

MOLECULAR ANALYSIS OF *HOB*O AND *HIMAR* 1 TRANSPOSITION

Karen Lipkow

Wadham College
Microbiology Unit
Division of Molecular Genetics
Department of Biochemistry
University of Oxford



A Dissertation Submitted for the Degree
Doctor of Philosophy

Oxford, Trinity Term 2003



*"Getting the award was nice,
but still the best reward is being allowed to do research."*

Eric Wieschaus

to K. L., two weeks after the
announcement of his Nobel Prize

Molecular Analysis of *hobo* and *Himar1* Transposition

Karen Lipkow, Wadham College, Oxford

D.Phil Thesis, Trinity Term 2003

Abstract

The aim of this study is to find out more about how eukaryotic transposons work. Transposons are widespread genetic elements that can autonomously change their position within a genome. Their translocation is mediated by the transposase enzyme, encoded by the transposon itself. Transposition is often associated with mutations and recombination in the host and may promote evolutionary change. Most knowledge of the principles and molecular mechanisms has so far come from prokaryotic transposons. Eukaryotic transposons have mostly been studied on the level of population dynamics. In this thesis, I studied two eukaryotic transposons on a molecular level. *hobo* and *Himar1* are members of very widespread transposon families, *hAT* and *mariner*-like elements, respectively. Their interest lies mainly in the proposition that the order of catalytic steps is the opposite of all prokaryotic transposons studied so far.

I developed several protocols for the purification of *hobo* transposase. These preparations were assayed by several methods *in vitro*. DNA hairpin structures, which had been postulated to arise in the flanking DNA upon *hobo* excision, were assayed for with a newly developed very sensitive method, frayed duplex PCR. Here, and in reconstructed *in vivo* transposition systems, no specific transpositional activity, apart from a high toxicity for the host cells, was detected. This suggests a requirement for as yet unidentified host factors.

For *Himar1*, I improved the published transposase purification protocol and optimised the *in vitro* reaction conditions, achieving a c. 10-fold increase in specific activity. Specific transposase-DNA binding is shown, representing the earliest intermediates in transposition. Later complexes studied are target capture and strand transfer complexes, which are the final stage of the transposition reaction. Hydroxyl radical footprinting and band shift assays suggest that *Himar1* transposase binds a single DNA end as multimers while occupying only a single DNA binding site. *Himar1* can bind and commit to target DNA before as well as after excision from the donor. Target commitment is only clearly visible under conditions of macromolecular crowding, and insertions were found to be more efficient if the target is supercoiled. These are both conditions closely resembling the situation in the eukaryotic nucleus.

Acknowledgements

In my experience, most readers do not get beyond this page in a thesis. I hope, of course, that mine will be the exception, but I hope to rise to the challenge and not forget any of the many people I am grateful to.

First and foremost, I would like to thank my parents, to whom I dedicate this thesis, for bringing me up the way they did and for their constant love and encouragement.

It was great to have a supervisor who is also a friend. I wish to thank Ronnie Chalmers for lots of ideas, an open door, the freedom to find my own solutions, sharing the joys and frustrations lab work invariably brings with it, and the occasional quote from Hamlet. Many thanks to the members of our lab, especially Nicolas Buisine, for help and discussions. David Sherratt and his group and visitors provided many inspirations and helped me to see the big picture. Thank you also for intermittent group meetings and office space. To all members of the Microbiology Unit, thank you for a nice atmosphere and environment. In the Biochemistry Department, Mary Gregoriou and Stuart Holt have been exceptionally helpful throughout.

I am indebted to Jill Featherstone for introducing me to the secrets of cell culture, and to Yuying Liu for her help with cell culture and cloning. For technical instructions, thank you especially to Ronnie, Sean Colloms, Isabelle Lucet and Said Ismail. David Lampe, Axel Imhof, Peter Becker, Bernard Moss and Nina Kotsopoulou kindly supplied me with plasmids, proteins and information.

Wadham College was a home in more than one way. I am especially grateful to Keith Dyke, who was always there for questions and treated me to enchanted evenings at High Table, and to Mike Sauvage, who was very helpful in providing housing, furniture and funds. Being involved in the Oxford University Biochemical Society might have distracted me from my research at times, but it clearly added to my development as a scientist. So thank you to Atlanta, James, Sven, Tom, John, Tony Watts, Gill McLure and all speakers for a truly enjoyable time. Thank you to my friends, family and in-laws here and abroad for enriching my life outside the lab, especially Tanja, Tobias, Irene, Paul, Laura, Ernesto, Eva, Steven, Stephanie, Regina, Detlef, Anne, Gesine, Nadia, Ben and selected members of the Wadham MCR.

A special thank you to Dennis Bray for graciously allowing me the time to finish this manuscript and for trusting me with new and exciting challenges. Dennis, Sergej, Matthew and Bernhard each in their way helped me in the final months.

My biggest special thank you to my favourite bench-neighbour, short-term research assistant, proofreader, Mac-wiz, cyclist, dance-partner, masseur, idea-bouncer, roommate, fellow mobile element and husband, Sven Sewitz. It is hard to imagine how I would have done it without you.

I gratefully acknowledge the financial support by the Marie Curie TMR Programme, the BBSRC, the Gottlieb Daimler- und Karl Benz-Stiftung, Oxford University and Wadham College.

Table of Contents

Abstract	2
Acknowledgements	3
Table of Contents	4
Table of Tables	9
Table of Figures	9
Abbreviations	11
CHAPTER 1 - INTRODUCTION	13
1.1 Transposition	13
1.1.1 Principle of catalysis	15
1.1.2 Order of strand transfers	19
1.1.3 Nucleoprotein complexes	21
1.1.4 Eukaryotic transposons	21
1.2 V(D)J recombination	22
1.2.1 V(D)J is like a transposon	22
1.2.2 The RAG recombinase	24
1.2.3 Molecular view of V(D)J recombination	26
1.2.3.1 Cleavage at the recombination signal sequences	26
1.2.3.2 Coding joint formation	28
1.2.3.2.1 Processing of coding ends	29
1.2.3.2.2 Coding end joining	29
1.2.3.3 Signal joint formation and cleavage	31
1.2.3.4 Transposition	31
1.3 <i>hAT</i> transposons	32
1.3.1 The mechanism of <i>hAT</i> transposition	32
1.3.1.1 Excision in <i>hAT</i> transposition	32
1.3.1.2 <i>hAT</i> transposon circles	33
1.3.1.3 <i>hAT</i> insertion	34
1.3.1.4 <i>hAT</i> transposase structure, function and activity	34
1.3.2 Study of the <i>hAT</i> transposition process	35
1.3.3 The <i>hobo</i> transposon	36
1.3.3.1 Distribution, genetics and molecular structure	36
1.3.3.2 The <i>hobo</i> HFL-1 element	37
1.3.3.3 Choice of the <i>hobo</i> transposon as an experimental system	37
1.4 <i>Tc1/mariner</i> transposons	38
1.4.1 Distribution and activity	38
1.4.2 Structure of <i>Tc1/mariner</i> transposons	39
1.4.2.1 Structure and function of the transposase	39
1.4.3 Current model for <i>Tc1/mariner</i> transposition	41
1.4.3.1 Target site specificity	41
1.4.3.2 <i>Tc1/mariner</i> complexes	43
1.4.4 Choice of the <i>Himar1</i> transposon	43
1.5 Goals of the project	44

CHAPTER 2 - MATERIALS AND METHODS	45
2.1 Materials	45
2.1.1 Strains and cell lines	45
2.1.2 Oligonucleotides, plasmids and other DNA substrates	47
2.1.3 Chemicals	65
2.2 General DNA manipulation	65
2.2.1 Cloning	65
2.2.1.1 Polymerase Chain Reaction	65
2.2.1.2 Restriction digest and end-treatment	65
2.2.1.3 Ligation	66
2.2.1.4 Transformation of <i>E. coli</i>	66
2.2.1.5 DNA purification	66
2.2.1.6 Sequencing	66
2.2.1.7 Analytical Polymerase Chain Reaction	67
2.2.2 Selected cloning strategies	67
2.2.2.1 Cloning of pKL2, pKL3 and pKL4	67
2.2.2.2 Cloning of pKL7, pKL8 and pKL9	67
2.2.2.3 Cloning of pKL10 and pKL11	67
2.2.3 Radiolabelling of DNA	69
2.2.3.1 Labelling transposon end substrate	69
2.2.3.2 Internal labelling by random priming	70
2.3 Protein expression and purification	70
2.3.1 <i>hobo</i> transposase expression and purification in <i>E. coli</i>	70
2.3.1.1 Expression of <i>hobo</i> transposase in <i>E. coli</i>	70
2.3.1.2 Western Blotting for tagged <i>hobo</i> transposase	71
2.3.1.3 Growth of cells for 6xHis- <i>hobo</i> transposase, preparation of lysate and storage	71
2.3.1.4 Purification of 6xHis- <i>hobo</i> transposase	72
2.3.2 Expression of GST- <i>hobo</i> transposase in <i>S. pombe</i>	72
2.3.3 Coupled transcription and translation of <i>hobo</i> transposase	72
2.3.4 MBP- <i>Himar1</i> transposase expression and purification	73
2.3.4.1 Expression of MBP- <i>Himar1</i> transposase in <i>E. coli</i>	73
2.3.4.2 Amylose column purification of MBP- <i>Himar1</i> transposase	73
2.3.4.3 Anion exchange column purification of MBP- <i>Himar1</i> transposase	73
2.4 Specialised PCR procedures	74
2.4.1 Generation of substrates by PCR	74
2.4.2 Generation of substrates from custom-built plasmids	74
2.4.3 Frayed-duplex PCR of <i>hobo</i> excision reactions	74
2.5 Transposition <i>in vitro</i>	75
2.5.1 <i>hobo</i> transposition <i>in vitro</i>	75
2.5.1.1 <i>hobo in vitro</i> excision reactions	75
2.5.1.2 <i>In vitro</i> plasmid-to-plasmid transposition with reticulocyte expressed <i>hobo</i> transposase	75
2.5.2 <i>Himar1</i> transposition <i>in vitro</i>	75
2.5.2.1 <i>Himar1</i> plasmid substrate reactions	76
2.5.2.2 <i>Himar1</i> insertion reactions	76
2.5.2.3 <i>Himar1</i> gel retardation	76
2.6 Transposition <i>in vivo</i>	76
2.6.1 Addition of a <i>Cam^R</i> marker to pOX38 for use in <i>in vivo</i> transposition	76
2.6.2 Cell culture	77
2.6.2.1 Stable transfection of SCGR11 and NIH 3T3 fibroblasts	77
2.6.2.2 Determination of copy number of integrated plasmids	78
2.6.2.2.1 Comparison of integrations with a single-copy mouse gene	78
2.6.2.2.2 Comparison of integrations with pure plasmid	79
2.6.2.3 Retroviral transduction	79

CHAPTER 3 - EXPRESSION AND PURIFICATION OF <i>HOB</i>O TRANSPOSASE	80
3.1 Introduction	80
3.2 Initial trials for transposase expression in <i>E. coli</i>	81
3.2.1 CBP-fusion in <i>E. coli</i>	81
3.2.1.1 Construction of the CBP- <i>hobo</i> expression vector	81
3.2.1.2 Expression of the CBP- <i>hobo</i> fusion protein	81
3.2.1.3 Western blot detection	83
3.2.2 Native and His-tagged transposase in <i>E. coli</i>	83
3.2.2.1 Expression of native and His-tagged <i>hobo</i> transposase	84
3.3 Expression and purification of 6xHis-transposase	84
3.3.1 Kinetics of 6xHis-transposase expression	84
3.3.2 Solubilisation and metal affinity chromatography of the transposase	86
3.3.3 Preparative purification of 6xHis-transposase	88
3.4 Additional <i>hobo</i> transposase expression trials in <i>E. coli</i>	88
3.4.1 Replacement of rare codons in the transposase	91
3.4.2 Increased expression with extra tRNA genes	91
3.4.3 Expression of transposase with a C-terminal His-tag	94
3.5 Expression of GST-transposase in <i>S. pombe</i>	96
3.5.1 Cloning of the GST- <i>hobo</i> expression plasmid and transformation into <i>S. pombe</i>	96
3.5.2 Expression of the GST- <i>hobo</i> fusion protein	96
3.6 <i>In vitro</i> transcription and translation in reticulocyte lysate	98
3.7 Discussion	100
3.7.1 Expression systems	100
3.7.2 Low level of expression	101
3.7.3 Protein truncations	101
3.7.4 Solubility	102
3.7.5 Alternative expression systems	103
3.7.6 Conclusion	103
CHAPTER 4 - DEVELOPMENT OF PCR ASSAYS FOR HAIRPIN DETECTION	104
4.1 Introduction	104
4.2 Direct detection of hairpins via PCR	105
4.2.1 Single-primer polymerase chain reaction (SPPCR)	105
4.2.2 Frayed-duplex polymerase chain reaction (FDPCR)	107
4.3 Generation of components for the assay and control experiments	108
4.3.1 Adapters for FDPCR	108
4.3.2 Generation of artificial hairpin templates as positive controls for FDPCR and SPPCR	108
4.3.3 Plasmids for generating artificial hairpins directly	110
4.3.4 Frayed-duplex templates	110
4.4 Results and optimisation of FDPCR and SPPCR on control template	110
4.4.1 Analysis of FDPCR products	110
4.4.2 Optimisation of SPPCR	112
4.4.3 Optimisation of PCR for FDPCR	113
4.4.4 Nested PCR	113
4.4.5 Optimisation of ligation for FDPCR	118
4.4.6 Size-selection of the substrate	119
4.4.7 Effect of background DNA	119
4.4.7.1 Effect of competitor DNA on SPPCR	120

4.4.7.2	Effect of competitor DNA on FDPCR	120
4.4.7.3	Effect of competitor DNA on ligation	120
4.4.8	FDPCR of short, pre-ligated templates	121
4.4.8.1	Optimisation of FDPCR on short templates	121
4.5	Discussion	123
4.5.1	Efficiency of the PCR reaction	124
4.5.2	Influence of competitor DNA on the specificity of FDPCR	124
4.5.3	Applications of the assay	125
CHAPTER 5 - <i>HOB</i>O TRANSPOSITION <i>IN VITRO</i> AND <i>IN VIVO</i>		126
5.1	Introduction	126
5.2	<i>hobo in vitro</i> experiments with purified 6xHis-transposase	126
5.2.1	Analysis of <i>hobo in vitro</i> reactions by gel electrophoresis and FDPCR	126
5.2.2	Gel retardation	127
5.3	<i>In vitro</i> experiments with <i>in vitro</i> translated transposase	130
5.4	Mating out assay in <i>E. coli</i>	130
5.4.1	Components of the different mating out assays	130
5.4.1.1	Donor and recipient strains	132
5.4.1.2	Construction of transposon and expression plasmids	132
5.4.1.3	Selection of conjugative plasmids	132
5.4.2	<i>In vivo</i> transposition and mating out	133
5.4.3	Analysis of recombinant plasmids	134
5.4.4	Detection of <i>in vivo</i> excisions by minipreps	135
5.5	Toxicity of <i>hobo</i> transposase	135
5.5.1	Toxicity of components of the mating-out assay	137
5.5.2	Morphology of expressing cells	139
5.6	Excision of genomic <i>hobo</i> copies in mouse fibroblasts	141
5.6.1	Stable transfectants	141
5.6.2	Transfection and expression of <i>hobo</i> transposase constructs	145
5.6.2.1	Chemical methods of transfection	145
5.6.2.2	Retroviral transduction	146
5.6.3	Analysis of S10.1 fibroblast DNA after expression of <i>hobo</i> -transposase	149
5.6.3.1	Analysis of excision by two-dimensional gel electrophoresis	149
5.6.3.2	Hairpin detection with FDPCR	149
5.7	Discussion	151
5.7.1	<i>In vitro</i> experiments	151
5.7.2	<i>hobo in vivo</i> activity in <i>E. coli</i>	153
5.7.3	Toxicity of <i>hobo</i> transposase to the host cells	154
5.7.4	<i>hobo</i> transposase expression and activity in mouse fibroblasts	155
5.7.5	Host specificity of <i>hobo</i>	156
5.7.6	Outlook	157
CHAPTER 6 - <i>HIMAR1</i> TRANSPOSITION <i>IN VITRO</i>		159
6.1	Introduction	159
6.2	Purification and activity of <i>Himar1</i> transposase	160
6.2.1	Further purification of MBP- <i>Himar1</i> transposase	160
6.2.2	Nicking activity	162
6.2.3	Activity tests for the purified transposase	164
6.2.4	Systematic variation of the <i>Himar1 in vitro</i> reaction conditions	164

6.3	Catalytic activities of the <i>Himar1</i> transposase	165
6.3.1	Effect of the conformation of the substrate DNA on excision and insertion	165
6.3.2	Analysis of excision products and intermediates	167
6.3.3	The <i>in vitro</i> insertion product retains a supercoiled conformation	169
6.3.4	Calcium supports strand transfer in <i>Himar1</i> transposition and stabilises the STC	172
6.3.5	Insertion of uncleaved ends results in more stable strand transfer complexes	174
6.4	Target commitment and the influence of macromolecular crowding	175
6.4.1	<i>Himar1</i> exhibits target commitment before and after donor cleavage	175
6.4.2	Macromolecular crowding enhances target commitment	177
6.4.3	Effects of PEG on <i>Himar1</i> transposition <i>in vitro</i>	179
6.5	Identification and analysis of nucleoprotein complexes	180
6.5.1	Specific <i>Himar1</i> transposase-DNA complexes can be formed <i>in vitro</i>	181
6.5.2	Transposase titration reveals additional complexes	181
6.5.3	Formation of strand-transfer and target-capture complexes	183
6.5.4	<i>Himar1</i> transposase displays exonucleolytic activity in the presence of magnesium	185
6.5.5	The mobility of complexes is predominantly determined by the protein component	185
6.5.6	Complexes 1 and 2 contain a single transposon end	186
6.5.7	The two single end complexes show identical hydroxyl radical protection patterns	188
6.6	Discussion	190
6.6.1	Single end complexes	190
6.6.1.1	Composition of single ends complexes	190
6.6.1.2	Specificity of transposon DNA-binding	192
6.6.1.3	Influence of DNA sequences beyond the ITR	193
6.6.1.4	The molecular basis for overproduction inhibition	194
6.6.2	The synaptic complex	194
6.6.3	The target capture complex	195
6.6.3.1	Influence of macromolecular crowding on the TCC	195
6.6.3.2	Target commitment	196
6.6.3.2.1	Genomic distribution and target commitment	196
6.6.3.2.2	Structural implications of target commitment for the TCC	196
6.6.3.3	Implications of target DNA conformation	197
6.6.4	The strand transfer complex	198
6.6.4.1	High stability of strand transfer complexes	198
6.6.4.2	The prevalence of supercoiling after strand transfer	199
6.6.4.3	Exonuclease activity of <i>Himar1</i> transposase	199
6.6.5	Conclusion	201
APPENDIX		202
A.1	Calculating the fraction of committed synaptic complexes	202
A.2	Target commitment measurements and results of calculations	203
A.3	ANOVA definitional formulas	204
A.4	DnaCalc for designing arbitrary primers	204
REFERENCES		205

Table of Tables

Table 2.1: <i>Escherichia coli</i> strains generated for this work	45
Table 2.2: Other <i>Escherichia coli</i> strains	45
Table 2.3: <i>Schizosaccharomyces pombe</i> strain	46
Table 2.4: Mammalian cell lines	46
Table 2.5: Oligonucleotides designed for this work	47
Table 2.6: Other oligonucleotides	54
Table 2.7: Plasmids created for this work	55
Table 2.8: Other plasmids	60
Table 2.9: Artificial templates for testing SPPCR and FDPCR	63
Table 2.10: Transposon end substrates	64
Table 3.1: Solubilisation and purification yields of 6xHis- <i>hobo</i> transposase in different reagents	87
Table 3.2: Codon usage	90
Table 3.3: Optimisation of <i>hobo</i> -transposase expression by coupled transcription/translation	99
Table 4.1: Optimisation of PCR conditions for template 10B-hf with primers KL87/KL88	114
Table 4.2: Optimisation of PCR conditions for template 42S-hf with primers KL87/KL88	115
Table 4.3: Optimised SPPCR and FDPCR conditions	118
Table 5.1: Mating out assay	134
Table 5.2: Influence of plasmids, antibiotics and inducer on <i>E. coli</i> viability	138
Table 6.1: Analysis of variance in a factorial design for influences on <i>Himar1</i> target commitment	179

Table of Figures

Figure 1.1: Transposon structure and hosts	14
Figure 1.2: The mechanism of replicative transposition	16
Figure 1.3: The mechanism of non-replicative transposition	17
Figure 1.4: Chemistry of the transesterifications	18
Figure 1.5: Order of catalytic steps in excision and insertion	20
Figure 1.6: Overview of V(D)J recombination and <i>hAT</i> transposition	23
Figure 1.7: Comparison of terminal sequences of eukaryotic transposons and RSS	25
Figure 1.8: Current model of V(D)J recombination and footprint formation	27
Figure 1.9: Model for <i>Tc1/mariner</i> transposition	42
Figure 2.1: Major <i>Himar1</i> transposon end substrates	64
Figure 2.2: Plasmid substrates pKL10 and pKL42	68
Figure 3.1: Expression of CBP- <i>hobo</i> transposase in <i>E. coli</i>	82
Figure 3.2: T7 Expression of His-tagged proteins in <i>E. coli</i> BL21(DE3)pLysS	85
Figure 3.3: Purification of 6xHis- <i>hobo</i> transposase using cobalt affinity resin	89
Figure 3.4: Comparison of protein expression of wildtype and RR <i>hobo</i> transposase	92
Figure 3.5: Expression of 6xHis- <i>hobo</i> transposase with extra copies of rare tRNA genes	93
Figure 3.6: Comparison of N- and C-terminal 6xHis-tags	95
Figure 3.7: GST-fusion protein expression in <i>S. pombe</i>	97
Figure 3.8: Coupled transcription/translation in reticulocyte lysate	99
Figure 4.1: PCR techniques for the detection of hairpin structures	106
Figure 4.2: Schematic of hairpin template generation	109
Figure 4.3: Generation of assay components and analysis of FDPCR products	111
Figure 4.4: Optimisation of FDPCR reactions	117
Figure 4.5: FDPCR of short, pre-ligated templates	122

Figure 5.1: <i>hobo in vitro</i> excision and FDPCR	128
Figure 5.2: Second PCR after <i>hobo in vitro</i> excision	129
Figure 5.3: Mating out assay	131
Figure 5.4: Analysis of insertions into R388	136
Figure 5.5: Influence of protein expression on <i>E. coli</i> morphology	140
Figure 5.6: <i>hobo</i> excision assay in mouse fibroblasts	142
Figure 5.7: Cell lines with stable <i>hobo</i> integrations	144
Figure 5.8: Transfected 293T cells producing recombinant retrovirus	147
Figure 5.9: Transduced S10.1 cells	148
Figure 5.10: 2D gels for the detection of hairpins from transduced S10.1 cells	150
Figure 5.11: FDPCR for the analysis of genomic DNA from mouse fibroblasts after <i>hobo</i> induction	152
Figure 6.1: Purification of MBP- <i>Himar1</i> C9 transposase	161
Figure 6.2: Activities of <i>Himar1</i> transposase purifications	163
Figure 6.3: <i>Himar1</i> transposition with supercoiled and linear plasmid substrates	166
Figure 6.4: <i>Himar1</i> excision	168
Figure 6.5: Intermolecular insertion reactions	170
Figure 6.6: Topology and stability of insertion products	171
Figure 6.7: Two-dimensional gel to investigate the conformation of the main insertion product	173
Figure 6.8: Target commitment of <i>Himar1</i> transposase	176
Figure 6.9: Influence of polyethylene glycol on <i>Himar1</i> insertion properties	178
Figure 6.10: Specific complexes between transposase and the left end of <i>Himar1</i>	182
Figure 6.11: Effect of metal ions and substrate sizes on <i>Himar1</i> complexes	184
Figure 6.12: Identification of complex stoichiometry	187
Figure 6.13: Footprints of <i>Himar1</i> transposase-transposon end complexes	189
Figure 6.14: Summary of complexes formed during <i>Himar1</i> insertion	191
Figure 6.15: Proposed structures of MLE nucleoprotein complexes	200

Abbreviations

'	minute(s)	EtBr	ethidium bromide
''	second(s)	EtOH	ethanol
*	radioactive label	FBS	foetal bovine serum
A	adenine	FDPCR	frayed-duplex polymerase chain reaction
A _x	absorption at x nm wavelength	FPLC	fast performance liquid chromatography
aa	amino acid(s)	G	guanidine
Ad 5	adenovirus type 5	G418	geneticin, G-418 sulphate
Amp	ampicillin	Gb	gigabase pairs
<i>Amp^R</i>	β-lactamase gene	GFP	green fluorescent protein
ATP	adenosine triphosphate	Glc	glucose
ATS	altered target specificity	GST	glutathione-S-transferase
bis	N'N'-bis-methylene-acrylamide	HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid
bp	base pair(s)	HFL-1	<i>hobo</i> full length clone no. 1
BSA	bovine serum albumin	HIV	human immunodeficiency virus
C	cytidine	hp	hairpin
Cam	chloramphenicol	HRP	horseradish peroxidase
CBD	chitin binding domain	HMG	high-mobility group protein
CBP	calmodulin-binding peptide	HPLC	high-pressure liquid chromatography
CIP	calf intestinal phosphatase	HSV	herpes simplex virus
CMC	critical micelle concentration	HU	histone U
CMV	human cytomegalovirus	IAA	isoamyl alcohol, 3-methylbutanol
<i>cPfu</i>	cloned <i>Pfu</i> polymerase	IE	immediate early (gene)
cps	counts per second, disintegrations per second	IHF	integration host factor
cpx	complex	IPTG	isopropyl-β-D-thiogalactopyranoside
CSC	cleaved signal complex	IR	inverted repeat
DAPI	4,6-diamidino-2-phenylindone	IS	insertion sequence
dATP	deoxyadenosine 5'-triphosphate	ITR	inverted terminal repeat
dCMP	deoxycytidine 5'-monophosphate	Kan	kanamycin
dCTP	deoxycytidine 5'-triphosphate	<i>Kan^R</i>	aminoglycoside phosphotransferase gene
DEB	double end break complex	kb	kilobase pairs
dGTP	deoxyguanosine 5'-triphosphate	LIC	ligation independent cloning
dIdC	poly-(deoxyinosinic-deoxycytidylic) acid	LMPCR	ligation-mediated polymerase chain reaction
DMEM	Dulbecco's modified Eagle medium	LTR	long terminal repeat
DNA	deoxyribonucleic acid	Mb	megabase pairs
ds	double strand(ed)	MBN	mung bean nuclease
DSBR	double strand break repair	MBP	maltose binding protein
DTT	dithiothreitol	MCS	multiple cloning site, polylinker
dTTP	deoxythymidine 5'-triphosphate	MEM	minimum essential medium
ECL	enhanced chemiluminescence	MES	2-(N-morpholino)ethanesulfonic acid
EDTA	ethylenediaminetetraacetic acid	MITE	miniature inverted-repeat transposable element
EGFP	enhanced green fluorescent protein	MLE	<i>mariner</i> -like element
EK	enterokinase	MLV	murine leukaemia virus
EMCV	encephalomyocarditis virus	MoMuLV	Moloney murine leukaemia virus
EMM	Edinburgh minimal medium	MoMuSV	Moloney murine sarcoma virus
EMS	ethyl methanesulfonate		

MW	molecular weight (Da)	SB3-14	N-tetradecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate
nDMP	n-decyl-β-D-maltopyranoside	SEB	single end break complex
NBS	newborn bovine serum	SEC	single end complex, signal end complex
NEB	New England BioLabs	SC	single end complex
NHEJ	non-homologous end joining	<i>scid</i>	severe combined immune deficiency
NLS	nuclear localisation sequence	SDS	sodium dodecyl sulphate, lauryl sulphate
nt	nucleotide(s)	SPPCR	single-primer polymerase chain reaction
NTS	non-transferred strand	ss	single strand(ed)
OD _x	optical density at x nanometres wavelength	STC	strand transfer complex
orf	open reading frame	Str	streptomycin
ori	origin of replication	SV40	simian virus 40
<i>p</i>	promoter	<i>t</i>	transcriptional terminator
PA	polyacrylamide	T	thymine
PAGE	polyacrylamide gel electrophoresis	TAE	Tris / acetate / EDTA
pBS	pBluescript™ vector	T _{ag}	large T antigen
PBS	phosphate buffered saline	tase	transposase
PBST	phosphate buffered saline + Tween-20	TBE	Tris / borate / EDTA
PBSTM	phosphate buffered saline + Tween-20 + milk	TCC	target capture complex
PC	paired complex	TCR	T-cell receptor
PCR	polymerase chain reaction	TdT	terminal deoxynucleotidyl transferase
PEC	paired ends complex	TE	Tris / EDTA
PEG	polyethylene glycol	Tet	tetracycline
Pen	penicillin	thi	thiamine
polyA	mRNA polyadenylation signal	TIR	terminal inverted repeat
poly(dIdC)	poly-(deoxyinosinic-deoxycytidylic) acid	TK	thymidine kinase
PVDF	polyvinylidene fluoride	TLE	TcI-like element
^R	resistant, resistance	Tm	melting temperature
RAG	recombination activating gene	Tn	transposon
RNA	ribonucleic acid	TNT	transcription-and-translation
rpm	rotations per minute	Tp	trimethoprim
RSS	recombination signal sequence	Tris	tris(hydroxymethyl)methylamine
RT	room temperature	TS	transferred strand
^s	sensitive, sensitivity	U	units
S1	S1 nuclease	UTR	untranslated region
SB3-10	N-decyl-N,N-dimethyl-3-ammonio-1-propanesulfonate	wt	wildtype
		Ψ ⁺	extended viral packaging signal (psi)

CHAPTER 1 - INTRODUCTION

1.1 Transposition

Transposons are widespread genetic elements discovered in all prokaryotes and eukaryotes examined, with a few notable exceptions such as the specialised intracellular parasite *Encephalitozoon cuniculi* (Katinka *et al.*, 2001). At least 35-50% of the human genome consists of transposable elements, or relics thereof (Lander *et al.*, 2001; Venter *et al.*, 2001). In contrast to all other genes, transposable elements have specific mechanisms to move to new locations within their host genomes. Transposition into coding sequences and the initiation of chromosome rearrangements can affect gene expression and provide substrates for genome evolution (see Kempken and Windhofer, 2001; Kidwell and Lisch, 2002 for reviews of the biological function).

Autonomous transposons encode a transposase protein, which catalyses all chemical steps of the recombination reaction (Figure 1.1A). The transposase binds to the ends of the transposon DNA and makes flush or staggered cuts. Almost all transposon ends are organised as inverted terminal repeats (ITR), which contain the recognition sites for the transposase. Just inside the ITR, there can be sites for DNA-binding proteins from the host, e.g. for IHF in Tn10, and Ku70 in *P* elements (Morisato and Kleckner, 1987; Beall *et al.*, 1994). The two terminal inverted repeats of a particular transposon are usually nearly identical, with small differences between the left end and the right end further inside. Efficient transposition sometimes requires one end of each polarity, e.g. in *piggyBac* (Li *et al.*, 2001).

LTR-retrotransposons (such as Ty elements) and retroviruses (such as HIV and MoMuLV) differ from the classical DNA transposons, because they have an RNA intermediate. Their RNA genomes are converted by reverse transcriptase into DNA recombination substrates. The ends of these substrates are organised as directly repeated long terminal repeat sequences (LTRs), which are bound, processed and integrated by the transposase (usually called integrase) (reviews: Craigie, 2002; Voytas and Boeke, 2002). In addition, there is the separate class of non-LTR retrotransposons (including LINEs and SINEs), in which RNA is the substrate for recombination (review: Moran and Gilbert, 2002).

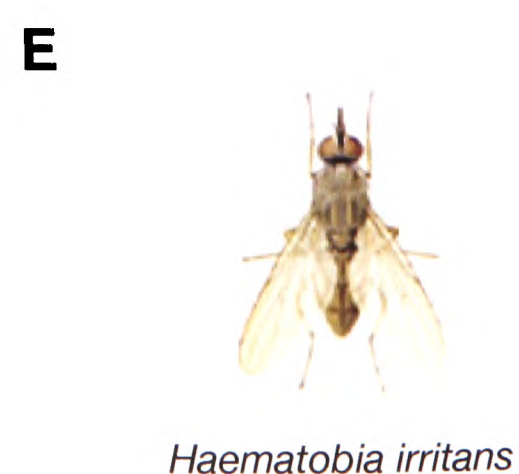
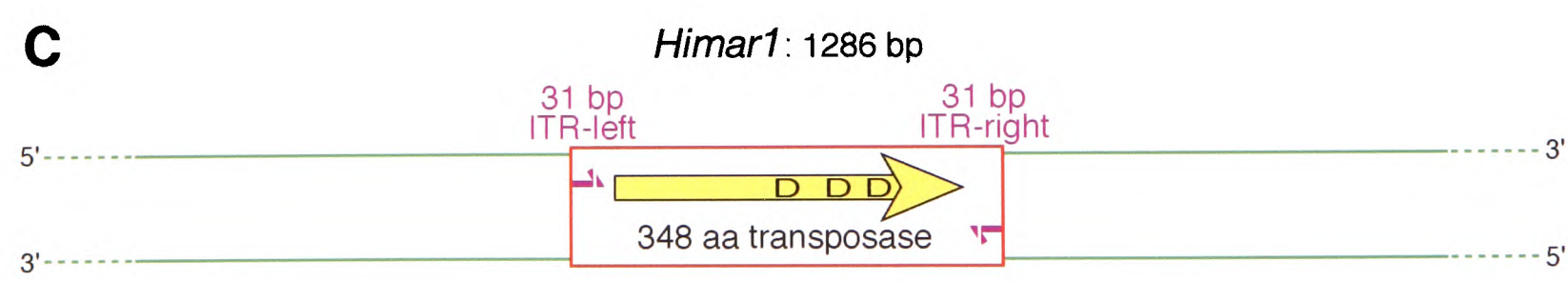
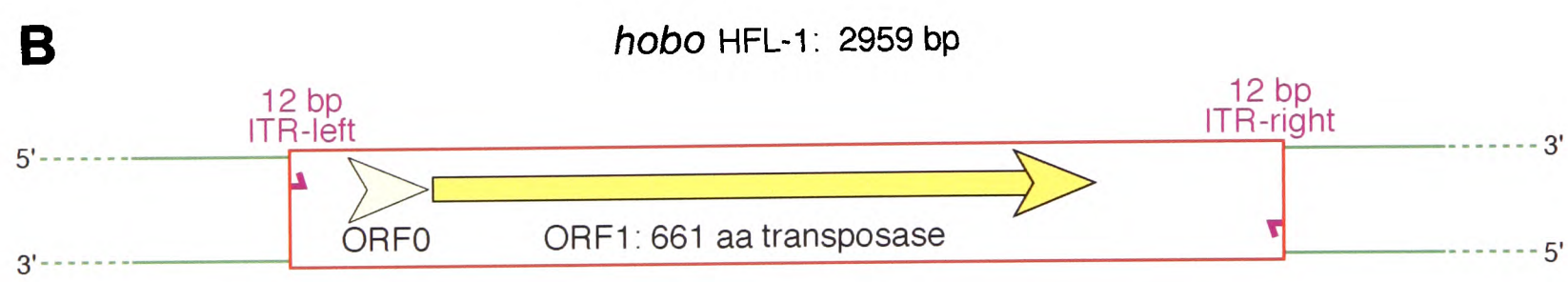
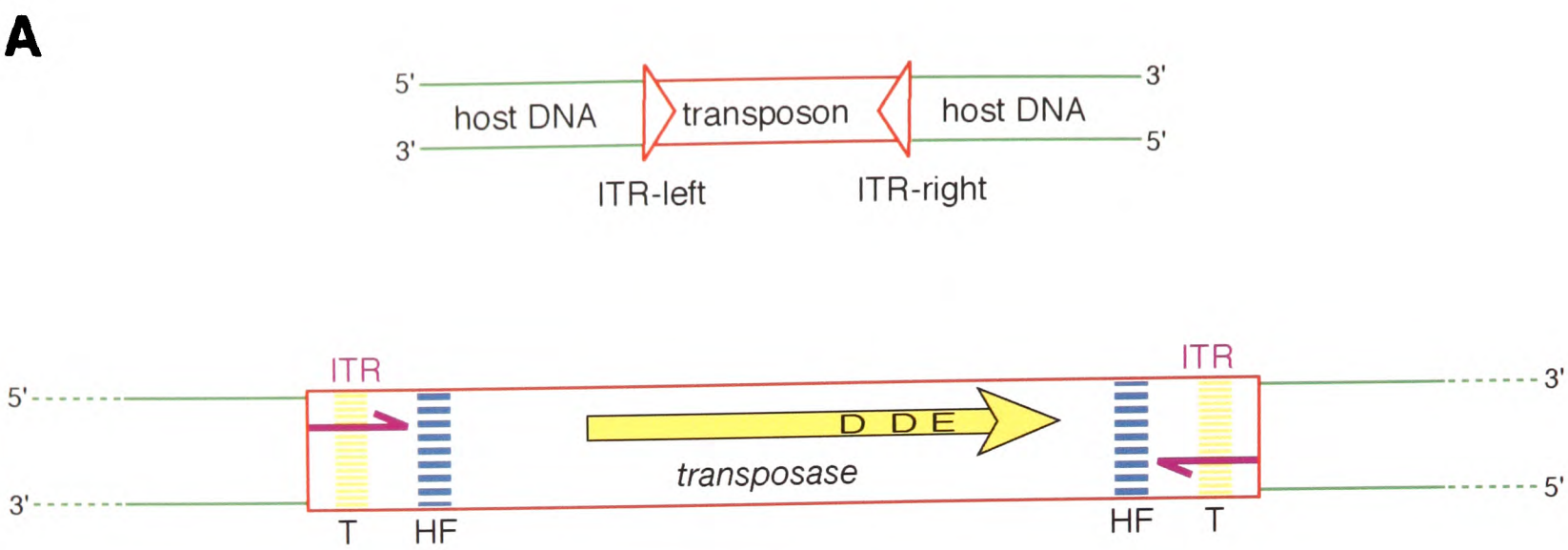


Fig. 1.1

Figure 1.1: Transposon structure and hosts

- (A) General structure of a transposon. The transposon has at least one orf, encoding the transposase. Horizontal lines symbolise DNA strands, vertical lines demark segments. Red, transposon sequence; green, flanking donor DNA; red triangles or purple half-arrows, inverted terminal repeats (ITRs); T, transposase binding site; HF, host factor binding site (note that not all transposons have host factor requirements); DDE, catalytic residues.
- (B) Structure of the *hobo* HFL-1 element. In the literature, the inverted terminal repeats are often denoted as “IR left” and “IR right”.
- (C) Structure of the reconstituted *Himar1* transposon. The white line inside the ITRs represents bp28, which is different in the ITR-left and ITR-right. In the literature, the inverted terminal repeats are often denoted as “5'ITR” and “3'ITR”.
- (D) A male fruitfly *Drosophila melanogaster*, the natural host of *hobo*. (Drawing from <http://flybase.bio.indiana.edu/anatomy/Drosophilidae>; Patterson, 1943.)
- (E) A female hornfly *Haematobia irritans*, the natural host of *Himar1*. (Drawing from <http://res2.agr.ca/ecorc/apss/fly/horn.htm>.)

Non-autonomous transposons are common in eukaryotes and often arise by loss of the transposase gene by point mutation or DNA recombination. Recombinant mini-transposons are also widely used as genetic tools. For historical reasons, some natural non-autonomous transposons have different names than the corresponding autonomous transposon: Non-autonomous *Activator* (*Ac*) elements are called *Dissociation* (*Ds*) (McClintock, 1950); a non-autonomous *Mos1* is called *peach* (Jacobson *et al.*, 1986). Aided by bioinformatics, many small non-autonomous transposons of very high copy number were discovered in the past ten years, the miniature inverted-repeat transposable elements (MITEs) (Wessler *et al.*, 1995; Bureau *et al.*, 1996).

The classical DNA transposons use either a replicative or a non-replicative transposition mechanism. Replicative transposons, e.g. Tn3 and phage Mu, capture a target while still attached to the donor site and replicate to form a cointegrate structure in which the donor and target molecules are joined with a copy of the transposon at each junction (Figure 1.2) (reviews: Grindley, 2002; Chaconas and Harshey, 2002). Non-replicative transposons transpose through a cut-and-paste mechanism (Figure 1.3): They are excised from the donor, leaving behind a double strand break, and are inserted into the target. In some, such as Tn7, the transposase binds to the target before excision; others, such as Tn10, form free transposon intermediates (reviews: Craig, 2002a, b; Haniford, 2002).

During insertion of all DNA transposons, phosphate groups on opposite strands of the target DNA are attacked by free 3'-hydroxyl groups on the transposon ends. The phosphate groups are usually 2-9 bp apart, and the distance is fixed for each transposon. The staggered insertion produces single stranded DNA on each end of the transposon. These stretches are filled in by host repair mechanisms and form characteristic direct repeats in the flanking DNA.

1.1.1 Principle of catalysis

DNA transposition and retroviral integration occurs by a series of polynucleotidyl transfer steps, catalysed by the transposase. These are all direct transesterifications. Unlike conservative site-specific recombination by the tyrosine and serine recombinases, transposition does not involve the formation of a covalent protein-DNA intermediate. For hydrolysis of the phosphodiester bond (nicking), the scissile phosphate is transferred to water; for strand transfer, it is transferred to the 3'OH group at the end of the transposon (Figure 1.4). In all transposases tested, the phosphoryl transfer reactions proceed by an S_N2 inline nucleophilic substitution mechanism with an inversion of the stereoconfiguration of

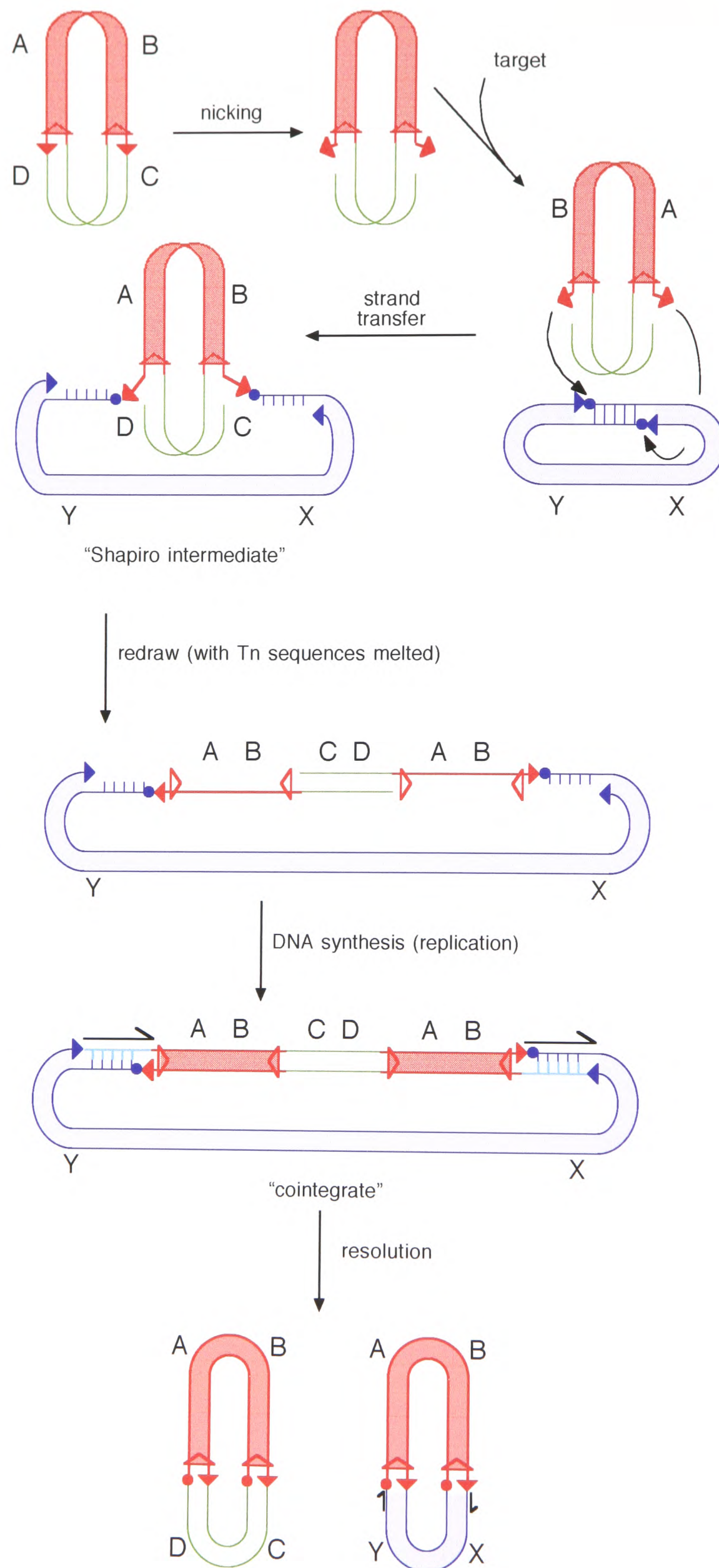


Fig. 1.2

Figure 1.2: The mechanism of replicative transposition

The DNA components of the reactions are shown, but supercoiling and protein components have been omitted for clarity. The transposon and donor (red and green, respectively) are fused with the target molecule (blue) to form a Shapiro intermediate, which is converted to a cointegrate by replication. Symbols: open triangles, transposon ends; solid triangles, 3'-hydroxyl groups; solid circles, 5'-phosphate groups; black half arrows, direct repeats created by repair of target sequences after insertion. (Modified from Chalmers and Blot, 1999.)

