei, F., M. Radman, et al. (1997). "Role of mutator alleles in adaptive evolution." *Nature* **387**: 700-702.
- Taghian, D. G., H. Hough, et al. (1998). "Biased short tract repair of palindromic loop mismatches in mammalian cells." *Genetics* **148**(3): 1257-68.
- Taghian, D. G. and J. A. Nickoloff (1997). "Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells." *Mol Cell Biol* **17**(11): 6386-93.
- Thio, C. L., M. Carrington, et al. (1999). "Class II HLA alleles and hepatitis B virus persistence in African Americans." *J Infect Dis* **179**(4): 1004-6.
- Thompson, C. B. (1992). "Creation of immunoglobulin diversity by intrachromosomal gene conversion." *Trends Genet* **8**(12): 416-22.
- Thursz, M. R., H. C. Thomas, et al. (1997). "Heterozygote advantage for HLA class-II type in hepatitis B virus infection." *Nat Genet* **17**(1): 11-2.
- Torres, R. M. and R. Kühn (1997). *Laboratory Protocols for Conditional Gene Targeting*. New York, Oxford University Press.
- Trowsdale, J. (1995). "'Both man & bird & beast': comparative organization of MHC genes." *Immunogenetics* **41**: 1-17.
- Tyagi, S. and F. R. Kramer (1996). "Molecular Beacons: Probes that Fluoresce upon Hybridization." *Nature Biotechnology* **14**: 303-308.
- Valen, L. V. (1973). "A New Evolutionary Law." *Evolutionary Theory* **1**: 1-30.
- Wagner, R., P. Debbie, et al. (1995). "Mutation detection using immobilized mismatch binding protein (MutS)." *Nucleic Acids Res* **23**(19): 3944-8.

- Wakeland, E. K., S. Boehme, et al. (1990). "The generation and maintenance of MHC class II gene polymorphism in rodents." *Immunol Rev* **113**: 207-26.
- Wakeland, E. K., S. Boehme, et al. (1990). "Ancestral polymorphisms of MHC class II genes: divergent allele advantage." *Immunol Res* **9**(2): 115-22.
- Waldman, A. S. and R. M. Liskay (1987). "Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells." *Proc Natl Acad Sci U S A* **84**(15): 5340-4.
- Waldman, A. S. and R. M. Liskay (1988). "Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology." *Mol Cell Biol* **8**(12): 5350-7.
- Wedekind, C., M. Chapuisat, et al. (1996). "Non-random fertilization in mice correlates with the MHC and something else." *Heredity* **77**(Pt 4): 400-9.
- Weill, J. C. and C. A. Reynaud (1987). "The chicken B cell compartment." *Science* **238**(4830): 1094-8.
- Weill, J. C. and C. A. Reynaud (1996). "Rearrangement/hypermutation/gene conversion: when, where and why?" *Immunol Today* **17**(2): 92-7.
- Widmer, M. B. and H. R. Macdonald (1980). "Cytolytic T lymphocyte precursors reactive against mutant Kb alloantigens are as frequent as those reactive against a whole foreign haplotype." *J. Immunol.* **12**: 48-55.
- Wilgenbus, K. K. and P. Lichter (1999). "DNA chip technology ante portas." *J Mol Med* **77**(11): 761-8.
- Williams, R. L., D. J. Hilton, et al. (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." *Nature* **336**(6200): 684-7.
- Winkler, H. (1930). *Die Konversion der Gene*. Jena, Gustav Fischer Verlag.
- Winter, R. B. and P. H. v. Hippel (1981). "Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids
2. The *Escherichia coli* Repressor-Operator Interaction: Equilibrium Measurements." *Biochemistry* **20**: 6948-6960.
- Wylie, C. C. (1993). "The Biology of Primordial Germ Cells." *Eur Urol* **23**: 62-67.
- Youil, R., B. W. Kemper, et al. (1995). "Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII." *Proc Natl Acad Sci U S A* **92**(1): 87-91.
- Yuhki, N. and S. J. O'Brien (1994). "Exchanges of short polymorphic DNA segments predating speciation in feline major histocompatibility complex class I genes." *J Mol Evol* **39**(1): 22-33.
- Zangenberg, G., M. M. Huang, et al. (1995). "New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm [see comments]." *Nat Genet* **10**(4): 407-14.
- Zhong, X. B., P. M. Lizardi, et al. (2001). "Visualization of oligonucleotide probes and point mutations in interphase nuclei and DNA fibers using rolling circle DNA amplification." *Proc Natl Acad Sci U S A* **98**(7): 3940-5.
- Zingsheim, H. P., N. Geisler, et al. (1977). "Complexes of *Escherichia coli* lac-Repressor with Non-operator DNA Revealed by Electron Microscopy: Two Repressor Molecules Can Share the Same Segment of DNA." *Journal of Molecular Biology* **115**: 565-570.

ERKLÄRUNG

Ich versichere, dass ich die von mir vorlegte Dissertation selbständig angefertigt, die benutzen Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlauf oder dem Sinn nach entnommen sind, im jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Jonathan C. Howard betreut worden.

Teilpublikationen :

The gene conversion hypothesis of MHC evolution . a review

Jann Th. Martinsohn, Ana B. Sousa, Lisbeth A. Guethlein, Jonathan C. Howard

Immunogenetics 50: 168-200 (1999)

Jann Th. Martinsohn

CURRICULUM VITAE

Name: Jann Thorsten Martinsohn
Date of Birth: 29.02.1968
Place of Birth: Heidelberg
Nationality: German

Education

1974-1978 Primary School (Mönchhofgrundschule Heidelberg)
1978-1987 High School (Hölderlingymnasium Heidelberg)
1987 Abitur (Allgemeine Hochschulreife)
1987-1989 Civil Service at the Red Cross Germany (Rettungssanitäter RSI and RSII)
1989-1991 Undergraduate studies in Biology at the University of Cologne
1992 Practical training in Immunology (Advisor: Prof. J.F.T. Griffin – University of Otago; New Zealand)
1992 Studies in Marine Science at the University of Otago (New Zealand)
1992 Diploma of Science with credits in Marine Science (Thesis: “Sharks of Otago Region”; Advisor: Prof. P. Mladenov)
1993 Practical Training in Molecular Biology (Advisor: Prof. D. Stéhelin – Laboratoire d’Oncologie Moléculaire; Institut Pasteur Lille; France)
1993-1995 Postgraduate Studies in Biology (Genetics; Biochemistry; Developmental Biology) at the University of Cologne
1995-1996 Masters of Science (Diplomarbeit) in Genetics (Advisor: Prof. J.C. Howard)
1997-2001 Ph.D. at the University of Cologne (Advisor: Prof. J.C. Howard)

ACKNOWLEDGEMENTS

I would like to express my thankfulness

To Jonathan Howard, who stimulated my interest into gene conversion and never ceased to encourage me to carry on.

And also to Libby Guethlein, who always was there when needed.

I am happy and consider myself lucky that Ana Sousa accompanied me throughout these years as a supporting colleague but much more as a great friend and companion.

Ich habe das Glück, Sandra, Wera, und Martin an meiner Seite zu wissen, die niemals aufgeben, mich an das „Leben diesseits“ zu erinnern.

Mit großer Dankbarkeit denke ich an Christine, die mich immer unterstützt hat, und die mich lehrte zu leben.

Auch Gert schulde ich großen Dank, für seine Geduld und seinen großartigen Beistand!

Et tout d'abord je pense tendrement à Sophie qui, malgré tout, n'a jamais cessé d'être à mes côtés...