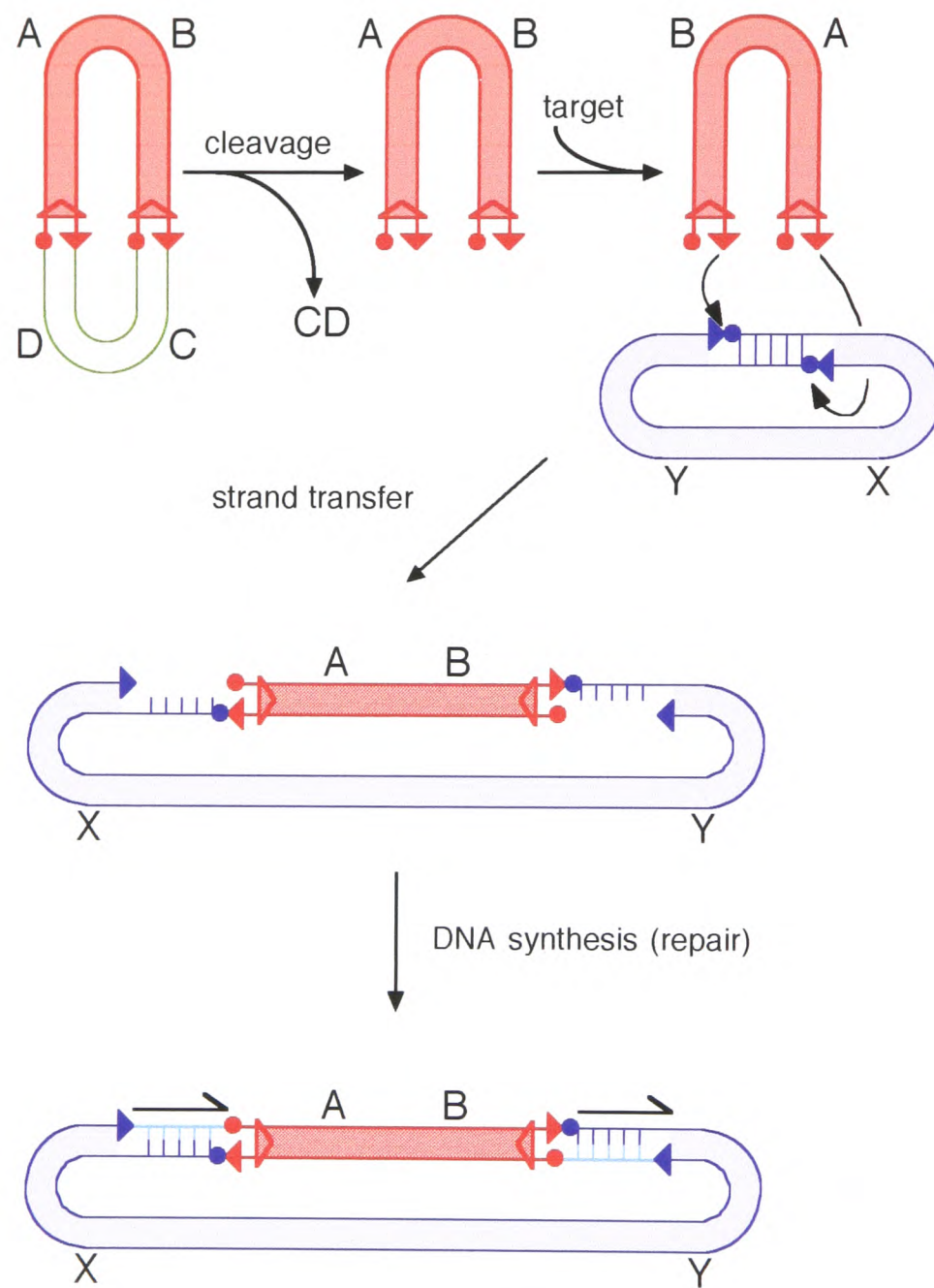


Fig. 1.3

Figure 1.3: The mechanism of non-replicative transposition

In the cut-and-paste mechanism, the first catalytic step is the cleavage of both strands at both transposon ends, thereby separating the transposon (red) from the donor DNA (green). The double strand breaks create 5'-phosphate and 3'-hydroxyl groups at the ends of the transposon. Next, the transposase attaches to a target site, and the 3'-hydroxyl groups at each transposon end undergoes strand transfer to 5'-phosphate groups of the target DNA (blue). The strand bearing the inserting 3'OH is therefore called the “transferred strand” (TS) of each transposon end, the opposite strand, which ends on the 5'-phosphate, is called the “non-transferred strand” (NTS). The attacked 5'-phosphate groups of the target DNA are separated by a few basepairs. The length of this stagger, here depicted as 5 bp, varies among transposon species but is specific to each. This overlap is converted to single stranded gaps at both sides of the transposon, which are repaired by the host to characteristic direct repeats in the flanking DNA. Symbols as in Figure 1.2. (Modified from Chalmers and Blot, 1999.)

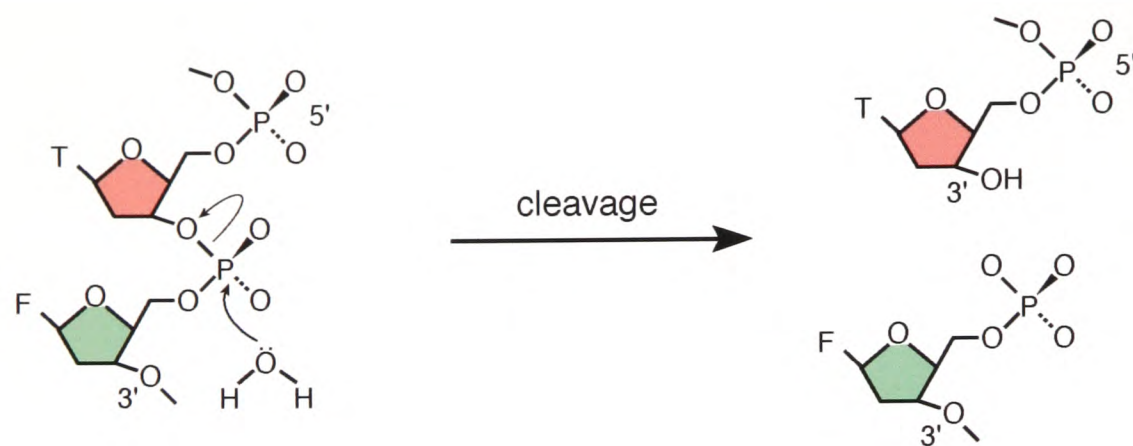
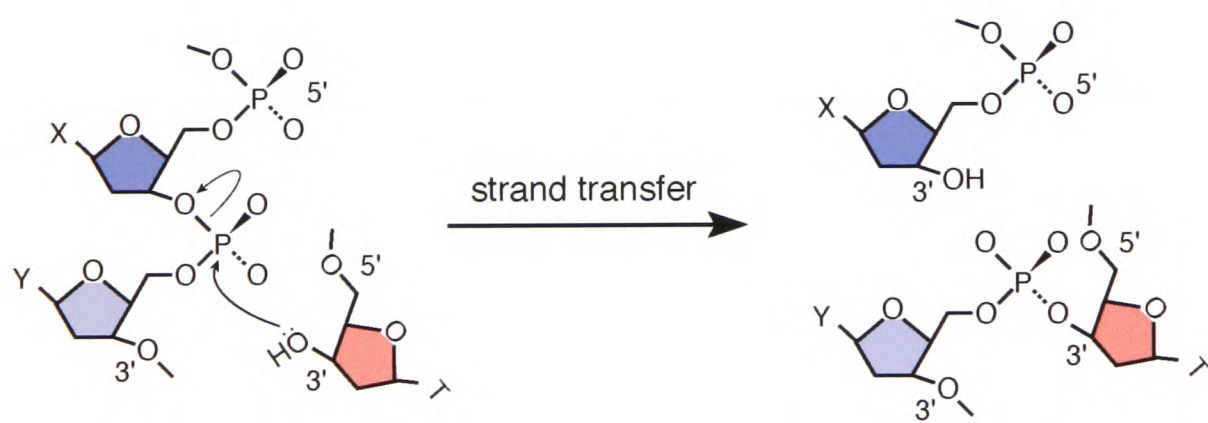
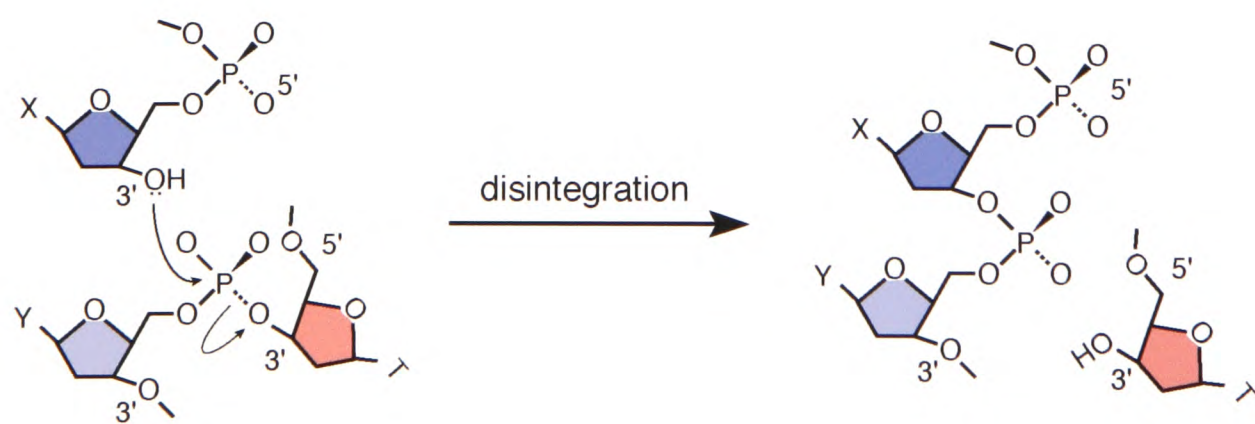
A**B****C**

Fig. 1.4

Figure 1.4: Chemistry of the transesterifications

T, F, X, Y, bases of nucleotides. T, terminal 3' nucleotide of the transposon end; F, first 5' nucleotide of the flanking DNA; X, Y, target DNA.

- (A) Cleavage: Nucleophilic attack of the transposon phosphate backbone by a water molecule, generating a free 3'OH group on the transposon end.
- (B) Strand transfer. Nucleophilic attack of the target phosphate backbone by the 3'OH at the transposon end, joining the transposon to a 5' target end (Y) and creating a free 3'OH group on the neighbouring target base (X).
- (C) Disintegration, the reverse reaction of (B). (Modified from Chandler and Mahillon, 2002.)

the phosphorous centre at each reaction step (Kennedy *et al.*, 2000; Mizuuchi and Baker, 2002).

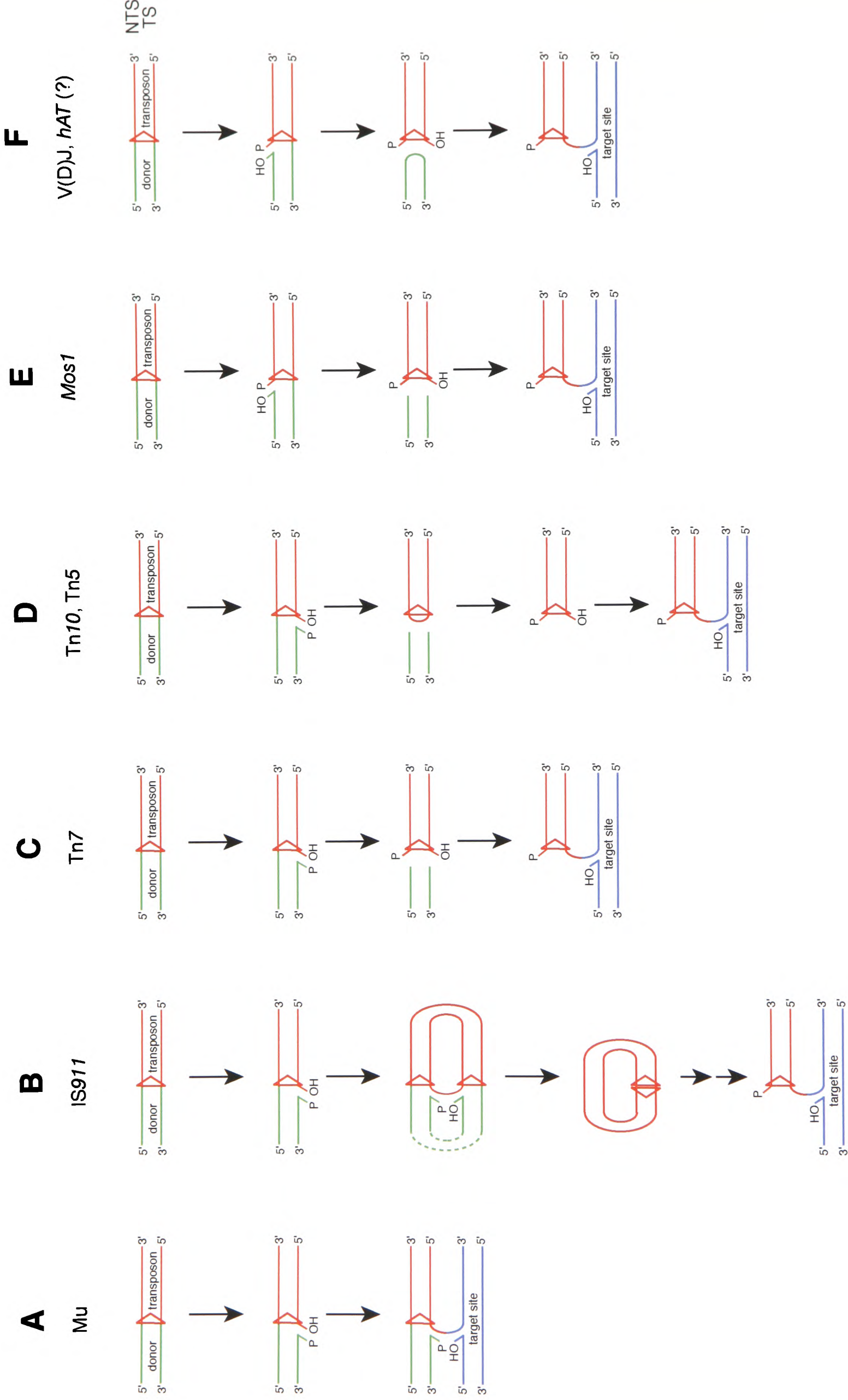
In the proposed catalytic mechanism, divalent metal ions in the active site activate water as a nucleophile by lowering its pK_a . The metal ions also polarise the P—O bonds of the DNA to make the phosphorous centre a better target for nucleophilic attack (Mizuuchi and Baker, 2002).

The active sites of all of the classical DNA transposases contain the so-called DDE motif. This catalytic triad of the amino acids D (asp), D (asp) and E (glu) is shared with several bacterial IS elements, RAG1 V(D)J recombinase, bacteriophage Mu, long terminal repeat (LTR) retrotransposons, retroviruses and other proteins such as DNA polymerases, RuvC and RNaseH (Fayet *et al.*, 1990; Khan *et al.*, 1991; Landree *et al.*, 1999; Doak *et al.*, 1994; Capy *et al.*, 1996). The motif consists of two distant aspartic acid residues (typically more than 90 residues apart), followed by a glutamic acid - or rarely aspartic acid - residue at a distance of approximately 35 residues. The DDE motif coordinates the two divalent cations (Mg^{2+} or Mn^{2+}) necessary for catalysis (Kulkosky *et al.*, 1992; Mizuuchi, 1992b; Rice *et al.*, 1996). Mg^{2+} is usually the physiological metal ion. In some systems, other metal ions such as Mn^{2+} or Ca^{2+} can support some or all of the steps. However, they often produce aberrant reaction products (Hiom *et al.*, 1998; Savilahti *et al.*, 1995; Sakai and Kleckner, 1997; Mizuuchi, 1997).

1.1.2 Order of strand transfers

Many different variations on the order of catalytic steps during excision are employed by different transposons (Figure 1.5). In most transposons, the first nick cleaves the transferred strand, generating the 3'OH at the end of the transposon (Figure 1.5A-D). However, in V(D)J recombination, and probably *hAT* and *Tc1/mariner* transposition, the non-transferred strand is nicked first (Figure 1.5E,F). In non-replicative transposition, the transposon is released either by a second nick on the opposite strand (Figure 1.5C,E), or by hairpin formation through attack of the opposite strand by the free 3'hydroxyl (Figure 1.5D,F). Depending on the polarity of the first nick, the hairpin is formed on either the transposon end or the flanking DNA (Figure 1.5D,F, respectively). The transferred strand (TS) is generally cleaved precisely at the transposon end, while the non-transferred strand (NTS) is cleaved at different positions, leading to a blunt end (*Tn10*), a staggered end with 5' overhang (*Tn7*), or a 5' recess (*Tc1/mariner*).

Fig. 1.5



replicative _____ non-replicative _____ eukaryotic _____

prokaryotic

Figure 1.5: Order of catalytic steps in excision and insertion

Only one transposon end is shown. Red, transposon DNA; red open triangle, ITR or RSS; green, donor DNA; blue, target; P, phosphate group.

- (A) Replicative transposition, cointegrate formation. The transposon end is cleaved at only one strand. (Based on Mizuuchi, 1992a.)
- (B) Intramolecular strand transfer, resulting in “figure-eight” formation in a circular substrate (third from top) and circularisation of the transposon by replication for release from the donor DNA. (Based on Ton-Hoang *et al.*, 1999; Rousseau *et al.*, 2002.)
- (C) Direct cleavage of both strands, each with a different protomer. In Tn7, the first (3') and second (5') nick are carried out by TnsB and TnsA, respectively. The second nick is usually 5' of the transposon end, creating a 5' overhang (not shown). (Based on Sarnovsky *et al.*, 1996; Biery *et al.*, 2000; Lu and Craig, 2000.)
- (D) Cleavage by hairpin formation on the transposon end. One transposase protein catalyses all steps at one end: nicking, hairpin formation, resolution and insertion. (Based on Kennedy *et al.*, 1998.)
- (E) Direct cleavage of both strands, first nick on the NTS. The first nick is usually 3' of the transposon end, creating a 3' staggered end (not shown). (Based on Dawson and Finnegan, 2003.)
- (F) Cleavage by hairpin formation on the flanking DNA; first nick on the NTS. (Based on van Gent *et al.*, 1996a.)

Insertion, on the other hand, is very uniform: The 3'OH on the transferred strand attacks the target DNA making a covalent bond. Cleaved target DNA is not detected in the absence of strand transfer (Chandler and Mahillon, 2002).

1.1.3 Nucleoprotein complexes

A key event early in transposition is the recognition and synapse of the transposon ends into a high order nucleoprotein complex. This reaction defines the DNA segment that forms the transposon, and permits coordinated cleavage and joining reactions at the transposon termini (Haren *et al.*, 1999). The synaptic complexes must be assembled completely before the transposases become catalytically active. Once the complex is assembled, transposase maintains the synaptic complex until the multiple steps of the reaction have been completed. This reduces the occurrence of incomplete recombination reactions (Mizuuchi and Baker, 2002).

Some complexes have been studied in detail, and important differences have been found. For example, the number of recombinase molecules per synaptic complex varies: In the simplest case, there are two monomers of transposase, e.g. in Tn5 and Tn10 (Bhasin *et al.*, 2000; Kennedy, 1999). Phage Mu and *Sleeping Beauty* complexes contain four monomers (Mizuuchi *et al.*, 1992; Izsvak *et al.*, 2002). In complete contrast, Tn7 forms a multiprotein complex with a transposase heterodimer (TnsA and TnsB) at each end plus two additional proteins, which mediate target interaction (Craig, 2002a). The V(D)J precleavage synaptic complex is made of 3-4 RAG1 and 2 RAG2 molecules (see Section 1.2.3.1) (Swanson and Desiderio, 1999; Landree *et al.*, 2001; Mundy *et al.*, 2002).

1.1.4 Eukaryotic transposons

The best-known families of eukaryotic DNA transposons are the *P* element family, the *Tc1/mariner* superfamily, the *hAT* family, the CACTA (*En/Spm*) superfamily, and plant *MuDR/Mu* elements. Most of these have terminal inverted repeats 8-50 bp long and encode a transposase with an N-terminal ITR-binding domain and a C-terminal catalytic domain (Augé-Gouillou *et al.*, 2000). Many eukaryotic transposons have been studied genetically, but only a few are understood on a molecular level. However, compared to the best-studied model transposons from prokaryotes, very little is known about their molecular mechanism of transposition. Most detailed knowledge is available from the V(D)J recombination system, which, although not a transposon in the narrow sense, shares many fundamental molecular aspects with transposition.

1.2 V(D)J recombination

V(D)J recombination is the only known site-specific recombination system in higher eukaryotes. It is now widely accepted that it originates from a transposon (Section 1.2.1). It shares all important aspects of the transposition mechanism and appears to be related to both *hAT* and *Tc1/mariner* transposons. Due to its medical relevance, V(D)J recombination has been studied thoroughly. This has led to a very detailed knowledge about its molecular mechanism of recombination. Far less details are known for other eukaryotic systems. The V(D)J reaction mechanism is presented here to demonstrate important aspects of eukaryotic transposition.

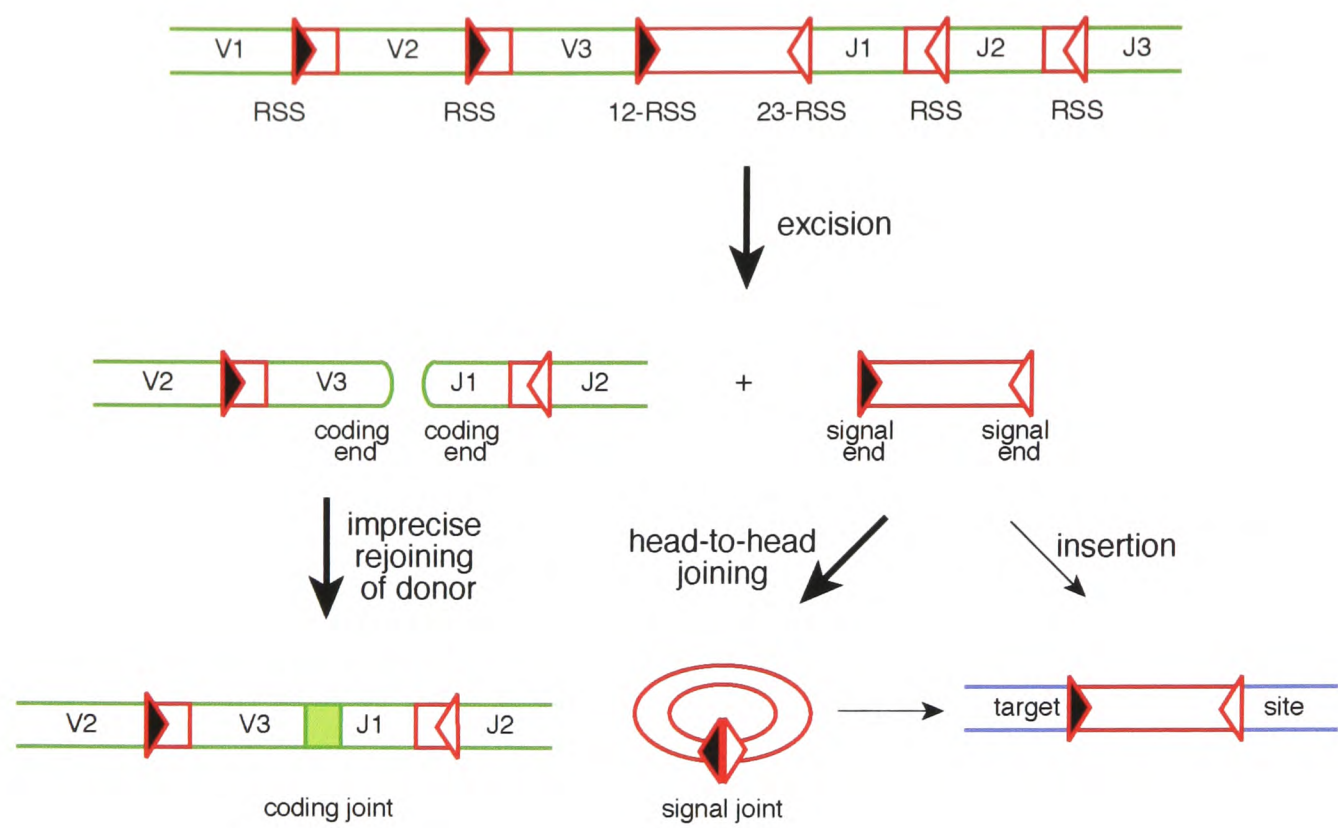
V(D)J recombination is the mechanism that assembles immunoglobulin (Ig) and T cell receptor (TCR) genes in developing pre-B and pre-T cells from separate gene segments, the variable (V), joining (J) and sometimes diversity (D), elements (Figure 1.6A). Variability is created because there are several different V, D, and J elements, which can be joined in almost every combination, and because the joining *per se* is usually imprecise. Recombination occurs at the recombination signal sequences (RSS), which flank each segment. The main enzymes to catalyse the recombination events are RAG1 and RAG2 (recombination activating gene 1 and 2) (Oettinger *et al.*, 1990).

1.2.1 V(D)J is like a transposon

In all species tested, the RAG1 and RAG2 genes are nearest neighbours, convergently transcribed, and in most species, they lack introns. The level of conservation of both protein sequences among species is very high: between 50 and 90% at the amino acid level. This, together with their highly unusual genetic structure, suggests that the RAG genes arrived in vertebrates by horizontal transfer (Oettinger *et al.*, 1990; Thompson, 1995).

The arrangement of RSSs in the Ig and TCR loci resembles the inverted repeats at transposon ends (Sakano *et al.*, 1979). They are organised as interrupted inverted repeats, each consisting of a conserved heptamer and a conserved nonamer motif separated by non-conserved spacer regions of 12 ± 1 bp and 23 ± 1 bp. Efficient recombination requires one 12-spacer and one 23-spacer RSS, analogous to the pairing of one left and one right transposon end (Tonegawa, 1983; Eastman *et al.*, 1996). RSSs are highly conserved, and the same recognition sequences are used in all species which have V(D)J recombination. Slight variations in the consensus sequence are tolerated, but reduce the efficiency of recombination (Akamatsu *et al.*, 1994; Feeney *et al.*, 2000; Lampe *et al.*, 2001). The

A V(D)J recombination



B *hAT* transposition

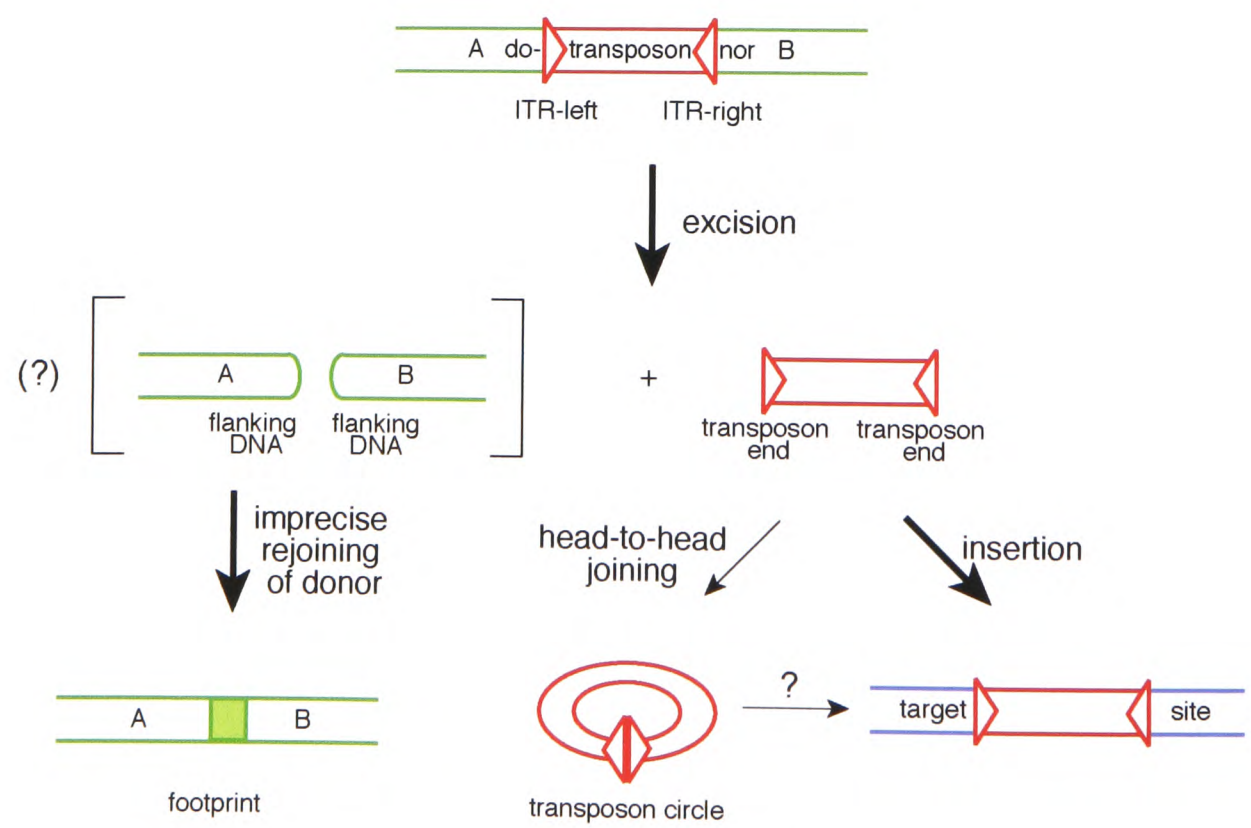


Fig. 1.6

Figure 1.6: Overview of V(D)J recombination and *hAT* transposition

The two recombination systems are very similar in overall structure, and in the mechanism of excision and repair of the empty donor site. In both cases, the DNA is cut just outside recognition sequences organised as inverted repeats (RSSs or ITRs). As postulated for *hAT* transposition and shown for V(D)J recombination, excision proceeds via a hairpin intermediate in the flanking DNA. Repair of the empty donor sites creates coding joints or footprints of very similar structure. The excised piece is either joined head-to head or inserted into a new target site. Black filled triangles, 12-RSS; white filled triangles, 23-RSS or ITRs.

heptamer motif of the RSSs is homologous to the conserved sequence motifs in the ITRs of both the *hAT* family and the *Tc1/mariner* superfamily of transposons (Figure 1.7). Significantly, the conservation is at the very end where the chemistry of the reaction takes place. The conserved bases probably interact directly with functional groups in the active site of the transposase or the RAG proteins.

Mechanistically, V(D)J recombination appears to be most closely related to *hAT* transposition. Although the identity of some of the intermediates and proteins involved remain uncertain, the overall DNA strand exchange events in both systems are quite similar (Figure 1.6). In both cases, DNA sequences lying between the two transposon ends, or the RSSs in V(D)J, is excised. In V(D)J, hairpin ends are formed on the flanking DNA and these are resolved and joined imprecisely to create the coding joint, and hairpins have also been postulated for *hAT* excision. This mechanism of hairpin resolution contributes to the diversity in V(D)J recombination and *hAT* footprints. The major difference between V(D)J and transposition is the fate of the excised piece of DNA. In V(D)J recombination, it is usually lost after head-to-head joining of the RSS to form signal joints (Fujimoto and Yamagishi, 1987; Okazaki *et al.*, 1987; Roth *et al.*, 1992b), whereas in transposition, the transposon is inserted into a new target site. However, both types of product have been detected in both *hAT* transposition and V(D)J recombination (Sections 1.2.3.4; 1.3.1.2), only the efficiency of the respective reactions differ by several orders of magnitude.

As for transposases and integrases, the stereochemistry of the reaction shows that V(D)J hairpin formation is by a direct transesterification mechanism without covalent intermediates (van Gent *et al.*, 1996a) (see Section 1.1.1). Furthermore, the first nicking step mediated by both RAG recombinase and HIV integrase can use alcoholysis instead of hydrolysis, further substantiating the evolutionary relatedness (Vink *et al.*, 1991; van Gent *et al.*, 1996a). Additionally, HIV integrase inhibitors interfere with the DNA cleavage and disintegration activities of RAG1/2, suggesting a close mechanistic relationship (Melek *et al.*, 2002).

1.2.2 The RAG recombinase

In V(D)J recombination, RAG1 and RAG2 recognise the cleavage site and catalyse the nicking and strand transfer steps of the excision reaction (McBlane *et al.*, 1995; Schatz, 1997). Due to low activity and solubility of the full-length proteins, *in vitro* experiments have been carried out with truncated proteins, comprising aa 384-1008 of the 1040 aa RAG1 (Sadofsky *et al.*, 1993) and aa 1-383 of the 527 aa RAG2 (Cuomo and Oettinger,

Fig. 1.7

family	element	terminal inverted repeat / RSS 5' → 3'	host	target site duplication
hAT	<i>Tam3</i>	TAAAGATGTGAA	<i>Anterrhinum majus</i>	8 bp
	<i>Ac</i>	CAGGATGAAA	<i>Zea mays</i>	
	hobo	CAGAGAACTGCA	<i>Drosophila melanogaster</i>	
	<i>Hermes</i>	CAGAGAACAA ^{AA} CAACAAG	<i>Musca domestica</i>	
	<i>Ascot-1</i>	CAGTGTCTCAA ^{CA} AGTC ^{AA} GTCCGGC	<i>Ascobolus immersus</i>	
	<i>restless</i>	CAGAGTGCGTAATCAACCAA	<i>Tolypocladium inflatum</i>	
Tc1/mariner	<i>V(D)J</i>	CACAGTG (12 / 23) ACAAAACC	vertebrates	5 bp
	<i>Tc1</i>	t _a CAGTGCTGGCCAAAAGATATCCACTTTTGGT + 22bp	<i>Caenorhabditis elegans</i>	2 bp
	<i>Tc3</i>	t _a CAGTGTTGGGAAAGTTCTATAGGACCCCCCTA + 430bp	salmonid fish	
	<i>SB</i>	t _a CAGTTGAAGTCGGAAAGTTTACATACAC ^T T ^A AG	vertebrates	
	<i>impala</i>	t _a CAGTGGGTGCAATAAGTTTGAAATACA ^T CT	<i>Fusarium oxysporum</i>	
	Himar1	t _a ACAGGTTGGCTGATAAGTCCCCGGTCT ^G A ^A CA	<i>Haematobia irritans</i>	8 bp
P	<i>Mos1</i>	t _a ^C _T CAGGTGTACAAAGTA ^G _A ^G AAATGTCG ^G _T ^T T	<i>Drosophila mauritiana</i>	
	<i>P</i>	CATGATGAAATAACATAA	<i>Drosophila melanogaster</i>	

Figure 1.7: Comparison of terminal sequences of eukaryotic transposons and recombination signal sequences

The alignment shows the conservation of the terminal basepairs within transposon families, and that V(D)J RSSs show homologies to the ITRs of both *hAT* and *Tc1/mariner* transposons. Upper case letters, 5'→3' sequence of inverted terminal repeats (or recombination signal sequences); double letters: top, left end; bottom, right end; outlined lower case letters, conserved target site duplication. Shading of bases: dark colour (red, blue), 100% conserved in transposon family; light colour (orange, light blue), 50-99% conserved. *SB*, *Sleeping Beauty*. Origin of sequences: *Tam3* (Hehl *et al.*, 1991); *Activator* (Pohlman *et al.*, 1984a, b); *hobo* (Streck *et al.*, 1986; Calvi *et al.*, 1991); *Hermes* (Warren *et al.*, 1994; O'Brochta *et al.*, 1996); *Ascot-1* (Colot *et al.*, 1998); *restless* (Kempken and Kück, 1996); V(D)J (Hesse *et al.*, 1989); *Tc1* (Rosenzweig *et al.*, 1983); *Tc3* (Collins *et al.*, 1989); *Sleeping Beauty* (Cui *et al.*, 2002); *impala* (Langin *et al.*, 1995); *Himar1* (Lampe *et al.*, 1996); *Mos1* (Jacobson *et al.*, 1986; Augé-Gouillou *et al.*, 2001b); *P* (O'Hare and Rubin, 1983).

1994; Sadofsky *et al.*, 1994). Although these core proteins cause mild defects *in vivo* (Akamatsu *et al.*, 2003), they are sufficient to mediate the whole reaction *in vitro* (Steen *et al.*, 1999).

In V(D)J, the RAG1 protein contains the catalytic functions. The RAG1 core region contains an N-terminal domain for sequence-specific DNA-binding (Difilippantonio *et al.*, 1996; Spanopoulou *et al.*, 1996), a C-terminal DDE motif (Landree *et al.*, 1999; Kim *et al.*, 1999) and a C-terminal region for contacting the flanking donor DNA (Mo *et al.*, 2001; Arbuckle *et al.*, 2001). A number of further protein motifs in RAG1 and RAG2 have been identified, but their attributed functions remain hypothetical (Freemont *et al.*, 1991; Rodgers *et al.*, 1996; Callebaut and Mornon, 1998; Aravind and Koonin, 1999). The function of RAG2 is less well defined. However, it is noteworthy that some point mutations in RAG2 abolish the catalytic activity, implying that the full active site of the RAG recombinase involves parts of both RAG1 and RAG2 (Qiu *et al.*, 2001).

1.2.3 Molecular view of V(D)J recombination

V(D)J recombination can be efficiently reconstructed with plasmid or oligonucleotide substrates, which shows that it is not strictly dependent on any unique chromosomal context or structure (Hesse *et al.*, 1987; McBlane *et al.*, 1995). There is, on the other hand, a bias of recombination frequency based on chromosomal location (Williams *et al.*, 2001) and accessibility (Kwon *et al.*, 1998), which is altered in a developmentally regulated fashion (see Bassing *et al.*, 2002 for a review).

1.2.3.1 Cleavage at the recombination signal sequences

Two dimers of RAG1 and two monomers of RAG2 assemble on a single 12-RSS to form single end complexes (SC), aided by the DNA-bending proteins HMGB1/2 (Figure 1.8A). Upon binding to the RSS, RAG1 undergoes a major conformational change (Ciubotaru *et al.*, 2003; Godderz *et al.*, 2003). For paired complex (PC) formation, the 23-RSS is probably recruited into the SC as naked DNA (Landree *et al.*, 2001; Swanson, 2002; Mundy *et al.*, 2002; Jones and Gellert, 2002)(see Figure 1.8A). The distance between RSS does not seem to greatly effect the efficiency of V(D)J recombination, as recombination between loci that are 2-3 Mb apart are utilised reasonably often (Gellert, 2002).

As mentioned earlier (Section 1.1.2; Figure 1.5F), the first nick by RAG1/2 creates a 5'-phosphate group at the heptamer end and a 3'-hydroxyl group in the flanking DNA

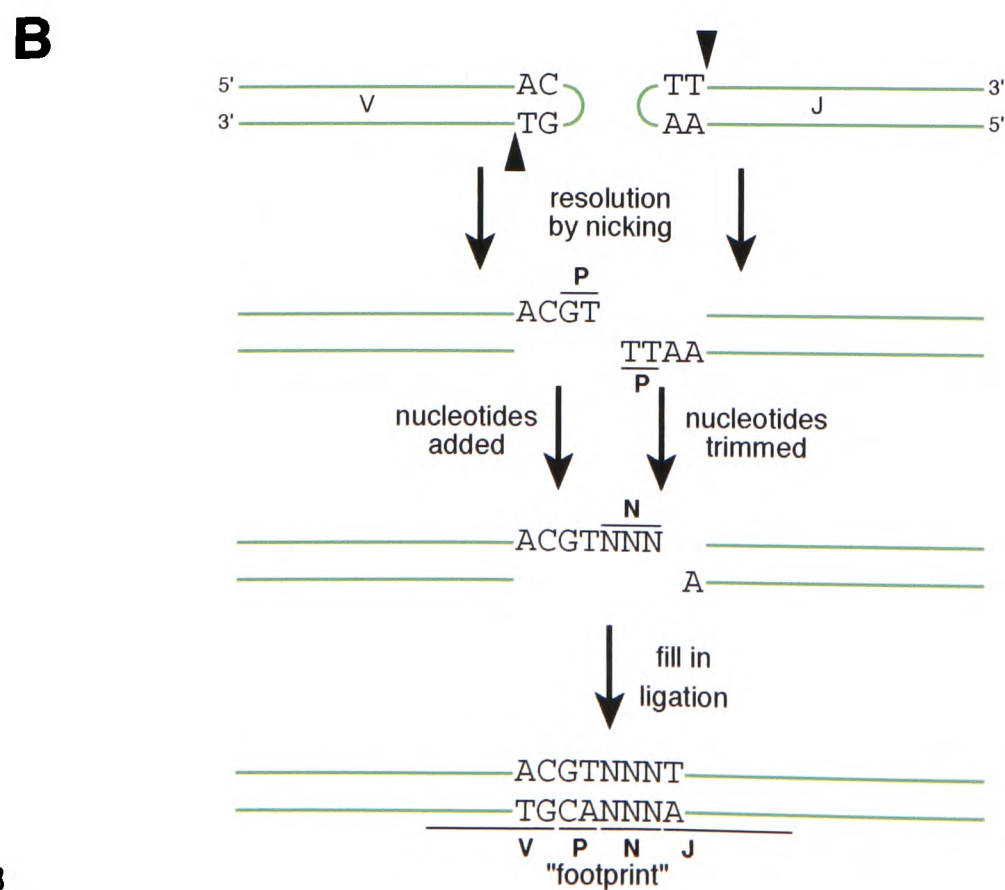
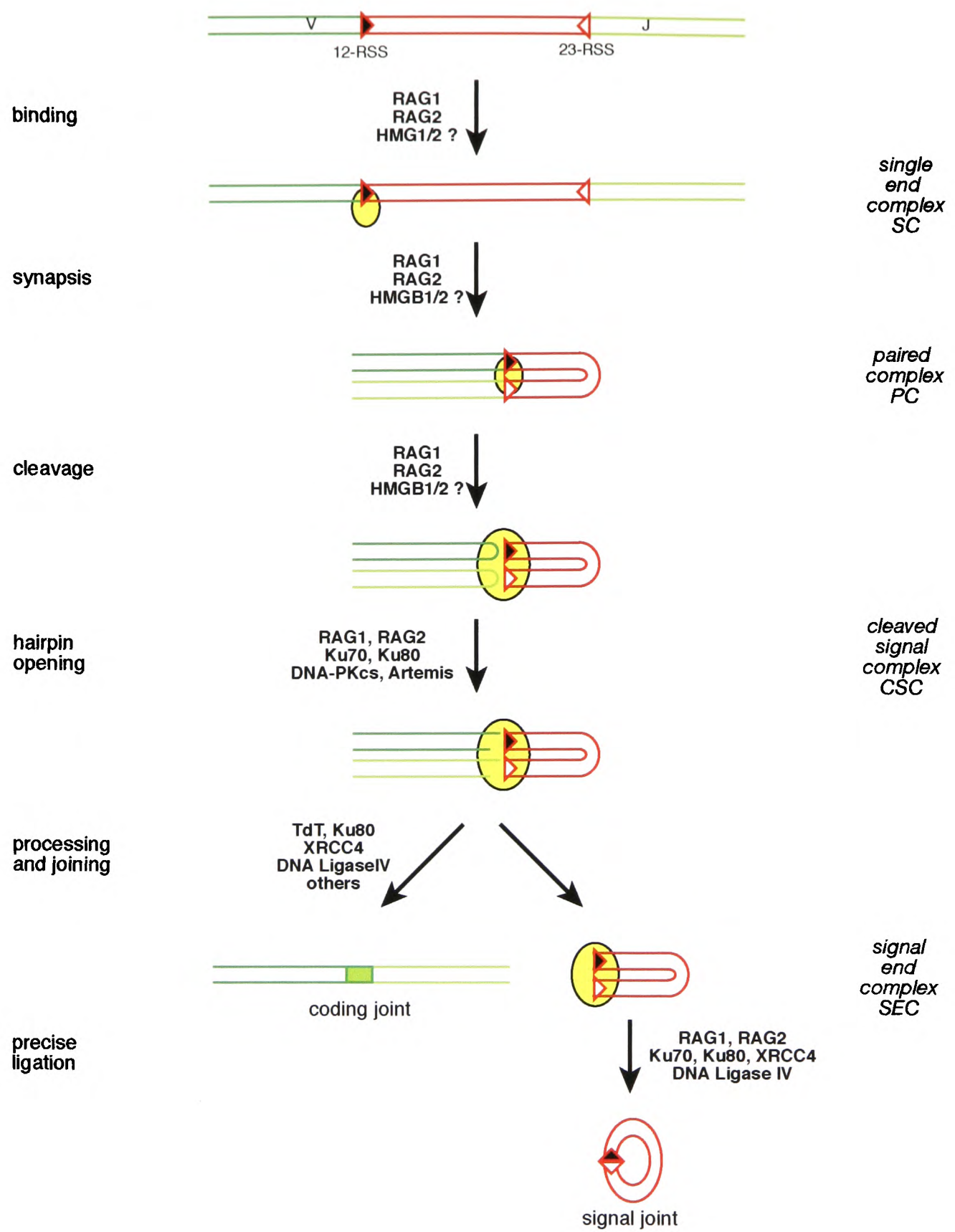


Fig. 1.8

Figure 1.8: Current model of V(D)J recombination and footprint formation

- (A) Schematic model of protein-DNA complexes. The RAG-proteins first recognise the 12-RSS to form stable complexes referred to as the 12-SC (single end complexes). Synapsis with a 23-RSS leads to the formation of the paired complex (PC), within which DNA cleavage takes place. After cleavage, the four ends are held in a cleaved signal complex (CSC), which is likely the complex in which much or all of coding joint formation occurs. This leaves behind the signal end complex (SEC), which is the precursor for signal joint formation. Nicking and hairpin processing are not depicted. Black and white triangles, 12-RSS and 23-RSS respectively; green, coding segments; yellow ovals, protein complexes. (Based on Fugmann *et al.*, 2000; Jones and Gellert, 2001; Schlissel, 2002. Nomenclature from Hiom and Gellert, 1998.)
- (B) Model for the generation of footprints in *hAT* transposition and coding joints in V(D)J recombination. P, palindromic P-nucleotides; N, non-templated N-nucleotides. (Based on Roth *et al.*, 1992a). See text for details.

(McBlane *et al.*, 1995). The 3'-OH group is now free and attacks the opposite strand at the same nucleotide position by direct transesterification, creating a hairpin in the coding end (flanking DNA) and a blunt signal end (transposon end) (Roth *et al.*, 1992a). Hairpinning is very sensitive to the correct RSS sequence and spacer length, whereas nicking is more tolerant of variation (Cuomo *et al.*, 1996; Ramsden *et al.*, 1996).

Cleavage occurs *in trans*, i.e. a RAG1 subunit that is bound sequence-specifically to one RSS, cleaves the opposite RSS. A single DDE triad is responsible for both nicking and hairpin formation (Swanson, 2001).

In the presence of Mg^{2+} , cleavage of the second strand occurs only after synapse formation (coupled cleavage), and there is a preference for mixed 12/23 pairs (van Gent *et al.*, 1996b). Coupling through paired complex formation is not required for nicking or if Mg^{2+} is exchanged for Mn^{2+} (McBlane *et al.*, 1995).

As hairpin formation requires a significant bend in one or both DNA strands at the coding flank/heptamer border, it is to be expected that DNA flexibility at this site is mandatory. In fact, there is evidence that the nucleotides at this border may become partly unpaired during binding and cleavage. Substrates with a double stranded coding sequence and a single stranded RSS are efficiently cleaved and hairpinned, and unpairing the first few nucleotides of the coding sequence enhances cleavage (Cuomo *et al.*, 1996; Ramsden *et al.*, 1996; Kale *et al.*, 2001). Furthermore, some RAG1 mutants are capable of nicking but defective in hairpin formation; this defect is rescued by mismatched bases at the cleavage site (Kale *et al.*, 2001). The unwinding might be aided by the sequence of the first 4 bp of the RSS. The CACA/GTGT sequence has an intrinsically distorted structure in both solution and crystals (Cheung *et al.*, 1984; Timsit *et al.*, 1991).

1.2.3.2 Coding joint formation

Following cleavage, there is evidence that the two coding and two signal ends stay together in a cleaved signal complex (CSC) (Lewis *et al.*, 1988; Lewis and Hesse, 1991; Agrawal and Schatz, 1997; Melek *et al.*, 1998) (Figure 1.8A). Whereas the blunt-ended signal ends can potentially be ligated directly, the hairpinned coding ends need to be processed before coding joint formation can occur.

The hairpin processing steps that produce the sequence variations in the coding joints are illustrated in Figure 1.8B. It is identical to the model first proposed to explain the footprints left behind after *hAT* transposition. The hairpins are resolved asymmetrically *in vivo*, generally 1-5 bp from the tip, resulting in 3' overhangs and the creation of the

palindromic P-nucleotides (Lafaille *et al.*, 1989; McCormack *et al.*, 1989; Aguilar and Belmont, 1991; Engler *et al.*, 1992; Gauss and Lieber, 1996; Schlissel, 1998). Before being joined, the coding ends are further modified by addition of non-templated N-nucleotides and/or by deletion of nucleotides (Lewis *et al.*, 1985; Schlissel, 1998). These activities contribute to at least 90% of the diversity between TCRs (Arstila *et al.*, 1999; Cabaniols *et al.*, 2001).

1.2.3.2.1 Processing of coding ends

The hairpin-opening nuclease has recently been established as Artemis, which is responsible for radiation-sensitive severe combined immune deficiency (RS-SCID) in humans (Moshous *et al.*, 2001, 2003; Rooney *et al.*, 2002, 2003; Mansilla-Soto and Cortes, 2003). It forms a complex with the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs). Artemis has a 5' to 3' exonuclease activity until phosphorylated by DNA-PK, when it is able to resolve hairpin ends and cleave 3' and 5' overhanging flaps (Ma *et al.*, 2002). The resolved hairpins contain the 3' overhanging ends similar in structure to those observed in purified DNA from lymphoid progenitors (Ma *et al.*, 2002). In addition, there is evidence of some involvement of the RAG proteins in hairpin opening (Schlissel, 2002; Huye *et al.*, 2002).

The terminal deoxynucleotidyl transferase (TdT) is responsible for modifying the opened hairpins. Tracts of up to 15 bp non-templated N-nucleotides are added by the short splice form, TdTS (Gilfillan *et al.*, 1993; Komori *et al.*, 1993). Ku80 is required for this activity (Bogue *et al.*, 1997; Purugganan *et al.*, 2001). Added nucleotides are not completely random, as TdT has a preference for G residues (Basu *et al.*, 1983) and strings of purines or pyrimidines (Gauss and Lieber, 1996).

The long spliceform of TdT (TdTL) acts as a 3'→5' exonuclease both *in vitro* and *in vivo* and is capable of deleting nucleotides from coding ends (Thai *et al.*, 2002). Other as yet unidentified exo- or endonucleases are probably involved in nucleotide loss, as deletions are also observed in TdT^{-/-} mice (Komori *et al.*, 1993; Gilfillan *et al.*, 1993). Expression of both TdT isoforms is normally limited to early lymphoid cells where V(D)J recombination is active (Thai *et al.*, 2002).

1.2.3.2.2 Coding end joining

From analysis of broken DNA ends at antigen receptor loci and from plasmid substrates, it emerged that coding ends are very quickly opened, processed and joined

(Ramsden and Gellert, 1995; Zhu and Roth, 1995), probably within a postcleavage complex that holds all four ends together (Lewis *et al.*, 1988; Lewis and Hesse, 1991).

There are two possibilities for the joining of the processed coding joints (Grawunder and Harfst, 2001). Either the overhangs produced by hairpin opening are filled in by a polymerase and blunt-ligated (as depicted in Figure 1.8B), or the ends are annealed using short microhomologies (Gerstein and Lieber, 1993). This would create 5' and 3' flap structures, which must be trimmed, e.g. by RAG1/2, which has a sequence-non-specific 3' flap endonuclease activity (Santagata *et al.*, 1999). Other factors for trimming might be the 3' to 5' exonuclease activity of Mre11 (Paull and Gellert, 2000) or the 5' flap endonuclease activity of FEN-1 (Harrington and Lieber, 1994).

Aberrant intermolecular coding joint formation can be the cause of lymphoid tumours and thus needs to be tightly controlled. While intermolecular signal joints are produced quite effectively, coding joint formation is up to 1000 fold more efficient between recombination sites on the same molecule (chromosome or plasmid) than intermolecular recombination. (Lipkowitz *et al.*, 1990; Han *et al.*, 1999; Tevelev and Schatz, 2000). The mechanism, however, is not yet clear.

After processing, the double strand break in both coding and signal joints is repaired by non-homologous end-joining (NHEJ). This last step requires the end-binding protein XRCC4, DNA LigaseIV and DNA-dependent protein kinase, consisting of the catalytic subunit DNA-PKcs and the heterodimer Ku70/Ku80 (Frank *et al.*, 1998; Han *et al.*, 1998; Gao *et al.*, 1998; Ramsden and Gellert, 1998). Ku proteins have a high affinity to DNA ends (reviewed in Dynan and Yoo, 1998). Ku70 and Ku80 form a ring that encircles duplex DNA, a structure which seems well designed to structurally support broken DNA ends and to bring the DNA helix into phase across the junction during end processing and ligation (Walker *et al.*, 2001). The final ligation of coding and signal joints is catalysed by DNA LigaseIV in complex with tetrameric XRCC4 (Gao *et al.*, 1998; Grawunder *et al.*, 1998; Junop *et al.*, 2000; Grawunder and Harfst, 2001). This complex interacts with Ku at DNA ends (Modesti *et al.*, 1999; Teo and Jackson, 2000; Chen *et al.*, 2000; Nick McElhinny *et al.*, 2000) and might be regulated by Ku or the DNA-PK holoenzyme (Grawunder and Harfst, 2001).

RAG1 and RAG2 are also involved in joining, as shown by defects of separation-of-function mutants of the recombinase (Schultz *et al.*, 2001; Qiu *et al.*, 2001; Fugmann and Schatz, 2001). Some of the RAG1/2 mutants result in both defective signal and coding joints, others are specifically impaired at coding joint formation. Some of the latter are also

defective for hairpin opening *in vitro*. Other factors, such as a gene mapped to human chromosome 10, are likely involved in coding joint formation (Moshous *et al.*, 2000).

1.2.3.3 Signal joint formation and cleavage

With the normal orientation of RSSs, signal end joining results in extrachromosomal circles. Signal joints are usually precise ligations of the blunt signal ends (Schlissel *et al.*, 1993); they can contain non-templated insertions, but nucleotide loss from the signals is very rare (Lieber *et al.*, 1988).

Virtually all of the cleaved signal ends remain associated with each other and RAG1/2 in the stable nuclease- and heparin-resistant signal end complex (SEC) (Hiom and Gellert, 1998; Agrawal and Schatz, 1997; Jones and Gellert, 2001). Only in SECs, but not in cleaved signal complexes (CSCs), which contain the RSSs, can the signal ends be joined by NHEJ (Jones and Gellert, 2001)(see Figure 1.8A).

In wt mouse thymocytes and bone marrow cells, signal ends are much more abundant than coding ends, which are processed quickly. Furthermore, signal joints only appear when RAG expression is downregulated as the cells mature (Ramsden and Gellert, 1995; Zhu and Roth, 1995; Livak and Schatz, 1996). It appears that ligated signal joints are efficiently cleaved by RAG1/2 (Neiditch *et al.*, 2002). Thus, signal joints are probably formed quickly, but immediately resolved again until the down-regulation of RAG1/2. Signal joint cleavage proceeds *in cis* and employs a nick-nick mechanism which does not involve synaptic complex or hairpin formation: RAG proteins bound to one RSS introduces a single-strand nick on one strand, while RAG proteins on the other RSS either simultaneously or sequentially nick the other strand, forming two blunt signal ends. These signal ends are relatively efficiently used as substrates for transposition (double-ended insertion) *in vitro* (Neiditch *et al.*, 2002).

1.2.3.4 Transposition

In vitro, RAG1/2 bound signal ends transpose at a low frequency into unrelated target DNA (Agrawal *et al.*, 1998; Hiom *et al.*, 1998). This transposition is strongly dependent on the presence of HMGB1/2. It typically requires both types of RSSs, although a low level of strand transfer is observed with 12-RSSs only. Transposition reactions have been reconstituted with oligonucleotides, plasmids, uncleaved and precleaved substrate. Both single-ended and double-ended insertions are observed (Melek and Gellert, 2000).

In RAG-mediated transposition, the RSSs are cut exactly at the heptamer border and inserted into target. In double-ended insertions, the target 5'-phosphates are 5 bp apart. However, more variation (3-5 bp) is observed during intramolecular transposition (Hiom *et al.*, 1998; Agrawal *et al.*, 1998). No specific target sequence has been identified, but there is a modest preference for sequences with a high GC content (Hiom *et al.*, 1998). Furthermore, similar to Tn7 (Rao *et al.*, 2000; Rao and Craig, 2001; Peters and Craig, 2001), there is a strong preference for insertion into distorted DNA structures, and RAG1/2 targets especially hairpin ends (Lee *et al.*, 2002).

RAG-mediated transposition events into unrelated sequences have only very recently been confirmed in living cells (Messier *et al.*, 2003). Earlier, several chromosomal rearrangements observed in cell culture and lymphoid tumours were explained by single ended insertions (Roth and Craig, 1998; Lee *et al.*, 2002). It appears that the C-terminus of full-length RAG2 and physiological concentrations of GTP inhibit transposition by preventing the formation of target capture complexes (Elkin *et al.*, 2003; Tsai and Schatz, 2003).

1.3 *hAT* transposons

The *hAT* superfamily is one of the largest groups of eukaryotic transposons and includes the first transposon identified, *Activator/Dissociation* (*Ac/Ds*) (McClintock, 1948, 1949). The family was named after its first three members identified: *hobo* (from *Drosophila melanogaster*), *Activator* (from *Zea mays*) and *Tam3* (from *Antirrhinum majus*) (Calvi *et al.*, 1991). Sequencing projects and analysis of examples of hybrid dysgenesis and other types of genetic instability have identified other members of the family in the three eukaryotic kingdoms plants, animals and fungi (listed in Kempken and Windhofer, 2001). The transposons of this group are categorised primarily on the basis of sequence similarities between their ITRs (Figure 1.7), similarities among the transposases and the formation of 8 bp target site duplications (Rubin *et al.*, 2001; McGinnis *et al.*, 1983; Streck *et al.*, 1986).

1.3.1 The mechanism of *hAT* transposition

1.3.1.1 Excision in *hAT* transposition

hAT transposition is non-replicative (Atkinson *et al.*, 1993). After excision of the transposon, the broken DNA ends at the donor site are often rejoined by illegitimate

recombination, and characteristic sequence variations, the “footprint”, are left behind. Most of these footprints include self-complementary and non-templated insertions like the P- and N-nucleotides in the coding joints of V(D)J recombination (see Figure 1.8B).

An excision mechanism involving 5' nicks and hairpins in the flanking DNA (as was later shown for V(D)J recombination, see Section 1.2.3; Figure 1.5F) was first proposed to explain the footprints of *Tam3*. This model is consistent with the footprints found after excision of *Tam3* and *Ascot-1* from chromosomes (Coen *et al.*, 1986; Colot *et al.*, 1998), *hobo* from plasmids within fly embryos (Atkinson *et al.*, 1993; Handler and Gomez, 1995) and *Ac* from plasmids in budding yeast (Weil and Kunze, 2000). On the other hand, chromosomal excision footprints of *Ac* lack palindromic insertions and have been explained by the “template switching model”, which involves nicks on both strands at each end of the 8 bp target site duplication, coupled with DNA-synthesis by DNA polymerase (Peacock *et al.*, 1984; Saedler and Nevers, 1985). Although the hairpin model appears to be by far the most likely, it has not been proven yet for any *hAT* transposon. Attempts to detect hairpins in empty donor sites after *Ac* excision from maize chromosomes have not been successful (Gorbunova and Levy, 2000). The formation of double strand breaks without hairpin formation, at least part of the time or under certain conditions, cannot be excluded with the present data.

1.3.1.2 *hAT* transposon circles

The similarities of *hAT* transposition and V(D)J recombination include the formation of transposon circles, joined head-to-head as for V(D)J signal joints (Figure 1.6). Extrachromosomal circles of the *hAT* transposons *Ac/Ds* and *restless* were found by PCR and plasmid rescue (Gorbunova and Levy, 1997a, 2000; Kempken and Kück, 1998). On the assumption that these circles were transposition intermediates, as in the transposition of the bacterial transposons Tn4555 and IS911 (Smith and Parker, 1993; Ton-Hoang *et al.*, 1998), both groups have unsuccessfully searched for the reinsertion of these transposon circles. The negative result does not prove the case, but it seems likely that the circles are not intermediates, but rather a dead end and an alternative to insertion. The same seems to apply for the *C. elegans* transposon Tc1 (Radice and Emmons, 1993).

Unlike signal joints in V(D)J, the majority of *Ac* circles contain insertions or deletions at the junction (Gorbunova and Levy, 2000). Similar modifications were observed on free linear transposons and after circularisation of non-transposon linear extrachromosomal DNA (Gorbunova and Levy, 1997b, 2000). It is therefore likely that the

Ac circles are generated by the host DNA repair machinery from linear transposition intermediates. The insertions and deletions observed at the junctions of the transposon circles support the hairpin model of excision, with a 1 bp 5' overhang on the transposon end (Gorbunova and Levy, 2000).

1.3.1.3 *hAT* insertion

Ac, *Tam3* and *hobo* all show a preference for short-range transposition (Jones *et al.*, 1990; Dowe *et al.*, 1990; Hudson *et al.*, 1990; Ladevèze *et al.*, 2001). This, and the structure of a chromosomal inversion flanked by *Tam3* elements suggests an association of donor and target sites during transposition (Robbins *et al.*, 1989; Kunze and Weil, 2002). The latter study also concludes that the 3' end of *Tam3* probably inserts before the 5' end (Robbins *et al.*, 1989). The structure of another *Tam3*-flanked inversion suggested that the two transposon ends remain closely associated in a complex throughout transposition (Lister *et al.*, 1993). This is similar to the cleaved signal complex in V(D)J recombination (Section 1.2.3.2).

1.3.1.4 *hAT* transposase structure, function and activity

So far, only the transposase of *Ac* has been studied in any detail (reviewed in Kunze, 1996; Kunze and Weil, 2002). The N-terminus contains a DNA-binding domain that partially overlaps with three nuclear localisation sequences (NLSs) (Boehm *et al.*, 1995). The C-terminus contains a 50 aa dimerisation domain, which is highly conserved among *hAT* transposases, an additional multimerisation domain, and probably the catalytic DDE residues (Essers *et al.*, 2000). Transposase binding to the terminal and subterminal ITRs has been shown (Feldmar and Kunze, 1991; Kunze and Starlinger, 1989; Becker and Kunze, 1996, 1997), but cleavage of the DNA substrate has not yet been observed. *Hermes* transposase requires its N-terminal 110 bp for efficient nuclear localisation. This region is different from the *Ac* NLSs, but highly homologous to the N-terminus of *hobo* transposase (Michel and Atkinson, 2003).

The plant transposons *Ac* and *Tam3* are inactivated by methylation of the transposase-binding sites and the promoter region (Wang and Kunze, 1998; Ros and Kunze, 2001; Kitamura *et al.*, 2001). Overexpression of *Ac* transposase was found to reduce the efficiency of transposition, probably through aggregation of the protein (Kunze *et al.*, 1993; Scofield *et al.*, 1993; Heinlein *et al.*, 1994; Weil and Kunze, 2000).

1.3.2 Study of the *hAT* transposition process

The structure of *hAT* transposons and their donor and acceptor sites are well characterised (Coen *et al.*, 1986; Calvi *et al.*, 1991; Baran *et al.*, 1992; Atkinson *et al.*, 1993; Warren *et al.*, 1994; Scott *et al.*, 1996; Colot *et al.*, 1998; Yamashita *et al.*, 1999), as is their frequency and specificity of transposition in different hosts and tissues (Smith *et al.*, 1993; O'Brochta *et al.*, 1994; Yannopoulos *et al.*, 1994a, b; Pinkerton *et al.*, 1996; Ladevèze *et al.*, 2001). Many of the transposases have been identified and cloned (Blackman *et al.*, 1989; Kunze and Starlinger, 1989). However, almost all of the DNA intermediates and protein/DNA complexes involved in the reaction remain obscure.

Although *hAT* transposition appears to share many characteristics with V(D)J recombination, it cannot be concluded that the mechanism of recombination is identical. The regulation of the steps during the reaction will probably be different: V(D)J recombination has been adapted to its role in the immune system and is maximised for excision at a specific developmental stage and minimised for potentially harmful insertions. *hAT* transposition, on the other hand, is presumably adapted for maximum spread and persistence of the transposon in so far as this is not in conflict with the fitness of the host. On the molecular level, there is plenty of scope for differences. The hairpin model of excision fits the available data best, but there is not enough data available to exclude other models, and no intermediate, including nicked or hairpinned DNA, has yet been identified. Attempts to detect a hairpin, the defining intermediate, which would clearly identify the reaction mechanism, have, as mentioned earlier, been unsuccessful (Gorbunova and Levy, 2000). Most other molecular details, such as the composition and behaviour of nucleoprotein complexes, the action of host proteins and the exact mechanism of insertion, are still completely unknown.

If the mechanism of *hAT* transposition was indeed identical to V(D)J recombination in all important aspects, this would be a further indication that both have the same evolutionary origin. At the beginning of this study on the mechanism of *hobo* transposition in 1998, RAG1/2-mediated transposition had not been shown yet. A further objective of this study was therefore to establish a well-functioning transposition system, which would also help the understanding of recombination in the immune system.

Identification of the molecular mechanism requires a simple and controllable system, ideally an *in vitro* system. Here, conditions and components can be chosen, for example, by excluding the confounding effects of DNA repair systems. Also, intermediates appear

in stoichiometric amounts and can thus be analysed and their order in the pathway determined.

1.3.3 The *hobo* transposon

1.3.3.1 Distribution, genetics and molecular structure

hobo elements are found in all members of the *Drosophila melanogaster* species complex: *D. melanogaster*, *D. simulans*, *D. mauritiana*, and *D. sechellia*, as well as in some members of the *melanogaster* subgroup and the related *montium* subgroup. *hobo*-containing strains are designated as “H strains”. The presence of empty (E) strains in most of these species and the low level of divergence of full-length elements among species suggests that active *hobo* elements have been transferred horizontally into or among these species within the last 60 years (Boussy and Daniels, 1991; Simmons, 1992). In *D. melanogaster*, there is evidence for two stages of invasion (Bonnivard *et al.*, 2000).

Crosses between H and E strains of *D. melanogaster* can result in a hybrid dysgenesis syndrome, in which the first generation (F1) offspring show an increased number of mutations due to a burst of *hobo* activity during early development. The observed changes include thermosensitive sterility due to gonadal atrophy, chromosomal breaks and rearrangements, mutations and male recombination (Blackman *et al.*, 1987; Yannopoulos *et al.*, 1987). The molecular basis for *hobo* induced hybrid dysgenesis is not completely understood yet, but is clearly different from the well-studied situation in *P* elements (Bazin and Higuete, 1996; Bazin *et al.*, 1999; Rubin *et al.*, 1982; Corish *et al.*, 1996).

In a new host, *hobo* inserts into genomic hotspots and then transposes mainly locally (Ladevèze *et al.*, 1994, 1998, 2001). 50 generations after the first insertion, amplification slows down (Ladevèze *et al.*, 2001) and stabilises at about 50-60 copies per individual in natural populations (Vieira *et al.*, 1999). The insertion consensus sequence (and 8 bp target site duplication) is GTTTAAAC with a fidelity of 30-90% per base (O'Brochta *et al.*, 1994). Excision of *hobo* results in typical *hAT* footprints of 8-20 additional nucleotides at the empty donor site (Atkinson *et al.*, 1993).

hobo was used as a tool to create mutants in *D. melanogaster* by local transposition (Newfeld and Takaesu, 1999) and deletion between two copies of the transposon (Huet *et al.*, 2002). It has also been employed for transformation in *D. melanogaster* (Blackman *et al.*, 1989; Simmons *et al.*, 2002a, b), *D. virilis* (Lozovskaya *et al.*, 1996; Gomez and

Handler, 1997) and the lepidopteran *Helicoverpa zea* (DeVault *et al.*, 1996), and for enhancer trap screens in *D. melanogaster* (Smith *et al.*, 1993; Newfeld and Takaesu, 2002).

1.3.3.2 The *hobo* HFL-1 element

The ends of *hobo* transposons are perfect 12 bp ITRs (McGinnis *et al.*, 1983). HFL-1, the active *hobo* full-length element used here, is 2959 bp long (Blackman *et al.*, 1989). It has a transposase-coding orf of 1983 bp (ORF1) and a small 96 bp upstream ORF0 of no identified function (Figure 1.1B)(Calvi *et al.*, 1991). Removal of all material 3' of ORF1 does not disrupt transpositional activity in embryos (Calvi *et al.*, 1991). There is no evidence for splicing of the *hobo* transcript (Calvi and Gelbart, 1994). The expression product from an *hsp70* promoter located upstream of the orfs is active for mobilisation of a *hobo* mini-transposon in fly embryos (Calvi and Gelbart, 1994; Lozovskaya *et al.*, 1996).

1.3.3.3 Choice of the *hobo* transposon as an experimental system

The *hobo* transposon from *D. melanogaster* is a typical, yet relatively simple member of the *hAT* family and thus has several advantages as an experimental system. In contrast to the multiple inverted repeats of *Ac*, for example, *hobo* has simple, perfect terminal inverted repeats of 12 bp. The transposase is among the smallest (Calvi *et al.*, 1991) and presumably least complicated in the *hAT* family. Although *hobo* transposase is naturally only transcribed in the germline, protein activity is not restricted to specific tissues (Calvi and Gelbart, 1994) or to only one fly species (O'Brochta *et al.*, 1994), suggesting that the reaction is independent of unusual accessory factors. Since transposase is the only protein encoded by *hobo*, and all autonomous transposons examined to date encode the minimal set of proteins needed for their activity, reconstituting the transposition reaction from a system containing only DNA substrate and purified transposase seemed feasible.

Another advantage of *hobo* is that its *Drosophila* host is a well established system for molecular biology and genetics, and its complete genome sequence has been determined (Adams *et al.*, 2000). It also allows comparison with the *P* elements, the best characterised *Drosophila* transposons. For *P* elements, the donor site is usually repaired by gene conversion without leaving footprints (Engels *et al.*, 1990). Since both *hobo* and *P* elements are naturally active in the *Drosophila* germline, the different fates of the empty donor site must be caused by a different *recruitment* of repair proteins to the breakage point, not by a different protein expression pattern in the host cell. An important question to be answered is how this is regulated and what its significance is.

1.4 Tc1/*mariner* transposons

The Tc1/*mariner* superfamily consists of the closely related *mariner*-like elements (MLE), Tc1-like elements (TLE), maT elements, IS630-like elements and the more distant *pogo*-like elements (Robertson, 1995; Hartl *et al.*, 1997a; Plasterk, 1996; Claudianos *et al.*, 2002; Doak *et al.*, 1994; Smit and Riggs, 1996). The ITmD37E, ITmD37D and ITmD39D families and miniature inverted-repeat transposable elements (MITEs) of the *Stowaway*-type have been proposed to belong to the same superfamily (Shao and Tu, 2001; Turcotte and Bureau, 2002; Feschotte *et al.*, 2003). Members of this superfamily contain a single transposase gene, duplicate a TA dinucleotide pair upon insertion, transposase via a DNA intermediate, and their transposase is expressed in both the germline and the soma.

1.4.1 Distribution and activity

MLEs and TLEs are probably the most widespread DNA transposons in nature. Members of the superfamily have been identified in bacteria (Henikoff, 1992), ciliates (Doak *et al.*, 1994), fungi (Daboussi *et al.*, 1992; Langin *et al.*, 1995; Maurer *et al.*, 1997), plants (Jarvik and Lark, 1998; Tarchini *et al.*, 2000; Feschotte and Wessler, 2002) and most animal phylae from porifera (sponges) to humans (Liao *et al.*, 1983; Robertson, 1995; Robertson and Lampe, 1995b; Robertson *et al.*, 1996; Arkhipova and Meselson, 2000; Augé-Gouillou *et al.*, 1995; Morgan, 1995; Oosumi *et al.*, 1995; Hartl, 1996; Smit and Riggs, 1996). MLEs have now even been found in bdelloid rotifers, a taxon of pseudocoelomata which gave up sexual reproduction (including males, hermaphrodites and meiosis) many millions of years ago, and, until recently, had been thought of being completely devoid of transposon-like sequences (Arkhipova and Meselson, 2000; Welch and Meselson, 2000). Phylogenetic studies of MLE sequences among different hosts demonstrated numerous examples of horizontal gene transfer, suggesting that these elements function autonomously with minimal requirements for host cell function (Lidholm *et al.*, 1991; Maruyama and Hartl, 1991; Kidwell, 1993; Robertson, 1993; Robertson and MacLeod, 1993; Capy *et al.*, 1994; Garcia-Fernández *et al.*, 1995; Robertson and Lampe, 1995a; Robertson, 1995; Lohe *et al.*, 1995b; Plasterk, 1996; Hartl *et al.*, 1997a).

The independence from host-specific accessory factors has been demonstrated *in vivo* by the ability to mobilise several elements in hosts highly divergent from the original. *Mos1* (“mosaic factor”, a particular copy of *Dmmar1*, from *Drosophila mauritiana*), *Himar1* (reconstructed from *Haematobia irritans*), Tc1 and Tc3 (from *Caenorhabditis elegans*), *Sleeping Beauty* (reconstructed from salmonid fishes) and *Minos* (from

Drosophila hydei) were made to transpose in other insects, nematodes, vertebrates, mammalian and avian cell lines, and even in protozoans, archaea and bacteria (listed in Plasterk and van Luenen, 2002). Functional *in vitro* systems were developed for *Tc1* (Vos *et al.*, 1996), *Himar1* (Lampe *et al.*, 1996), and *Mos1* (Tosi and Beverley, 2000), all requiring no other protein but transposase. (This of course does not exclude non-obligatory involvement of other proteins in the natural host.)

Cladistic methods of phylogenetic analysis of the MLEs have produced vastly different trees, dependent on the domain used for the alignment (Robertson and MacLeod, 1993; Robertson, 1995, 1997; Lohe *et al.*, 1995a; Robertson and Asplund, 1996; Gomulski *et al.*, 1997; Augé-Gouillou *et al.*, 2000). This suggests that gene conversions and recombination events have played an important part in the evolution of these transposons, resulting in mosaic elements that could be represented as a network, but not as a simple tree (Augé-Gouillou *et al.*, 2000). However, TLEs and MLEs are clearly separated, and *Mos1* and *Himar1*, the two best-studied MLEs, have been placed consistently into different subfamilies (named after them the *mauritiana* and the *irritans* group, respectively). Common characteristics may therefore apply to the entire *mariner* family.

1.4.2 Structure of *Tc1/mariner* transposons

All elements of the superfamily contain a single gene coding for the transposase and are flanked by terminal inverted repeats. The terminal inverted repeats in *Tc1/mariner* transposons range from 26 bp (*pogo*) to almost 500 bp (*Tc3*), containing one or two transposase binding sites. However, most MLEs identified so far have ITRs of c. 30 bp length (Leroy *et al.*, 2003) (Figure 1.7). In *Sleeping Beauty*, each inverted repeat contains two directly repeated transposase binding sites that have been suggested to form synaptic complexes with tetramers of transposase (Cui *et al.*, 2002; Izsvak *et al.*, 2002).

1.4.2.1 Structure and function of the transposase

Although highly diverse in sequence, the N-terminal domains of both MLE and TLE transposases are responsible for DNA binding. In TLEs, it consists of two helix-turn-helix (HTH) subdomains (PAI + RED = PAIRED), similar to the paired domain of the Pax family of transcription factors (Franz *et al.*, 1994; Vos and Plasterk, 1994; Ivics *et al.*, 1996; van Pouderooyen *et al.*, 1997; Breitling and Gerber, 2000). Between the two DNA binding domains is a conserved GRPR-like motif, characteristic of homeodomain-proteins (Gehring *et al.*, 1994), which is also shared with RAG1 recombinase (Spanopoulou *et al.*, 1996; Difilippantonio *et al.*, 1996) and Hin invertase (Feng *et al.*, 1994). Both subdomains

and the GRPR-motif contribute to DNA binding, but the N-terminal PAI subdomain is most sequence-specific (Izsvak *et al.*, 2002). The RED domain increases this specificity (Fischer *et al.*, 2001). For *Tc1* it has been demonstrated that the GRPR homeo-like domain interacts with the 5' part of the binding site and the paired-like domain with the 3' part (Vos and Plasterk, 1994).

In contrast, the N-terminal domains of MLEs only have one HTH motif and two α -helices, which together provide specific ITR binding (Zhang *et al.*, 2001; Augé-Gouillou *et al.*, 2001a; C. Augé-Gouillou, personal communication). In *Mos1*, the sequence-specific binding domain lies within the first 120 aa of the transposase (Augé-Gouillou *et al.*, 2001a; Zhang *et al.*, 2001). Changing a basic residue within the putative second helix of the HTH motif abolishes DNA binding (Zhang *et al.*, 2001). A variable motif upstream of the HTH motif might correspond to a GRPR-like motif (Robertson, 1995).

Both TLE and MLE transposases have a bipartite nuclear localisation sequence (NLS) in the N-terminal domain, albeit at different positions (Ivics *et al.*, 1996; Augé-Gouillou *et al.*, 2001a). A single amino acid mutation in the *Mos1* NLS abolishes transposase function (Lohe *et al.*, 1997).

The C-terminal catalytic domain always contains a DDE motif which has the form D, D (34) E in TLEs and D, D (34) D in MLEs. Interestingly, if the unusual third D is mutated to E in *Mos1*, the transposase is inactivated (Lohe *et al.*, 1997). Active MLE transposases share the highly conserved amino acid motifs WVPHEL and YSPDLAP, separated by c. 150 amino acid residues in the C-terminal domain, a finding which has been used to identify new MLEs by PCR (Robertson, 1993).

Residues required for interaction between *Mos1* transposase monomers lie along the length of the protein (but predominantly in the N-terminal half), i.e. there is no single interaction domain (Zhang *et al.*, 2001).

Similar to *Ac* (see Section 1.3.1.4), *Tn10* (Chalmers and Kleckner, 1994) and *Tn5* (Wiegand and Reznikoff, 1992; Weinreich *et al.*, 1994a), MLE transposition activity is optimal for a very narrow range of transposase concentration, a phenomenon that has been termed “overexpression inhibition”. This phenomenon has been observed for the *in vitro* insertion of short, labelled *Himar1* ends into supercoiled plasmid (Lampe *et al.*, 1998), for *in vivo* excision of *Mos1* (Lohe and Hartl, 1996b), and for *Mos1 in vitro* plasmid-to-plasmid transposition (Zhang *et al.*, 2001). However the first publication of the *Mos1 in vitro*

system reported a plateau value of plasmid-to-plasmid transposition activity for increasing transposase concentration (Tosi and Beverley, 2000).

1.4.3 Current model for *Tc1/mariner* transposition

Tc1/mariner transposition is non-replicative (Lampe *et al.*, 1996; Izsvak *et al.*, 2000). A model for the chemical steps has first been proposed for *Tc3* (van Luenen *et al.*, 1994). Results obtained since then are in agreement with the model being applicable to the entire superfamily. A refined version is shown in Figure 1.9.

In this model, the transposon binds to the transposon ITRs and makes double-stranded DNA breaks at each end by hydrolysis of phosphodiester bonds. The mechanism of second strand cleavage has not been conclusively established (Figure 1.9F), but it was recently shown for *Mos1* that excision can occur without the formation of hairpin ends on either side of the break (Dawson and Finnegan, 2003). The transferred strand (3') is cleaved exactly at the transposon end, whereas cleavage of the non-transferred strand (5') is inexact with a range of cleavage sites centring around a 2-3 bp staggered cut inside the transposon (van Luenen *et al.*, 1994; Lampe *et al.*, 1997). The NTS is cleaved before the TS (Lohe *et al.*, 2000; Dawson and Finnegan, 2003). The 3'-hydroxyl group at each transposon end attacks the phosphodiester bond that flanks the TA target with a 2 bp stagger. This creates initially a c. 4-5 bp single stranded gap at each terminus, and after repair by yet unidentified host proteins, a target duplication of the TA dinucleotide. End-to-end joining of the empty donor site results in characteristic footprints (Figure 1.9H). However, *in vivo*, most breaks are probably repaired by templated double-stranded break repair from the sister chromatid or the homologous chromosome, resulting in restoration of the intact transposon in homoallelic, and precise restoration of the empty site in heteroallelic cells (Plasterk, 1991; Plasterk and Groenen, 1992).

1.4.3.1 Target site specificity

Apart from the TA dinucleotide, there is no obvious target site sequence specificity for *Mos1* (Bryan *et al.*, 1990; Tosi and Beverley, 2000), or *Himar1* (Lampe *et al.*, 1996, 1998). By replacing Mg^{2+} with Mn^{2+} , even the requirement for TA is relaxed, so that *Himar1* and *Mos1* insert at a frequency of 50% into other A/T containing sites (i.e. AT, AA or TT) (Lampe *et al.*, 1996; Zhang *et al.*, 2001). For *Tc1* and *SB*, larger consensus sites have been identified (Eide and Anderson, 1988; Korswagen *et al.*, 1996; Vigdal *et al.*, 2002).

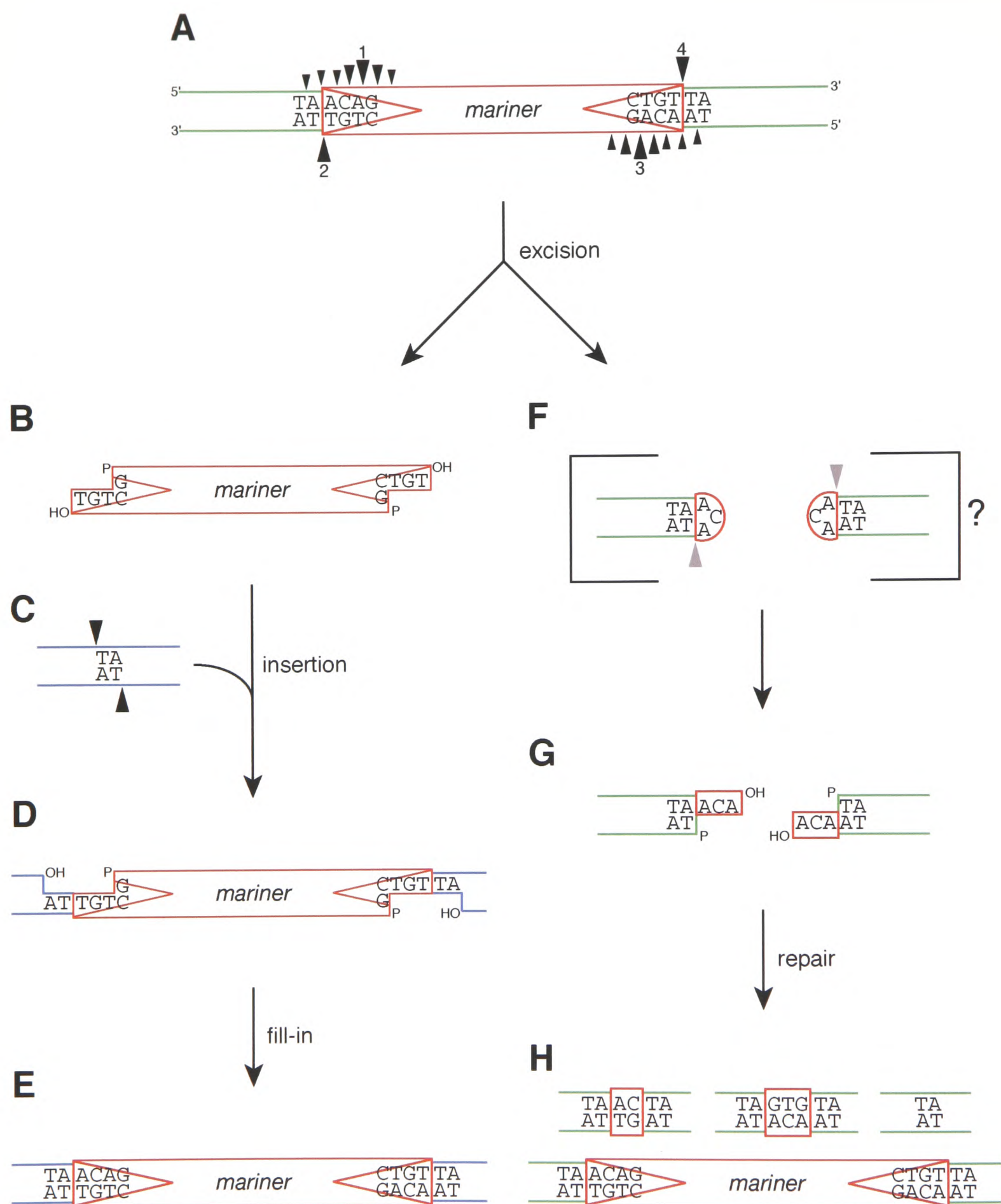


Fig. 1.9

Figure 1.9: Model for *Tc1/mariner* transposition

Model developed from the analysis of insertion sites (A) and (E), free linear transposon (B), empty donor sites (H) and *in vitro* cleavage reactions. Arrowheads indicate the major positions of cleavage sites with larger sizes meaning higher frequencies; numbers are the postulated sequence of cuts. The nucleotides depicted are those of *Himar1*. (Based in part on van Luenen *et al.*, 1994; Lampe *et al.*, 1996, 1997, 2001; Lohe *et al.*, 2000).

For excision, the flanking TA dinucleotides are absolutely required for *Tc1* (Vos *et al.*, 1996), dispensable for *Tc3* (van Luenen *et al.*, 1994) and important but not essential for *peach/Mos1* (Lohe *et al.*, 2000).

1.4.3.2 *Tc1/mariner* complexes

By analogy to other transposons, it seems likely that synaptic complexes are formed during excision and that the target DNA is attacked with both ends in a coordinated fashion. However, very little has been determined about the structure and stoichiometry of active transposition complexes (Plasterk and van Luenen, 2002).

There is indirect evidence for synaptic complex formation by *Tc1/mariner* transposases. The DNA-binding domain of *Tc3* transposase forms a dimer in co-crystals with DNA (van Pouderooyen *et al.*, 1997). *In vivo*, more than half of the mutant *Mos1* transposase proteins which had been found in an EMS screen inhibit the function of wildtype transposase in heterozygous mutant / non-mutant genotypes (Lohe and Hartl, 1996b). This dominant-negative effect points towards a multimerisation of transposase monomers. *In vitro*, *Mos1* transposase shows coordinated cleavage of both ends, as concluded from the formation of two-sided excision products even at very low transposase concentrations (Zhang *et al.*, 2001). Furthermore, *Mos1* transposase monomers interact in a yeast two-hybrid assay (Lohe *et al.*, 1996). A point-mutant (L124S), which inhibits this interaction but is catalytically active and able to bind *Mos1* ends, shows reduced transposition activity and cuts the two ends independently of each other (Zhang *et al.*, 2001). Gel retardation analyses of *Mos1* transposase to (single and paired) DNA ends have only been published after the completion of experimental work in this study (Zhang *et al.*, 2001; Augé-Gouillou *et al.*, 2001a, b; Dawson and Finnegan, 2003) (see Discussion 6.6.1.3).

1.4.4 Choice of the *Himar1* transposon

Himar1 has been reconstructed from the consensus sequence of c. 17,000 inactive copies, found in the *Haematobia irritans* (hornfly) genome (Figure 1.1C,E) (Robertson and Lampe, 1995a; Lampe *et al.*, 1996). The high number of copies indicates a high activity of the element before it ceased to be active, and indeed the reconstructed element has a high activity both *in vivo* and *in vitro*. *Himar1* was the first MLE, for which an *in vitro* system was established, consisting of fairly standard buffer conditions, DNA substrate and transposase as the only enzyme (Lampe *et al.*, 1996, 1998). The efficiency of the system has been further improved by the isolation of hyperactive transposase mutants (Lampe *et*

al., 1999). The most active of these, C9, has two point mutations in the putative non-specific DNA binding domain (Q131R, E137K). Compared to the wt transposase, the activity of this double mutant is increased by c. 50-fold in *E. coli*, 7-fold in *in vitro* insertion assays and 4-fold in human tissue culture (Lampe *et al.*, 1999; Fischer *et al.*, 2001). N-terminal fusions to maltose binding protein (MBP) also appear to behave like wildtype protein, which allows for rapid purification. MBP-C9 transposase has been successfully used for *in vitro* mutagenesis of bacterial DNA (Pelicic *et al.*, 2000).

1.5 Goals of the project

The purpose of this project is to study the molecular mechanism of eukaryotic transposition. To this end, the primary aim was to set up a *hobo in vitro* system, to detect the putative hairpin intermediates, and to show whether the mechanism is similar to V(D)J recombination. Technical difficulties confounded this project. Therefore, I changed course at a late date to look at *Himar1* target interactions and complex formations, and to compare this eukaryotic transposon with V(D)J, phage Mu and Tn10 transposition.

CHAPTER 2 - MATERIALS AND METHODS

2.1 Materials

2.1.1 Strains and cell lines

Table 2.1: *Escherichia coli* strains generated for this work

name	genotype	use
KL1001	AB1157 pOX38::Tn10(<i>Cam</i> ^R)	<i>in vivo</i> transposition
KL1002	HMS174(DE3) pOX38::Tn10(<i>Cam</i> ^R)	<i>in vivo</i> transposition
KL1003	MM294(DE3) pOX38::Tn10(<i>Cam</i> ^R)	<i>in vivo</i> transposition
KL1004	BL21(DE3) pOX38::Tn10(<i>Cam</i> ^R)	<i>in vivo</i> transposition
KL1005	ER2566 pOX38::Tn10(<i>Cam</i> ^R)	<i>in vivo</i> transposition
KL1043	BL21(DE3) R388	<i>in vivo</i> transposition
KL1046	ER2566 R388	<i>in vivo</i> transposition
KL1048	MM294(DE3) R388	<i>in vivo</i> transposition
KL1050	HMS174(DE3) R388	<i>in vivo</i> transposition

Table 2.2: Other *Escherichia coli* strains

name	genotype	reference or source	use
AB1157	<i>thr-1, leu-6, thi-1, lacY1, i, ara-14, xyl-5, mtl-1, proA-2, his-4, grg-3, str-31, tsx-33, λ⁻ (Str^R)</i>	Bachman, 1972	mating out (recipient)
BL21(DE3)	<i>E. coli</i> B; F ⁻ , <i>dcm, ompT, hsdS</i> (rB ⁻ mB ⁻), <i>gal</i> , (λcIts857, <i>ind1, S am7, nin5, lacUV5-T7 gene 1</i>)	Stratagene	<i>in vivo</i> transposition
BL21(DE3) pLysS	<i>E. coli</i> B; F ⁻ , <i>dcm, ompT, hsdS</i> (rB ⁻ mB ⁻), <i>gal</i> , (λcIts857, <i>ind1, S am7, nin5, lacUV5-T7 gene 1</i>), pLysS (<i>Cam</i> ^R)	Stratagene	expression from pKL5,7,8,9
BL21-CodonPlus (DE3)-RIL	<i>E. coli</i> B; F ⁻ ; <i>ompT, hsdS</i> (rB ⁻ mB ⁻), <i>dcm⁺ Tet^R gal λ</i> (DE3) <i>endA, Hte [argU ileY leuW Cam^R]</i>	Stratagene	expression from pKL8, 25, 54, 88
BMH71-18	<i>thi, supE, Δ(lac-proAB), [mutS::Tn10] [F⁺proAB, lacI^q ZΔM15]</i>	Clontech	site-directed mutagenesis
DH5α	<i>supE44, ΔlacUI69, (φ80 lacZΔM15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1</i>	Hanahan, 1983; Grant <i>et al.</i> , 1990	cloning
DS675	AB1157 R388	D. J. Sherratt	source of R388 for <i>in vivo</i> transposition
ER2566	F ⁻ , λ ⁻ , <i>fhuA2, [lon], ompT, lacZ::T7 gene 1, gal, sulA11, Δ(mcrC-mrr)]114:: IS10R (mcr-73::mini Tn10)2, R(zgb-210::Tn10) 1 endA1 [dcm] (Tet^R)</i>	NEB	<i>in vivo</i> transposition
HMS174(DE3)	F ⁻ , <i>recA, hsdR</i> (rK12 ⁻ mK12 ⁺), <i>Rif^R, (λcIts857, ind1, S am7, nin5, lacUV5-T7 gene 1)</i>	Novagen	<i>in vivo</i> transposition

LC836	pOX38		source of pOX38 for <i>in vivo</i> transposition
MM294	<i>supE44, hsdR, endA1, pro, thi</i>	Meselson and Yuan, 1968	expression from pKL5
MM294(DE3)	<i>supE, hsdR, endA1, pro, thi, (λcIts857, ind1, Sam7, nin5, lacUV5-T7 gene 1)</i>	R. M. Chalmers	<i>in vivo</i> transposition
XL-1 Blue	<i>recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac[F⁺ proAB, lacI^q ZΔM15, Tn10(Tet^R)</i>	Stratagene	cloning pKL5
XL-10 Gold	<i>Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac Hte [F⁺ proAB, lacI^q, Z, ΔM15, Tn10 (Tet^R), Amy Cam^R]</i>	Stratagene	cloning pKL6

Table 2.3: *Schizosaccharomyces pombe* strain

name	relevant genotype	source	use
SP-Q01	<i>leu1-32h⁻</i>	Stratagene	expression from pKL6

Table 2.4: Mammalian cell lines

cell line	293T	COS-7	NIH 3T3	SCGR11	S10.1
cell type	Ad 5 / SV40-transformed human embryonic kidney fibroblasts	SV40-transformed monkey kidney fibroblasts	mouse embryonic fibroblasts	mouse embryonic fibroblasts	mouse embryonic fibroblasts
genotype	SV40 T antigen	SV40 T antigen	wt	DNA-PKcs ⁻	<i>hobo</i> (pKL42), DNA-PKcs ⁻
relevant phenotype		large nucleus	wt (contact-inhibited)	scid	<i>hobo</i> , scid
medium	DMEM, 1 mM glutamine, 1% Pen Str	DMEM, glutamax I, high glc, Na pyruvate	DMEM, glutamax I, high glc, Na pyruvate	DMEM, glutamax I, high glc, Na pyruvate	DMEM, glutamax I, high glc, Na pyruvate
serum	10% FBS	10% FBS	10% NBS	5% FBS	5% FBS
selection	-	-	1 mg/ml G418	0.2 mg/ml G418	-
use	production of recombinant retrovirus	visualising expression and localisation of GFP-fusion proteins	transfection	transfection, control for <i>hobo</i> transposition	<i>hobo</i> transposition
source	A. J. Kingsman	A. J. Kingsman	A. J. Kingsman	P. A. Jeggo	this work
reference	DuBridge <i>et al.</i> , 1987	Gluzman, 1981	Jainchill <i>et al.</i> , 1969	Li <i>et al.</i> , 1995	this work

2.1.2 Oligonucleotides, plasmids and other DNA substrates

Table 2.5: Oligonucleotides designed for this work

name		sequence 5' → 3'	length	T _m	use
KL1	LIC 5'	<u>...gacgacgacaag</u> ATGGCGCCA TACATAATGATTG	34 nt	53.6°C	adds LIC site 5' of <i>hobo</i> ORF for cloning into pCAL-n-EK and pESP-2
KL2	ESP-LIC 3'	caggacagagcatcaCTATTGCGA GTTGTTTAGGTT	36 nt	45.6°C	adds LIC site 3' of <i>hobo</i> ORF for cloning into pESP-2
KL3	CAL-LIC 3'	ggaacaagaccgctCTATTGCGAG TTGTTTAGGTT	35 nt	45.6°C	adds LIC site 3' of <i>hobo</i> ORF for cloning into pCAL-n-EK
KL4	hobo+Nde 5'	agtcgagtc <u>cat</u> ATGGCGCCATACA TAATGATTG	33 nt	53.6°C	adds <i>Nde</i> I site 5' of <i>hobo</i> ORF; for cloning into pET11a
KL5	hobo+Nde 3'	ataag <u>catatg</u> CTATTGCGAGTTG TTTAGG	33 nt	43.0°C	adds <i>Nde</i> I site 3' of <i>hobo</i> ORF; for cloning into pET11a
KL6	X Nde	GCCGCCTTATGCCAgATGCTTAGA TTTGCC	30 nt	66.3°C	destroys <i>Nde</i> I site in <i>hobo</i> ORF (site directed mutagenesis, mutagenic primer)
KL7	Kpn-Bam	GGGGCCCGGat <u>CCC</u> AGCTTTTG	22 nt	67.6°C	converts <i>Kpn</i> I to <i>Bam</i> HI site in MCS of pBS-KS (site directed mutagenesis, selection primer)
KL8	Seq X Nde	CATCTAGCTCACCATCTATATGC	23 nt	47.4°C	sequencing <i>hobo</i> across <i>Nde</i> I site (bp 1577-1599)
KL9	Seq hobo	CGAAATCAGCCGCTTCTGCC	20 nt	59.7°C	sequencing <i>hobo</i> , from orf to 5' (bp 484-465)
KL10	Seq hobo	GGTCCAGTGTCTCGGC	16 nt	48.5°C	sequencing <i>hobo</i> (bp 356-371)
KL11	Seq hobo	GGTTGTCCAAGATTGTCGG	19 nt	51.2°C	sequencing <i>hobo</i> (bp 762-780)
KL12	Seq hobo	GAGGGCAATACCCG	14 nt	43.1°C	sequencing <i>hobo</i> (bp 1231-1244)
KL13	Seq hobo	CGCAAACCTTATACCCTCAG	19 nt	45.4°C	sequencing <i>hobo</i> , from orf to 3' (bp 2097-2115)
KL14	ESP-LIC 3'	caggacagagcatcaCTATTGCGA GTTGTTTAGGTTTTTGTAAATAG	46 nt	53.5°C	cloning the <i>hobo</i> ORF into pESP-2
KL15	N-IS-5'	GAAGGAGATATACATATGTGCGAA CTCG	28 nt	55.5°C	cloning the N-terminus of IS10 transposase from pET3a into pTYB1 (<i>Nde</i> I site)
KL16	N-IS-3'	actgagc <u>tcttccgcagcc</u> TCTCG CTTTGGTTGGCAGG	38 nt	57.1°C	cloning the N-terminus of IS10 transposase from pET3a into pTYB1 (<i>Sap</i> I site)
KL17	D.m. 5'	GTGAGCGGCACCCGCGACGC	20 nt	70.1°C	sequencing pHFL-1 genomic regions and <i>in vivo</i> insertions (bp 190-171)
KL18	D.m. 3'	CGGCTCTTGCGCAGCAGGCC	20 nt	67.1°C	sequencing pHFL-1 genomic regions and <i>in vivo</i> insertions (bp 2782-2801)
KL19	Not-Hpa-left	aggtctaggcgcgcccgcgatgtta <u>a</u> CGACCTAAAAGTAGGCAATACAG CGC	51 nt	58.2°C	cloning the left <i>hobo</i> end into pBS; adds <i>Not</i> I and <i>Hpa</i> I sites
KL20	left-Mlu	aggtcacgcgctGACTCGACTACCT ACGAGACCACTCG	37 nt	57.4°C	cloning the left <i>hobo</i> end to <i>Kan</i> ^R insert; adds <i>Mlu</i> I site
KL21	Mlu-right	aggtcacgcgctCTTATACCCGTCA CGCTTCCACC	34 nt	57.7°C	cloning the right <i>hobo</i> end to <i>Kan</i> ^R insert; adds <i>Mlu</i> I site

KL22	right-Nhe-Not	aggtctaggtg <u>cgggccgc</u> atgcta gcTTCGGTTGACGTGTTAGAGAAT TTGTC	53 nt	57.0°C	cloning the right hobo end into pBS; adds <i>NotI</i> and <i>NheI</i> sites
KL23	BspE-PfIM 1	CCGGAttCCAAAAGTTGG	18 nt	n.a.	adapter between <i>BspEI</i> in pKL7-9 and <i>PfIMI</i> in pACYC184
KL24	BspE-PfIM 2	TTGGaaT	7 nt	n.a.	adapter between <i>BspEI</i> in pKL7-9 and <i>PfIMI</i> in pACYC184
KL25	PfIM-PpuM 2	GTCCTccCCAATTC	14 nt	n.a.	adapter between <i>PpuMI</i> in pKL7-9 and <i>PfIMI</i> in pACYC184
KL26	PfIM-PpuM 1	TTGGccAG	8 nt	n.a.	adapter between <i>PpuMI</i> in pKL7-9 and <i>PfIMI</i> in pACYC184
KL27	Bsa-hobo	tactggtctctATGGCGCCATACA TAATGATTGTTGAG	38 nt	59.0°C	cloning <i>hobo</i> ORF into pMAL (<i>BsaI</i> site)
KL28	BamH-hobo	attcggatcctCTATTGCGAGTTG TTTAGGTTTTTGTAATAGGAATGC	48 nt	60.4°C	cloning <i>hobo</i> ORF into pMAL (<i>BamHI</i> site)
KL29	Seq hobo	GCATATAGATGGTGAGCTAGATG	23 nt	47.4°C	sequencing <i>hobo</i> (bp 1599-1577)
KL30	Seq hobo	GCCGAGACACTGGACC	16 nt	48.5°C	sequencing <i>hobo</i> (bp 371-356)
KL31	Seq hobo	CCGACAATCTTGGACAACC	19 nt	51.2°C	sequencing <i>hobo</i> (bp 780-762)
KL32	Seq hobo	CGGGTATTGCCCTC	14 nt	43.1°C	sequencing <i>hobo</i> (bp 1244-1231)
KL33	Seq hobo	CTGAGGGTATAAGTTTGCG	19 nt	45.4°C	sequencing <i>hobo</i> (bp 2115-2097)
KL34	Seq hobo	CCTGTGACGGAAGATCACTTCGC	23 nt	59.5°C	sequencing <i>hobo</i> ORF in pKL13 from pACYC184
KL35	hobo RR f	GATGAAATTGAACGATATATTcGt CAAcGtGTTCCATTGTCTCAA	45 nt	65.6°C	site directed mutagenesis of two rare Arg codons in <i>hobo</i> ORF (aa 576 & 578), forward primer, adds <i>HindII</i> site
KL36	hobo RR r	TTGAGACAATGGAACaCgTTGaCg AATATATCGTTCAATTCATC	45 nt	65.6°C	site directed mutagenesis of two rare Arg codons in <i>hobo</i> ORF (aa 576 & 578), reverse primer, adds <i>HindII</i> site
KL37	SCM1 f	GAACGATATATTcGtCAAAGAGTT CCA	27 nt	52.2°C	site directed mutagenesis of one rare Arg codons in <i>hobo</i> ORF (aa 576), forward primer
KL38	SCM1 r	TGGAACCTCTTTGaCgAATATATCG TTC	27 nt	52.2°C	site directed mutagenesis of one rare Arg codons in <i>hobo</i> ORF (aa 576), reverse primer
KL39	Bam f	ATGCTCGAGGAcCCGGCTGCTA	22 nt	60.8°C	destroys <i>BamHI</i> site in pKL8 backbone (site directed mutagenesis, selective primer), forward
KL40	Bam r	TAGCAGCCGGgTCCTCGAGCAT	22 nt	60.8°C	destroys <i>BamHI</i> site in pKL8 backbone (selective primer), reverse
KL41	hobo LT 1	gcgccTCAGGACAGAGAACTGCAA GGGTGGCgt	33 nt	52.8°C	<i>in vitro</i> band-shift assays (left <i>hobo</i> end, top strand)
KL42	hobo LB 2	cgacGCCACCCTTGCAGTTCTCTG TCCTGAg	31 nt	52.8°C	<i>in vitro</i> band-shift assays (left <i>hobo</i> end, bottom strand)
KL43	hobo RT 3	gcgccCGAGTGGCGGGCTGCAGTT CTCTGATCAGGgt	37 nt	58.6°C	<i>in vitro</i> band-shift assays (right <i>hobo</i> end, top strand)

KL44	hobo RB 4	cgacCCTGATCAGAGAACTGCAGC CCGCCACTCGg	35 nt	58.6°C	<i>in vitro</i> band-shift assays (right <i>hobo</i> end, bottom strand)
KL45	hobo LTbl 5	CAGAGAACTGCAAGGGTGGCgt	22 nt	44.7°C	<i>in vitro</i> insertion assays
KL46	hobo LBbl 6	cgacGCCACCCCTTGCACTTCTCTG	24 nt	44.7°C	<i>in vitro</i> insertion assays
KL47	hobo RTbl 7	gcgcccGAGTGGCGGGCTGCAGTT CTCTG	29 nt	53.4°C	<i>in vitro</i> insertion assays
KL48	hobo RBbl 8	CAGAGAACTGCAGCCCGCCACTCG g	25 nt	53.4°C	<i>in vitro</i> insertion assays
KL49	hobo LT 9	gcgcccTCAGGACAGAGAACTGCAA GGGTGGCACTTTTTTACCACTCGA CTCACACCCTACAATTTTGTgt	70 nt	61.3°C	<i>in vitro</i> band-shift assays (left <i>hobo</i> end, top strand)
KL50	hobo LB 10	cgacACAAAATTGTAGGGTGTGAG TCGAGTGGTAAAAAGTGCCACCC TTGCAGTTCTCTGTCTCTGAg	68 nt	61.3°C	<i>in vitro</i> band-shift assays (left <i>hobo</i> end, bottom strand)
KL51	hobo RT 11	gcgcccAGTACCGAGTGTTTATCGG GTGGACGTAGAGTGCAGTGGCGG GCTGCAGTTCTCTGATCAGGgt	70 nt	65.8°C	<i>in vitro</i> band-shift assays (right <i>hobo</i> end, top strand)
KL52	hobo RB 12	gcacCCTGATCAGAGAACTGCAGC CCGCCACTCGCACTCTACGTCCAC CCGATAAACACTCGGTACTg	68 nt	65.8°C	<i>in vitro</i> band-shift assays (right <i>hobo</i> end, bottom strand)
KL53	hobo front Eco	cgaattctgATGGCGCCATACATA ATGATTG	31 nt	53.6°C	adds <i>Eco</i> RI site to <i>hobo</i> ORF
KL54	hobo back Eco	gcagaattcgTTGCGAGTTGTTTA GG	26 nt	40.3°C	adds <i>Eco</i> RI site to <i>hobo</i> ORF
KL55	hobo arb 5'	CACCCATGTGTTTAACGGTATACC C	25 nt	55.8°C	arbitray PCR, first specific primer for left end (bp 252 - 228)
KL56	hobo arb 3'	GACTGTGCGTCCACTCACCC	20 nt	56.8°C	arbitray PCR, first specific primer for right end (bp 2759 - 2778)
KL57	arbit 1	GTAGATCCATGCCATATCCTATCC G	25 nt	55.3°C	arbitrary - nested PCR, binds to tail of KL58-60
KL58	arbit 2	gtagatccatgccatatacctatcc gnnnnnnnnntatcc	40 nt	varies	arbitrary amplification of mini- <i>hobo</i> insertions in R388
KL59	arbit 3	gtagatccatgccatatacctatcc gnnnnnnnnnncttcg	40 nt	varies	arbitrary amplification of mini- <i>hobo</i> insertions in R388
KL60	arbit 4	gtagatccatgccatatacctatcc gnnnnnnnnntccgg	40 nt	varies	arbitrary amplification of mini- <i>hobo</i> insertions in R388
KL61	Bam-hobo 5'	cgggatccaATGGCGCCATACATA ATGATTG	31 nt	53.6°C	cloning the <i>hobo</i> ORF into pEGFP-N1, adds <i>Bam</i> HI site
KL62	hobo-Bam 3'	gcggatccgtTTGCGAGTTGTTTA GGTTTTTGTAATAGG	39 nt	55.4°C	cloning the <i>hobo</i> ORF into pEGFP-N1, adds <i>Bam</i> HI site
KL63	tetX-Sac	atctgagctcttAGGTAAAGTATG TACCAATTTC	34 nt	41.2°C	cloning <i>tetX</i> into pMAL, adds <i>Sac</i> I site
KL64	Kpn-hobo	ggggtaccATGGCGCCATACATAA TGATTG	30 nt	53.6°C	cloning the <i>hobo</i> ORF into pEGFP-N1, adds <i>Kpn</i> I site
KL65	hobo-Not	aaggacgtaagcggccgcttaCTA TTGCGAGTTGTTTAGGTTTTGTA ATAG	52 nt	53.5°C	cloning the <i>hobo</i> ORF into pEGFP-N1 without EGFP, adds <i>Not</i> I site
KL66	Seq flank	CCGAAATAATTATAATTCCG	20 nt	43.4°C	sequencing right flanking DNA in pHFL-1
KL67	Bsa-flank R	tatccgggtctccatATCAGGATA TCGTAAATCCCAAATGG	41 nt	55.5°C	making hairpins from DNA flanking the right <i>hobo</i> end in pKL3; adds <i>Bsa</i> I site
KL68	flank R-SacII	cttcgaccgcgGATTTACTGACTT ACTCGTACGAAACGG	39 nt	54.7°C	making hairpins from DNA flanking the right <i>hobo</i> end in pKL3; adds <i>Sac</i> II site

KL69	pBS Seq	CGCTATTACGCCAGCTGGCG	20 nt	60.2°C	sequencing of DNA flanking the right <i>hobo</i> end from pBS
KL70	Bsa-flank L	tatccgggtctccatATCCTGATG CTGCGCTGTATTGC	38 nt	59.6°C	making hairpins from DNA flanking the left <i>hobo</i> end in pKL10, pKL42, pKL47; adds <i>Bsa</i> I site
KL71	flank L-Bgl 42	AGCTCAGAGGCCGAGGCG..	18 nt	59.1°C	making hairpins from DNA flanking the left <i>hobo</i> end in pKL42; includes <i>Bgl</i> II site
KL72	flank R-Bgl 1047	GTCCCATTCGCCATTCAGGC	20 nt	59.4°C	making hairpins from DNA flanking the right <i>hobo</i> end in pKL10, pKL47; includes <i>Bgl</i> II site
KL73	Y top Bgl 42	cgtccatcagccacactgatgcaa ttgctATGAATGTGCATGAAGGCT AGG	51 nt	n.a.	top strand of frayed duplex to ligate to <i>Bgl</i> II site in left flanking DNA of pKL42
KL74	Y bot Bgl 42	PAGCCTTCATGCACATTCATAtta cacggctgcttgcgagattccaa gB	48 nt	n.a.	bottom strand of frayed duplex to ligate to <i>Bgl</i> II site in left flanking DNA of pKL42
KL75	top Y LM	cgtccatcagccacactgatgc..	22 nt	59.9°C	binds to top duplex primer; top primer for LMPCR: forms <i>Nsi</i> I site with both blunt flanking ends
KL76	bot Y	cttggaatctgcgcaagcagc	21 nt	59.7°C	binds to bottom duplex primer
KL77	LM bot	..Pgcatcagtggtgctg	15 nt	n.a.	bottom primer for LMPCR: forms <i>Nsi</i> I site with both blunt flanking ends
KL78	R for LM	TTCGGTTGACGTGTTAGAGAATTT GTC	27 nt	57.0°C	binds within right hairpin to amplify with LMPCR
KL79	R2 for LM	CCGATCTGCGTTCCATTTGG	20 nt	58.1°C	binds further towards right hairpin to amplify with nested LMPCR
KL80	high melt primer 1	GGCCTCGGCCTCTGCATAAATAAA	24 nt	88.6°C =59.6+29	direct amplification of hairpin
KL81	L for LM42	CTCGCCACGTTGCGCG	16 nt	59.3°C	binds within left hairpin of pKL42 to amplify with LMPCR
KL82	L2 for LM42	GTTCCGATTTAGTGCTTTACGGCA CC	26 nt	59.7°C	binds further towards left hairpin of pKL42 to amplify with nested LMPCR
KL83	Y top 42	cgtccatcagccacactgatgcaa ttgctATGAATGTGCATGAAGCCT AGGCGGCCTCGGC	61 nt	72.5°C	top strand of frayed duplex to ligate to left flanking DNA of pKL42, incl. <i>Bgl</i> II site
KL84	Y bot 42	GCCGAGGCCGCTAGGCTTCATGC ACATTCATAttacacggctgcttg cgcagattccaagB	61 nt	72.5°C	bottom strand of frayed duplex to ligate to left flanking DNA of pKL42, incl. <i>Bgl</i> II site
KL85	Y top 10-2	Batgaccgggatattcacgctcag CGAAAACAATCTAATCATAACTAC CGCGAATACTCAATCATCTCGGCT TC	73 nt	n.a.	top strand of frayed duplex to ligate to <i>Bgl</i> II site in right flanking DNA of pKL10 / 47
KL86	Y bot 10-2	PGCCGAGATGATTGAGTATTCGCG GTAGTTATGATTAGATTGTTTTCG accgggcaaaaccttcgcttccat	71 nt	n.a.	bottom strand of frayed duplex to ligate to <i>Bgl</i> II site in right flanking DNA of pKL10 / 47
KL87	top Y2	ATGACCGGGATATTCACGC	19 nt	52.9°C	binds to frayed end of KL85
KL88	bot Y2	ATGGAAGCGAAGGTTTTGC	19 nt	52.3°C	binds to frayed end of KL86
KL89	LM top 2	gcggtgactcgggagatctgaatg c..	25 nt	64.7°C	top primer for LMPCR: forms <i>Nsi</i> I site with both blunt flanking ends

KL90	LM bot 2	<u>..Pgcattcagatc</u>	11 nt	n.a.	bottom primer for LMPCR: forms <i>NsiI</i> site with both blunt flanking ends
KL91	LM top 3	gctatgtactacccgggaattatg <u>c..</u>	25 nt	54.3°C	top primer for LMPCR: forms <i>NsiI</i> site with both blunt flanking ends
KL92	LM bot 3	<u>..Pgcataattccc</u>	11 nt	n.a.	bottom primer for LMPCR: forms <i>NsiI</i> site with both blunt flanking ends
KL93	Y top 42-2 Sfi	Fatgaccgggatattcacgctcag CGAAAACAATCTAATCATAACTAC CGCGAATACTCAATCATCTCGGCA GG	73 nt	n.a.	top strand of frayed duplex to ligate to <i>SfiI</i> / <i>BglI</i> site in left flanking DNA of pKL42
KL94	Y top 42-3 DraIII	Fatgaccgggatattcacgctcag CGAAAACAATCTAATCATAACTAC CGCGAATACTCAATCATCTC <u>CACGG</u> A	72 nt	n.a.	top strand of frayed duplex to ligate to <i>DraIII</i> site in left flanking DNA of pKL42
KL95	Y bot 42-2	PGCCGAGATGATTGAGTATTCGCG GTAGTTATGATTAGATTGTTTTCG Accgggcaaaaccttcgcttccat	71 nt	n.a.	bottom strand of frayed duplex for KL 85, 93, 94 = KL86
KL96	seq flank l	GACTCGACTACCTACGAGACC	21 nt	47.5°C	sequencing into flanking DNA from <i>hobo</i> /pKL10 (bp 313 - 293)
KL97	seq flank r	CTTATACCCGTCACGCTTCC	20 nt	51.5°C	sequencing into flanking DNA from <i>hobo</i> /pKL10 (bp 2659 - 2678)
KL98	Y bot 42-3	PGTGAGATGATTGAGTATTCGCGG TAGTTATGATTAGATTGTTTTCGa ccgggcaaaaccttcgcttccat	70 nt	n.a.	bottom strand of frayed duplex to for KL94, 100
KL99	Y top 42-2 Sfi	Batgaccgggatattcacgctcag CGAAAACAATCTAATCATAACTAC CGCGAATACTCAATCATCTCGGCA GG	73 nt	n.a.	top strand of frayed duplex to ligate to <i>SfiI</i> / <i>BglI</i> site in left flanking DNA of pKL42
KL100	Y top 42-3 DraIII	Batgaccgggatattcacgctcag CGAAAACAATCTAATCATAACTAC CGCGAATACTCAATCATCTC <u>CACGG</u> A	72 nt	n.a.	top strand of frayed duplex to ligate to <i>DraIII</i> site in left flanking DNA of pKL42
KL101	flank R Dra	CATTTGCCCCACATGCAATC	20 nt	56.9°C	making hairpins from DNA flanking the right <i>hobo</i> end in pKL42 including the <i>DraIII</i> site
KL102	flank L Sfi	CCTCCTCACTACTTCTGGAATAGC TCAGAGG	31 nt	60.1°C	making hairpins from DNA flanking the left <i>hobo</i> end in pKL42 including the <i>SfiI</i> site
KL103	Mlu URA3	gacgacgcgtTCGTCTTGAGTCCA ACCCGG	30 nt	58.1°C	creation of <i>hobo</i> reporter for <i>S. cerevisiae</i>
KL104	URA3 Mlu	tactacgcgtTTAGTTTTGCTGGC CGCATCTTC	33 nt	58.3°C	creation of <i>hobo</i> reporter for <i>S. cerevisiae</i>
KL105	BssH ColE1	cgtagcgcgcTTCTCGCTCACTG ACTCGC	30 nt	56.6°C	creation of <i>hobo</i> reporter for <i>S. cerevisiae</i>
KL106	ColE1 BssH	tgctgcgcgcTTGAGAAAGCGCCA CGCTTC	30 nt	58.2°C	creation of <i>hobo</i> reporter for <i>S. cerevisiae</i>
KL107	flank R Dra2	GGCGAATCAATAGCGCAGC	19 nt	56.3°C	making hairpins from DNA flanking the right <i>hobo</i> end in pKL42 including the <i>DraIII</i> site
KL108	top Y2-2	GGATATTCACGCTCAGCG	18 nt	50.2°C	nested primer to bind on frayed duplex, top strand
KL109	bot Y2-2	AGGTTTTGCCCGGTCG	16 nt	53.1°C	nested primer to bind on frayed duplex, bottom strand

KL110	pKL42 probe R	GTGAACCATCACCTAATCAAG	22 nt	50.3°C	amplification of left flanking DNA for making radioactive probe
KL111	pKL42 probe S 42 L-451	TAACTGACACACATTCCACAGC	22 nt	50.8°C	amplification of left flanking DNA for making radioactive probe; SPPCR of left hairpin (451 bp)
KL112	42 L-376	CAGGGTTATTGTCTCATGAGCGG	23 nt	56.9°C	SPPCR of left hairpin (376 bp)
KL113	42 L-228	GCGTGACCGCTACACTTGCCAGCG	24 nt	67.3°C	SPPCR of left hairpin (228 bp)
KL114	42 R-252	GGCCAAATAGTTCAATCTGGCAGG G	25 nt	60.9°C	SPPCR of right hairpin (252 bp)
KL115	42 R-216	GATGTTCTGCTTAAGGGATGAG	23 nt	53.1°C	SPPCR of right hairpin (216 bp)
KL116	42 L-836	GAACCTGCGTGCAATCCATCTT	22 nt	57.7°C/6 5.3°C (Tth5x)	SPPCR of left hairpin (836 bp) before digestion with <i>Sfi</i> I
KL117	42 L-967	CCCAGTCATAGCCGAATAGCCTCT CC	26 nt	61.1°C/6 8.3°C (Tth5x)	SPPCR of left hairpin (967 bp) before digestion with <i>Sfi</i> I
KL118	Bio-hobo Rsa	BtttttACCCGCGATGTGCGTGCG GAGTAGC	30 nt	70.4°C	removal of non-hairpinned pKL42 (top strand)
KL119	hobo Rsa-Bio	BtttttCACACCCTACAATTTTGT GTGCGGG	30 nt	61.8°C	removal of non-hairpinned pKL42 (bottom strand)
KL120	Bio-Rsa HP	BtttttACGTGCTCGCTCGATGCG ATGTTTCGC	32 nt	70.6°C	selection of hairpins flanking the pKL42 left end
KL121	NBhobo F	acttcagttaacCAGAGAACTGCA AGGGTGGC	32 nt	55.6°C	amplification of <i>hobo</i> and addition of <i>Hpa</i> I site 5'
KL122	NBhobo R	gaccaagttaacCAGAGAACTGCA GCCCCCA	32 nt	61.8°C	amplification of <i>hobo</i> and addition of <i>Hpa</i> I site 3'
KL123	Nco-his	AGAAGGAGATATACCATGGGCAGC	24	55.4°C	amplification of 6xHis-tase from 5' end in pKL8 (pET14), includes <i>Nco</i> I site
KL124	hobo-PvuII	tatgcgcagcTGCAGTTGTTTAG GTTTTTGTAATAGG	38	55.7°C	adds <i>Pvu</i> II site 3' of <i>hobo</i> tase (w/o stop codon) for fusing to intein
KL125	Nco-hobo	TGCAGTCGACGGTACCATGG	20	57.8°C	amplifies <i>hobo</i> tase w/ <i>Nco</i> I site and Kozak sequence from pKL39 for cloning into pTM1
KL126	polyA-Xho	attgctcGAGTTTGGACAAACCAC AACTAGAATGC	35	57.4°C	adds <i>Xho</i> I site to poly-A-tail signal in pKL39 for cloning into pTM1
KL127	Rsa HP	ACGTGCTCGCTCGATGCGATGTTT CGC	27 nt	70.6°C	amplification of hairpins flanking the pKL42 left end after purification away from genomic DNA
KL128	SgrAI-pT7	CGGTGATGTCGGCGATATAGG	21 nt	56.7°C	amplification of pT7-6xHis- <i>hobo</i> from pKL8 (5' of <i>Sgr</i> AI)
KL129	hobo-Xho	aaggaactcgagTTGCGAGTTGTT TAGGTTTTTGTAATAGG	41 nt	55.4°C	introducing <i>Xho</i> I site 3' of <i>hobo</i> ORF w/o stop codon for cloning into pET-32a
KL130	Bsa-Himar hp	tatccgggtctcctaTATAATCTG CAGAATTCAGCACACTGG	43 nt	58.6°C	making hairpins from DNA flanking the 5' <i>Himar</i> I end in p27fh-5'; adds <i>Bsa</i> I site
KL131	Xho-Sfi-Himar flank 50	acttcactcgaggcccaggtggCC ACTAGTAACGGCCGCCAGTG	44 nt	59.9°C	cloning flanking DNA 50 bp from <i>Himar</i> I 5' ITR; adds overlapping <i>Xho</i> I and <i>Sfi</i> I site

KL132	Himar end-Bsa-Xho 50	acttcactcgcgaggtctcTACTAGC GACGCCATCTATGTGTCAGAC	45 nt	59.0°C	cloning <i>Himar1</i> end, 50 bp from 5' ITR; adds overlapping <i>BsaI</i> and <i>XhoI</i> site
KL133	Xho-Sfi-hobo flank 50	acttcactcgcgagggccaggtggcC GGCCGCATGTTAACGACC	43 nt	59.2°C	cloning flanking DNA 50 bp from <i>hobo</i> left end in pKL10; adds overlapping <i>XhoI</i> and <i>SfiI</i> site
KL134	hobo end-Bsa-Xho 65	acttcactcgcgaggtctCGATGTGC GTGGCGAGTAGC	36 nt	59.8°C	cloning <i>hobo</i> left end, 65 bp IR; adds overlapping <i>BsaI</i> and <i>XhoI</i> site
KL135	Y top Sfi	atgaccgggatattcacgctcagC GAAAACAATCTAATCATAACTACC GCGAATACTCAATCATCTCGGCAG G	73 nt	n.a.	top strand of frayed duplex to ligate to <i>SfiI</i> / <i>BglI</i> site =KL99 w/o biotin
KL136	Y bot Sfi	PGCCGAGATGATTGAGTATTCGCG GTAGTTATGATTAGATTGTTTTCG Accggggcaaaaccttcgcttccat	71 nt	n.a.	bottom strand of frayed duplex for KL 85, 93, 94, 135 = KL86, KL95
KL137	seq ins 5'	GATTTTGAGACACAACGTGGC	21 nt	52.7°C	sequencing from 5' of <i>Kan^R</i> gene in pKL10 into flanking DNA (bp 654-634)
KL138	seq ins 3'	ACGCTGACTTGACGGGACGG	20 nt	60.5°C	sequencing from 3' of <i>Kan^R</i> gene in pKL10 into flanking DNA (bp 1596-1615)
KL139	Bam 5' ITR	CagtctggatcccTATAATCGTTGT ATCGCTCTTGAAGGG	39 nt	55.8°C	cloning <i>Himar1</i> ends (and backbone) around new insert; 109 bp of 3' end plus <i>BamHI</i> site
KL140	BamKpn 3' ITR	cagtctggatccggtaccGACACG AATCATTTGAAGGTTGG	41 nt	54.7°C	cloning <i>Himar1</i> ends (and backbone) around new insert; 101 bp of 5' end plus <i>KpnI</i> and <i>BamHI</i> site
KL141	5' ITR blunt top	ACAGGTTGGCTGATAAGTCCCCGG TCTGACACATA	35 nt	n.a.	blunt cleaved <i>Himar1</i> end for insertion, top strand
KL142	5' ITR blunt bot	aggTATGTGTCAGACCGGGGACTT ATCAGCCAACCTGT	38 nt	n.a.	cleaved <i>Himar1</i> end for insertion, bottom strand
KL143	5' ITR+backb top	GTGTGATGGATTATAACAGGTTGG CTGATAAGTCCCCGGTCTGACACA TA	50 nt	n.a.	uncleaved <i>Himar1</i> end for insertion, top strand
KL144	5' ITR+backb bot	aggTATGTGTCAGACCGGGGACTT ATCAGCCAACCTGTTATAATCCAT CACAC	53 nt	n.a.	uncleaved <i>Himar1</i> end for insertion, bottom strand
KL145	5' ITR stag top	AGGTTGGCTGATAAGTCCCCGGTC TGACACATA	33 nt	n.a.	staggered cleaved <i>Himar1</i> end for insertion, top strand
KL146	Xho BsrDI 5' ITR	actgtcctcgcagcaatgACAGGTT GGCTGATAAGTCCCCGG	41 nt	61.8°C	adds <i>BsrDI</i> site to <i>Himar1</i> ITR which allows generation of staggered end; <i>XhoI</i> for cloning into pBS -> pKL105
KL147	pBS Sac	CGCCAAGCGCGCAATTAACC	20 nt	61.8°C	amplifies MCS of pBSII-SK+ from downstream of <i>SacI</i> site; cloning of pKL105 with staggered end
KL148	hp top	atgaccgggatattcacgctcagC GAAAACAATCTAATCATAACTACC GCGAATACTCAATCATCTCGGTTA TATA	76 bp	n.a.	top strand of frayed hairpin, <i>Himar1</i> flanking DNA
KL149	hp bot	PTAACCGAGATGATTGAGTATTCG CGGTAGTTATGATTAGATTGTTTT CGaccggggcaaaaccttcgcttcc at	73 bp	n.a.	bottom strand of frayed hairpin, <i>Himar1</i> flanking DNA

KL150	flank top	atgaccgggatattcacgctcagC GAAAACAATCTAATCATAACTACC GCGAATACTCAATCATCTCGGTTA TAac	76 bp	n.a.	top strand of frayed duplex adapter to ligate to <i>Bsr</i> DI cut pKL105 <i>Himar</i> 1 end
KL151	flank bot	PTATAACCGAGATGATTGAGTATT CGCGGTAGTTATGATTAGATTGTT TTCGaccgggcaaaaccttcgctt ccat	75 bp	n.a.	bottom strand of frayed duplex adapter to ligate to <i>Bsr</i> DI cut pKL105 <i>Himar</i> 1 end
KL152	flank top for hp	atgaccgggatattcacgctcagC GAAAACAATCTAATCATAACTACC GCGAATACTCAATCATCTCGGTTA TA	74 bp	n.a.	top strand of frayed duplex adapter to ligate to <i>Hga</i> I-cut pRC251/252 <i>Himar</i> 1 end
KL153	flank bot for hp	PgttccTATAACCGAGATGATTGA GTATTCGCGGTAGTTATGATTAGA TTGTTTTTCGaccgggcaaaacctt cgcttccat	81 bp	n.a.	bottom strand of frayed duplex adapter to ligate to <i>Hga</i> I-cut pRC251/252 <i>Himar</i> 1 end

Oligonucleotides designed for this work.

Upper case letters indicate sequences that are complementary to the PCR template or complementary adapter oligonucleotide, lower case letters stand for other sequence, such as overhangs and mutations. Relevant restriction and recognition sites are underlined. n, degenerate sequence; P, 3' or 5' phosphorylation; B, biotin; F, fluorescent label (Cy5); A, 2-amino-dA; C, C-5-propynyl-dC. The melting temperatures (T_m) were calculated for matching bases on <http://alces.med.umn.edu/rawtm.html> (Breslauer *et al.*, 1986) with the parameters 500 nM oligo, 10 mM salt concentration. KL15, KL16 and KL63 were for contributions to (Chalmers *et al.*, 2000).

Table 2.6: Other oligonucleotides

name	sequence 5' → 3'	length	use	source
ARB2	GGCCACGCGTCGACTAGTAC	20 nt	arbitrary - nested PCR	C. Tang, IMM, Oxford
ARB4	ggccacgcgctcgactagtcnnnnnnnnnnntgacg	34 nt	arbitrary PCR	C. Tang
ARB5	ggccacgcgctcgactagtcnnnnnnnnnnxxxxx	34 nt	arbitrary PCR	C. Tang
ARB6	ggccacgcgctcgactagtcnnnnnnnnnnnacgcc	34 nt	arbitrary PCR	C. Tang
ESP-f	GTACTTGAAATCCAGCAAGTATATAGC	27 nt	confirmation of pKL6 insert in <i>S. pombe</i> transformants	Stratagene
ESP-r	CAAAATCGTAATATGCAGCTTGAATGGGCTTCC	33 nt	confirmation of pKL6 insert in <i>S. pombe</i> transformants	Stratagene
NG61	GCATTACGCTGACTTGACGGGACG	24 nt	nested PCR, binds in <i>Kan^R</i> gene	C. Tang
NG62	TTGGTTAATTGGTTGTAACACTGG	24 nt	arbitrary PCR, binds in <i>Kan^R</i> gene	C. Tang
T3	AATTAACCCTCACTAAAGGG	20 nt	sequencing into 5' junction from vector	Stratagene, sequencing facility
T7	GTAATACGACTCACTATAGGGC	22 nt	sequencing into 3' junction from vector	Stratagene, sequencing facility

Upper case letters indicate sequences that are complementary to the template, lower case letters stand for other sequence, such as overhangs. n, degenerate sequence; x, unknown sequence.

Table 2.7: Plasmids created for this work

name	relevant features	parent plasmids	primers used for cloning	enzymes used for cloning	primers for sequencing	size (bp)
pKL1	dimer of pHFL-1	pHFL-1	-	-	KL13	c.12600
pKL2	dimer, pHFL-1 without <i>NdeI</i> site in <i>hobo</i> ORF1	pHFL-1	KL6, KL7	(<i>Bam</i> HI)	KL8 - KL13, KL66	c.12600
pKL3	monomer of pKL1	pHFL-1	-	<i>KpnI</i>	KL13, KL17, KL18, KL66, KL69, T3, T7	c.6300
pKL4	monomer of pKL2	pKL2	-	<i>KpnI</i>	KL8, KL12	c.6300
pKL5	expresses CBP- <i>hobo</i> tase fusion protein	pHFL-1 pCAL-n-EK	KL1, KL3 -	(Enterokinase)	KL8 - KL13	7779
pKL6	expresses GST- <i>hobo</i> tase fusion protein in <i>S. pombe</i>	pHFL-1 pESP-2	KL1, KL14 -	(Enterokinase)	-	11705
pKL7	expresses native <i>hobo</i> tase	pKL4 pET11a	KL4, KL5 -	<i>NdeI</i> <i>NdeI</i>	-	7657
pKL8	expresses 6xHis- <i>hobo</i> tase	pKL7 pET14b	- -	<i>NdeI</i> <i>NdeI</i>	-	6651
pKL9	expresses 10xHis- <i>hobo</i> tase	pKL7 pET16b	- -	<i>NdeI</i> <i>NdeI</i>	-	7691
pKL10	<i>hobo</i> minitransposon: left and right end flanking <i>Kan^R</i> , pBluescript backbone, <i>Amp^R</i>	pKL3 pNK1182 pBSII-SK+	KL19 - KL22 - -	<i>MluI</i> , <i>NotI</i> <i>MluI</i> <i>NotI</i>	-	5285
pKL11	<i>hobo</i> minitransposon: two left ends flanking <i>Kan^R</i> , pBluescript backbone, <i>Amp^R</i>	pKL3 pNK1182 pBSII-SK+	KL19 - KL22 - -	<i>MluI</i> , <i>NotI</i> <i>MluI</i> <i>NotI</i>	-	5205
pKL12	<i>hobo</i> minitransposon: two right ends flanking <i>Kan^R</i> , pBluescript backbone; <i>Amp^R</i>	pKL3 pNK1182 pBSII-SK+	KL19 - KL22 - -	<i>MluI</i> , <i>NotI</i> <i>MluI</i> <i>NotI</i>	-	5365
pKL13	expresses native <i>hobo</i> tase; p15A ori; <i>Tet^R</i>	pKL7 pACYC184	KL23/KL24 ^a , KL25/KL26 ^a	<i>BspEI</i> , <i>PpuMI</i> <i>PfI</i> MI	KL8 - KL13, KL29 - KL34	7038
pKL14	expresses 6xHis- <i>hobo</i> tase; p15A ori; <i>Tet^{Rr}</i>	pKL8 pACYC184	KL23/KL24 ^a , KL25/KL26 ^a	<i>BspEI</i> , <i>PpuMI</i> <i>PfI</i> MI	-	5841
pKL15	expresses 10xHis- <i>hobo</i> tase; p15A ori; <i>Tet^R</i>	pKL9 pACYC184	KL23/KL24 ^a , KL25/KL26 ^a	<i>BspEI</i> , <i>PpuMI</i> <i>PfI</i> MI	-	7072
pKL16	pOX38 with mini-Tn10- <i>Cam^R</i>	pOX38 pNK2884	- -	- -	-	?
pKL17	pEF-BOS with mini- <i>hobo</i> , <i>Kan^R</i> and <i>p_{EF1-α}</i> divergent orientation	pEF-BOS pKL10	- -	<i>NotI</i> ^K <i>SapI</i> ^K	-	8106
pKL18	pEF-BOS with mini- <i>hobo</i> , <i>Kan^R</i> and <i>p_{EF1-α}</i> same orientation	pEF-BOS pKL10	- -	<i>NotI</i> ^K <i>SapI</i> ^K	-	8106
pKL19	pEF-BOS expressing native <i>hobo</i> tase; mini- <i>hobo</i>	pKL17 pKL7	- -	<i>XbaI</i> ^K <i>NdeI</i> ^K	-	9654
pKL20	pEF-BOS; native <i>hobo</i> tase in reverse orientation; mini- <i>hobo</i>	pKL17 pKL7	- -	<i>XbaI</i> ^K <i>NdeI</i> ^K	-	9654
pKL21	pEF-BOS expressing 6xHis- <i>hobo</i> tase; mini- <i>hobo</i>	pKL17 pKL8	- -	<i>XbaI</i> ^K <i>NcoI</i> ^K , <i>XhoI</i> ^K	-	9721

pKL22	pEF-BOS; 6xHis- <i>hobo</i> tase reverse; mini- <i>hobo</i>	pKL17 pKL8	- -	<i>XbaI</i> ^K <i>NcoI</i> ^K , <i>XhoI</i> ^K	-	9721
pKL23	pEF-BOS expressing 10xHis- <i>hobo</i> tase; mini- <i>hobo</i>	pKL17 pKL9	- -	<i>XbaI</i> ^K <i>NcoI</i> ^K , <i>XhoI</i> ^K	-	9724
pKL24	pEF-BOS; 10xHis- <i>hobo</i> tase reverse; mini- <i>hobo</i>	pKL17 pKL9	- -	<i>XbaI</i> ^K <i>NcoI</i> ^K , <i>XhoI</i> ^K	-	9724
pKL25	pKL8, 2 rare Arg codons in 6xHis- <i>hobo</i> -tase exchanged	pKL8	KL35, KL36	(<i>HindII</i>)	KL8	6651
pKL26	inducible <i>hobo</i> tase expression in <i>S. cerevisiae</i> ; <i>URA3</i> ; <i>ARS/CEN</i>	pKL7 pYX113	- -	<i>NdeI</i> ^K <i>EcoRI</i> ^K	KL9, KL13	7801
pKL27	<i>hobo</i> tase reverse orientation; <i>URA3</i> ; <i>ARS/CEN</i>	pKL7 pYX113	- -	<i>NdeI</i> ^K <i>EcoRI</i> ^K	-	7801
pKL28	inducible <i>hobo</i> tase expression in <i>S. cerevisiae</i> ; <i>URA3</i> ; 2 μ	pKL7 pYX213	- -	<i>NdeI</i> ^K <i>EcoRI</i> ^K	KL9, KL13	9483
pKL29	<i>hobo</i> tase reverse orientation; <i>URA3</i> ; 2 μ	pKL7 pYX213	- -	<i>NdeI</i> ^K <i>EcoRI</i> ^K	-	9483
pKL30	inducible <i>hobo</i> tase expression in <i>S. cerevisiae</i> ; mini- <i>hobo</i> (<i>Kan</i> ^R) (reverse to pKL31-34); <i>URA3</i> ; <i>ARS/CEN</i>	pKL10 pKL26	- -	<i>NotI</i> ^K <i>AhdI</i> ^T	-	10128
pKL31	inducible <i>hobo</i> tase expression in <i>S. cerevisiae</i> ; mini- <i>hobo</i> (<i>Kan</i> ^R); <i>URA3</i> ; <i>ARS/CEN</i>	pKL10 pKL26	- -	<i>NotI</i> ^K <i>AhdI</i> ^T	-	10128
pKL32	<i>hobo</i> tase reverse; mini- <i>hobo</i> (<i>Kan</i> ^R); <i>URA3</i> ; <i>ARS/CEN</i>	pKL10 pKL27	- -	<i>NotI</i> ^K <i>AhdI</i> ^T	-	10128
pKL33	inducible <i>hobo</i> tase expression in <i>S. cerevisiae</i> ; mini- <i>hobo</i> (<i>Kan</i> ^R); <i>URA3</i> ; 2 μ	pKL10 pKL28	- -	<i>NotI</i> ^K <i>AhdI</i> ^T	-	11810
pKL34	<i>hobo</i> tase reverse; mini- <i>hobo</i> (<i>Kan</i> ^R); <i>URA3</i> ; 2 μ ; <i>Amp</i> ^R	pKL10 pKL29	- -	<i>NotI</i> ^K <i>AhdI</i> ^T	-	11810
pKL35	expresses <i>hobo</i> tase-EGFP fusion in mammalian cells	pKL3 pEGFP-N1,	KL61, KL62 -	<i>BamHI</i> <i>BamHI</i>	-	6716
pKL36	expresses <i>hobo</i> taseRR-EGFP fusion in mammalian cells	pKL25 pEGFP-N1	KL61, KL62 -	<i>BamHI</i> <i>BamHI</i>	-	6716
pKL37	expresses <i>hobo</i> tase-EGFP fusion in mammalian cells	pKL3 pEGFP-N1,	KL64, KL62 -	<i>KpnI</i> , <i>BamHI</i> <i>KpnI</i> , <i>BamHI</i>	KL8 – KL13, KL29 – KL33	6700
pKL38	expresses <i>hobo</i> taseRR-EGFP fusion in mammalian cells	pKL25 pEGFP-N1	KL64, KL62 -	<i>KpnI</i> , <i>BamHI</i> <i>KpnI</i> , <i>BamHI</i>	KL8 – KL13	6700
pKL39	expresses native <i>hobo</i> tase in mammalian cells	pKL3 pEGFP-N1,	KL64, KL65 -	<i>KpnI</i> , <i>NotI</i> <i>KpnI</i> , <i>NotI</i>	KL8 – KL13, KL29 – KL33	5964
pKL40	expresses native <i>hobo</i> taseRR in mammalian cells	pKL25 pEGFP-N1	KL64, KL65 -	<i>KpnI</i> , <i>NotI</i> <i>KpnI</i> , <i>NotI</i>	KL9, KL13, KL29, KL31 – KL33	5964
pKL41	<i>hobo</i> in plasmid with <i>Kan</i> ^R / <i>Neo</i> ^R	pKL1 pEGFP-N1	- -	<i>KpnI</i> ^T , <i>SacI</i> ^T <i>DraIII</i> ^T , <i>AflIII</i> ^T	-	c.6270
pKL42	<i>hobo</i> in plasmid with <i>Kan</i> ^R / <i>Tet</i> ^{Rr} , reverse	pKL1 pEGFP-N1	- -	<i>KpnI</i> ^T , <i>SacI</i> ^T <i>DraIII</i> ^T , <i>AflIII</i> ^T	-	c.6270

pKL43	<i>hobo</i> (without <i>NdeI</i> site) in plasmid with <i>Kan^R/Tet^{Rr}</i>	pKL2 pEGFP-N1	- -	<i>KpnI^T, SacI^T</i> <i>DraIII^T, AflIII^T</i>	-	c.6270
pKL44	<i>hobo</i> (without <i>NdeI</i> site) in plasmid with <i>Kan^R/Tet^{Rr}</i> , reverse orientation	pKL2 pEGFP-N1	- -	<i>KpnI^T, SacI^T</i> <i>DraIII^T, AflIII^T</i>	-	c.6270
pKL45	mini- <i>hobo</i> (<i>Kan^R</i>) in plasmid with <i>KanR/Tet^R</i>	pKL10 pEGFP-N1	- -	<i>KpnI^T, SacI^T</i> <i>DraIII^T, AflIII^T</i>	-	5228
pKL46	mini- <i>hobo</i> (<i>Kan^R</i>) in plasmid with <i>KanR/Tet^R</i> , reverse	pKL10 pEGFP-N1	- -	<i>KpnI^T, SacI^T</i> <i>DraIII^T, AflIII^T</i>	-	5228
pKL47	<i>Kan^R/Tet^R</i> in mini- <i>hobo</i> ; <i>Amp^R</i>	pKL10 pEGFP-N1	- -	<i>MluI^K</i> <i>DraIII^T, BsaI^T</i>	-	5650
pKL48	<i>Kan^R/Tet^{Rr}</i> in mini- <i>hobo</i> ; reverse orientation; <i>Amp^R</i>	pKL10 pEGFP-N1	- -	<i>MluI^K</i> <i>DraIII^T, BsaI^T</i>	-	5650
pKL49	expresses EGFP- <i>hobo</i> tase fusion in mammalian cells	pKL37 pEGFP-C1	- -	<i>KpnI, BamHI</i> <i>KpnI, BamHI</i>	-	6698
pKL50	mini- <i>hobo</i> (<i>sacB</i>); <i>Amp^R</i> ; <i>ColE1</i> ori	pKL10 pRC44	- -	<i>MluI^K</i> <i>PstI^T</i>	-	6386
pKL51	mini- <i>hobo</i> (<i>sacB</i>) (reverse to pKL50); <i>Amp^R</i> ; <i>ColE1</i> ori	pKL10 pRC44	- -	<i>MluI^K</i> <i>PstI^T</i>	-	6386
pKL52	mini- <i>hobo</i> (<i>sacB</i>); pEF-BOS expressing native <i>hobo</i> tase in mammalian cells	pKL19 pRC44	- -	<i>Eco47III</i> <i>PstI^T</i>	-	10837
pKL53	mini- <i>hobo</i> (<i>sacB</i>) (reverse to pKL50); pEF-BOS expressing native <i>hobo</i> tase in mammalian cells	pKL19 pRC44	- -	<i>Eco47III</i> <i>PstI^T</i>	-	10837
pKL54	<i>p_{tac}-hobo</i> tase; <i>Amp^R</i> ; M13 ori; <i>ColE1</i> ori; <i>rop</i> ; <i>lacI^q</i>	pKL8 pMAL-p2X	- -	<i>NdeI</i> <i>NdeI</i>	-	8703
pKL55	<i>p_{tac}-hobo</i> tase reverse orientation; <i>Amp^R</i> ; M13 ori; <i>ColE1</i> ori; <i>rop</i> ; <i>lacI^q</i>	pKL8 pMAL-p2X	- -	<i>NdeI</i> <i>NdeI</i>	-	8703
pKL56	mini- <i>hobo</i> (<i>Kan^R</i>); <i>p_{tac}-hobo</i> tase; <i>ColE1</i> ori; <i>rop</i> ; <i>lacI^q</i>	pKL10 pKL54	- -	<i>NotI^K</i> <i>BglII^K, NaeI</i>	-	8217
pKL57	mini- <i>hobo</i> (<i>Kan^R</i>) (reverse to pKL56); <i>p_{tac}-hobo</i> tase; <i>ColE1</i> ori; <i>rop</i> ; <i>lacI^q</i>	pKL10 pKL54	- -	<i>NotI^K</i> <i>BglII^K, NaeI</i>	-	8217
pKL58	<i>p_{MT}-hobo</i> -tase; <i>ColE1</i> ori; <i>Amp^R</i>	pKL39 pMT/V5-HisB	- -	<i>KpnI, NotI</i> <i>KpnI, NotI</i>	-	5478
pKL59	<i>p_{MT}-hobo</i> -tase-EGFP; <i>ColE1</i> ori; <i>Amp^R</i>	pKL37 pMT/V5-HisB	- -	<i>KpnI, NotI</i> <i>KpnI, NotI</i>	-	6214
pKL60	<i>p_{MT}-hobo</i> -tase-V5-6xHis; <i>ColE1</i> ori; <i>Amp^R</i>	pKL37 pMT/V5-HisB	- -	<i>KpnI, BamHI^K</i> <i>KpnI, SacII^T</i>	-	5452
pKL61	<i>p_{MT}-hobo</i> -tase-6xHis; <i>ColE1</i> ori; <i>Amp^R</i>	pKL37 pMT/V5-HisB	- -	<i>KpnI, BamHI^K</i> <i>KpnI, AgeI^K</i>	-	5398
pKL62	<i>p_{MT}-6xHis-hobo</i> -tase; <i>ColE1</i> ori; <i>Amp^R</i>	pKL8 pMT/V5-HisB	- -	<i>NcoI^K, XhoI</i> <i>EcoRV, XhoI</i>	-	5566
pKL63	<i>p_{MT}-10xHis-hobo</i> -tase; <i>ColE1</i> ori; <i>Amp^R</i>	pKL9 pMT/V5-HisB	- -	<i>NcoI^K, XhoI</i> <i>EcoRV, XhoI</i>	-	5569
pKL64	5' MoMuLV LTR; ψ^+ ; <i>Neo^R</i> ; <i>p_{CMV}-hobo</i> -tase; 3' MoMuLV LTR; pBR322 ori; <i>Amp^R</i>	pKL39 pLNCX	- -	<i>SacI^T, HpaI</i> <i>HpaI</i>	-	8751

pKL65	5' MoMuLV LTR; ψ^+ ; <i>Neo^R</i> ; <i>p_{CMV}</i> ; <i>hobo</i> -tase (reverse); 3' MoMuLV LTR; pBR322 ori; <i>Amp^R</i>	pKL39 pLNCX	- -	<i>SacI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	8751
pKL66	5' MoMuLV LTR; ψ^+ ; <i>Neo^R</i> ; <i>p_{CMV}</i> - <i>hobo</i> -tase-EGFP; 3' MoMuLV LTR; pBR322 ori; <i>Amp^R</i>	pKL37 pLNCX	- -	<i>SacI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9487
pKL67	5' MoMuLV LTR; ψ^+ ; <i>Neo^R</i> ; <i>p_{CMV}</i> ; <i>hobo</i> -tase-EGFP (reverse); 3' MoMuLV LTR; pBR322 ori; <i>Amp^R</i>	pKL37 pLNCX	- -	<i>SacI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9487
pKL68	5' MoMuLV LTR; ψ^+ ; <i>Neo^R</i> ; <i>p_{CMV}</i> -EGFP- <i>hobo</i> -tase; 3' MoMuLV LTR; pBR322 ori; <i>Amp^R</i>	pKL49 pLNCX	- -	<i>NheI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9514
pKL69	5' MoMuLV LTR; ψ^+ ; <i>Neo^R</i> ; <i>p_{CMV}</i> ; EGFP- <i>hobo</i> -tase (reverse) ; 3' MoMuLV LTR; pBR322 ori; <i>Amp^R</i>	pKL49 pLNCX	- -	<i>NheI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9514
pKL70	CMV RU5; ψ^+ ; <i>hobo</i> -tase; <i>p_{SV40}</i> - <i>Neo^R</i> ; pBR322 ori; <i>Amp^R</i>	pKL39 pRV583	- -	<i>SacI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	8666
pKL71	CMV RU5; ψ^+ ; <i>hobo</i> -tase (reverse orientation); <i>p_{SV40}</i> - <i>Neo^R</i> ; pBR322 ori; <i>Amp^R</i>	pKL39 pRV583	- -	<i>SacI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	8666
pKL72	CMV RU5; ψ^+ ; <i>hobo</i> -tase-EGFP; <i>p_{SV40}</i> - <i>Neo^R</i> ; pBR322 ori; <i>Amp^R</i>	pKL37 pRV583	- -	<i>SacI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9402
pKL73	CMV RU5; ψ^+ ; <i>hobo</i> -tase-EGFP (reverse orientation); <i>p_{SV40}</i> - <i>Neo^R</i> ; pBR322 ori; <i>Amp^R</i>	pKL37 pRV583	- -	<i>SacI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9402
pKL74	CMV RU5; ψ^+ ; EGFP- <i>hobo</i> -tase; <i>p_{SV40}</i> - <i>Neo^R</i> ; pBR322 ori; <i>Amp^R</i>	pKL49 pRV583	- -	<i>NheI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9429
pKL75	CMV RU5; ψ^+ ; EGFP- <i>hobo</i> -tase (reverse); <i>p_{SV40}</i> - <i>Neo^R</i> ; pBR322 ori; <i>Amp^R</i>	pKL49 pRV583	- -	<i>NheI^T</i> , <i>HpaI</i> <i>HpaI</i>	-	9429
pKL86	<i>p_{MT}</i> -10xHis- <i>hobo</i> -tase; in mini- <i>hobo</i> (<i>sacB</i>); <i>ColE1</i> ori; <i>Amp^R</i>	pKL50 pKL63	- -	<i>NotI^K</i> <i>HpaI</i>	-	7897
pKL87	<i>p_{tac}</i> -6xHis- <i>hobo</i> -tase Δ 7aa(C-term)-intein-CBD; M13ori; <i>ColE1</i> ori; <i>rop</i> ; <i>lacI</i> ; <i>Amp^R</i>	pKL8 pCYB3	KL123, KL124 -	<i>NcoI</i> , <i>PvuII</i> <i>NcoI</i> , <i>SapI^K</i>	-	8832
pKL88	<i>p_{T7}</i> -6xHis- <i>hobo</i> -tase Δ 7aa(C-term)-intein-CBD; M13ori; <i>ColE1</i> ori; <i>rop</i> ; <i>lacI</i> ; <i>Amp^R</i>	pKL8 pTYB3	KL123, KL124 -	<i>NcoI</i> , <i>PvuII</i> <i>NcoI</i> , <i>SapI^K</i>	-	9460
pKL89	<i>in vitro</i> transcription & translation in rabbit reticulocyte extract: <i>tk_L</i> ; <i>p_{T7}</i> ; EMCV UTR- <i>hobo</i> tase-polyA; <i>t_{T7}</i> ; <i>tk_R</i> ; <i>f1(+)</i> ori; <i>Amp^R</i> ; pUC19 ori	pKL39 pTM1	KL125, KL126 -	<i>NcoI</i> , <i>XhoI</i> <i>NcoI</i> , <i>XhoI</i>	KL8 – KL13	7527
pKL91	mini- <i>hobo</i> <i>Kan^R</i> ; pSC101 ori, Rep AT (ts)	pKL10 pLN135	- -	<i>NotI^K</i> <i>BsaHI^K</i>	-	4074

pKL93	mini- <i>hobo</i> <i>Kan^R</i> ; pSC101 ori, Rep AT (ts); <i>rpsL</i> (<i>Str^S</i>)	pKL10 pLN135	- -	<i>NotI</i> ^K <i>HaeI</i> ^T	-	4628
pKL94	<i>p_{T7}-hobo-tase-6xHis</i> ; <i>f1ori</i> ; pBR322 ori; <i>lacI</i> ; <i>Amp^R</i>	pKL8 pET32a(+)	KL128, KL129 -	<i>NdeI</i> , <i>XhoI</i> <i>NdeI</i> , <i>XhoI</i>	KL8 – KL13	7345
pKL96	<i>Himar1</i> hairpin end substrate in pBS	p27fh-5' pBSII-SK+	KL130, KL131 -	<i>BsaI</i> , <i>XhoI</i> <i>XhoI</i>	T3	3075
pKL97	<i>Himar1</i> 5'ITR substrate in pBS	p27fh-5' pBSII-SK+	KL131, KL132 -	<i>XhoI</i> <i>XhoI</i>	T3	3077
pKL98	<i>Himar1</i> 5'ITR substrate reverse in pBS	p27fh-5' pBSII-SK+	KL131, KL132 -	<i>XhoI</i> <i>XhoI</i>	T3	3077
pKL99	<i>Himar1</i> 5'ITR substrate 2x in pBS	p27fh-5' pBSII-SK+	KL131, KL132 -	<i>XhoI</i> <i>XhoI</i>	T3	3193
pKL100	<i>Himar1</i> 5'ITR substrate reverse & forward in pBS	p27fh-5' pBSII-SK+	KL131, KL132 -	<i>XhoI</i> <i>XhoI</i>	T3	3193
pKL101	<i>hobo</i> hairpin end substrate in pBS	pKL10 pBSII-SK+	KL70, KL133 -	<i>BsaI</i> , <i>XhoI</i> <i>XhoI</i>	T3	3091
pKL102	<i>hobo</i> left end substrate in pBS	pKL10 pBSII-SK+	KL133, KL134 -	<i>XhoI</i> <i>XhoI</i>	T3	3118
pKL103	<i>hobo</i> left end substrate reverse in pBS	pKL10 pBSII-SK+	KL133, KL134 -	<i>XhoI</i> <i>XhoI</i>	T3	3118
pKL104	pBSII-SK+ without MCS	pBSII-SK+	-	<i>PvuII</i>	-	2513
pKL105	<i>Himar1</i> 5' ITR with <i>BsrDI</i> site to produce staggered cut end in pBS	pKL97 pBSII-SK+	KL146, KL147 -	<i>XhoI</i> <i>XhoI</i>	-	3028
pKL106	<i>Himar1</i> 5' ITR <i>BsrDI</i> staggered reverse in pBS	pKL97 pBSII-SK+	KL146, KL147 -	<i>XhoI</i> <i>XhoI</i>	-	3028
pKL107	<i>Himar1</i> 5' ITR <i>BsrDI</i> staggered 2x in pBS	pKL97 pBSII-SK+	KL146, KL147 -	<i>XhoI</i> <i>XhoI</i>	-	3095
pKL108	<i>Himar1</i> 5' ITR <i>BsrDI</i> staggered 2x reverse in pBS	pKL97 pBSII-SK+	KL146, KL147 -	<i>XhoI</i> <i>XhoI</i>	-	3098
pKL109	<i>Himar1</i> 5' ITR <i>BsrDI</i> staggered 3x reverse in pBS	pKL97 pBSII-SK+	KL146, KL147 -	<i>XhoI</i> <i>XhoI</i>	-	3168

column 1

2

3

4

5

6

7

Plasmids created for this work.

^a, annealed adapter for ligation (all other primers are PCR-primers); ^K, treated with Klenow Fragment DNA Polymerase I for 5' overhang fill-in after digestion; ^T, treated with T4 DNA Polymerase for 3' overhang removal after digestion. For sequenced region, see Table 2.5 and Table 2.6. See Section 2.2.1 for cloning procedure and further explanation.

Table 2.8: Other plasmids

name	relevant features	parent plasmid	size (bp)	reference or source
p1182	Tn10 minitransposon: inside ends flanking Tn903 <i>Kan^R</i> ; <i>Amp^R</i>	pBR322	6608	Morisato and Kleckner, 1987
p27fh-5'	<i>Himar1</i> 5' ITR	pcDNAII	2789	Lampe <i>et al.</i> , 1996
pACYC184	<i>Tet^R</i> ; <i>Cam^R</i> , p15A ori		4245	NEB
pBluescriptII-SK+ = pBSII-SK+	<i>Amp^R</i> ; ColE1 ori; f1 ori		2961	Stratagene
pCAL-n-EK	expresses N-terminal CBP-tag, cleavable at EK-recognition site; <i>p_{T7}</i> , CBP, Thrombin target, EK recognition, MCS, <i>t_{T7}</i> ; <i>Amp^R</i> ; ColE1 ori; <i>lacI^q</i>	pET11	5825	Stratagene
pCRScript-mouse IF γ R	mouse interferon γ receptor gene	pCR-Script Amp SK(+)	c.3760	Kumar <i>et al.</i> , 1989; Julian Symons, Pathology Dept., Oxford
pCYB3	expresses intein-chitin binding domain-fusions; <i>p_{tac}</i> -intein-CBD; <i>Amp^R</i> ; M13 ori; ColE1 ori; rop; <i>lacI</i>		6846	NEB
pEF-BOS	expresses native protein in mammalian cells; <i>p_{EF1-α}</i> , MCS, polyA; SV40 ori; M13 Ig; <i>Amp^R</i> ; ColE1 ori	pUC119	c.5775	Mizushima and Nagata, 1990
pEGFP-C1	expresses N-terminal EGFP-fusion in mammalian cells; <i>p_{CMV}</i> , EGFP, MCS; SV40 polyA; f1 ori; <i>p_{bact}</i> , SV40 ori, <i>p_{SV40e}</i> , <i>Kan^R/Neo^R</i> , HSV TK polyA; pUC ori		4731	Clontech
pEGFP-N1 (=pRC97)	expresses C-terminal EGFP-fusion in mammalian cells; <i>p_{CMV}</i> , MCS, EGFP, SV40 polyA; f1 ori; <i>p_{bact}</i> , SV40 ori, <i>p_{SV40e}</i> , <i>Kan^R/Neo^R</i> , HSV TK polyA; pUC ori		4733	Clontech
pEPO	murine erythropoietin gene			Shoemaker and Mitsock, 1986; Andrew Thomson, IMM, Oxford
pESP-2	expresses N-terminal GST-tag in <i>S. pombe</i> , cleavable at EK-recognition site; <i>p_{nmt1}</i> ; <i>t_{nmt1}</i> ; <i>Amp^R</i> ; ColE1 ori; <i>ars-1</i> ; <i>LEU_{2-d}</i>	pESP-1	9760	Stratagene
pESP2-GST	expresses GST-GST in <i>S. pombe</i> ; <i>p_{nmt1}</i> ; <i>t_{nmt1}</i> ; <i>Amp^R</i> ; ColE1 ori; <i>ars-1</i> ; <i>LEU_{2-d}</i>	pESP-2		Stratagene
pET-11a	expresses C-terminal T7-tag; <i>p_{T7}</i> ; <i>t_{T7}</i> ; <i>lacI^q</i> ; <i>Amp^R</i> ; pBR322 ori		5677	Novagen
pET-14b	expresses N-terminal 6xHis-tag, cleavable with thrombin; <i>p_{T7}</i> ; <i>t_{T7}</i> ; <i>Amp^R</i> ; pBR322 ori		4671	Novagen

pET-16b	expresses N-terminal 10xHis-tag, cleavable with Factor Xa; p_{T7} ; t_{T7} ; $lacI^q$; Amp^R ; pBR322 ori		5711	Novagen
pET-32a(+)	expresses N-terminal Trx-tag, 6xHis-tag (cleavable w/ thrombin), S-tag (cleavable w/ enterokinase), C-terminal 6xHis-tag; p_{T7} ; t_{T7} ; $lacI$; Amp^R ; pBR322 ori; f1 ori		5900	Novagen
pGFP (=pRC95)	constitutive EGFP-reporter for mammalian cells to assay transfection efficiency: p_{CMV} -EGFP; BGH pA; f1 ori; SV40 ori; Neo^R ; SV40 pA; ColE1 ori; Amp^R	pcDNA3.1; pEGFP-N1	c. 7400	Chris Pugh, IMM, Oxford
pGFP-i-lacZ	retroviral delivery and expression of EGFP and lacZ	pRV583		Kingsman lab
pHFL-1	<i>hobo</i> transposon and genomic DNA from <i>D. melanogaster</i> in pBluescript-KS	pBS-KS	c.6300	Blackman <i>et al.</i> , 1989
pHIT60	codes for MoMuLV core : p_{CMV} - <i>gag-pol</i> ; SV40 ori; Amp^R			Soneoka <i>et al.</i> , 1995
pHIT123	codes for MoMuLV envelope protein to infect mouse cells: p_{CMV} - <i>env</i> ; SV40 ori			Soneoka <i>et al.</i> , 1995
pLN135	pSC101 ori, Rep AT (ts); ; <i>rpsL</i> (Str^S)		4409	sent by F. Cornet
pLNCX	retroviral delivery and expression; 5' MoMuSV LTR, Ψ^+ , Neo^R ; p_{CMV} , MCS; 3' MoMuLV LTR; ColE1 ori; Amp^R		6620	Miller and Rosman, 1989; Clontech
pLucT7 ("Luciferase T7 Control DNA")	<i>in vitro</i> expression of luciferase; p_{T7} - <i>luc</i> -poly(A) ₃₀ ; ori; Amp^R		5365	Promega
pLXSN	retroviral delivery and expression; 5' MoMuLV LTR, Ψ^+ , MCS; p_{SV40e} , Neo^R ; 3' MoMuLV LTR; ColE1 ori; Amp^R		5874	Miller and Rosman, 1989; Clontech
pMAL-p2X	p_{tac} - <i>malE</i> ; Amp^R ; M13 ori; ColE1 ori, rop; $lacI^q$		6721	NEB
pMarKan	mini- <i>Himar1</i> with Kan^R and truncated tase; ori		3836	Lampe <i>et al.</i> , 1996
pMAL-C9	p_{tac} , MBP- <i>Himar1</i> tase C9 (hyperactive); Amp^R	pMAL		D. Lampe
pMAL-Kam	p_{tac} , MBP- <i>Himar1</i> tase (wt); Amp^R	pMAL		D. Lampe
pMPL	286 bp from mouse <i>c-mpl</i> cDNA; Amp^R	pSPT18	3390	Thomson <i>et al.</i> , 1999; Andrew Thomson, IMM, Oxford
pMW324	expresses 6xHis- <i>S.pombe</i> Rec12 (345 aa, 40 kD)	pET14b	c. 5700	M.C. Whitby, unpublished
pMYB129	expresses MBP-intein-chitin binding domain-fusion; p_{tac} - <i>malE</i> -intein-CBD; Amp^R ; M13 ori; ColE1 ori; rop; $lacI$	pCYB	8185	NEB

pNK81	Tn10 (<i>Tet^R</i>) ; <i>Salmonella</i> his operon; <i>Amp^R</i> ; ori	pBR322	c.17000	Foster <i>et al.</i> , 1981; Way <i>et al.</i> , 1984
pNK2884	Tn10 minitransposon: IS10 R outside ends flanking <i>Cam^R</i> ; ATS-IS10 transposase; <i>Amp^R</i>	pBR322		Kleckner <i>et al.</i> , 1991
pNK3287	Tn10 minitransposon: IS10 R outside ends flanking <i>Kan^R</i> ; <i>Amp^R</i>			
pOX38	<i>IncFI</i> ; <i>Tra⁺</i> ; <i>RepFIA⁺</i> ; f1 <i>HindII</i> fragment of F	F		Guyer <i>et al.</i> , 1980
pRC44	<i>sacB</i> ; <i>Kan^R</i>			
pRC60	expresses IS10 transposase (402 aa, 45 kD); <i>p_{T7}</i>	pET3a		
pRC62	IS10 tase on pACYC184; <i>Tet^R</i>	pACYC184		
pRC98	mammalian Neo-resistance for co-transfection: <i>Neo^R</i> ; <i>Amp^R</i> ; ColE1 ori	pSV2-neo; pTZ18R	c. 5700	William Brown, Biochemistry Dept., Oxford
pRV583	retroviral delivery and expression; CMV RU5, Ψ^+ , MCS; <i>p_{SV40e}</i> , <i>Neo^R</i> ; 3' MoMuLV LTR; ColE1 ori; <i>Amp^R</i>	pLXSN	6535	Kingsman lab
pTM1	gives high expression in mammalian cells and extracts; <i>tk_L</i> ; <i>p_{T7}</i> ; EMCV leader sequence; <i>t_{T7}</i> ; <i>tk_R</i> ; f1(+) ori; <i>Amp^R</i> ; pUC19 ori		5357	Moss <i>et al.</i> , 1990
pTYB3	<i>p_{T7}</i> -Sce VMA Intein-CBD; <i>Amp^R</i> ; M13 ori; ColEI ori, <i>rop</i> ; <i>lacI</i>		7477	NEB
pYX113	expresses native protein in <i>S. cerevisiae</i> ; <i>p_{GAL}</i> ; polyA; f1 ori; <i>URA3</i> ; <i>ARS/CEN</i> ; <i>Amp^R</i> ; ColE1 ori		5815	Ingenius
pYX213	expresses native protein in <i>S. cerevisiae</i> ; <i>p_{GAL}</i> ; polyA; f1 ori; <i>URA3</i> ; <i>2μ</i> ; <i>Amp^R</i> ; ColE1 ori		7497	Ingenius
R388	conjugative plasmid, <i>Tp^R</i> ; i		33918	Amyes and Smith, 1976; Avila and de la Cruz, 1988

Other plasmids used in this work.

Table 2.9: Artificial templates for testing SPPCR and FDPCR

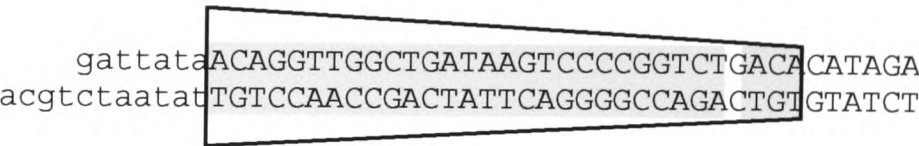
template structure	name	flanking IR	plasmid of origin	restriction site(s)	oligonucleotide primers			size (bp)		
					creation	adapters	amplific.	hairpin/ds ^b	1x-product	2x-product
hairpin	42D-h	<i>hobo</i> right	pKL42	<i>Dra</i> I	KL67, KL107	-	KL114	280	252	504
	42D-h	<i>hobo</i> right	pKL42	<i>Dra</i> I	KL67, KL107	-	KL115	280	216	432
	42S-h	<i>hobo</i> left	pKL42	<i>Sfi</i> I	KL70, KL102	-	KL80	708	704	1408
	42S-h	<i>hobo</i> left	pKL42	<i>Sfi</i> I	KL70, KL102	-	KL111	708	451	902
	42S-h	<i>hobo</i> left	pKL42	<i>Sfi</i> I	KL70, KL102	-	KL112	708	376	752
	42S-h	<i>hobo</i> left	pKL42	<i>Sfi</i> I	KL70, KL102	-	KL113	708	228	456
hairpin + frayed duplex	10B-hf	<i>hobo</i> right	pKL10	<i>Bgl</i> I	KL67, KL72	KL85, KL86	KL87, KL88	411	474	948
	42D-hf	<i>hobo</i> right	pKL42	<i>Dra</i> I	KL67, KL107	KL100, KL98	KL87, KL88	280	350	700
	42S-hf	<i>hobo</i> left	pKL42	<i>Sfi</i> I	KL70, KL102	KL99, KL95	KL87, KL88	708	779	1557
	42S-hf	<i>hobo</i> left	pKL42	<i>Sfi</i> I	KL70, KL102	KL99, KL95	KL108, KL109	708	771	1540
	96S-hf	<i>Himar</i> I left	pKL96	<i>Sfi</i> I	- ^a	KL135, KL136	KL87, KL88	50	121	241
	101S-hf	<i>hobo</i> left	pKL101	<i>Sfi</i> I	- ^a	KL135, KL136	KL87, KL88	58	129	257
open double strand + frayed duplex	42BB-df	<i>hobo</i> left	pKL42	<i>Bgl</i> II, <i>Bst</i> Z17I	-	KL99, KL95	KL87, KL88	941 ^c (708+233)	(1012) ^d	(2023) ^e
	99SB-df	<i>Himar</i> I left	pKL99	<i>Sfi</i> I, <i>Bsa</i> I	- ^a	KL135, KL136	KL87, KL88	100 ^c (50+50)	(171) ^d	(341) ^e
	102SB-df	<i>hobo</i> left	pKL102	<i>Sfi</i> I, <i>Bsa</i> I	- ^a	KL135, KL136	KL87, KL88	147 ^c (58+89)	(218) ^d	(435) ^e
column 1	2	3	4	5	6	7	8	9	10	11

^a, generated by restriction digest from plasmids especially created for this purpose; ^b, size without adapter; ^c, structure of negative control templates, e.g. 42BB-df: adapter + 708 bp left flanking sequence (as in 42S-h) + 233 bp *hobo* end sequence, not hairpinned. Note that due to overlapping recognition sequences, all *Sfi*I sites are also cut by *Bgl*II. ^d, linear amplification only, or would arise if strand switching occurred exactly at template end. ^e, theoretical length, would arise if strand switching occurred exactly at template end.

Plasmid restriction fragments

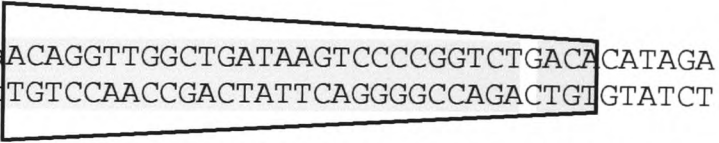
97PS : pKL97PstI - SalI (11 + 67 bp) 

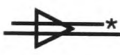
TGGCGTCGCTAGTAGAGACCTCGAGG.*..
ACCGCAGCGATCATCTCTGGAGCTCCAGCT



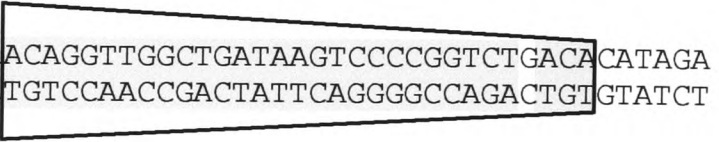
97SX : pKL97 SfiI - XbaI (50 + 124 bp) 

TGGCGTCGCTAGTAGAGACCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCCGGGGGATCCACTAGTT*...
ACCGCAGCGATCATCTCTGGAGCTCCAGCTGCCATAGCTATTCGAACTATAGCTTAAGGACGTCGGGCCCCCTAGGTGATCAAGATC



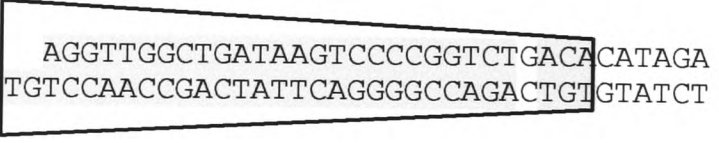
99SX : pKL99 SfiI - XhoI (50 + 111 bp) 

TGGCGTCGCTAGTAGAGACC.*..
ACCGCAGCGATCATCTCTGGAGCT



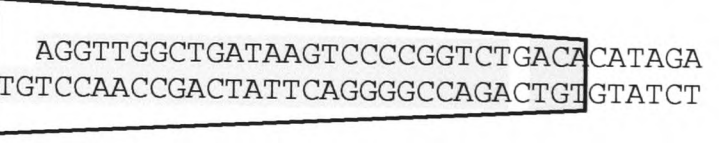
105BS : pKL105 BsrDI - SalI (69 bp) 

TGGCGTCGCTAGTAGAGACCTCGAGG.*..
ACCGCAGCGATCATCTCTGGAGCTCCAGCT



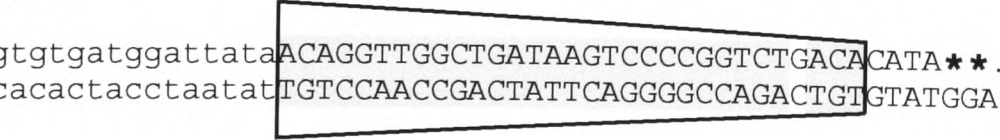
105BX : pKL105 BsrDI - XbaI (124 bp) 

TGGCGTCGCTAGTAGAGACCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCCTGCAGCCCCGGGGGATCCACTAGTT*...
ACCGCAGCGATCATCTCTGGAGCTCCAGCTGCCATAGCTATTCGAACTATAGCTTAAGGACGTCGGGCCCCCTAGGTGATCAAGATC



Annealed oligonucleotides

KL143 (15 + 38 bp) 
KL144



KL145 (38 bp) 
KL142

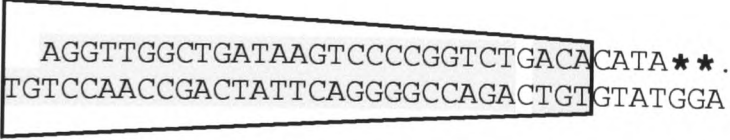


Fig. 2.1

Figure 2.1: Major *Himar1* transposon end substrates

Lower case, flanking DNA; upper case, transposon; asterisk, α -³²P-dCMP; grey boxes, inverted terminal repeats.

Table 2.10: Transposon end substrates

origin		substrate	IR	cleaved	total	length (bp) flanking transposon		restriction site total transposon		labelled end
hobo	plasmid pKL	102BX	5'	no	152	58	94	BglII	XhoI	transposon
		102SX	5'	no	152	58	94	SfiI	XhoI	transposon
	oligos	41/42	5'	no	35	11	24	-	-	both
		43/44	3'	no	39	10	29	-	-	both
		45/46	5'	blunt	24	-	24	-	-	transposon
		47/48	3'	blunt	29	-	29	-	-	transposon
		49/50	5'	no	72	11	61	-	-	both
		51/52	3'	no	72	10	62	-	-	both
HimarI	plasmid pKL	27SP	5'	no	228	41	187	SpeI	PvuII	flanking
		97AS	5'	no	223	74	149	Acc65I	SacI	flanking
		97BD	5'	no	168	108	60	BssHI	DrdI	flanking
		97KB	5'	no	187	74	113	KpnI	BamHI	transposon
		97NS	5'	no	827	397	430	NaeI	SapI	transposon
		97PS	5'	no	78	11	67	PstI	SalI	transposon
		97SS	5'	no	480	50	430	SfiI	SapI	transposon
		97SX	5'	no	174	50	124	SfiI	XbaI	transposon
		99SX	5'	no	111	50	61	SfiI	XhoI	transposon
		99X	5'	no	120	59	61	XhoI	XhoI	both
		105BS	5'	-2 nt	69	-	67	BsrDI	SalI	transposon
		105BX	5'	-2 nt	124	-	124	BsrDI	XbaI	transposon
	oligos	141/142	5'	blunt	38	-	38	-	-	transposon
		143/144	5'	no	53	15	38	-	-	transposon
		145/142	5'	-2 nt	38	-	38	-	-	transposon
Tn10	plasmid pPKC	3XH	5'	no	147	62	85	XbaI	HincII	flanking

hobo and *Himar1* substrates were excised from pKL plasmids with the matching number, e.g. 102BX was created by a *Bgl*I-*Xho*I digest of pKL102. Oligo ends are annealed from two oligonucleotides of the corresponding KL numbers (Table 2.5). -2 nt, staggered cut with a 2 nt 5' recess. The stated length (bp) includes both paired and unpaired (single-stranded) nucleotides.

2.1.3 Chemicals

Chemicals were generally of the highest quality available. Most chemicals were purchased from Sigma, unless stated otherwise. Restriction enzymes, T4 DNA ligase, T4 DNA polymerase, DNA polymerase I Klenow fragment for filling-in of overhangs and BSA for cloning and molecular weight markers (DNA and protein) were from New England BioLabs. BSA used in the *hobo* and *Himar1* transposition reactions and Klenow Fragment used for radioactive labelling were from Roche. Cloned *Pfu* and *PfuTurbo* DNA polymerases were from Stratagene. *Tth* polymerase, *Tfl* polymerase and *MasterAmp* PCR enhancer were from EpicentreTechnologies. Radiochemicals (α - ^{32}P -dATP or α - ^{32}P -dCTP) were from Amersham Biosciences at a specific activity of 6000 Ci/mmol. Polyacrylamide (30% 29:1 and 40% 19:1 acrylamide/bis) was from BioRad or National Diagnostics. Cell culture media and selective reagents were from GibcoBRL. SYBR Green and SYBR Gold are from Cambridge BioScience.

DNA-preparation and purification kits (maxipreps, QIAquick) were from QIAGEN. The AffinityTM LIC Cloning and Protein Purification Kit by Stratagene was used for ligation independent cloning.

Bacterial growth media and antibiotics were prepared as described (Sambrook *et al.*, 1989). Agarose and polyacrylamide gel-electrophoresis was done as described (Sambrook *et al.*, 1989) at the varying gel concentrations given in TBE or TAE.

2.2 General DNA manipulation

2.2.1 Cloning

2.2.1.1 Polymerase Chain Reaction

For cloning, DNA was amplified with cloned *Pfu* or *PfuTurbo* polymerase (Stratagene) in 1x cPfu buffer with 500 nM of each primer (Table 2.7, column 4; Table 2.5), 200 nM of each dNTP, 0.1 ng/ μl template and 5 units of enzyme per 100 μl reaction volume. Any EDTA present, as in TE, was neutralised by the addition of equimolar amounts of magnesium chloride. After an initial 2 min denaturation step at 94°C, the DNA was denatured 45 s at 94°C, annealed 45 s at 5.0°C below the calculated T_m (see Table 2.5), and elongated at 72°C for 2 min per kb of product, for 20-30 cycles, followed by a final 10-20 min elongation step. An aliquot was analysed by agarose gel electrophoresis. The DNA was separated from other reaction components with the QIAquick PCR-purification kit.

2.2.1.2 Restriction digest and end-treatment

2-5 μg plasmid DNA or purified PCR product was digested with the chosen restriction enzyme(s) (Table 2.7, lane 5) in the buffer supplied for 4 h at the recommended temperature. Restriction enzymes were usually from NEB. If the DNA was to be blunt-ended, 5' overhangs were

filled in with Klenow polymerase (NEB), and 3' overhangs were removed with T4 DNA polymerase (NEB) according to the manufacturer's instructions (Where used, blunt-ending of digested DNA is indicated by the indices ^K or ^T, respectively, in Table 2.7, lane 5.) Of the two fragments to be joined, the one with the origin of replication and/or the selectable marker (usually the vector backbone) was dephosphorylated by two sequential 1 h incubations with calf intestinal phosphatase (CIP, NEB). The restriction fragments were separated on 1x TAE, agarose gels at 30-100 V, excised and purified with the QIAquick gel extraction kit.

2.2.1.3 Ligation

For sticky-end ligations, 50 ng linear vector DNA and different ratios of insert DNA were incubated with 400 units T4 DNA ligase (NEB) in the supplied buffer at 16°C overnight.

Blunt-end ligations were performed in 50 mM Tris-HCl (pH7.5), 10 mM MgCl₂, 10 mM DTT, 0.05 mM ATP, 25 µg/ml BSA and 1.25 mM hexamine cobalt chloride. Up to 500 ng of linear vector and insert DNA in two or three different ratios were mixed with 800-2000 units T4 DNA ligase. The ligation was performed in a thermocycler for 30 min at 16°C, followed by 80 cycles of 2 min at 4°C and 4 min at 16°C, and several hours at 16°C.

2.2.1.4 Transformation of *E. coli*

Chemically competent bacteria were prepared with RbCl and transformed as described (Glover, 1985-1987). Electrocompetent cells were prepared as described in the BioRad Gene Pulser manual and electroporated with a CellJect electroporation system (Flowgen).

2.2.1.5 DNA purification

For screening of constructs, plasmid DNA was prepared by boiling preps (modified from Clewell and Helinski, 1970). For most applications and for storage, plasmids were purified with the QIAGEN midiprep or maxiprep kit. Plasmids that were to be transfected into mammalian cells, were prepared with the QIAGEN EndoFree Plasmid Maxi kit. Alkaline lysis and CsCl gradient centrifugation (Sambrook *et al.*, 1989) were used for very large plasmids (R388- and pOX38-derivatives) and to obtain maximum purity of transposition substrates (pMarKan).

2.2.1.6 Sequencing

The sequence of orfs and junctions of plasmids that had been created with PCR were confirmed by automated sequencing of the DNA (Table 2.7, lane 6; Table 2.5) (Biochemistry Department Sequencing Facility). For manual sequencing, the T7 Sequenase DNA sequencing kit v.2.0 (Amersham Life Sciences) was used.

2.2.1.7 Analytical Polymerase Chain Reaction

For testing *S. pombe* transformants for rearrangements of pKL6, colony-PCR was performed with Taq Long Plus polymerase (Stratagene) in a buffer containing 50 mM KCl, 0.1% gelatin and 0.1% Tween-20. Primers ESP-f and ESP-r (Table 2.5), which bind on both sides of the insert, were annealed at 48°C and the insert amplified in 30 cycles. The PCR reactions were digested with proteinase K in the presence of SDS and EDTA, and analysed by agarose gel electrophoresis. Details are described in the manual for the ESP™ LIC Cloning and Purification Kit (Stratagene).

2.2.2 Selected cloning strategies

2.2.2.1 Cloning of pKL2, pKL3 and pKL4

Many of the pET vectors for bacterial expression have a unique *NdeI* site at the ATG start codon. In order to be able to clone the *hobo* open reading frame into this *NdeI* site, the internal *hobo NdeI* site (at bp 1728) was removed. This was done on pHFL-1 (pKL1) with the Transformer™ Site-Directed Mutagenesis Kit (Clontech), using mutagenic primer KL6, which induces a single A to C switch within the *NdeI* site without changing the coded amino acid (isoleucin, ATA → ATC, also a frequently used codon).

The integrity of the full-length plasmid, pKL2, was tested by comparative restriction digest of pKL2 and pHFL-1 (not shown). It demonstrated that pKL2 has lost the *NdeI* site, but retained nine other restriction sites of pHFL-1. Sequencing across the open reading frame confirmed the single nucleotide change in pKL2.

During this mutagenesis it became apparent that the pHFL-1 preparations received from W. Gelbart were plasmid-dimers. Both pHFL-1 (pKL1) and pKL2 were therefore monomerised by *KpnI* digestion, precipitation and self-ligation. The integrity of the resulting plasmid monomers (named pKL3 and pKL4, respectively) was checked by analytical restriction digests. The sequence of the mutagenised region in pKL4 was verified by sequencing.

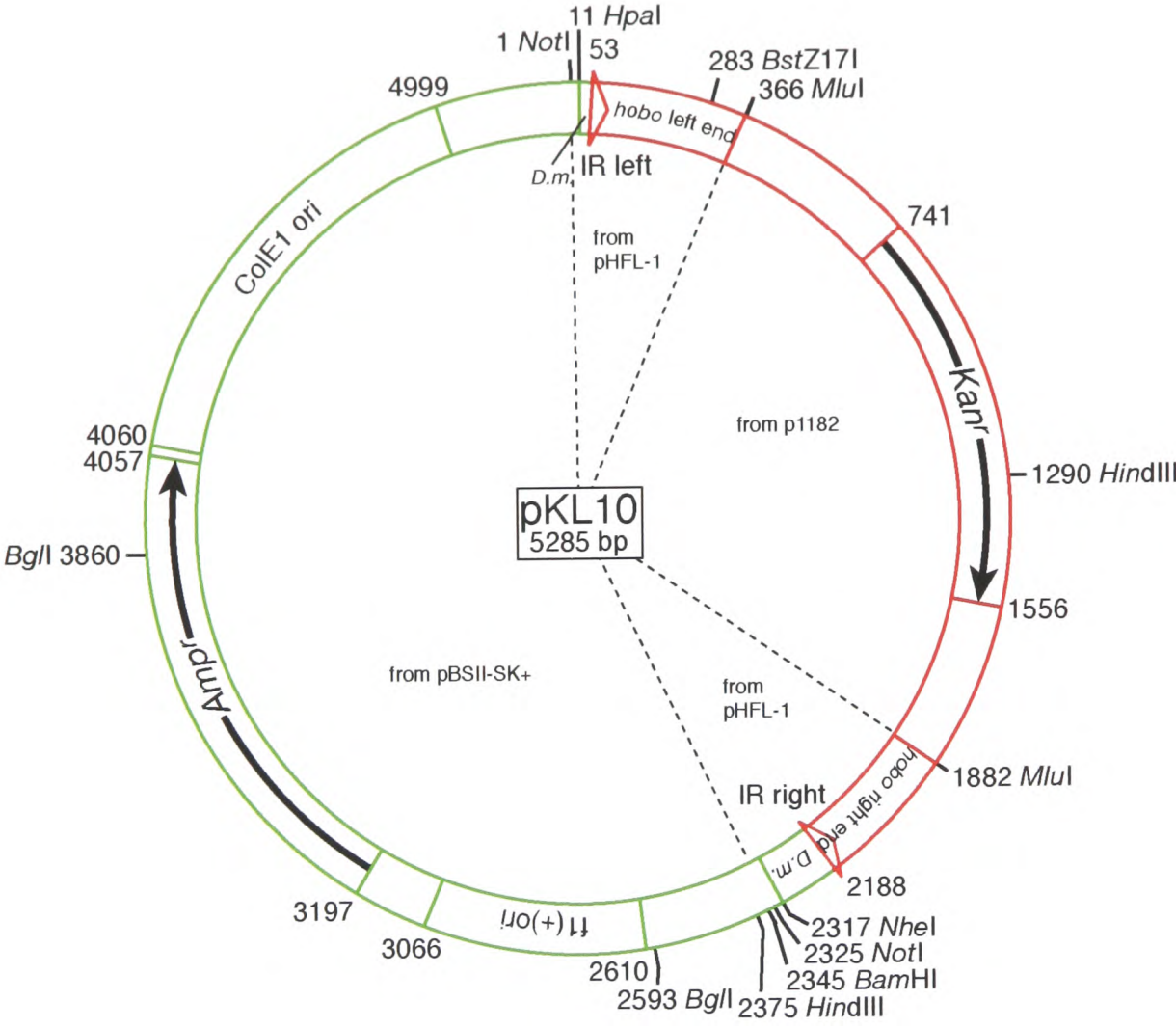
2.2.2.2 Cloning of pKL7, pKL8 and pKL9

The *hobo* transposase gene was amplified by PCR from pKL2 with primers KL4 and KL5, which add *NdeI* sites at the natural start codon and just 3' of the natural stop codon. The PCR product was cloned into *NdeI*-digested pET11a to give pKL7. The native transposase from pKL7 clones was cut out with *NdeI* and each ligated into His-tag vectors pET14b and pET16b, to produce pKL8 and pKL9, respectively (Table 2.7).

2.2.2.3 Cloning of pKL10 and pKL11

Mini-transposon plasmids pKL10 and pKL11 were created by PCR and four-piece-ligation (Table 2.7; Figure 2.2A). The left and right *hobo* transposon ends were amplified with primers KL19-KL22 to contain the terminal 300 bp of *hobo* (including the inverted repeats) and 50 or

A



B

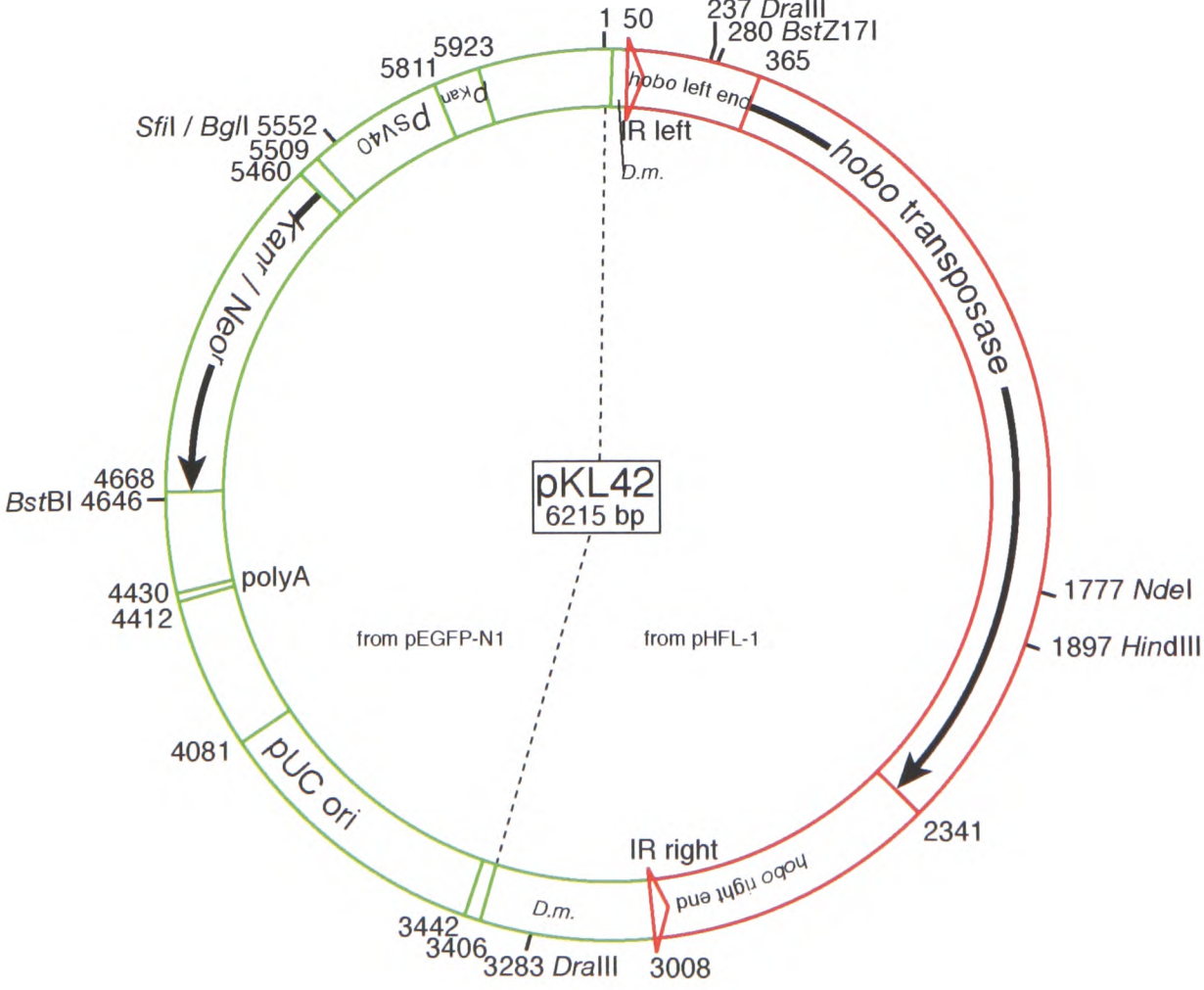


Fig. 2.2

Figure 2.2: Plasmid substrates pKL10 and pKL42

Maps of *hobo* reporter plasmids. Red, transposon; green, vector backbone and flanking DNA; *D.m.*, *Drosophila melanogaster* genomic DNA, which had flanked the original HFL-1 insertion; black arrows, orfs. Genes, promoters and origins of replication are in the same orientation as their labels. Numbers mark the basepair number of the first or last bp of the gene or marked sequence or the first bp of a restriction enzyme recognition sequence.

- (A) Map of pKL10, *hobo* reporter plasmid for mating out assay in *E. coli*, substrate for *in vitro* transposition, excision from stable integrations in mouse fibroblasts and development of the FDPCR assay.
- (B) Map of pKL42, *hobo* substrate plasmid for excision from stable integrations in mouse fibroblasts and development of the FDPCR assay.

140 bp of flanking genomic *D. melanogaster* DNA, respectively. *Mlu*I sites were added on the inside and *Not*I sites on the outside. In addition, an *Hpa*I site was added to the left flanking DNA and an *Nhe*I site to the right flanking DNA.

An excess of *Mlu*I/*Not*I digested, dephosphorylated PCR products was ligated to pBluescriptII-SK+, linearised with *Not*I, and a *Mlu*I fragment from p1182 containing the kanamycin resistance gene. Plasmid DNA recovered from transformants resistant to both ampicillin and kanamycin were test-digested with *Hind*III (which cuts in insert and backbone) for the presence and orientation of the *Kan^R* insert. Subsequent double-digests with *Alw*NI (which cuts in the backbone) and *Hpa*I or *Nhe*I established the identity of left or right ends. Mini-transposons with one left and one right end (pKL10) and with two left ends (pKL11) were obtained.

2.2.3 Radiolabelling of DNA

2.2.3.1 Labelling transposon end substrate

Radioactively labelled DNA was either derived from plasmids or from oligonucleotides. The desired restriction fragments were separated from the remaining plasmids either before or after labelling.

For rarely used plasmid derived substrate, 2.5 pmol (= usually 5 µg) plasmid was digested overnight in a 40 µl reaction and labelled in the restriction buffer with 50 µCi of α-³²P-dCTP and 2 U of Klenow polymerase (Roche). Unincorporated nucleotides were removed using Microspin G-50 columns (Amersham Pharmacia Biotech). The labelled DNA was loaded onto a 15 cm long 5% PA-gel (1x TAE) and run for 3 h at 250 V at 4°C. The gel was exposed for 10-30 min and this film was used to locate the restriction fragment. This was excised and the gel slice was incubated overnight in 400 µl STE (100 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA) at 37°C. The DNA was precipitated with ethanol in the presence of 10 µg glycogen, washed with 70% ethanol and dried. TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) was added to adjust the activity of the DNA to 1000 cps/µl, and the DNA was stored at 4°C.

For frequently used substrate (97 SX, 99 SX, 105 BX), 200 µg plasmid was digested overnight. The restriction fragments were separated by HPLC or FPLC. For HPLC, a Nucleogen DEAE 4000-7 column (Macherey-Nagel) was used. The DNA was heat-denatured to separate sticky ends, loaded in buffer A (6 M urea, 15 mM KH₂PO₄, 15 mM K₂HPO₄, pH 6.0 with H₃PO₄) and eluted in a 60 ml gradient to 100% buffer B (= buffer A with 1.2 M KCl). The peak fractions were dialysed against 3x 1 l STE and precipitated. FPLC was performed using a MonoQ anion exchange column (Pharmacia). The DNA was loaded in 600 to 650 mM NaCl in TE and eluted in a 20 ml gradient to 800 mM NaCl in TE. The peak fractions were precipitated and dissolved in TE, followed by photometric determination of concentration, and stored at -20°C. 2.5 pmol purified end fragment was labelled in DNA polymerase buffer with 50 µCi of α-³²P-dCTP and 2 U of

Klenow polymerase. It was repurified with a Microspin G-50 column, precipitated and adjusted with TE to 1000 cps/ μ l as above.

Synthetic oligonucleotides were annealed at a concentration of 100 nM each in 1x SSC (150 mM NaCl, 15 mM NaCitrate) in a thermocycler: 5 min each at 100°C, 95°C, 90°C, 85°C, 80°C, 75°C, 70°C, 4°C with a ramping speed of 0.05°C/second. The reactions were dialysed on Millipore 0.25 μ m nitrocellulose disks against 20 ml TE for 1-4 h, recovered and stored at 4°C. 2.5 pmol were labelled and purified as described above.

2.2.3.2 Internal labelling by random priming

For internal labelling of probes used in Southern blots, 60 ng of linear substrate DNA and 6 ng lambda *Bst*EII digest in a volume of 33.6 μ l were boiled for 5 min and placed on ice. Half of this denatured DNA was mixed with 1.2 μ l BSA (10 mg/ml, NEB), 6 μ l OLB (250 mM Tris-HCl, pH 8.0; 25 mM MgCl₂; 0.36% β -mercaptoethanol; 1 M HEPES, pH6.6; 27 OD units/ml random hexamers, GibcoBRL), 1 μ l cold nucleotide mix (1 mM each of dGTP, dTTP, and dATP or dCTP, Roche), 4 μ l (40 μ Ci) α -³²P-dCTP or α -³²P-dATP, 2 U Klenow fragment (labelling grade, Roche). The reaction was incubated for 1 h at 37°C. Unincorporated nucleotides were removed with a Nick Column (Pharmacia). Prior to use, the probe was denatured by adding 1/10 vol 2 M NaOH and incubating it for 10 min at RT.

2.3 Protein expression and purification

2.3.1 *hobo* transposase expression and purification in *E. coli*

2.3.1.1 Expression of *hobo* transposase in *E. coli*

Expression plasmids were transformed into CaCl₂ or RbCl-competent cells (modified after Sambrook *et al.*, 1989) of BL21(DE3)pLysS, unless stated otherwise, and plated on selective agar plates. Single colonies were picked and grown in 5 ml selective medium at 37°C to approximately OD₆₀₀ = 0.1. This preculture was transferred to an Erlenmeyer flask with 25 ml to 1 l prewarmed medium (1/10 of flask volume), grown with shaking at 37°C to OD₆₀₀ = 0.2-0.3 and transferred to a shaker at the induction temperature (37°C, 30°C, 25°C, RT or 16°C). At mid-to-late log phase (OD₆₀₀ = 0.3-0.7, usually 0.4), the cultures were induced with IPTG to a final concentration of 1 mM unless noted otherwise, and shaking was continued. At the time of induction and at different intervals afterwards, samples of the culture were taken for SDS-PAGE analysis, pelleted for 1 min at 14,000x g and resuspended in 1x reducing SDS-loading buffer (NEB).

To ensure that an equal amount of protein is loaded onto each lane, the cell pellets were resuspended in a volume of [(50 $\times$ OD₆₀₀) μ l 1x buffer / ml culture]. As 1 ml of *E. coli* culture at OD₆₀₀ = 1.0 contains approximately 75 μ g of total protein (Silhavy *et al.*, 1984), this formula gives a

final protein concentration of 1.5 µg/µl‡. 10 µl (15 µg) of this resuspension were loaded per lane onto 10-well (0.75 mM thick) mini-ProteanII gels; 6.7 µl (10 µg) were loaded onto 15-well gels.

$$\ddagger V_2 = c_1 \times V_1 / c_2 = [(75 \mu\text{g}/1000 \mu\text{l}) \times \text{OD}_{600} \times 1000 \mu\text{l}] / 1.5 \mu\text{g}/\mu\text{l} = 50 \mu\text{l} \times \text{OD}_{600}$$

2.3.1.2 Western Blotting for tagged *hobo* transposase

The blotting technique was optimised in order to be able to use Western blots in optimising expression and purification methods. First, the concentrations of SDS in the transfer buffer and of milk in the blocking solution were optimised to 0.037% and 7.5%, respectively. Blots of each half a gel loaded with expression cultures of transformants of pKL8 (induced 40 min and 3.5 h), pKL9 (induced 40 min), and 1/10 the volume of pMW324 (induced 40 min) were blocked, cut into strips and incubated in five antibody dilutions from 1:10,000 to 1:200,000, and in five concentrations of non-fat dried milk (Co-op brand, in PBST) from 5% to 20%. The 1:10,000 dilution and 7.5% dried milk had the highest signal to noise ratio; the control (no antibody, no milk) was blank.

The protein samples were routinely boiled for 5 min, centrifuged briefly and separated by SDS-PAGE (Laemmli, 1970). Separating gels were 8 or 10%, and gels were run at 100 to 200 V until the bromophenol blue dye had reached the bottom of the gel. The gels were equilibrated by shaking in 3 changes of 10 min in transfer buffer (25 mM Tris base, 192 mM glycine, 0.037% SDS). They were blotted onto PVDF membrane (Millipore) in a BioRad mini-TransBlot unit in an ice bucket for 3 h at 200 mA. Afterwards, the blots were blocked overnight at 4°C in 7.5% PBSTM (7.5% non-fat dried milk, 0.1% Tween in phosphate buffered saline).

The blots were washed briefly in PBST and incubated 1 h at RT in 20 ml 7.5% PBSTM with 1:10,000 HRP-coupled antibody [anti-GST peroxidase conjugate rabbit IgG (Sigma, Lot 77H4801) or monoclonal anti-polyhistidine clone HIS-1 peroxidase conjugate purified mouse immunoglobulin (Sigma, Lot 77H4880)]. The blots were then washed 6x for 15 min in PBST. They were developed for 1 min in a fresh 1:1 mixture of Solution 1 and 2 from the ECL-detection kit (Amersham). The blots were wrapped in cling film and exposed for initially 1 min in an autoradiography cassette on Hyperfilm ECL (Amersham).

2.3.1.3 Growth of cells for 6xHis-*hobo* transposase, preparation of lysate and storage

Cultures of pKL8-transformed *E. coli* BL21(DE3)pLysS were grown to $\text{OD}_{600} = 0.45$ and induced at room temperature for 30 minutes and at 37°C for 60 minutes. Cell lysate was prepared by disrupting the cells in three passages through a French Pressure Cell at 20,000 psi. The lysate was harvested by centrifugation and resuspended in 2 ml ice-cold Tris buffer (20 mM, pH 8.0) per gram of pellet. This lysate was quick frozen by dripping into liquid nitrogen and stored at -80°C. Aliquots were used for the subsequent experiments to optimise solubilisation and purification using TALON™ Metal Affinity Resin (Clontech).

2.3.1.4 Purification of 6xHis-*hobo* transposase

0.45-2 ml cell lysate of induced pKL8 transformant BL21(DE3)pLysS cells (Section 2.3.1.3) was thawed on ice, diluted with 9 vol cold Tris (50 mM, pH 8.0), and centrifuged for 2 h at 14,000x g at 4°C. The pellet was taken up in 10 lysate volumes of cold resuspension solution (50 mM Tris, 500 mM NaCl, 1 mM β -mercaptoethanol, 0.8%v/v Triton X-100, pH 8.0) and resuspended by pipetting with a cut 200 μ l pipette tip, and by swirling on a wheel for at least 12 h at 4°C. The suspension was centrifuged 2 h at 14,000x g at 4°C. The clarified sample (supernatant) was loaded onto 1.5 ml bed volumes TALON™ Metal Affinity Resin (Clontech), which had been equilibrated with 2 x 7.5 ml TTN-buffer (50 mM Tris, 0.8%v/v Triton X-100, 500 mM NaCl, pH 8.0). Following the manufacturer's protocol for batch / gravity flow purification, the protein was bound to the resin and washed in batch, then the resin was loaded onto a disposable gravity-flow column, washed with 15 mM imidazole and eluted with 100 mM imidazole. The entire purification procedure was performed at 4°C or on ice; all wash and elution solutions were based on TTN-buffer. Aliquots of the peak fractions were diluted with glycerol to a final concentration of 40% and stored at -80°C.

2.3.2 Expression of GST-*hobo* transposase in *S. pombe*

pKL6 transformant *S. pombe* SP-Q01 colonies were picked into precultures and grown overnight. These were diluted and grown to OD₆₀₀ = 0.7-1.0, washed, and resuspended in EMM (inducing conditions) or EMM + thiamine (repressing conditions). The cells were harvested after 20 h of growth and induction at 30°C. Cell pellets were ruptured by vortexing with acid-washed glass beads in the presence of protease inhibitors at 4°C. After brief centrifugation (5 min at 12,000x g), the clarified samples were analysed by SDS-PAGE. See the ESP™ LIC Cloning and Purification Kit manual (Stratagene) for details.

2.3.3 Coupled transcription and translation of *hobo* transposase

The Promega TNT Quick Coupled Transcription/Translation System (T7), based on rabbit reticulocyte extract, was used according to the manual and modified.

For optimising the system for *hobo* transposase expression, 10 μ l reticulocyte extract was mixed on ice with 5 μ Ci ³⁵S-methionine, pKL89 (or luciferase control plasmid), any additional components as specified, and nuclease free water to a final volume of 12.5 μ l and incubated at the specified temperature. For stopping the reaction, 1 μ l aliquots were taken into 49 μ l 1x SDS loading dye with DTT. After 2 min denaturation of the samples, 10 μ l were loaded onto 10% polyacrylamide mini-gels with SDS and electrophoresed for 55 min at 150 V, RT. The running front was removed, the gel fixed 30 min in 50% methanol, 10% acetic acid, equilibrated 30 min in 7% methanol, 7% acetic acid, 1% glycerol, dried for 1 h at 80°C onto Whatman 3MM paper and exposed overnight onto Kodak Biomax MR film.

For expressing *hobo* transposase for transposition reactions, 10 μ l reticulocyte extract, 1 μ l PCR enhancer from the TNT kit, 20 μ M unlabelled methionine, 125 ng supercoiled pKL89 and nuclease-free water were mixed for 12.5 μ l reactions and incubated for 4.5 h at RT.

2.3.4 MBP-*Himar1* transposase expression and purification

Initially, MBP-*Himar1* transposase was purified essentially as described in the New England BioLabs pMAL™ Protein Fusion and Purification System instruction manual and in Akerley *et al.*, 2002 (Sections 2.3.4.1 and 2.3.4.2). This purification was extended by an anion exchange column (Section 2.3.4.3).

2.3.4.1 Expression of MBP-*Himar1* transposase in *E. coli*

A preculture of the fusion plasmid (pMAL-C9 or pMAL-Kam) in *E. coli* DH5 α was diluted 1:100 into LB + Amp (100 μ g/ml) + glucose (0.2%) and grown with shaking at 37°C to OD₆₀₀ = 0.4-0.5. Protein expression was induced with 0.3 mM IPTG for 2 h and the cells were harvested by centrifugation at 4,000x g for 25 min. The pellets were resuspended in 1/50 culture volume ice cold TE and spun again at 4,000x g for 20 min in preweighed tubes. This pellet was resuspended step-wise in 9 ml cold lysis buffer (20 mM Tris pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) per gram of cells. Cells were lysed by one passage through a French pressure cell at 1,500 psi, frozen drop-wise in liquid nitrogen and stored at -80°C.

2.3.4.2 Amylose column purification of MBP-*Himar1* transposase

A miniature gravity flow column was packed with 1 ml of amylose resin and equilibrated with column buffer (10% glycerol, 25 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 2 mM DTT). The lysate (Section 2.3.4.1) was adjusted to the concentration of all column buffer components and clarified by ultracentrifugation at 83,000x g for 30 min at 4°C. 1 ml of the supernatant was diluted 1:5 with column buffer, loaded onto the amylose column and washed with 10 column volumes of column buffer. Transposase was eluted with elution buffer (column buffer + 10 mM maltose) and 0.3 ml fractions were collected. The protein concentration was determined by A₂₈₀ measurement and/or with a Bradford assay.

2.3.4.3 Anion exchange column purification of MBP-*Himar1* transposase

The pool of peak fractions from the amylose column (Section 2.3.4.2) was diluted with 3 vol column buffer lacking NaCl and adjusted to 10 mM n-decyl- β -D-maltopyranoside (nDMP). This was loaded onto a 1 ml MonoQ FPLC column (Pharmacia), equilibrated in buffer A (10% glycerol, 25 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA, 2mM DTT, 10 mM n-decyl- β -D-maltopyranoside) and eluted with a 30 ml gradient of 0.05-1 M NaCl. The peak fractions were adjusted with glycerol to 50% and stored at -20°C.

2.4 Specialised PCR procedures

Two customised PCR protocols were developed for detecting hairpin structures, single-primer polymerase chain reaction (SPPCR) and frayed-duplex polymerase chain reaction (FDPCR). See Chapter 4 for an explanation of the techniques.

2.4.1 Generation of substrates by PCR

Hairpinned substrates were generated as described in Figure 4.2. For the creation of non-hairpinned control substrates, 2.5 μ g pKL42 was digested with *Bst*Z17I and *Bgl*II. The 942 bp fragment was gel-purified. 500 ng of the restriction fragment was ligated to a 10-fold molar excess of annealed adapter KL99/KL95 with 200 U T4 DNA ligase. The 1013 bp ligation product was excised from an agarose gel and purified.

2.4.2 Generation of substrates from custom-built plasmids

For the generation of plasmids pKL96-pKL103, the required flanking DNA and/or transposon end was amplified by 25 PCR cycles with *Pfu* polymerase and primers as specified in Table 2.7. The product from 3 PCR reactions (50 μ l each) was purified with the QIAquick PCR purification kit. For hairpinned template (for pKL96, pKL101), the eluted DNA was digested with *Xho*I and *Bsa*I, for non-hairpinned template with *Xho*I only. The gel-purified digests were ligated to *Xho*I-digested, dephosphorylated, gel-purified pBluescriptII-SK+. Ligations were transformed into *E. coli* DH5 α . Plasmids were analysed by restriction digestion and native PAGE.

For purification of hairpinned DNA templates, 250 μ g of each plasmid pKL96 and pKL101 was digested with 600 U *Sfi*I. For non-hairpinned substrate, 500 μ g of each pKL99 and pKL102 was digested with 1000 U *Sfi*I and 750 U *Bsa*I. The small substrate fragments were separated from the large plasmid backbone by FPLC on a MonoQ column using a 20 ml gradient from 600-750 mM NaCl in TE. Fractions were analysed on 2%, TBE agarose gels. Fractions with the palindromic 2x fragment were denatured by boiling for 5 min and cooled down slowly to generate the hairpin conformation. All substrate-containing fractions were precipitated. Pooled pellets were resuspended in TE.

Purified hairpinned and non-hairpinned substrates were ligated to a 10-fold excess of annealed adapter KL135/KL136 and PAGE-purified.

2.4.3 Frayed-duplex PCR of *hobo* excision reactions

After completion of the *hobo in vitro* reactions (Sections 2.5.1.1, 5.2.1), aliquots containing 667 ng (0.2 pmol) of the substrate pKL10 were digested with 2 U *Bgl*II for 2 h at 37°C, followed by 20 min heat-inactivation of the restriction enzyme at 65°C. The digested DNA was then ligated to 2 pmol of annealed adapter KL85/KL86 with 200 U T4 DNA ligase for 12 h at 16°C. The ligations were ethanol precipitated and resuspended in 15 μ l TE. All 15 μ l were included in 50 μ l PCR

reactions (as detailed in Table 4.3 for template 10B-hf) with 1.25 mM of each primer KL87 and KL88. 20 μ l of the reactions were analysed by electrophoresis on a 1.5% agarose, 1x TBE gel.

The remaining 30 μ l of selected reactions were loaded into every third well of a 1.5% SeaKem gold agarose TBE gel and electrophoresed for 2 h 40 min at 100 V. The 0.5 kb band was excised and purified with the QIAquick gel extraction kit. The DNA was subjected to a second PCR amplification in 100 μ l reactions with parameters as above, except for an increased primer concentration of 5 μ M KL87 and KL88 and 65 cycles with annealing temperatures ramped down from 60 to 48°C and 3 min elongations. The reactions were analysed by electrophoresis on 1.5% SeaKem gold agarose gels and with SYBR Gold staining.

2.5 Transposition *in vitro*

2.5.1 *hobo* transposition *in vitro*

2.5.1.1 *hobo in vitro* excision reactions

Transposition reactions were performed in a volume of 25 μ l, containing 1 μ g pKL10 substrate, 100 mM Tris, pH 7.5, 5 mM MgCl₂, 50 μ g/ml BSA, 0.8 mM DTT, 1 mM GTP, either 35 mM KCl or 50 mM NaCl, and different combinations of 10 mM spermidine, 0.4 mg/ml tRNA and 4 mM ATP. 1-3 μ l of 6xHis-*hobo* transposase peak fraction (Section 2.3.1.4), neat or diluted 10-fold in reaction buffer plus 25% glycerol, were added to the reaction. The transposition reactions were incubated at RT for 12-16 h, the DNA precipitated with ethanol and resuspended in 12 μ l TE. The reaction products were analysed via agarose gel-electrophoresis and FDPCR.

2.5.1.2 *In vitro* plasmid-to-plasmid transposition with reticulocyte expressed *hobo* transposase

Completed transcription/translation reactions (Section 2.3.2) were diluted 4- to 80-fold into 20 μ l reactions containing 10%v/v glycerol, 25 mM HEPES, pH7.9, 100 mM NaCl, 10 mM MgCl₂, 250 μ g/ml BSA, 2 mM DTT, 250 ng pBluescriptII-SK+ tetramer, 70 ng pKL91, and 0, 1 or 10 μ l of *Drosophila melanogaster* nuclear extract or chromatin assembly extract (gift of P. Becker, Munich). The reactions were incubated at RT. 10 μ l of the reactions was stopped after 6 h and 20 h by dilution into 100 μ l 10 mM EDTA, 1x TE and heating for 10 min at 75°C. The reactions were phenol/chloroform extracted, precipitated, and the dried pellets resuspended in 4 μ l TE. This DNA was electroporated into competent *E. coli* DH5 α . 1/1000 of the transformation was plated onto LB-Amp, the rest onto LB-Amp-Kan agar plates and incubated overnight at 42°C.

2.5.2 *Himar1* transposition *in vitro*

Unless stated otherwise, *Himar1* transposition reactions were performed in 25 mM HEPES pH 7.9, 10% glycerol, 100 mM NaCl, 20 mM NH₄Cl, 250 μ g/ml BSA, 2 mM DTT, and 10 mM MgCl₂ or 2 mM CaCl₂ in a reaction volume of 15 μ l and incubated at 30°C for 1.5 h.

2.5.2.1 *Himar1* plasmid substrate reactions

In plasmid substrate reactions, 10 mM MgCl₂, 1 µg plasmid substrate (pMarKan) and up to 1.5 pmol (1 µM) purified transposase were used. The reactions were stopped by adding stop solution to a final concentration of 0.1% SDS, 10 mM EDTA and incubating 10 min at 75°C. Together with 8x loading dye, the reactions were loaded on a 1.1%, 250 ml, 25 cm long TBE agarose gel and electrophoresed for 12-16 h at 60 V, RT. The gel was stained in Ethidium bromide and analysed by Polaroid photography or on a fluorimager at 532 nm wavelength (filter: 580 nm).

2.5.2.2 *Himar1* insertion reactions

Himar1 insertion reactions were generally carried out with 64 fmol supercoiled target plasmid (125 ng pBluescriptII-SK+, pcDNA3 or 105 ng pKL104), 500 cps labelled transposon ends, and 1-10 nM purified transposase (absolute amounts given in figure legends). After incubation at 30°C, the reactions were stopped by adding 1/10 vol stop solution (1.5% SDS, 15 mM EDTA) and incubating for 10-15 min at 75°C. After addition of 8x gel-loading dye, the reactions were loaded on a 0.7%, 250 ml, 25 cm long 1x TAE or TBE agarose gel and electrophoresed for 14 h at 65 V, 4°C or for 3 h at 120 V, RT. The gel was placed on top of two layers of Whatman 3MM paper and one layer of Hybond-N+ nylon membrane and dried for 2 h at 70°C on a BioRad model 583 gel dryer with vacuum pump. It was then exposed overnight on Kodak Biomax MR film or a Fuji phosphorimager screen.

2.5.2.3 *Himar1* gel retardation

For gel shifts, 500 cps labelled transposon ends and the stated concentration of transposase were incubated in buffer containing 2 mM CaCl₂ for 1.5 h at 30°C. The reactions were loaded directly onto native 5% 29:1 acrylamide:bisacrylamide gels, poured and run in 1x TAE, and electrophoresed at 300V, 4°C for 2-16 h. Gels were 0.75 mm thick and 25 cm long. The gels were fixed for 10 min in 30% methanol, 10% acetic acid, dried onto Whatman 3MM paper at 80°C for 1 h and analysed by exposure to a phosphorimager screen or a Kodak Biomax MR X-ray film for 4 h or more as required.

2.6 Transposition *in vivo*

2.6.1 Addition of a *Cam*^R marker to pOX38 for use in *in vivo* transposition

LC836, an *E. coli* strain that contains the F-plasmid derivative pOX38, was transformed with pNK2884, a plasmid with both the altered-target-specificity (ATS) IS10 transposase (Bender and Kleckner, 1992) and a chloramphenicol-resistance-marked Tn10 mini-transposon. A culture of a transformant colony was grown to OD₆₀₀ = 0.3 at 37°C. One half of the culture was made 1 mM in IPTG to induce transposase expression. After 2 h continued growth to allow transposition, 10⁸ and 10⁹ cells of each donor culture were mated to 10⁹ AB1157(Str^R) recipient cells on a slow roller for

1 h. The mated cells were plated onto LB-Str-Cam agar plates for selection of recipients of a *Cam^R*-marked pOX38. The frequency of transposition onto pOX38, i.e. the ratio of *Str^R*, *Cam^R* colonies to the total number of mated recipients, was 0.1% (10^{-3}), irrespective of the number of donor cells and the presence of IPTG.

Two colonies from each of four matings were restreaked to obtain unique clones. For selecting a plasmid with a high mating efficiency, these eight AB1157-pOX38::mini-Tn10(*Cam^R*) strains were mated in liquid culture to HMS174(DE3). #2, 4, 6 and 8 were in addition mated to MM294(DE3), and #2 and 4 also to BL21(DE3) and ER2566. The mating efficiency varied between 10^{-6} and 2.5×10^{-1} of mated recipients. MM294(DE3) was the best recipient for all plasmids, HMS174(DE3) the second best, reaching 10% efficiency. pOX38-*Cam^R* #4, which came from an induced culture and a low donor : recipient mating, mated efficiently into all four strains. It was named pKL16 and used for all subsequent transposition and mating experiments. All insertions were saved as frozen stocks.

Plasmids R388 and pKL16 were mated from the donor, AB1157, into the host strains for *hobo in vivo* transposition, selecting on minimal medium, which does not allow growth of the donor.

2.6.2 Cell culture

2.6.2.1 Stable transfection of SCGR1 and NIH 3T3 fibroblasts

A calcium phosphate precipitate was formed with 5 μ g of transposon-containing plasmid and, if appropriate, a smaller amount of selection plasmid pRC98, which contains a *Neo^R* marker (Ausubel *et al.*, 1994). 1/8 of this precipitate was transfected into target cells (SCGR1 or NIH 3T3) growing at 50-70% confluency in 6- or 24-well plates. After 48 h, the cells were transferred to T25 flasks and selection was started by adding an optimised concentration of G418 (Table 2.4) and 0.1 M HEPES to the growth medium. After 2 weeks of selection (and 5 days after all non-DNA control cultures had died off), a 1:3,000 dilution of the cultures was seeded into 6-well plates. 8 days later, single colonies were transferred to new wells. To ensure single-cell-lineage, single colonies were again picked from these and amplified.

To test the newly established cell lines for the presence and abundance of inserted plasmid, the genomic DNA of confluent cultures in T25 flasks was isolated with DNeasy spin-columns (QIAGEN). The concentration of linearised genomic DNA was determined by measuring the A_{532} of diluted 20 μ l drops, mixed 1:10,000 with SYBR Green, with a fluorimeter and NIH Image software, and comparing the measurements with the linear range of equally prepared dilution series of calf thymus and *E. coli* genomic DNA. For the Southern blots, 10 μ g of cell DNA was digested with appropriate restriction enzymes and electrophoresed on a 25 cm, 1x TBE, 0.8% agarose gel for 14 h at 80 V, 4°C, alongside 100 pg of likewise digested transposon and selection plasmid and

10 ng lambda digest. Southern blots of these gels were hybridised with random-primed radioactive linear plasmid and lambda DNA.

For analysis of larger amounts, the genomic DNA was extracted after proteinase K digestion in SDS (Ausubel *et al.*, 1994) or with QIAGEN maxiprep columns for genomic DNA.

2.6.2.2 Determination of copy number of integrated plasmids

2.6.2.2.1 Comparison of integrations with a single-copy mouse gene

To determine the number of plasmids integrated per cell in stably transfected SCGR11 cell lines, their genomic DNA was isolated and c. 15 µg each digested with *Hind*III, *Bam*HI or *Ppu*MI. Together with 1 ng each of *Hind*III and *Bst*EII digested lambda DNA, these digests were electrophoresed for 13 h at 65 V, 4°C on a 0.8% agarose gel in TBE. The gel was Southern blotted. The blot was hybridised sequentially with two radioactive probes, made fresh on the day of hybridisation. Both probes were labelled by random priming (see Section 2.2.3.2) using α -³²P-dCTP. In both, a total of 33 ng of linear DNA substrate was labelled, 3 ng of which was *Bst*EII digested lambda DNA. The substrate for the first probe contained 12 ng of the 1132 bp *Pst*I fragment with murine erythropoietin (EPO) cDNA from pEPO (Table 2.8). The substrate for the second probe contained 30 ng of linear pKL10, which is also equal to 14.8 ng of DNA that is homologous to the backbone of pRC98. Both hybridisations were performed overnight, and both times the blot was exposed to Kodak Xomat-AR film for 25 h with an intensifying screen.

Both films were scanned and displayed with Adobe Photoshop software. The intensity of bands was quantified with NIH Image 1.61. After subtraction of the background, the quantification results of the second film were divided by the results of the corresponding lanes or fragments of the first film. For the eleven lambda-fragments measured as a calibration, the mean of this quotient was 1.33. To normalise the differences in labelling- and hybridisation efficiency of the two hybridisations, the quotients of the genomic digest lanes was divided by this number. Furthermore, they were multiplied by 0.81 = 12 ng (pEPO fragment) / 14.8 ng (pRC98 fragment). As both plasmids have a GC content of c. 50%, this number is equal to the relative amount of α -³²P-dCMPs in the hybridising part of each probe. To obtain the copy number of pRC98 per cell, the quotients were finally multiplied by 2, which is the copy number of EPO per diploid genome. In summary: $2 \times 0.81 / 1.33 = 1.22 \rightarrow \text{Copy number of pRC98} = 1.22 \times (\text{signal strength after pKL10 hybridisation}) / (\text{signal strength after pEPO hybridisation})$.

With 8 days exposure, both pKL10 and pRC98 signals were detectable on film exposed to the pKL10 hybridised blot. Taking into account that pKL10 hybridises to 5285 bp of pKL10, but only to 2608 bp of pRC98, the copy number of pKL10 was calculated from the signals of the two bands in each lane.

2.6.2.2.2 Comparison of integrations with pure plasmid

To estimate the copy number of pKL42 in the S10.1 genome, 1 μ g plasmid pKL42 and 20 μ g aliquots of each purified genomic DNA from cell lines SCGR11 and S10.1 were digested with *Bst*BI, precipitated and resuspended in TE.

All 20 μ g of the digested genomic DNA, dilutions of digested pKL42 and 10 ng lambda *Bst*EII digest were loaded onto a 0.8% agarose, 1x TBE gel and electrophoresed for 16.6 h at 60 V, 4°C. A Southern blot of the gel was hybridised with probe made from 3 ng lambda *Bst*EII digest and 30 ng of the 3257 bp *Apa*I-*Eco*RV of pKL42, which includes the backbone and flanking DNA. The blot was analysed by autoradiography with exposures of 5 days (not shown) and 6 weeks (Figure 5.7B). The strength of signals in the genomic and plasmid lanes was compared and copy numbers calculated.

2.6.2.3 Retroviral transduction

Retroviral vector stocks were produced according to the three-plasmid transient-transfection protocol described by Soneoka *et al.* (1995). In short, 10 μ g each of pHIT123 (mouse-cell specific envelope), pHIT60 (*gag-pol*), and the appropriate genomic plasmid (pKL64-pKL74 or pGFP-i-lacZ) were transfected by overnight calcium phosphate treatment (Promega ProFection mammalian transfection system) into 293T cells at 70-100% confluence in 10-cm-diameter dishes. The following morning, the cell medium was replaced with medium containing sodium butyrate (10 mM final concentration) for 12 h. Subsequently, this medium was replaced with 5 ml of fresh medium, and the viral supernatants were harvested 12 h later (48 h post-transfection). The cells were photographed.

Target cells (S10.1) were subcultured into 10-cm-diameter dishes 24 h prior to virus harvest so that they were 30 to 40% confluent on the day of transduction. For transduction, target cell medium was replaced with the viral supernatants. After 2 h, an additional 5 ml of fresh medium was added to each plate. After 24 h, the transduced cells were photographed and harvested.

CHAPTER 3 - EXPRESSION AND PURIFICATION OF *HOB*O TRANSPOSASE

3.1 Introduction

The main component required for an *in vitro* transposition system is an active preparation of purified transposase. Towards this end, *hobo* transposase was overexpressed in several systems to screen for differences in the level of expression, solubility and activity.

Expression in *Escherichia coli* is relatively simple and often generates high yields of fully active protein, even for eukaryotic proteins. On the other hand, prokaryotes do not perform eukaryotic-specific splicing and post-translational modification, such as phosphorylation and glycosylation, which may be necessary for the activity of the protein. However, nuclear proteins such as transposases are not glycosylated, and many eukaryotic proteins are active after prokaryotic expression. Furthermore, even if the protein is inactive, prokaryotically expressed protein can still be used for the production of antibodies.

The fission yeast *Schizosaccharomyces pombe* has also been commonly used for protein expression. Although it produces lower yields and grows more slowly than *E. coli*, post-translational modification is similar to that in higher eukaryotes, and the expressed proteins are often more active.

hobo transposase was expressed in these systems as various fusion products. Some of the tags used increase solubility, and all of them allow detection and affinity purification. All of the expression vectors used have inducible promoters: potentially toxic recombinant protein might otherwise kill the cells prematurely.

As a higher eukaryotic expression system, rabbit reticulocyte lysates were used. These cell extracts do not produce sufficient protein levels for subsequent purification, but provide a very natural environment, which can be tightly controlled. Thus produced transposase can be used directly in transposition assays.

Our starting point for expression was the HFL-1 clone of *hobo*, which is highly active *in vivo* (Blackman *et al.*, 1989). HFL-1 had been cloned along with some natural flanking sequence into pBluescript to give pHFL-1. The 2959 bp transposon contains a 1976 bp open reading frame (ORF1), which encodes a protein with a calculated molecular weight of

75 kD. This sequence was used to construct all of the expression vectors described in the following sections.

3.2 Initial trials for transposase expression in *E. coli*

3.2.1 CBP-fusion in *E. coli*

hobo transposase was expressed in *E. coli* as a fusion protein to the calmodulin-binding peptide (CBP). This small (4 kD) N-terminal tag increases the solubility of most proteins, allows easy purification with calmodulin affinity resin, and can be cleaved off with enterokinase, leaving no extra amino acids behind.

3.2.1.1 Construction of the CBP-*hobo* expression vector

The *hobo* transposase open reading frame from pHFL-1 was cloned by ligation independent cloning (Affinity™ LIC Cloning and Protein Purification Kit, Stratagene) into pCAL-n-EK behind its T7 promoter (see Table 2.7 and Table 2.8). The integrity of the resulting pCAL-*hobo* plasmid (pKL5) was confirmed by DNA sequencing.

3.2.1.2 Expression of the CBP-*hobo* fusion protein

pKL5 was transformed into BL21(DE3)pLysS cells, resulting in transformant colonies of very variable size. Protein expression was induced by addition of IPTG to a growing culture, and the results evaluated by SDS-PAGE and Coomassie staining. Despite the high activity of the T7 expression system, there was no clear *hobo* transposase band visible in the crude cell extract. (Figure 3.1A,C). The parallel expression of IS10 transposase from pRC60 as a control resulted in high transposase levels of c. 20-30% of the total cell protein (Figure 3.1B). This shows that the expression system was working correctly. Apart from the 45 kD IS10 transposase, all protein bands coincided in their position and temporal pattern with the *hobo* transposase expression culture. They are thus all formed by endogenous *E. coli* proteins. Those bands that increased in strength after induction are presumably formed by proteins that are normally expressed at the transition to stationary phase in *E. coli*, and not truncations of any of the transposase-fusions.

To increase levels of CBP-*hobo* transposase expression, a series of different conditions were tested: Four different induction temperatures from room temperature to 37°C; inducer concentrations of 200 µM and 1 mM IPTG; and induction times from 30 min to 44 h (and ten different intervals in between). Furthermore, small, medium and large

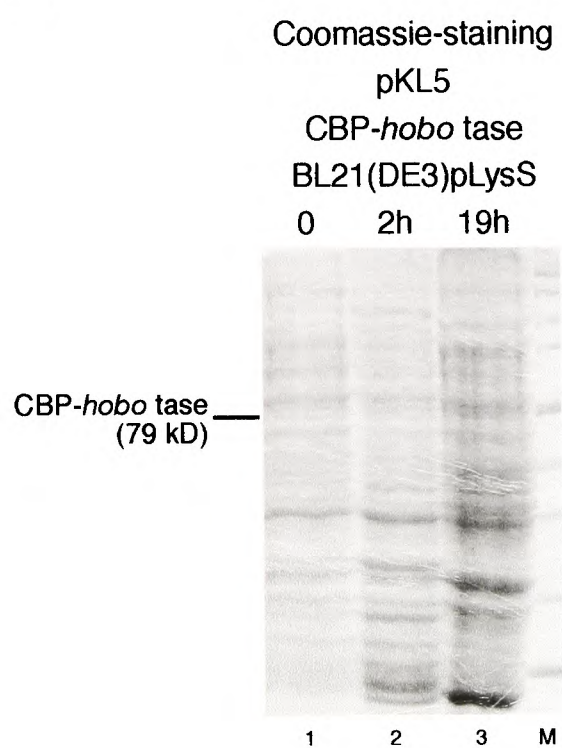
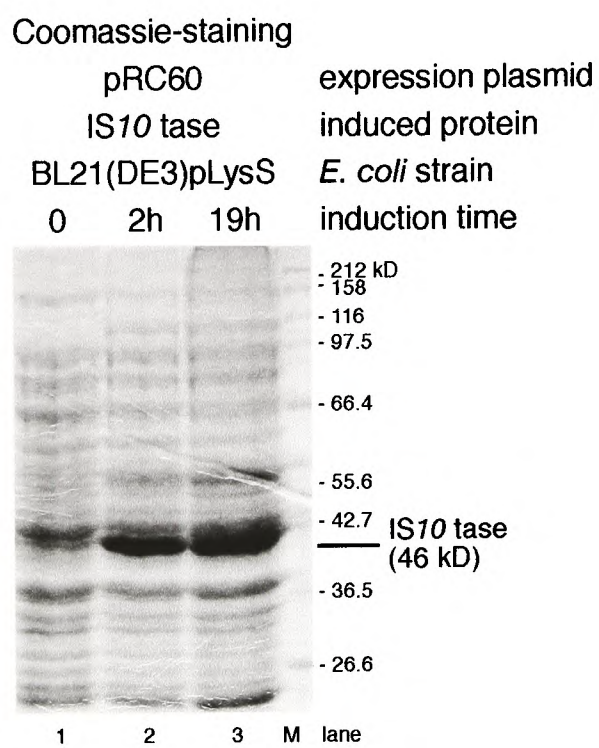
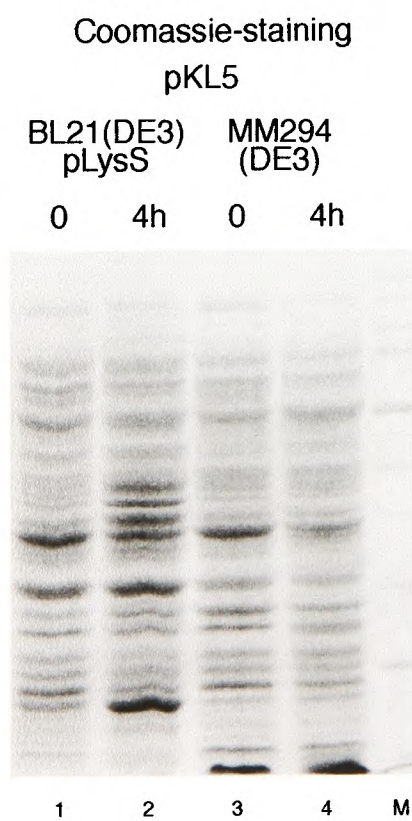
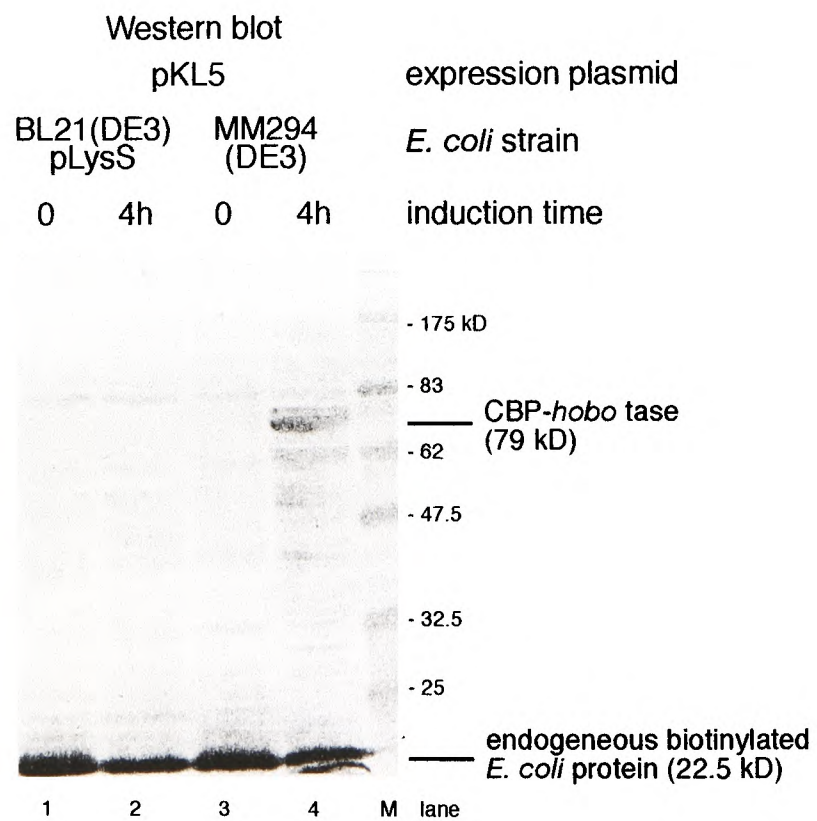
A**B****C****D**

Fig. 3.1

Figure 3.1: Expression of CBP-*hobo* transposase in *E. coli*

SDS-PAGE of whole cell extracts from *E. coli* expression cultures.

- (A)** CBP-*hobo* transposase expression from pKL5 in *E. coli* BL21(DE3)pLysS. The calculated molecular weight of the putative fusion protein is 79 kD. Coomassie-stained 10% PA gel, run 45 min at 200 V.
- (B)** IS10 transposase expression from pRC60 in *E. coli* BL21(DE3)pLysS. The protein has a molecular weight of 46 kD. Coomassie-stained 10% PA gel, run 45 min at 200 V.
- (C)** CBP-*hobo* transposase expression in different *E. coli* strains. Coomassie stained 10% PA gel, run 65 min at 200 V.
- (D)** Western blot of a gel identical to (C), probed with biotinylated calmodulin and streptavidin-coupled alkaline phosphatase.

transformant colonies; five different plasmid clones (before the sequencing was completed); different culture volumes, vessels and shaker speeds; and two bacterial expression strains: BL21(DE3)pLysS and MM294(DE3) were tested. None of these alterations brought about an increase in the expression of the CBP-transposase to a level detectable by Coomassie staining (not shown).

3.2.1.3 Western blot detection

As CBP-*hobo* transposase expression was not detected on Coomassie stained gels, Western blots were performed in order to increase the sensitivity and specificity of detection. Unfortunately, there is no antibody against the calmodulin binding peptide commercially or otherwise available locally. The only option was the Affinity™ CBP Fusion Protein Detection Kit by Stratagene, in which the main components are biotinylated Calmodulin (which binds to CBP) and Streptavidin alkaline phosphatase. Protein expression was performed as usual, with different plasmid clones and in two different *E. coli* strains (Figure 3.1C). After blotting and detection, a very strong band was visible at 22.5 kD, which represents a naturally biotinylated endogenous *E. coli* protein (Figure 3.1D). Strikingly, the second strongest band appeared at c. 79 kD in the induced sample of MM294(DE3). Full-length-fusion protein therefore appears to be produced. However, the signal-to-background ratio was low, and the signal did not reproduce in a second expression-blotting experiment. Due to the high price (£ 90 for a 6 × 8 cm² blot) and the extremely low sensitivity and reproducibility of this kit, expression of the CBP-fusion protein was abandoned at this stage.

3.2.2 Native and His-tagged transposase in *E. coli*

hobo transposase was cloned into a selection of pET vectors under the control of the strong bacteriophage T7 promoter (Sections 2.2.2.1, 2.2.2.2). Transposase was produced as a fusion to cleavable poly-histidine tags, which allow immunochemical detection and purification via metal-affinity chromatography. In pKL8 (pET14b-*hobo*), the N-terminal tag consists of six histidines and ten additional amino acids, which contain a thrombin cleavage site, enabling subsequent removal of all but the last three amino acids. pKL9 (pET16b-*hobo*) codes for ten N-terminal histidines and nine additional amino acids, all of which but one can be removed by Factor Xa cleavage. In addition, *hobo* transposase was expressed in its native form from pKL7 (pET11-*hobo*), as the tag might impede activity (Table 2.7).

3.2.2.1 Expression of native and His-tagged *hobo* transposase

In order to test the levels of transposase expression, pKL7, pKL8 and pKL9 were transformed into BL21(DE3)pLysS. LB cultures were induced at 37°C. In samples taken 40 min to 44 h after induction, there was an increase of full-length recombinant protein expression just detectable on Coomassie-stained gels (Figure 3.2A, lanes 5-14 and not shown). As a control, another eukaryotic protein cloned in pET-14b, 6xHis-Rec12, was expressed from pMW324. The 40 kD 6xHis-Rec12 fusion protein amounted to c. 60% of cellular protein after overnight induction (Figure 3.2A, lanes 1-4).

Samples from induced cultures of pMW324, pKL8 and pKL9 transformants (which express His-tagged proteins) were Western blotted and probed with anti-polyhistidine HRP-conjugate (Sigma) (Figure 3.2B). In pMW324 transformants, the major band is full-length recombinant protein, which is present even in the uninduced sample, and increases in strength after induction. In pKL8 and pKL9 transformants, no signal is detected before induction. This indicates that the detection is very specific to induced His-tagged proteins. After induction, the strongest signal is of the correct size to be the full-length *hobo* transposase fusion. It reaches its maximum strength after 40 minutes of induction. By comparison of Western blots and stained gels it was possible to identify the Coomassie stained band corresponding to the full-length transposase (c. 77 kD) on previous gels.

In the Western blot, there are a few other significant bands, which are all shorter and thus arise probably either from proteolysis at specific sites, or from prematurely terminated translation. Longer induction times produced a larger number of truncated bands, a weaker signal for the full-length protein, and a stronger signal at the running front and at the top of the gel (just below the border of stacking and separating gel). The pattern is identical for 6xHis- and 10xHis-transposase, but the overall strength of signal is much weaker with the 10xHis-tagged protein.

3.3 Expression and purification of 6xHis-transposase

3.3.1 Kinetics of 6xHis-transposase expression

Western blots (see Section 2.3.1.2) were used to optimise expression and purification of 6xHis-*hobo* transposase. A time course shows that expression of transposase from pKL8 begins immediately after induction (Figure 3.2C). The full-length signal appears at once (within the first five minutes), and the shorter truncated products appear shortly after. After 30 minutes, the full-length signal has reached its maximum level, which it maintains for

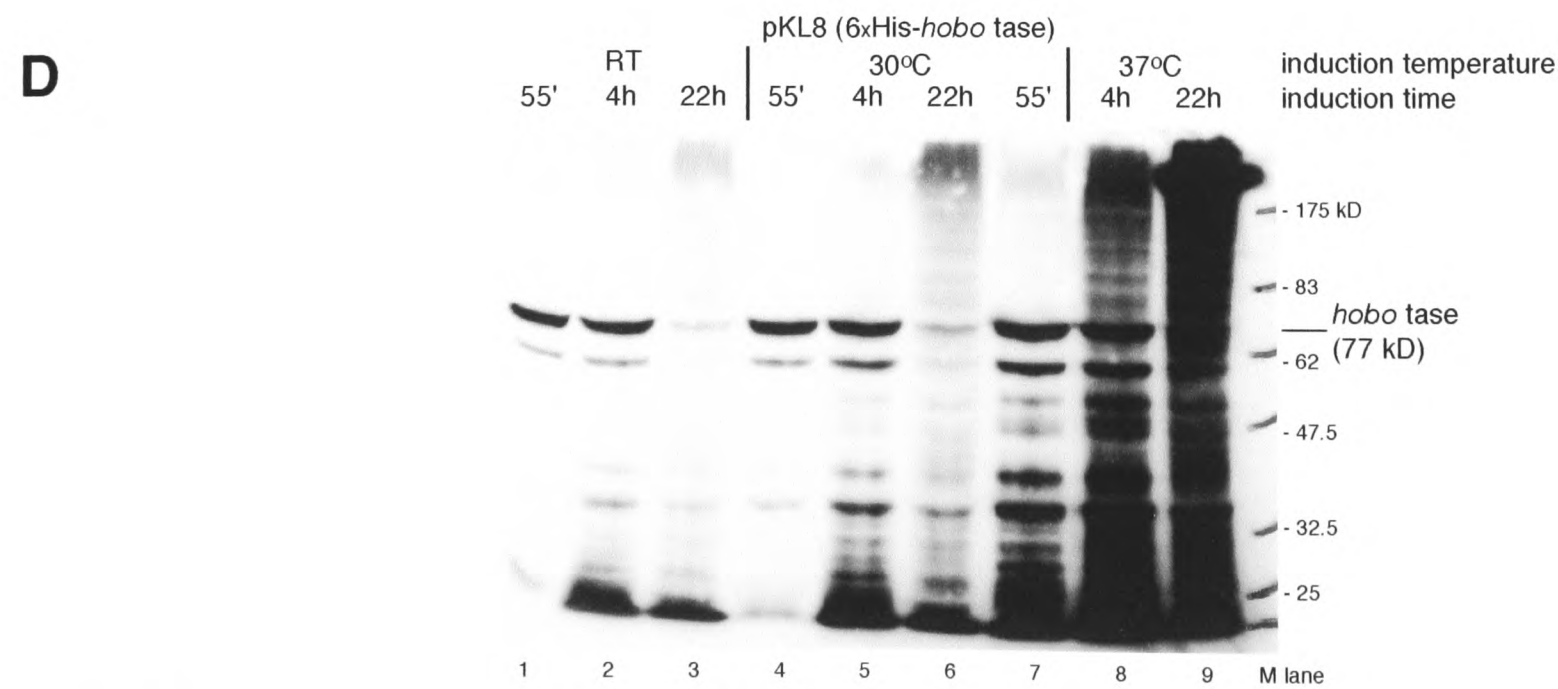
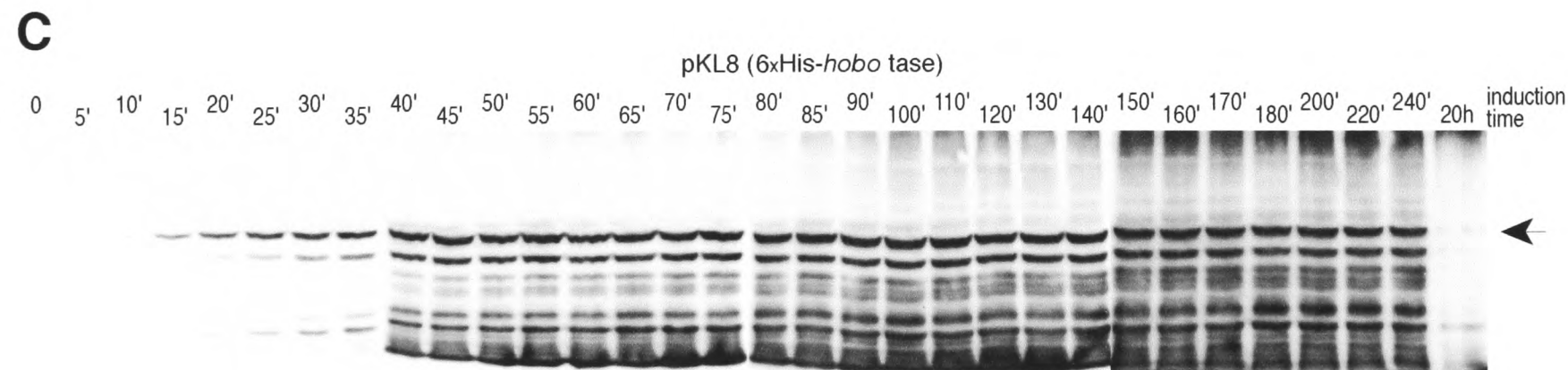
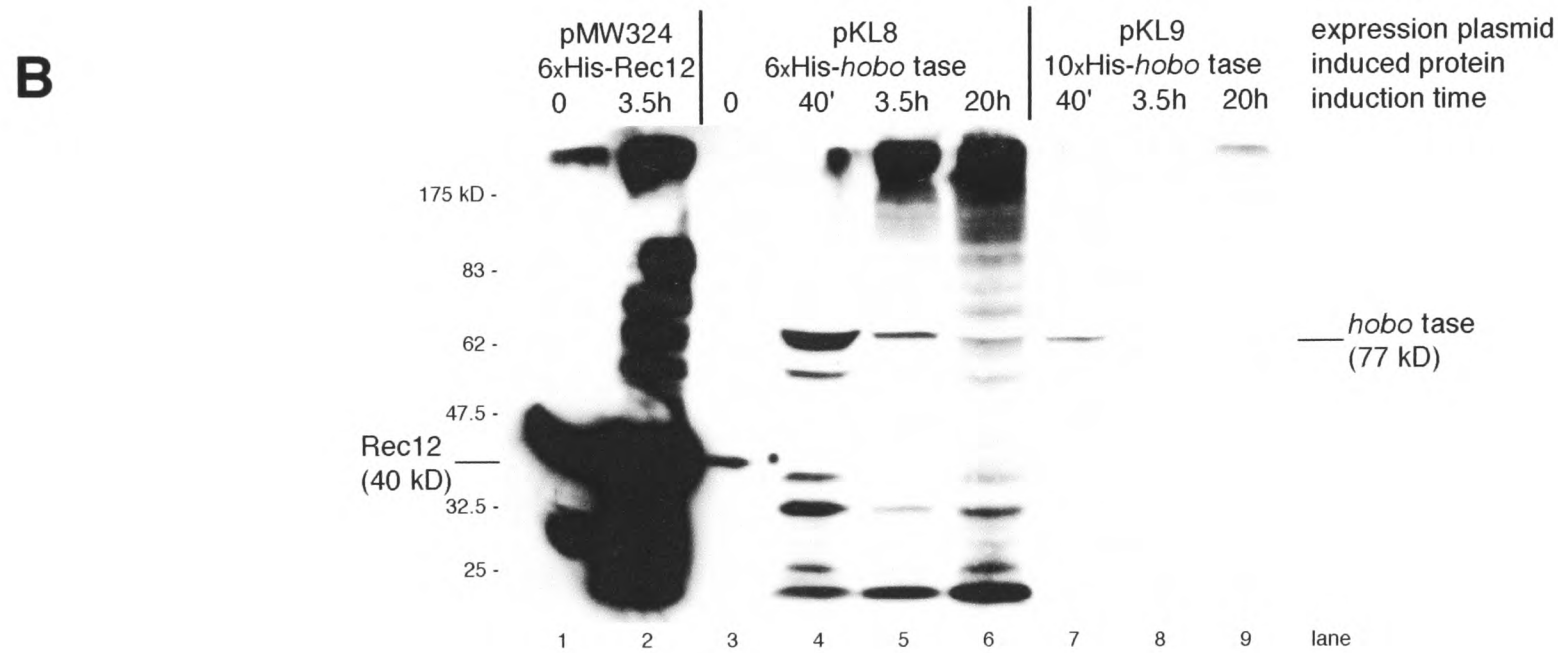
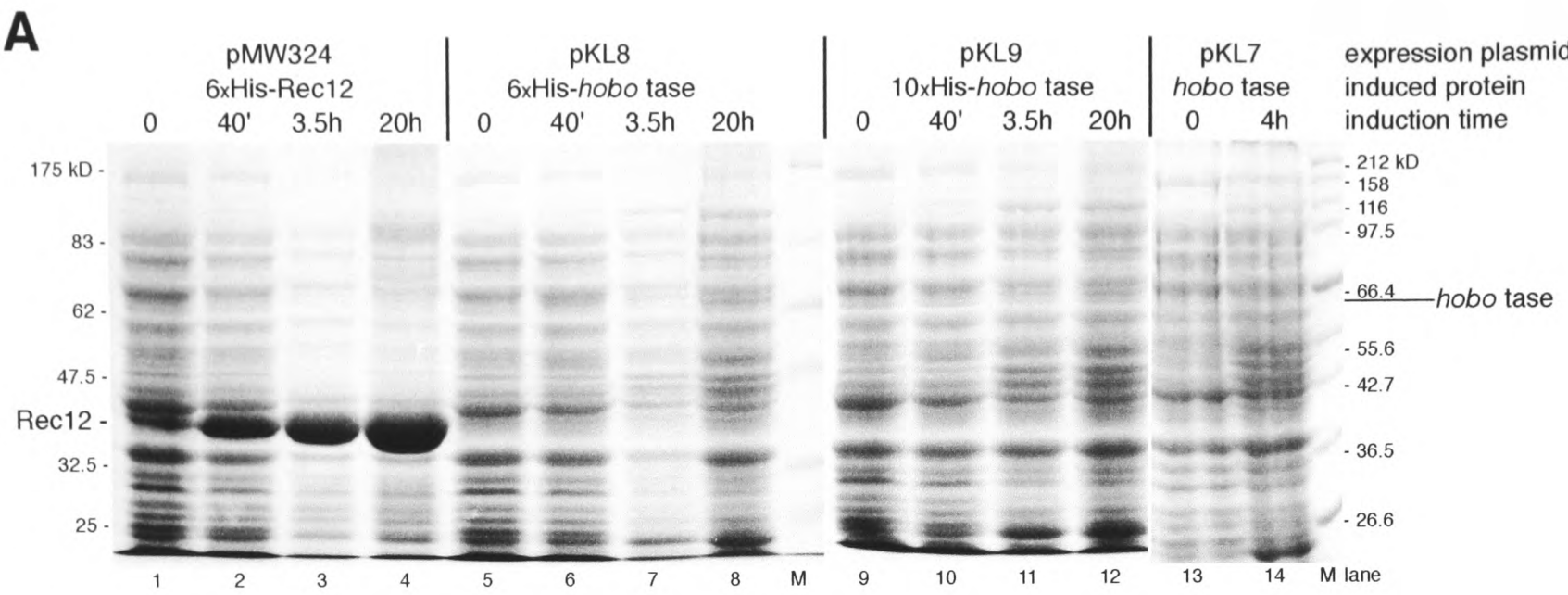


Fig. 3.2

Figure 3.2: T7 Expression of His-tagged proteins in *E. coli* BL21(DE3)pLysS

- (A) Expression of different His-tagged proteins in *E. coli* BL21(DE3)pLysS. Coomassie-stained SDS-PAGE-gel.
- (B) Western blot, developed with monoclonal anti-polyhistidine antibody. 6xHis-Rec12 is expressed very well; 6xHis- and 10xHis-*hobo* transposase are just above the level of detection.
- (C) Kinetics of 6xHis-*hobo* transposase expression in BL21(DE3)pLysS at 37°C; time course over 20 hours. Western blots, probed with anti-poly-histidine. The arrowhead points towards the full-length 6xHis-*hobo* transposase (77 kD).
- (D) A Western blot showing different temperatures during induction. Times are post-induction with 1 mM IPTG.

more than four hours. After the first half hour, the other bands and the smear continue to increase in strength. The best yield of relatively pure full-length fusion protein can therefore be obtained after only 30 minutes of induction. After overnight expression, most of the protein has disappeared, probably due to degradation.

Expression of 6xHis-transposase from pKL8 was carried out at different temperatures. On a Western blot, the level of the full-length signal was virtually identical at all temperatures, but the appearance of additional bands increased with increasing temperature (Figure 3.2D). The highest yield of full-length protein was achieved after 30 min induction at room temperature.

3.3.2 Solubilisation and metal affinity chromatography of the transposase

For optimising the solubilisation and purification of *hobo* transposase, a large expression culture was grown, lysed in a French pressure cell to release the proteins, and frozen in aliquots (see Section 2.3.1.3).

When the lysate is spun at 14,000x g, practically all 6xHis-transposase is found in the pellet, as determined by Western blot (not shown). As only soluble protein can bind to the affinity resin for purification, a number of chemicals and conditions were tested for their ability to keep the protein in solution (Table 3.1, column 5).

The components were tested in parallel at pH 8.0 and 6.0. The lower pH reduced the solubility consistently by c. 50% (not shown). Solubilisation by the denaturing reagents urea, guanidine and SDS was virtually complete. Triton X-100 was almost as effective in solubilisation as urea, and therefore preferred, since it allows purification under native conditions and prevents the need of subsequent renaturation. Magnesium chloride, glycerol, ethanol and DNaseI treatment had no positive effect, for either solubilisation or yield after purification with the metal affinity resin.

In the absence of detergent, the fusion protein is completely insoluble. In a first purification step, transposase can therefore be separated from all soluble proteins by simple centrifugation. The soluble proteins are discarded with the supernatant while the transposase is recovered from the pellet. When this pellet is resuspended in the solutions listed in Table 3.1, the same percentage of transposase can be solubilised as from the original lysate.

Table 3.1: Solubilisation and purification yields of 6xHis-*hobo* transposase in different reagents

component	charge of detergent	CMC of detergent	concentration	solubilisation	yield after purification
buffer: Tris pH 8.00			50 mM	0	<1%
urea			8 M	100%	n.d.
urea & NaCl			8 M & 0.1 M	100%	15%
guanidine-hydrochloride			6 M	100%	n.d.
NaCl			0.5 M	20%	5%
			2 M	10%	3%
			4 M	0	n.d.
MgCl ₂ & NaCl			0.01 M & 0.1 M	10%	1%
glycerol			20%	20%	n.d.
glycerol & NaCl			20% & 0.5 M	25%	n.d.
ethanol			12%	5%	n.d.
			30%	0 {5%}	n.d.
ethanol & NaCl			12% & 0.5 M	0	n.d.
ethanol & glycerol			12% & 20%	20%	n.d.
DNaseI			100 µg/ml	0	n.d.
DNaseI & NaCl			100 µg/ml & 0.5 M	10%	n.d.
SDS	anionic	8.27 mM	0.5%	100%	n.d.
			1%	100%	n.d.
			4.5%	80%	n.d.
deoxycholic acid & NaCl	anionic	5 mM	0.01 M & 0.5 M	0	n.d.
			0.1 M & 0.5 M	0	n.d.
cetylpyridinium chloride & NaCl	cationic	0.90 mM	0.01 M & 0.5 M	0	n.d.
			0.1 M & 0.5 M	0	n.d.
decanoyl-N-methylglucamide & NaCl	non-ionic	6-7 mM	0.01 M & 0.5 M	3%	n.d.
			0.1 M & 0.5 M	0	n.d.
Triton X-100	non-ionic	0.24 mM	0.8%	90%	n.d.
			5%	0 {80%}	5%
			10%	0 {60%}	n.d.
Triton X-100 & DNaseI	non-ionic	0.24 mM	0.8% & 0.1 mg/ml	80%	n.d.
Triton X-100 & NaCl	non-ionic	0.24 mM	0.8% & 0.5 M	95%	50%
			1% & 0.1 M	95%	15%
			5% & 2 M	0 {60%}	n.d.
Triton X-100 & DNaseI & NaCl	non-ionic	0.24 mM	0.8% & 0.1 mg/ml & 0.5 M	85%	n.d.
Triton X-100 & ethanol	non-ionic	0.24 mM	0.8% & 12%	25%	n.d.
Tween-20	non-ionic	0.06 mM	5%	0 {20%}	n.d.
Tween-20 & DNaseI	non-ionic	0.06 mM	0.8% & 0.1 mg/ml	10%	n.d.
Tween-20 & NaCl	non-ionic	0.06 mM	0.8% & 0.5 M	50%	n.d.
SB3-10 & NaCl	zwitterionic	25-40 mM	0.01 M & 0.5 M	20%	n.d.
			0.1 M & 0.5 M	95%	40%
SB3-14 & NaCl	zwitterionic	0.1-0.4 mM	0.01 M & 0.5 M	100%	50%
			0.1 M & 0.5 M	100%	n.d.

column 1 2 3 4 5 6

Solutions were prepared in 50 mM Tris-HCl or 50 mM MES and carefully adjusted to pH 8.0 or pH 6.0, respectively. For optimisation of solubilisation of 6xHis-*hobo* transposase, aliquots of the lysate (Section 2.3.1.3) were diluted 1:10 in the prepared solutions, incubated for 1 h on ice, and centrifuged for 2 h at 14,000x g, 4°C. The untreated lysate, the diluted lysate, the pellet and the supernatant were separated by SDS-PAGE and the His-tagged protein visualised by Western blot. The signal from protein remaining in the supernatant was compared to the signal of protein from the untreated whole lysate. This ratio was noted as percentage of solubilisation (column 5).

For optimisation of the purification yield, 6xHis-*hobo* transposase was purified from differently diluted lysates on a mini-scale with the use of 50 µl TALON™ Metal Affinity Resin each (Clontech, User Manual). Diluted lysate, flowthrough, washes and elution fractions were assayed on Coomassie-stained gels and Western blots. The purification yield was estimated by comparing the signal strength of loaded solubilised lysate to that of combined elution fractions, taking into account that the latter were ten times more concentrated than the earlier fractions (column 6).

The effect of different buffer components on the yield of protein after purification with cobalt affinity resin was tested similarly (Table 3.1, column 6; Figure 3.3A). Another criterion for selection of optimal conditions was the lack of background on Coomassie-stained gels. The best overall yield was obtained with a combination of low detergent concentration (Triton X-100 or SB3-14) and a relatively high salt concentration of 500 mM NaCl. The optimum imidazole concentrations of 15 mM for washes and 100 mM for elution were determined. For storage, the protein was found to be most stable in 25% glycerol and at a temperature of -80°C, as determined by Western blot after two months storage (not shown).

3.3.3 Preparative purification of 6xHis-transposase

His-tagged *hobo* transposase was purified by centrifugation, solubilisation and cobalt-affinity chromatography as optimised above and described in Section 2.3.1.4. Lysate, clarified sample, flowthrough, several wash fractions and the first two resin volumes of eluate were examined by Coomassie-stained SDS-PAGE and Western blot (Figure 3.3). Most protein, (c. 90%) eluted in 3-5 peak fractions. The peak fractions were stored at -80°C in the presence of glycerol and used for *in vitro* experiments (see Section 5.2).

3.4 Additional *hobo* transposase expression trials in *E. coli*

Western blots of 6xHis-*hobo* transposase from pKL8 in *E. coli* revealed several bands corresponding to truncated protein. Three of these bands are the most prominent and

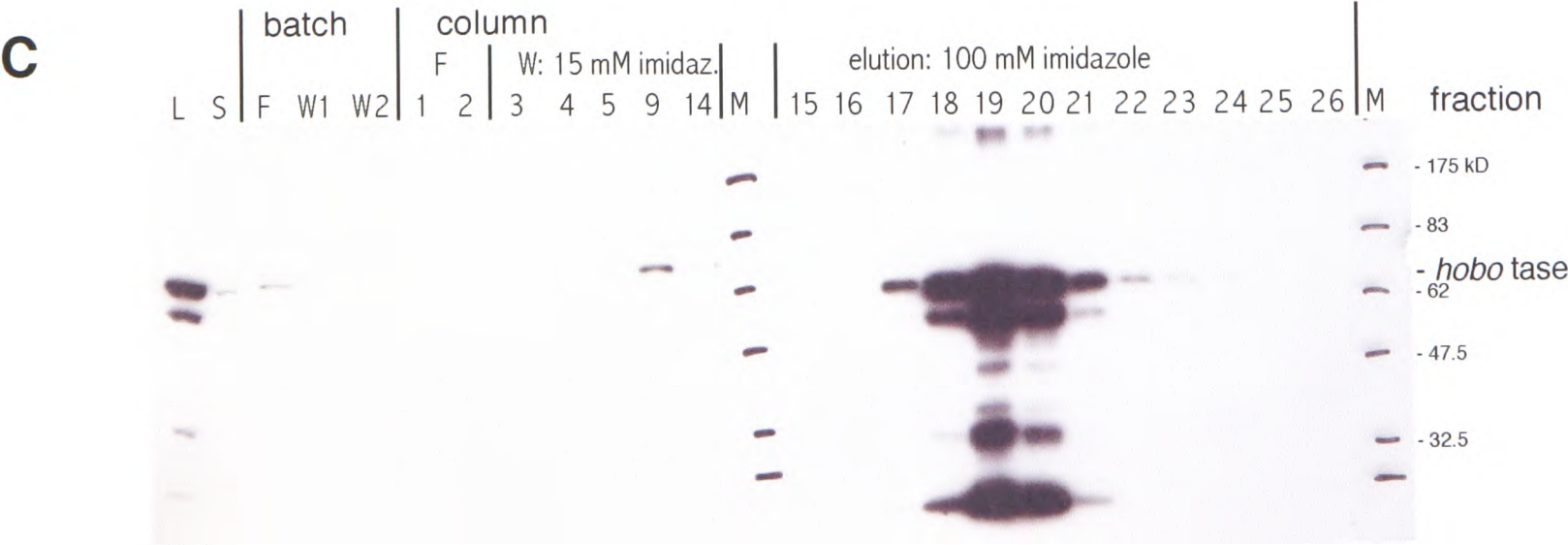
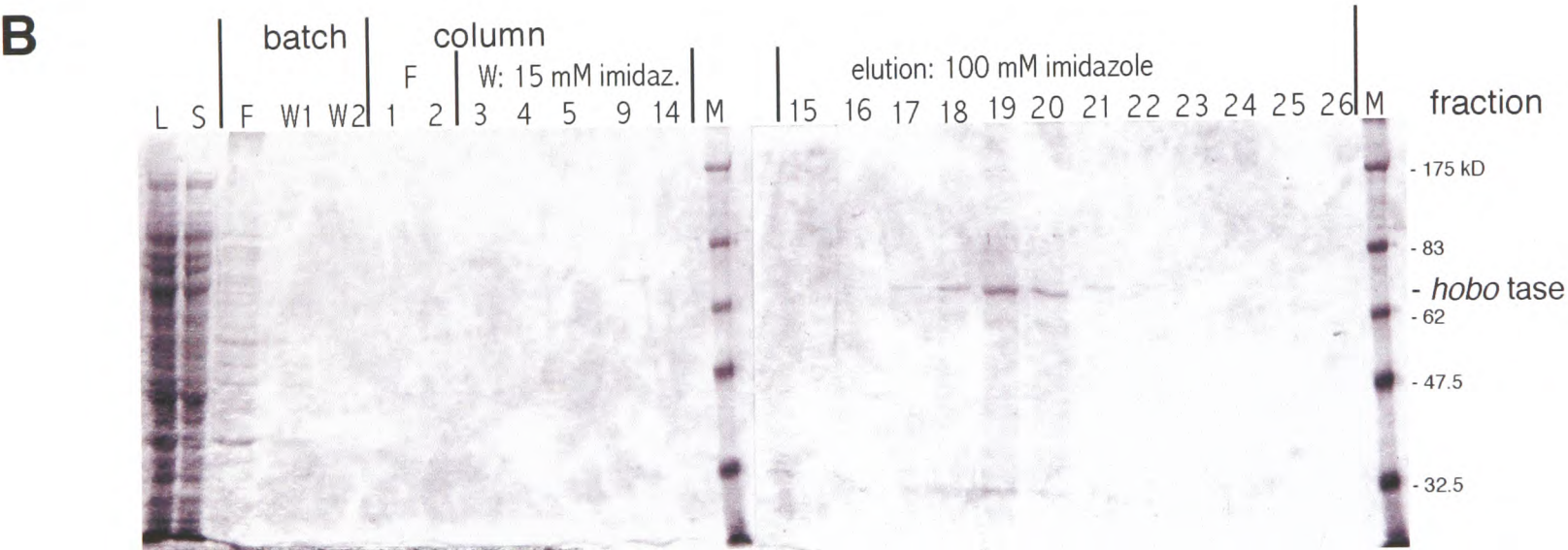
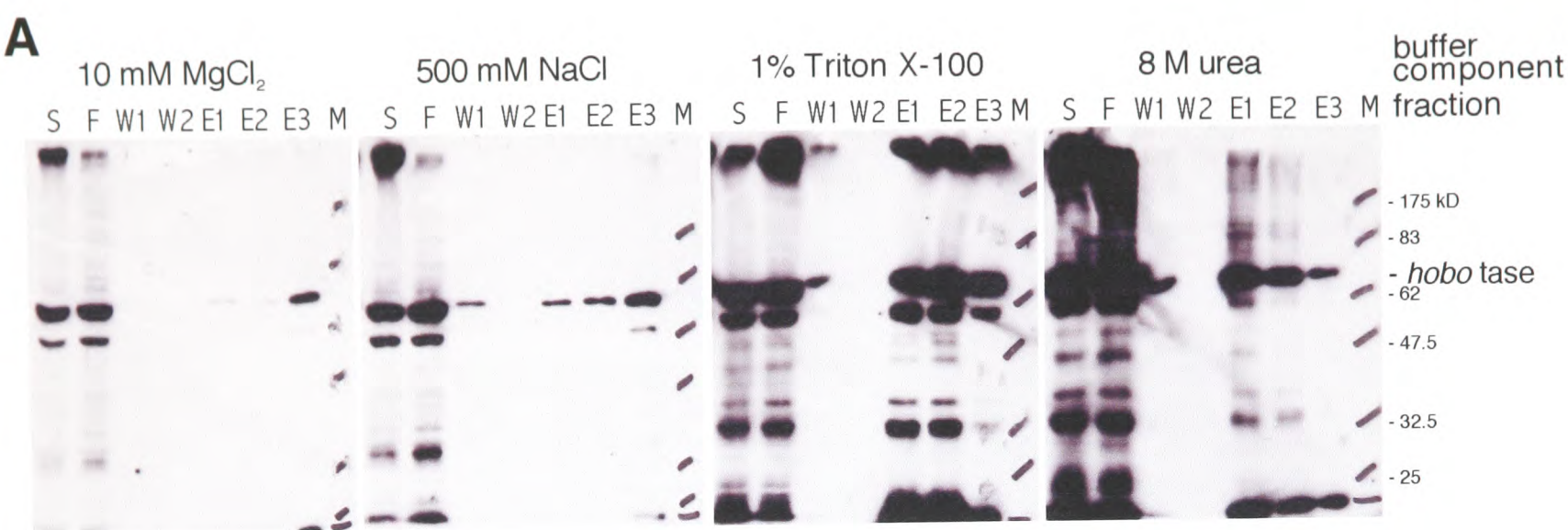


Fig. 3.3

Figure 3.3: Purification of 6xHis-*hobo* transposase using cobalt affinity resin

- (A) Mini-scale batch purification with TALON metal affinity resin. All four Western blots were exposed for 80 min. The buffers for resuspension of the lysate, washes and elution all contained 20 mM Tris, pH 8.0, 100 mM NaCl plus the reagent indicated at the top of the gels. S, supernatant, lysate clarified after 10 min incubation in the resuspension buffer; F, flowthrough; W1+2, wash; E1+2, elution with 100 mM imidazole; E3, elution with 500 mM imidazole, 1% SDS; M, molecular weight marker.
- (B) Full scale purification. Coomassie stained SDS-PA gel. Fractions: L, untreated cell lysate; S, supernatant after second centrifugation in TTN-buffer, loaded onto the metal affinity resin; F, flowthrough; W, wash; 1-26, fractions collected from the column; M, molecular weight marker. All washes and elutions were with TTN-buffer (50 mM Tris, 0.8%v/v Triton X-100, 500 mM NaCl, all pH 8.0) plus the indicated concentration of imidazole.
- (C) Western blot of a gel identical to (B), probed with anti-poly-His antibody.

appear in every experiment (Figure 3.2; Figure 3.3). The proteins in these bands bind to the anti-poly-histidine antibody and only appear after induction. They therefore probably contain the N-terminal 6xHis-tag and are truncated at the C-terminus.

The occurrence of rare codons in a gene can be the cause of premature termination of translation. As codon usage differs between species, *Drosophila* codon usage can cause problems in *E. coli*. For comparison, the rarest codons in *E. coli* and their frequency in *hobo* and some relevant species are listed in Table 3.2. Note that if all 64 codons were used equally, each would have a frequency of 15.6 per 1000. Codon usage of *hobo* ORF1 and the average *E. coli* gene differs especially for AGA (arg), AUA (ile) and CUA (leu).

Analysis of the *hobo* open reading frame using the GCG software package identified the position of several codons that are very rarely used in *E. coli* (not shown). At the C-terminus of *hobo* ORF1, at positions 576 and 578, two arginine codons (AGA), which are extremely rare in *E. coli*, lie only one amino acid apart. If translation stops at codon 575, the truncated protein has a calculated molecular weight of 67.7 kD. This is in good agreement with the mobility of the largest of the three major truncations. The assumption that the truncations are caused by prematurely terminated translation at rare codons is further supported by the observation that the full-length protein reaches its maximum concentration within an hour, after which time only the truncations increase in concentration. This is because the core tRNAs would be exhausted.

Table 3.2: Codon usage

Codon	amino acid	<i>E.coli</i> gene	Frequency per 1000 triplets					
			<i>E. coli</i>	<i>hobo</i> tase	<i>Himar1</i> tase	<i>S. pombe</i>	<i>D. mela-nogaster</i>	<i>M. musculus</i>
AGG	R	<i>argU</i>	2.1	4.6	4.2	5.1	6.4	11.7
AGA	R	<i>argU</i>	3.7	15.2	4.2	11.2	5.1	11.5
CGA	R	<i>argNI</i>	3.8	6.1	8.4	8.0	8.5	6.7
CUA	L	<i>leuW</i>	4.2	16.7	0.0	8.7	8.3	7.8
UGU	C	?	5.2	12.1	12.6	9.0	5.3	10.9
CCC	P	<i>proL</i>	5.4	3.0	16.7	8.1	18.0	18.4
CGG	R	<i>argX</i>	5.9	3.0	12.6	3.0	8.4	10.4
AUA	I	<i>ileY</i>	7.0	19.7	4.2	13.5	9.2	7.0
column 1	2	3	4	5	6	7	8	9

Frequency of triplets per 1,000 of the rarest codons in *E. coli* (excluding stop codons) and in other organisms and transposons. ?, the gene for this tRNA^{Cys} has not yet been identified (Kredich, 1996). Data for *Escherichia coli*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Mus musculus* protein coding gene sequences from: <http://www.kazusa.or.jp/codon/> (Nakamura *et al.*, 2000) (data source: NCBI-GenBank, updated 15/08/02).

3.4.1 Replacement of rare codons in the transposase

The two rare Arg codons at positions 576 and 578 in pKL8 were replaced by the Arg codon most common in *E. coli* (CGT), using the QuickChange™ Site-Directed Mutagenesis Kit (Stratagene) and primers KL35 and KL36, which concurrently introduce a new *Hind*II restriction site. The mutant plasmid (pKL25) was sequenced across the mutagenised site to confirm that no additional mutations have been introduced by PCR.

Protein was expressed from pKL25 and pKL8 in parallel according to the optimised protocol (Section 2.3.1.1). At both temperatures tested, the original transposase sequence and the one with exchanged codons give rise to a virtually identical expression pattern (Figure 3.4). This suggests that truncations are due to other rare codons or perhaps proteolysis. In both cultures, truncated proteins and aggregates accumulate very quickly, whereas the level of full-length transposase decreases again after its peak at 45 minutes.

3.4.2 Increased expression with extra tRNA genes

Eliminating rare codons is one approach to improve expression. Alternatively, expression from genes with rare codons can be increased by supplementing the number of rare tRNA molecules. During this work, the CodonPlus™ strains became commercially available (Stratagene). The BL21-CodonPlus(DE3)-RIL strain contains a *Cam*^R-marked plasmid with the three rare tRNA genes *argU* (AGG and AGA), *leuW* (CUA), and *ileY* (AUA). As these codons are found frequently in the *hobo* open reading frame (Table 3.2) it was chosen over the -RP strain, which contains extra copies of *argU* and *proL* (CCC), which is less common in the *hobo* transposase orf.

E. coli BL21-CodonPlus(DE3)-RIL were transformed with pKL8, and expression of 6xHis transposase was induced (Section 2.3.1.1). There was a dramatic increase of the band corresponding in size to full-length transposase on Coomassie-stained gels (Figure 3.5A,C), to c. 5-10% of total cell protein. Otherwise, the expression pattern is very similar to that in BL21(DE3)pLysS (Figure 3.4). Full-length product increased up to 4 h at RT and had a maximum at c. 1 h at 37°C. Western blots showed quick accumulation of the full-length protein, but also of truncations and aggregates (lower and higher bands) (Figure 3.5B,D).

The ratio of full-length to truncated protein is probably better than the signal strength of the Western blot suggests: Ten minutes after exposure to the chemiluminescent solution, the signal of the strongest bands is already bleached (not shown). This indicates that the high amount of protein contained in these bands depletes the luminescent reagent.

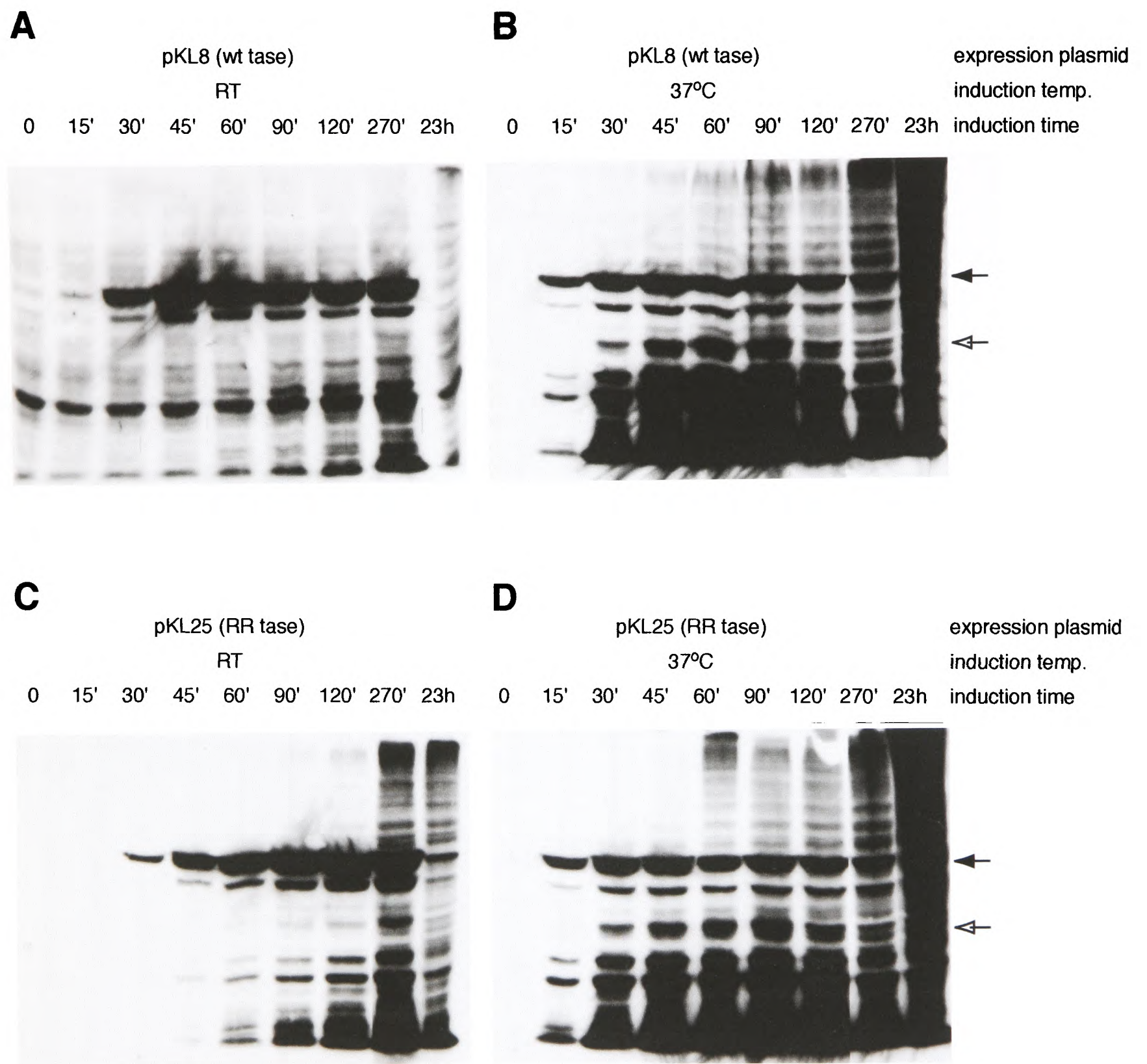


Fig. 3.4

Figure 3.4: Comparison of protein expression of wildtype and RR *hobo* transposase

Expression from pKL8 and pKL25 in *E. coli* BL21(DE3)pLysS. In RR transposase (expressed from pKL25), two rare Arg codons are replaced by the most frequently used one. Western blots of 10% SDS-PAGE, probed with α -poly-His antibodies. Filled arrows, full-length 6xHis-*hobo* transposase; open arrows, major truncation that is only present at high expression temperatures. Blot (A) was exposed longer than (B-D).

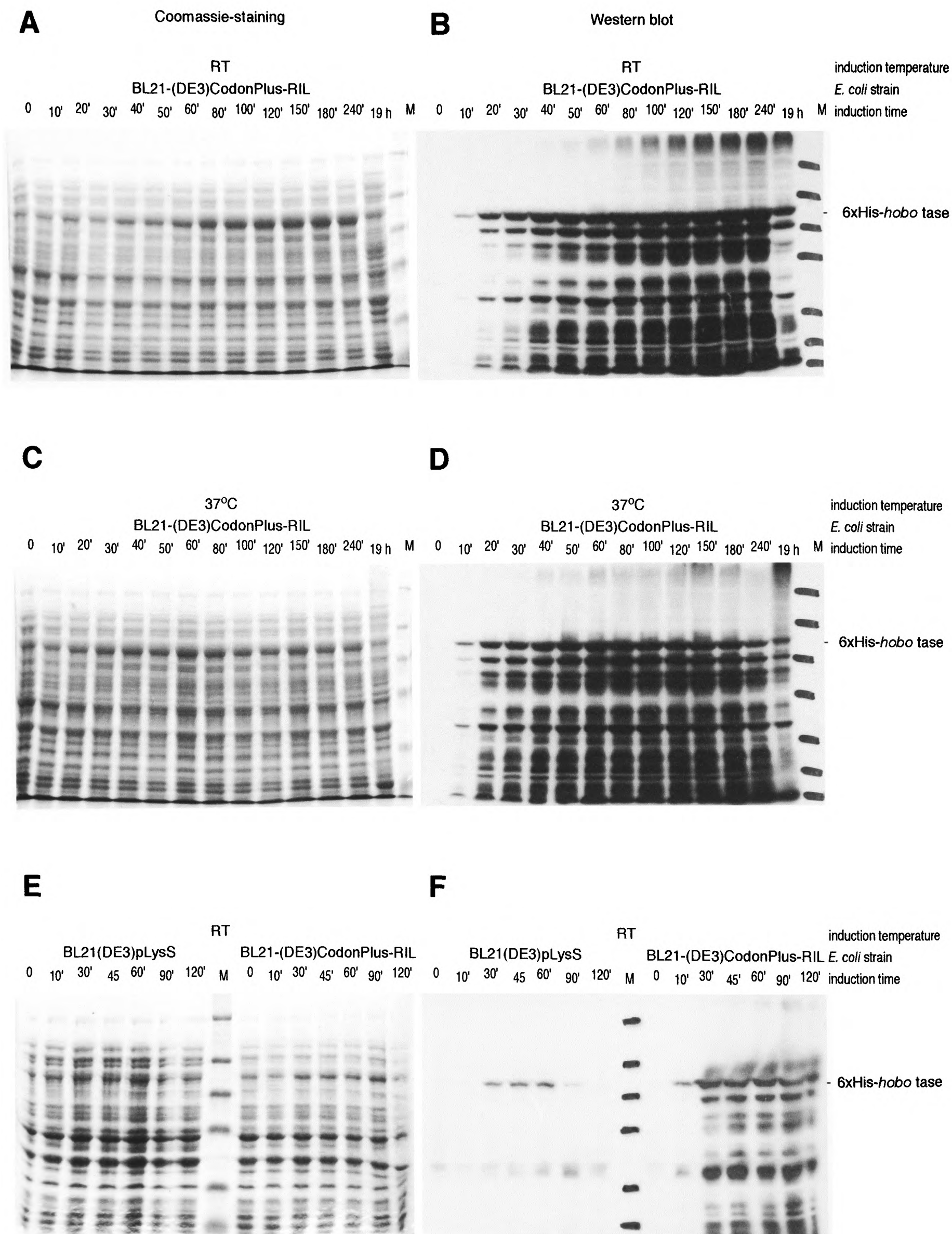


Fig. 3.5

Figure 3.5: Expression of 6xHis-hobo transposase with extra copies of rare tRNA genes

- (A) *E. coli* BL21-CodonPlus(DE3)-RIL were transformed with pKL8, grown and induced at RT as described. Coomassie stained gel.
- (B) Western blot of a gel identical to (A), probed with HRP-coupled α -poly-His.
- (C) *E. coli* BL21-CodonPlus(DE3)-RIL, transformed with pKL8, grown and induced at 37°C. Coomassie stained gel.
- (D) Western blot of a gel identical to (C).
- (E) *E. coli* BL21(DE3)pLysS and *E. coli* BL21-CodonPlus(DE3)-RIL, transformed with pKL8, grown in parallel and induced at RT. Coomassie stained gel.
- (F) Western blot of a gel identical to (E).

When 6xHis-transposase is expressed at 16°C, overall yield is lower, but the ratio of full-length to truncated protein is higher (Figure 3.6A,B).

The expression pattern looks similar in the wt and CodonPlus strains. For direct comparison, pKL8, pKL25 and pKL88 were transformed into both BL21(DE3)pLysS and BL21-CodonPlus(DE3)-RIL. Protein was expressed and analysed on the same gel for both strains (shown for pKL8 in Figure 3.5E,F). Although the amount of total protein is lower in the CodonPlus lanes, the amount of His-tagged protein is greatly increased. The largest band visible on a Western blot is of the same length in both strains. This confirms that this is full-length transposase.

In conclusion, expression in a CodonPlus-strain results in increased levels of His-tagged transposase protein of all lengths. Neither expression in a CodonPlus-strain nor the exchange of rare codons increases the ratio of full-length to truncated protein.

Due to the late availability of the strain, the expressions in the CodonPlus strain were carried out when all other experiments on *hobo* had been completed. This is why the higher expression was not utilised for purification or *in vivo* experiments. In addition, it does not seem very likely that transposase purified from the CodonPlus strain would have performed differently in *in vitro* transposition experiments.

3.4.3 Expression of transposase with a C-terminal His-tag

Premature translation termination results in peptides that are truncated at the C-terminus. They all contain an intact N-terminus and therefore copurify with the N-terminally tagged full-length 6xHis-transposase. Full-length protein may therefore be obtained by placing the purification tag on the C-terminus.

The *hobo* open reading frame was cloned into pET32a(+) to create a plasmid (pKL94) that is identical to pKL8, apart from the C-terminal position of the 6xHis-tag (see Table 2.6, Table 2.7). Expression in *E. coli* BL21-CodonPlus(DE3)-RIL gave a similar expression profile for pKL8 and pKL94 on Coomassie-stained gels (Figure 3.6A,C). On a Western blot, α -polyHis antibody gave rise to clear signals for the pKL8 product while failing to detect any product of pKL94 (Figure 3.6B,D).

According to the manufacturer's manual, the Sigma α -polyHis antibody can detect both N- and C-terminal His-tags. However, this is not the case for every individual protein. If a tag is not detected, it can mean that it is either not present or folded incorrectly. In these cases, purification with metal affinity resin is not feasible.

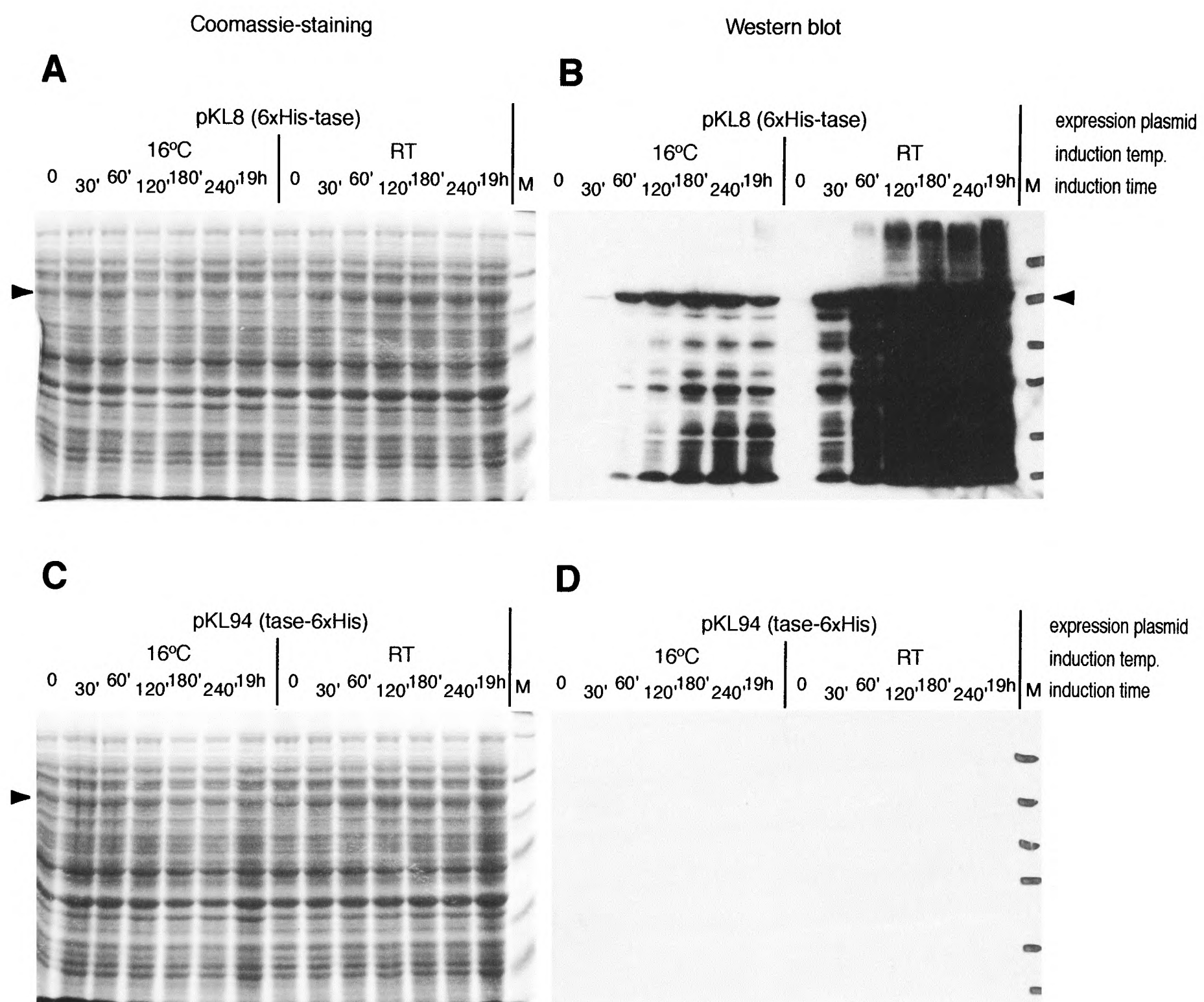


Fig. 3.6

Figure 3.6: Comparison of N- and C-terminal 6xHis-tags

- (A) *E. coli* BL21-CodonPlus(DE3)-RIL transformed with pKL8 and induced to express 6xHis-*hobo* transposase. Coomassie-stained gel.
- (B) *E. coli* BL21-CodonPlus(DE3)-RIL transformed with pKL94 and induced to express *hobo* transposase-6xHis. Coomassie-stained gel.
- (C) Western blot of a gel identical to (A).
- (D) Western blot of a gel identical to (B).

DNA sequencing of pKL94 did not identify any mutations. The majority of the expressed protein should therefore include the His-tag, although it is still possible that translation never reaches the extreme C-terminus due to rare codons or other reasons. As nothing more could be done to improve this system, no further experiments were conducted with the C-terminal (His-tag) fusion protein.

3.5 Expression of GST-transposase in *S. pombe*

In *Schizosaccharomyces pombe*, the *hobo* transposase was expressed as an N-terminal glutathione-S-transferase (GST) fusion, which can be entirely cleaved off by enterokinase digestion. Many proteins have been successfully purified as GST fusions, because GST tags often help to get otherwise insoluble proteins into solution.

3.5.1 Cloning of the GST-*hobo* expression plasmid and transformation into *S. pombe*

Three independent clones of pESP2-*hobo* (pKL6; Table 2.7) and a control plasmid (pESP2-GST; Table 2.8) were transformed into SP-Q01 *S. pombe* competent cells. Frozen stocks were made from four colonies of each transformation. Since plasmids freshly transformed into *S. pombe* sometimes suffer rearrangements, these stocks were tested for the presence of the transposase orf by colony-PCR (Section 2.2.1.7). One third of the transformants had indeed lost the insert due to recombination, but at least one colony per transformation had the correct plasmid.

3.5.2 Expression of the GST-*hobo* fusion protein

Cultures from one colony of each transformation were grown in repressing and inducing conditions (Section 2.3.2). Cells were ruptured, clarified and analysed by SDS-PAGE (Figure 3.7A). In the control cultures, a protein of the correct size for the GST-GST control protein is clearly induced (Figure 3.7A, lane 8). However, no induction of any protein in the pKL6 transformants was detectable (Figure 3.7A, lanes 2,4,6).

A Western blot was carried out and probed with anti-GST mouse monoclonal antibody (2TL 237 + 2TL 188, gift from Alison Banham, John Radcliffe Hospital, Oxford) and HRP-conjugated anti-mouse IgG (Amersham) (Figure 3.7B). The control protein gave rise to two very close bands, which were quite weak. There was no background, and no other band appeared. The pESP2-*hobo* lanes were without any signal.

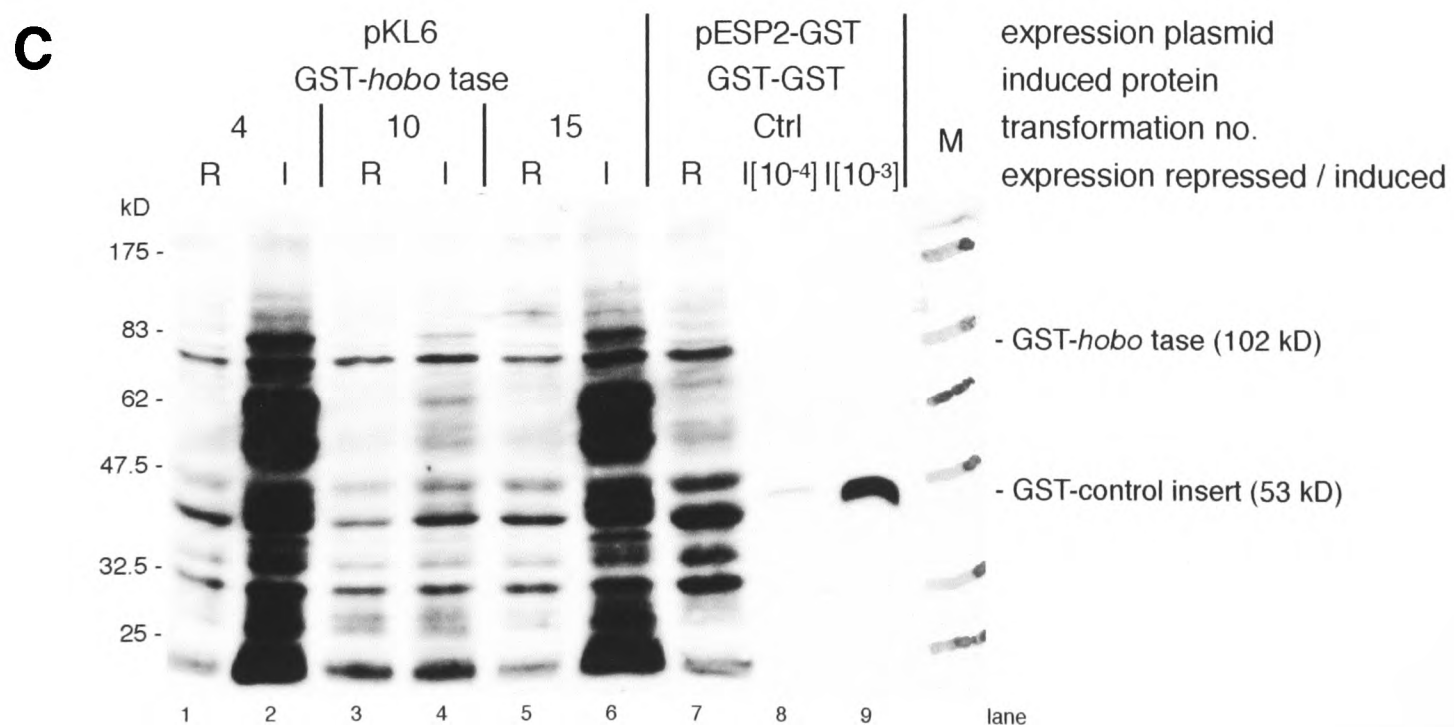
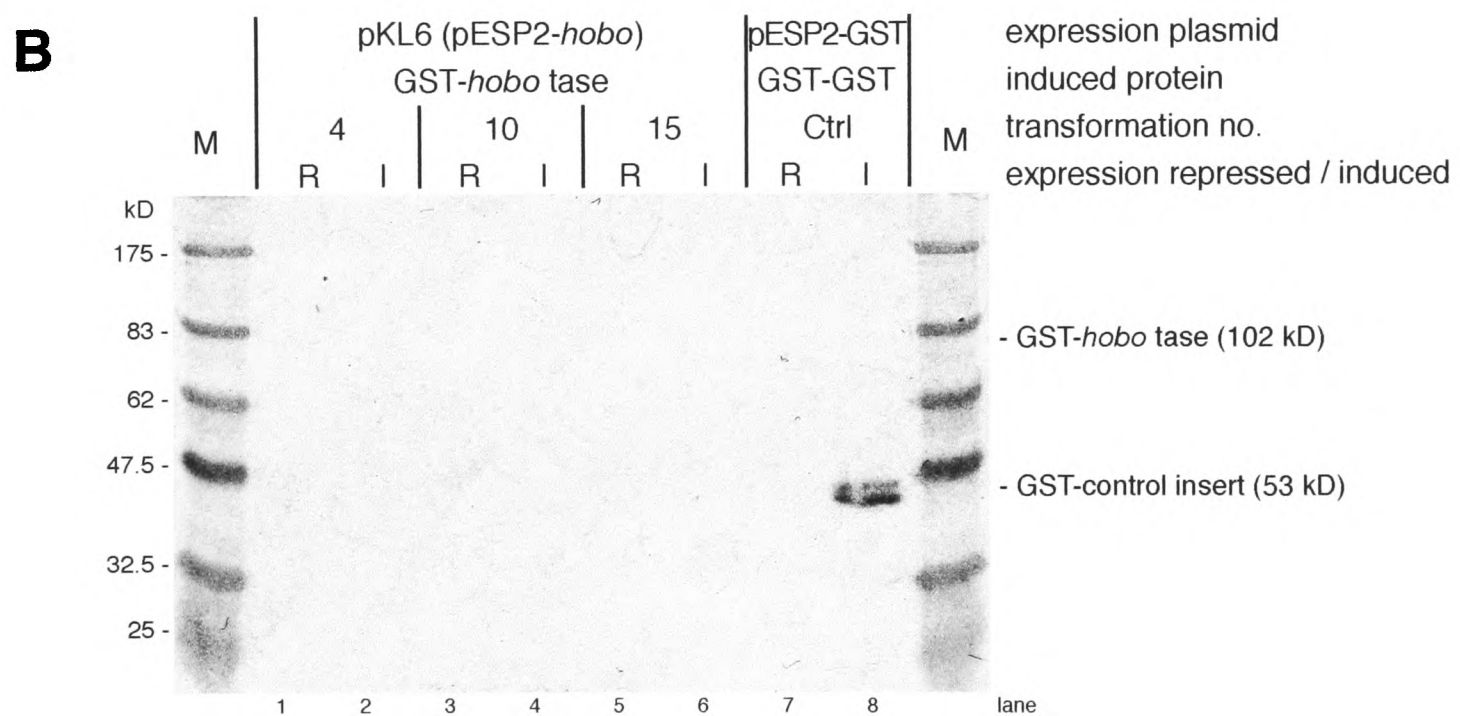
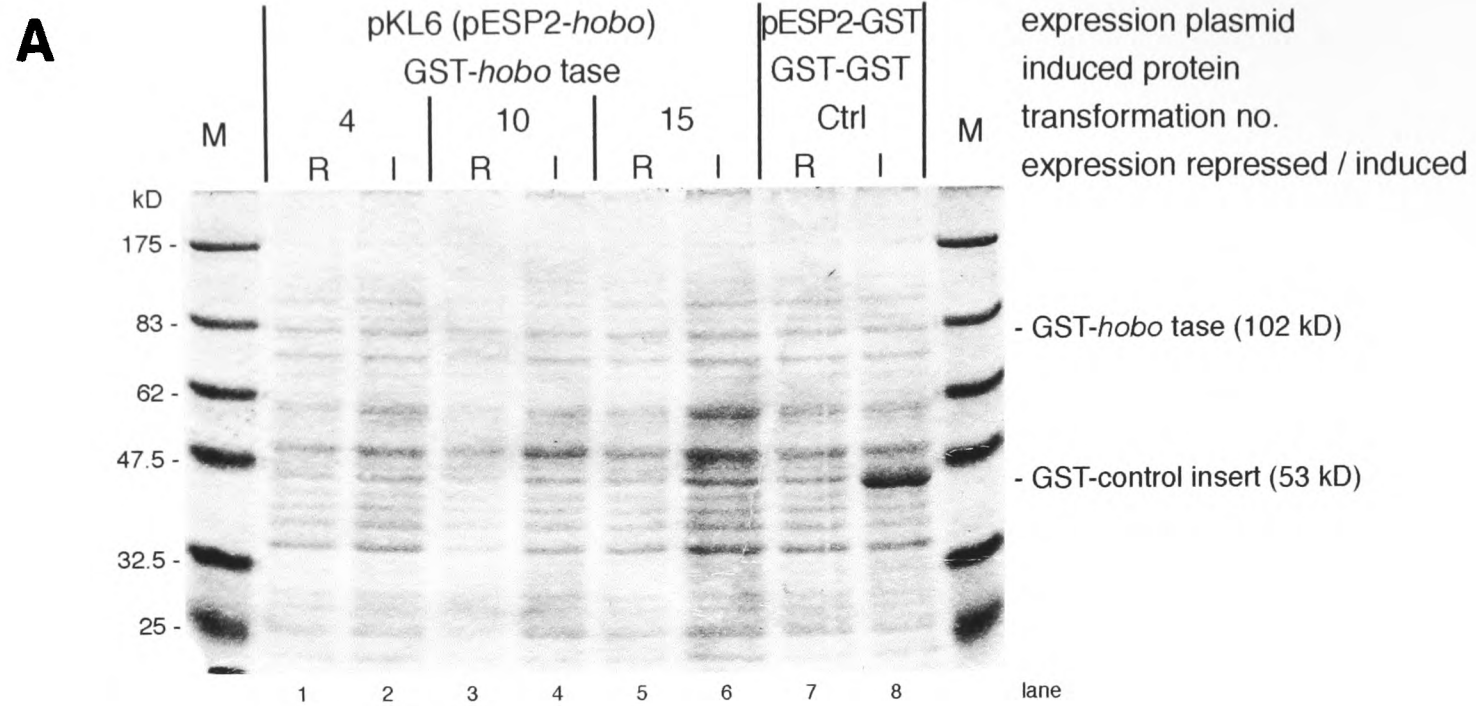


Fig. 3.7

Figure 3.7: GST-fusion protein expression in *S. pombe*

Numbers 4, 10, 15, different pESP2-*hobo* (pKL6) clones; Ctrl, pESP2-control plasmid; R, repressed with thiamine; I, induced (no thiamine). The pictured gels were loaded with the soluble fractions after a clarifying centrifugation.

(A) 8% SDS-PAGE, stained with Coomassie blue.

(B) Western blot, incubated with monoclonal mouse anti-GST and HRP-coupled anti-mouse antibodies.

(C) Western blot, incubated with HRP coupled rabbit anti-GST; the induced control samples are diluted. A band of the correct size to be the GST-*hobo* transposase and some shorter bands increase in strength after induction of the pESP2-*hobo* transformants. Soluble fraction, 8% SDS-PAGE, 1.5 h at 100 V.

To see if GST-*hobo* transposase was insoluble, cell extract was loaded onto the gel, as well as supernatant and pellet after the clarifying spin. The control protein was the strongest in the pellet fraction, adding up to almost 30% of total protein. But in none of the three pESP-*hobo* fractions did a band increase detectably in any of the lanes (not shown).

Western blots of gels with whole cell extract, supernatant and pellet were probed with rabbit anti-GST peroxidase conjugate (Sigma). The control protein gave a signal in all fractions, even when the cell fraction was diluted 10,000 fold (Figure 3.7C and not shown). Exposures up to one minute did not give rise to any signal in any of the other lanes. After 4.5 hours exposure, bands were visible in all lanes that had been loaded with undiluted cell extract. A number of bands appeared only in the induced cells. One of these extra bands has approximately the size of the full-length fusion protein (102 kD). Most of the additional bands were shorter, the strongest running at approximately 60 and 80 kD. These could represent the N-terminally tagged fusion protein, truncated either by proteolysis or by incomplete translation. However, the presence of full-length GST-transposase, how ever little, was promising.

3.6 *In vitro* transcription and translation in reticulocyte lysate

hobo transposase was produced *in vitro* in a rabbit reticulocyte transcription-translation system (Promega TNT® T7 Quick Coupled Transcription/Translation System). A transposase expression plasmid (pKL89; Table 2.7) was created from pTM1 (Moss *et al.*, 1990; Table 2.8). In this plasmid, the T7 promoter is followed directly by the encephalomyocarditis virus (EMCV) 5'UTR, which enables Cap-independent translation and had been found to increase expression 5- to 10-fold (Elroy-Stein *et al.*, 1989). In pKL89, the initiation codon is in a Kozak consensus context (A/gCCAUGG) and the transposase orf is followed by an SV40 polyadenylation signal to ensure optimal translation.

hobo transposase and the control protein luciferase were expressed in TNT T7 Quick Master Mix in the presence of ³⁵S-methionine (Section 2.3.3). The production of newly synthesised protein was assayed by SDS-PAGE and autoradiography (Figure 3.8, lanes 1-3). Protein of the correct size for *hobo* transposase and a few shorter species were produced. The expression level was within one or two orders of magnitude of the control protein. The experimental conditions were optimised (Table 3.3) such that a very high full-length-to-background ratio was achieved (Figure 3.8, lanes 4,5).

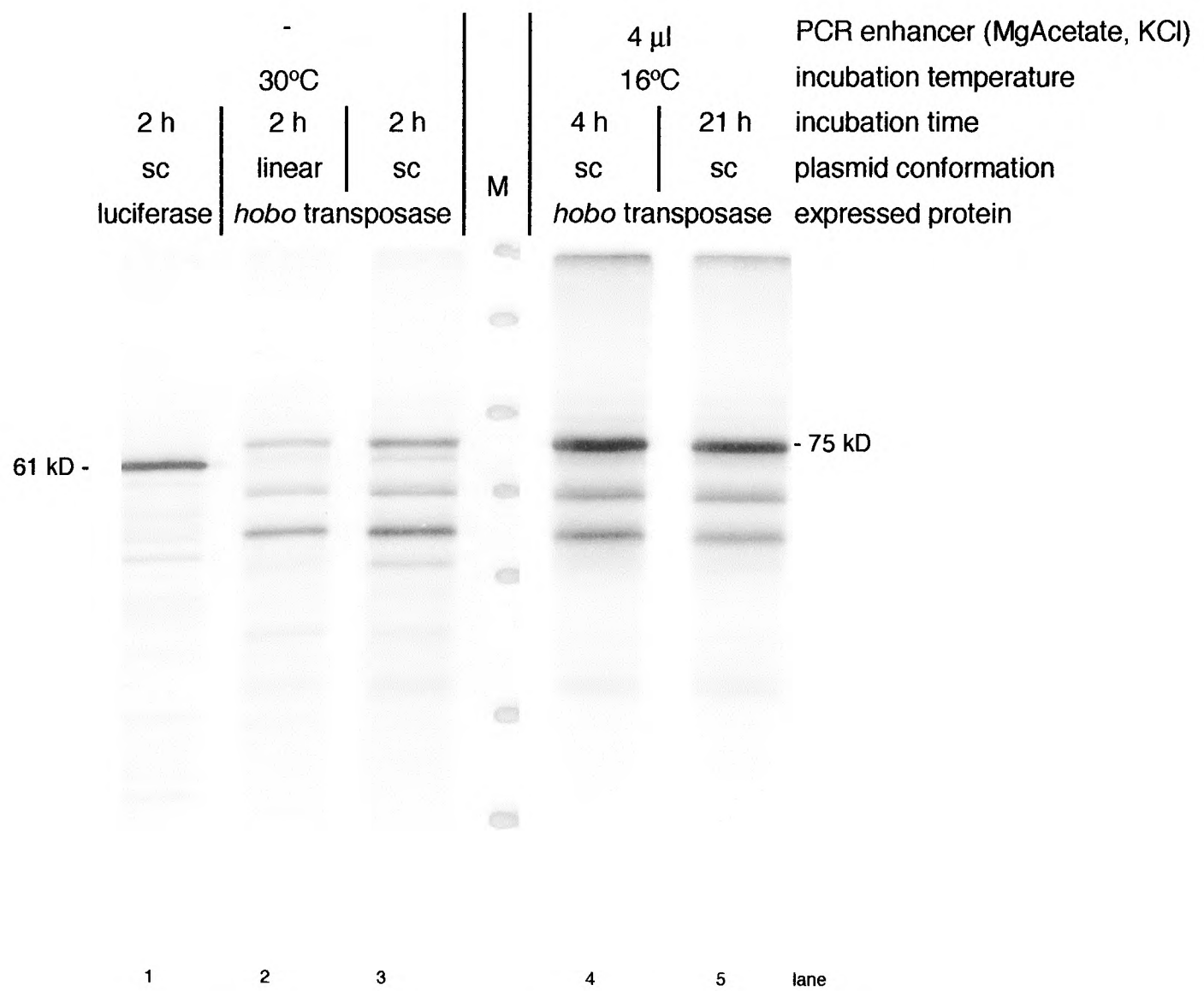


Fig. 3.8

Figure 3.8: Coupled transcription/translation in reticulocyte lysate

Autoradiogram of SDS-PAGE of *in vitro* expression reactions with incorporation of ³⁵S-Met. Luciferase expression from pLucT7 (Table 2.8), *hobo* transposase expression from pKL89 (Table 2.7). See Section 2.3.3 for details. Lanes 1-3, first expression reactions; lanes 4-5, optimised reactions.

Table 3.3: Optimisation of *hobo*-transposase expression by coupled transcription/translation

parameter	value or specification	signal strength	signal : noise	initial	final
pKL89 conformation	supercoiled	1	1	✓	✓
	linear	-	-		
pKL89 amount	50 ng	1	+		
	125 ng	1	+		✓
	250 ng	1	1	✓	
	500 ng	-	-		
PCR enhancer (MgAcetate, KCl)	0	1	1	✓	
	0.25 µl	+	+		
	0.5 µl	+	+		
	1 µl	++	++		✓
incubation time	0.5 h	--	-		
	1 h	-	1		
	2 h	1	1	✓	
	3 h	+	+		
	4 h	+	++		✓
	6 h	+	+		
	22 h	1	-		
incubation temperature	16°C	--	+		
	20°C	-	+		✓
	30°C	1	1	✓	
	37°C	-	--		
protease inhibitors	-	1	1	✓	✓
	Aprotinin, 1 µg/µl	1	1		
	Chymostatin, 1 µg/µl	+	1		
	Leupeptin	1	1		
<i>Drosophila</i> egg extract*	-	1	1	✓	✓
	nuclear extract, 0.5 µl	---	1		
	nuclear extract, 2 µl	0	(n.a.)		
	nuclear extract, high NH ₄ SO ₄ , 0.5 µl	0	(n.a.)		
	nuclear extract, high NH ₄ SO ₄ , 2 µl	0	(n.a.)		
	chromatin assembly extract, 0.5 µl	--	1		
	chromatin assembly extract, 2 µl	0	(n.a.)		

Master Mix, PCR enhancer and the basic protocol are from the TNT® T7 Quick Coupled Transcription/Translation System (Promega). See Section 2.3.3 for details. All results are for 12.5 µl reactions. 1, standard for comparison per group; +, better than standard; -, worse than standard; 0, not detectable. * *D. melanogaster* embryonic extracts: gift from Peter Becker, Munich.

Linearising the expression plasmid at the 3' end of the transcriptional terminator is often useful as it prevents a continued false transcription; in this case it did not have a positive effect. Maximum full-length translation is not always correlated with maximum transcription. This is manifested by higher full-length expression from lower template concentrations. Magnesium acetate is known to sometimes boost transcription and translation. Potassium chloride increases the preference of the translation machinery for the Kozak consensus sequence. These salts were found to have a positive effect on the expression level and specificity. As in *E. coli* (Section 3.3.1), a higher ratio of full-length transposase was obtained at lower expression temperatures. *Drosophila melanogaster* embryonic nuclear extracts, which were tested because they might contain components that aid *hobo* activity, were found to have a very detrimental effect on expression.

For transposition experiments (Section 5.3), the transposase was produced with unlabelled methionine.

3.7 Discussion

3.7.1 Expression systems

A large variety of protein expression systems are available. Each system has its advantages and disadvantages, and it is usually difficult to predict which will give the best results for a particular protein. Furthermore, each system can be altered in an almost infinite number of ways to optimise expression level, proportion of full-length protein, solubility and activity. Similarly, there are countless methods of protein-purification, and each can be optimised for yield, purity and activity of the purified protein.

A subset of systems from prokaryotes, lower eukaryotes and higher eukaryotes has been tested here for the expression of *hobo* transposase. The focus has been on the expression and purification of N-terminally 6xHis-tagged transposase in *E. coli* and the *in vitro* expression of native transposase in rabbit reticulocyte extract. Other constructs in *E. coli* and a GST-fusion in *S. pombe* were tested briefly, but due to problematic detection, the study of these was not expanded.

So far, no *in vitro* assay for *hobo* transposition has been established. It was therefore not possible to test the activity of the purified transposase. Instead, expression and purification were optimised for yield, solubility and purity of full-length protein.

3.7.2 Low level of expression

In all systems tested, the initial levels of expression of *hobo* transposase were very low. The necessity for Western blots or body-labelling with ^{35}S -Met for detection, significantly slowed the optimisation experiments. For bacterial expression of His-tagged protein, the biggest improvement of expression levels was achieved by the use of an *E. coli* strain with extra copies of three rare tRNA genes (BL21-CodonPlus(DE3)-RIL by Stratagene, Figure 3.5F). Thus, a limiting factor for expression was clearly the difference in codon usage between host bacterium and expressed protein. Further improvement might be achieved with the new Rosetta strains by Novagen, which supply a total of six tRNAs, including the three present in the RIL strain.

3.7.3 Protein truncations

In all systems analysed with highly sensitive detection methods (Western blot or incorporation of radioactive amino acids), not only full-length transposase, but several truncations were obtained. For both expression of 6xHis-transposase in *E. coli* (Figure 3.2B) and coupled transcription/translation in rabbit reticulocyte extract (Figure 3.8), the truncations were specific to the presence of *hobo* transposase expression plasmid and inducer (if applicable).

Truncations due to premature termination of translation can be caused by differences in codon usage (Section 3.4). This problem was addressed by exchanging rare codons in the protein codon sequence (Section 3.4.1), and supplementing rare tRNAs (Section 3.4.2) in *E. coli*. However, these methods did not increase the ratio of full-length to truncated protein. Although the rare codons reduce the overall expression level (Section 3.7.2) they are apparently not the cause for truncation.

In both closely studied systems, a significant reduction of protein truncations (and an increase in the ratio of full-length to truncated protein) was obtained by reducing the expression temperature (Figure 3.2D, Figure 3.4, Figure 3.6C; Table 3.3). One possibility is that truncations which appear *only* at higher expression temperatures are produced by proteases, which are not active at lower temperatures. An example is indicated by an open arrow in Figure 3.4.

Other truncations, which increase gradually in strength with temperature, might be caused by premature termination of translation. At each position along the mRNA, there is an equilibrium for the ribosomes between staying on and falling off. At stop codons, rare codons and some secondary structures, the equilibrium is biased towards detachment. The

translation machinery might be less stable at higher temperatures, causing the ribosomes to fall off the mRNA at difficult positions more frequently than at lower temperatures.

Another cause for truncations might be the different generation time. At lower temperatures, practically all enzymes slow down, but not all to the same extent. The cells divide less frequently, and the transcription/translation machinery is not as occupied by the expression of structural and housekeeping genes. For a badly expressing protein like *hobo* transposase, this could reduce the frequency of premature disassembly of transcription or translation complexes.

In vitro, the appearance of truncations decreased when the concentration of expression plasmid was reduced (Table 3.3). Reducing the template decreases the amount of transcription product produced. Thus, a larger proportion of the translation machinery is available for each transcript, which results in more successful and complete translation products. A reduction in the number of transcripts can be achieved in *E. coli* by reducing the copy number and promoter strength of the expression plasmid.

3.7.4 Solubility

Transposases in general tend to be insoluble, and *hobo* transposase is no exception (Table 3.1). Several approaches have been taken to increase solubility and obtain active protein. Tn10 transposase pellets can be solubilised by high salt and detergents (Chalmers and Kleckner, 1994). *Himar1* and *Tc1* transposase were purified as inclusion bodies, denatured and refolded by gel filtration (Lampe *et al.*, 1996; Garcia-Saez and Plasterk, 2000). Fusions to a relatively large, soluble affinity tag such as maltose binding protein (MBP) may also keep the transposase in solution (e.g. Tavakoli *et al.*, 1997).

Here, similar to Tn10, His-tagged *hobo* transposase was solubilised from an insoluble pellet. Optimal results were obtained with the non-ionic detergent Triton X-100 and a high concentration of NaCl. The observation of inclusion bodies after long expression times (see below in Section 5.5.2; Figure 5.5) could be explored further. After denaturation and renaturation, these might yield detergent-free protein.

A number of transposases and recombinases, such as Tn5, RAG1 and RAG2 were found to gain solubility while retaining their activity if truncated at the N- and/or C-terminus (Davies *et al.*, 1999; Sadofsky *et al.*, 1993, 1994; Silver *et al.*, 1993; Cuomo and Oettinger, 1994; Kirch *et al.*, 1996). In *Ac* transposase, an N-terminal truncation of 102 aa even increases the activity (Houba-Herlin *et al.*, 1990; Li and Starlinger, 1990; Becker *et al.*, 1992; Kunze *et al.*, 1993; Heinlein *et al.*, 1994; Weil and Kunze, 2000).

The C-terminal truncations of 6xHis-*hobo* transposase detected on Western blots are just as insoluble as the full-length protein (not shown). However, N-terminal truncations might increase solubility. Before trying to establish an *in vitro* system with the truncated transposases, however, their activity should be tested in *Drosophila* embryos for the ability to catalyse excision and transposition (as in Atkinson *et al.*, 1993; O'Brochta *et al.*, 1994; Lozovskaya *et al.*, 1996). The same approach could be used to test the activity of different fusion proteins.

3.7.5 Alternative expression systems

Among the established expression systems, *Drosophila* Schneider S2 tissue culture cells provide the most natural environment for *hobo* transposase. Expression of *P* element transposase in Schneider cells yielded active protein (Kaufman *et al.*, 1989). Expression of different tagged *hobo*-transposases from the inducible *Drosophila* metallothionein promoter had been planned with plasmids pKL58-pKL63 (Table 2.7). As a first step, it was planned to test the activity of S2-produced *hobo* transposase *in vivo* by scoring for excisions of a mini-transposons from the expression plasmid (such as pKL86). At this point, because of lack of time available, this part of the project was taken over by Nicholas Buisine (see Section 5.7.6). This system was not used from the beginning, because we had no access to a cell culture lab, and because of intellectual property rights demanded by Invitrogen.

Another alternative is the baculovirus expression system in ovarian cells of the moths *Spodoptera frugiperda* or *Trichoplusia ni*. These insects are not the natural hosts of *hobo*, but their pattern of protein processing is probably very similar to *Drosophila*. Baculovirus expression often delivers very high levels of recombinant protein. It has been used successfully for the production of active TcI transposase (Abad *et al.*, 1993) and RAG1/2 (McBlane *et al.*, 1995) among many others.

3.7.6 Conclusion

From these experiments we can conclude that the best expression of *hobo* transposase is from BL21-CodonPlus-RIL at room temperature after 30 minutes of induction. Also, the best solubility is obtained in buffers containing 0.8% Triton X-100 and 500 mM NaCl. The most highly purified protein was obtained by His-purification in this buffer.

CHAPTER 4 - DEVELOPMENT OF PCR

ASSAYS FOR HAIRPIN DETECTION

4.1 Introduction

Of the many details still unknown of the mechanism of *hobo* transposition, this work focuses on the detection of the predicted hairpin structures in the flanking DNA after *hobo* excision. Hairpin structures are the defining intermediates for the proposed *hAT* cleavage mechanism (see Section 1.3.1.1). Their detection in *hobo* transposition would finally permit us to draw firm conclusions about the order of strand transfer steps and further substantiate their evolutionary and mechanistic relationship to V(D)J recombination.

The most direct method to detect DNA hairpin structures is to analyse the potentially hairpinned DNA on a denaturing gel. If a hairpin is present, the denatured single stranded DNA is exactly twice the size of a denatured non-hairpinned molecule. Depending on the starting material, either a two-dimensional gel, with a native first and a denaturing second dimension (Roth *et al.*, 1992a), or a sequencing gel (Kennedy *et al.*, 1998; Bhasin *et al.*, 1999) is better suited. Even if the signal is enhanced by a Southern blot, however, the success of this method depends on a relatively large amount of starting material. A higher sensitivity can be achieved by amplifying any hairpin structures before electrophoresis.

In the past, ligation-mediated polymerase chain reaction (LMPCR) has been used to amplify and detect hairpinned DNA ends (Zhu and Roth, 1995). This LMPCR assay, developed for the study of the structure of coding and signal ends in V(D)J recombination, involves ligation of a double-stranded oligonucleotide adapter to the resolved hairpin end. The ligation is followed by PCR amplification using a primer that hybridises to the ligated adapter and a second primer that hybridises to the DNA end of interest (see also Roth *et al.*, 1993). Covalently sealed hairpin ends cannot be ligated to the adapter, and they are only amplified if opened by endonucleolytic cleavage prior to ligation (Figure 4.1A).

Zhu *et al.* (1995) claim that their LMPCR assay is at least 1,000-fold more sensitive than Southern blotting. Assuming that a Southern blot has an optimal sensitivity of 1 pg (Sambrook *et al.*, 1989), which translates to 2.3×10^6 molecules of the detected 2.5 kb fragment, this would translate to the LMPCR assay being able to detect 2.3×10^3 molecules

of hairpin DNA. As all experiments so far indicate that *hobo* transposase has a very low activity, a higher sensitivity is essential.

A drawback of the LMPCR is the reliance on blunt-ended ligations, which are c. 50x less efficient than sticky end ligations (personal observation). Furthermore, the blunt oligonucleotide adapters have to be dephosphorylated to prevent self-ligation, further reducing the ligation efficiency. As a result, many hairpinned DNA molecules will pass undetected.

Most importantly, this method of LMPCR relies on the complete absence of double strand breaks at the coding ends prior to endonucleolytic cleavage. If a significant proportion of the hairpin structures are resolved or broken by other means, the assay cannot prove the presence of the remaining hairpins, unless very sensitive quantitative PCR is used.

Detection of hairpins at the V(D)J coding ends by LMPCR has only been applied successfully to DNA isolated from *scid* mouse thymocytes, not from wildtype cells (Zhu and Roth, 1995). The *scid* mutation (DNA-PKcs⁻) impairs hairpin processing (see Section 1.2.3.2.1), which increases the lifetime of these intermediates and thus facilitates detection. We devised our *hobo* transposition experiments so as to limit the activity of any hairpin-opening nucleases. Nevertheless, additional hairpin-opening activities, such as by the transposase itself, cannot be excluded. A method, which allows detection of hairpin intermediates within a background of processed hairpins, is therefore essential.

Since *hobo*-mediated hairpin formation is expected to be very inefficient, we developed a very efficient assay, which can detect the hairpins themselves, as opposed to their absence in LMPCR.

4.2 Direct detection of hairpins via PCR

Two methods were used to directly amplify the hairpinned DNA template, using primers that hybridise upstream on the same molecule and that are extended through the hairpin loop.

4.2.1 Single-primer polymerase chain reaction (SPPCR)

Conceptionally, the simplest method is single-primer PCR. Here, the primer binding site is directly on the stem of the hairpin. Due to the special structure of the template, PCR amplification only requires one primer (Figure 4.1B).

Fig. 4.1

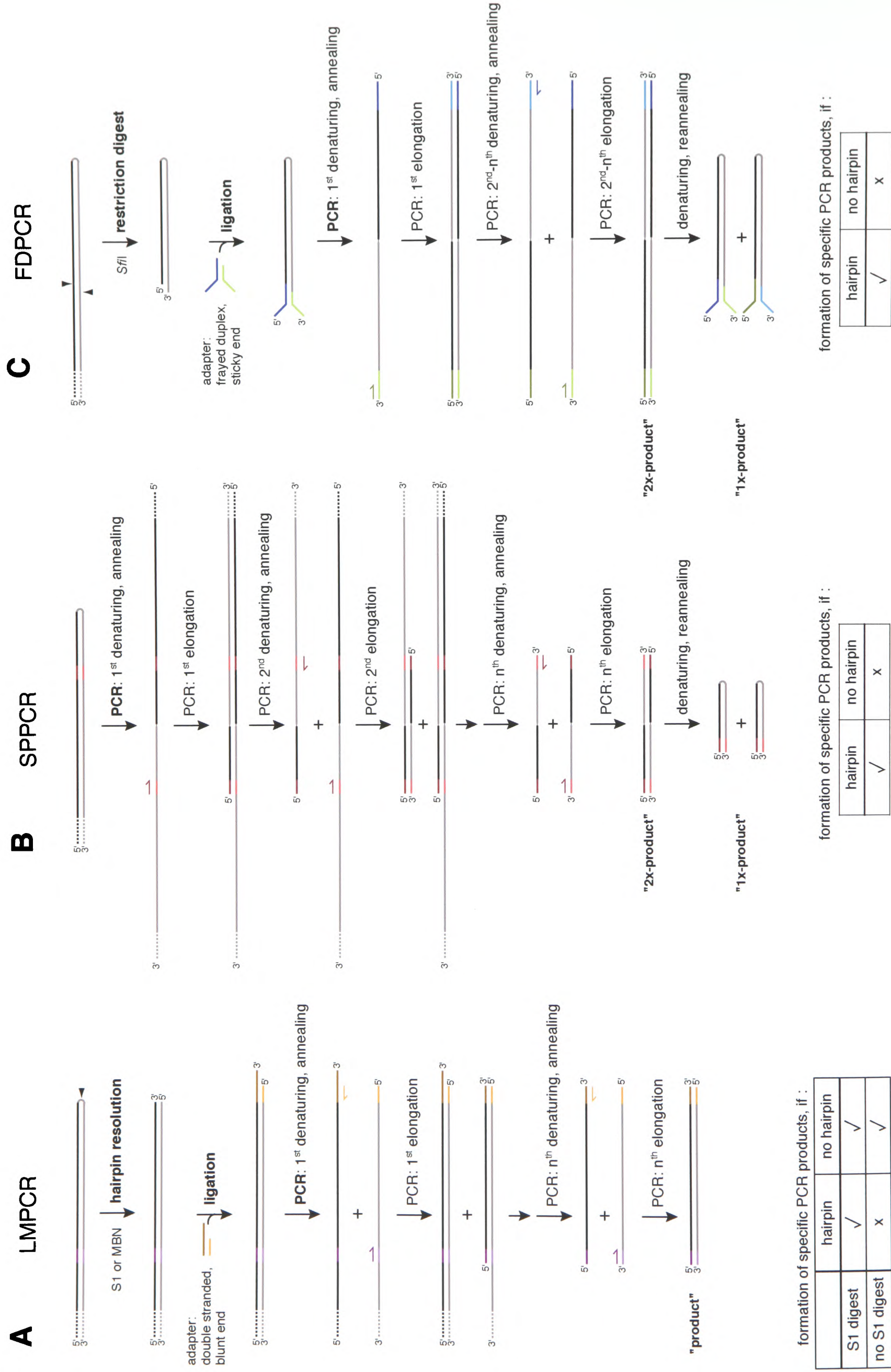


Figure 4.1: PCR techniques for the detection of hairpin structures

Schematic of reactions. Black or dark colour, top strand (5'→3'); grey or light colour, bottom strand (3'→5'); coloured, primer binding sites and/or oligonucleotide adapter; thin coloured half-arrows, PCR primers 5'→3'; black arrowheads, nuclease cleavage site.

- (A) Ligation-mediated polymerase chain reaction (LMPCR). Ligation of adapter and PCR are only successful if hairpin structures are resolved or have never been present. After ligation to the adapter, standard PCR between a template-specific primer (purple) and an adapter-specific primer (orange) amplifies a double-stranded template. (Based on Zhu and Roth, 1995.)
- (B) Single-primer polymerase chain reaction (SPPCR). The primer (auburn) binds directly to the denatured template and initiates amplification through the hairpin. This first elongation creates a double-stranded molecule in which the template sequence is joined head-to-head at the site of the former hairpin. This head-to-head template has a primer binding site for the same primer at both ends. After denaturing, a large proportion of it is converted into two identical hairpins. If the template is not hairpinned at the outset, the second primer binding site is not created, thus only linear, not exponential, amplification is possible.
- (C) Frayed-duplex polymerase chain reaction (FDPCR). The template is digested upstream of the hairpin end with a restriction enzyme that leaves a non-palindromic overhang. A frayed duplex adapter is ligated. Amplification from the first adapter-specific primer (green) through the hairpin creates a double-stranded head-to-head molecule and the binding site for the second primer (blue) in the correct orientation. During the following PCR cycles, the template can be amplified exponentially. If denatured, most of the double-stranded product is converted to hairpin conformation. (See Figure 4.2 and text for more details.)

However, this technique has a major disadvantage: The long, continuous template strand, which is self-complementary, can out-compete and displace the short primer. This reduces the availability of a template for the PCR primer drastically and decreases the efficiency of the amplification and with it the sensitivity. An additional disadvantage is that PCR reactions with only one primer species are prone to a high background of non-specific products.

One way the displacement of the primer can be minimised, is by the use of a primer with a high melting temperature. The stability of hydrogen bonds can be increased by incorporating modified bases such as 2-amino-dA or C-5-propynyl-dC, enabling the annealing step of the PCR to be performed at a temperature where the hairpin stem is preferentially unfolded. Results from low- and high-melt primers are presented in Section 4.4.2.

4.2.2 Frayed-duplex polymerase chain reaction (FDPCR)

The efficiency of primer hybridisation can be further improved by ligating a frayed-duplex structure with a stretch of non-compatible sequence to the non-hairpinned side of the hairpin stem (Figure 4.1C). This provides sites to which two distinct primers can hybridise effectively without competition from the self-complementary hairpin stem. This frayed duplex adapter is specifically designed for each transposition substrate to be complementary to an overhang created by restriction digest close to the putative hairpin end.

The specificity of the frayed-duplex PCR assay relies upon the adapter being ligated preferentially to those DNA fragments that form the hairpins. To achieve this, the choice of restriction enzyme to digest the potentially hairpinned substrate DNA is of utmost importance. To ensure efficient ligations, the digest must create DNA overhangs of maximum length. For maximum specificity, the number of restriction ends compatible to the adapter must be kept low within the background DNA present in the sample. The total number of restriction fragments is minimised by using an enzyme with a long recognition site. The number of overhangs with the same sequence as the intended ligation target is moreover reduced by choosing an enzyme that cuts outside its recognition sequence. Furthermore, if the overhang is non-palindromic, this prevents self-ligation of the adapters. Finally, the restriction site should be placed within a few hundred base pairs of the expected hairpin structure to allow efficient amplification during the following PCR.

The best commercially available enzyme to fulfil these requirements is *SfiI* (GGCCnnnn/nGGCC). On average, its interrupted 8 bp recognition sequence is expected to appear approximately every 65 kb (4^8 bp). Of these restriction sites, one in 32 (2 per 4^3) have the same 3 bp overhang in the non-specific centre. For non-palindromic overhangs, this results in a frequency of $2 \div 4^{11} = 4.8 \times 10^{-7}$. If one assumes a completely random distribution of nucleotides within the genome, this means that the *SfiI* digest of a mouse genome of 2.9×10^9 bp (NCBI, 2002) would create 1380 fragments with the correct overhang on at least one end. This gives an estimate of the level of the background. (When no *SfiI* site was available, *BglI* (GCCnnnn/nGGC) or *DraIII* (CACnnn/GTG) were used as second best alternatives with shorter recognition sequences but equivalent overhangs.)

4.3 Generation of components for the assay and control experiments

4.3.1 Adapters for FDPCR

Oligonucleotide adapters were designed and optimised with DNASTAR PrimerSelect and Amplify software packages. They have a double-stranded stem, which is non-repetitive and anneals well, and two single-stranded flares, which have no homology either to each other or to any other sequence within the hairpin, the plasmid substrate or the *E. coli* genome. The flares allow the binding of two non-repetitive primers with a similar melting temperature. The adapter stem has a non-palindromic overhang matching the overhang created by restriction digest near the putative hairpin end. To increase the ligation efficiency between adapter and target DNA, the adapter stem is 5' phosphorylated.

4.3.2 Generation of artificial hairpin templates as positive controls for FDPCR and SPPCR

In order to develop and optimise the assay, hairpin structures are required as positive controls. These were created from the DNA substrate used in transposition reactions, as outlined below in Figure 4.2 and documented in Figure 4.3.

The sequence flanking one *hobo* end in the plasmid substrate was amplified with a primer, which binds just outside the restriction site for ligation (*SfiI*), and a second primer, which amplifies from the first bases outside the inverted repeat. The second primer has a 5' extension including a *BsaI* recognition site [GGTCTC (1/5)], which is placed such that cleavage creates a 4 bp overhang, including the first two bases of the flanking DNA and a

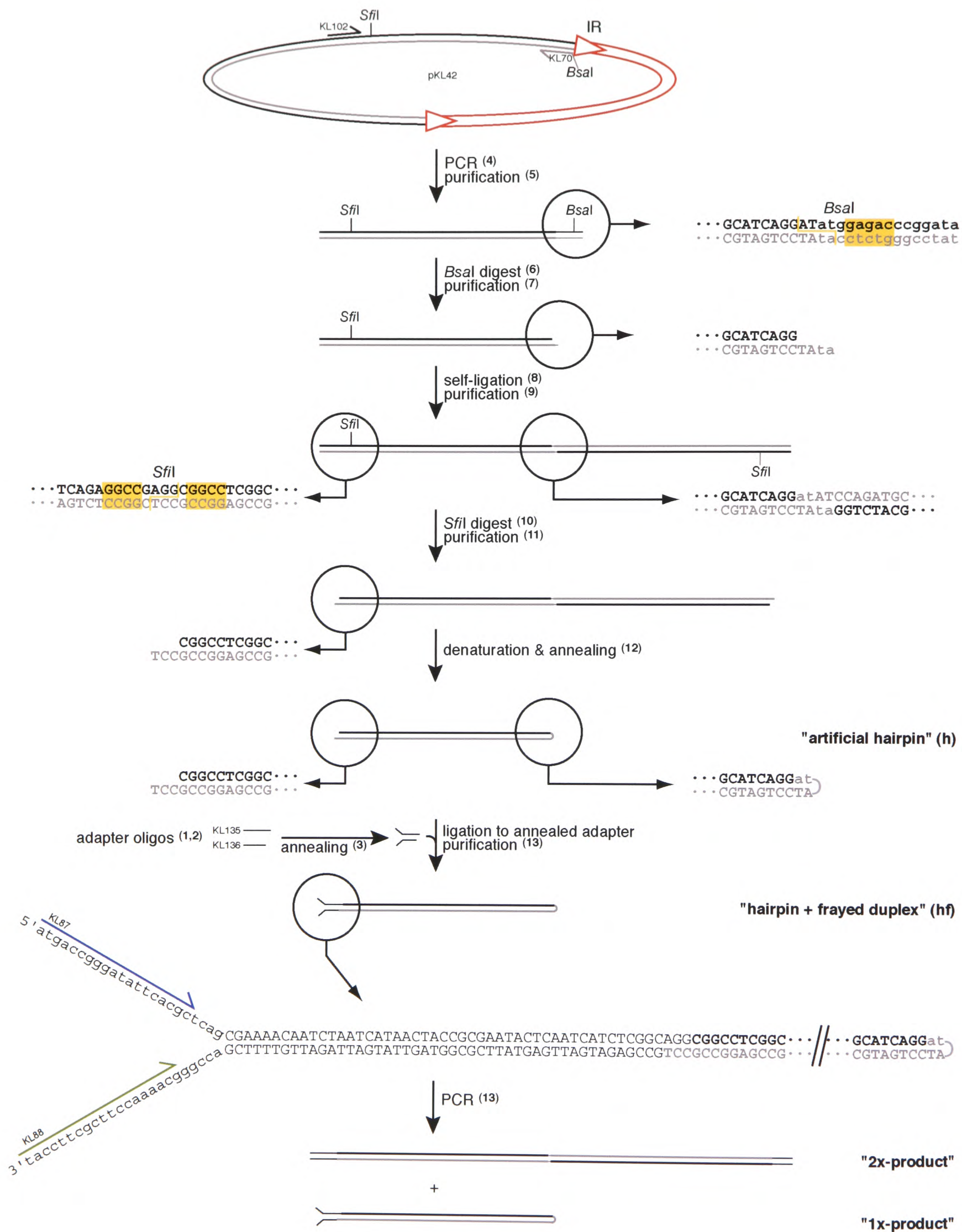


Fig. 4.2

Figure 4.2: Schematic of hairpin template generation

Lines and letters: red, transposon; bold black, flanking DNA, top strand; bold grey, flanking DNA, bottom strand; thin black, synthetic oligonucleotide adapter; blue and green half-arrows, PCR primers. Asymmetric restriction sites: yellow box, enzyme recognition site; yellow lines, cleavage position. Numbers in parentheses correspond to lane numbers in Figure 4.3A. The sequence shown is for template 42S (see also Figure 2.2).

(1-3): To anneal oligonucleotide adapters, 0.5 or 5 nmol of each PAGE purified oligonucleotide (Table 2.9, column 7) was added to a final volume of 50 μ l 1x SSC, 1x TE. They were denatured for 5 min at 99.9°C in a TGradient thermocycler and annealed by cooling to 4°C at 0.05°C per second. Excess salt was removed by drop dialysis on MilliporeVS filter disks (0.25 μ m pores) against 40 ml TE for 1–4 h. (4-5): The required flanking DNA was amplified by 30 PCR cycles with cloned *Pfu* polymerase and with two primers, as specified in Table 2.9, column 6. The products from 10-20 PCR reactions (50 μ l each) were excised from an agarose gel and purified with the QIAquick gel extraction kit. (6-7): The DNA was digested with 45 U *Bsa*I for 4 h at 50°C, followed by purification with the QIAquick PCR purification kit. (8-9): The eluted fragments were joined by overnight ligation at 16°C with 2,600 U T4 DNA ligase. The ligations were purified with the QIAquick PCR purification kit. (10-11): To create the specific sticky end for adapter ligation, the purified DNA was digested with the appropriate enzyme (e.g. *Bgl*I for 10B or *Sfi*I for 42S) and separated by overnight electrophoresis at 50 V, 4°C on a 1.5% SeaKem gold agarose gel. The major, upper band, which is twice the size of the sought hairpin, was excised and gel-purified. (12): The hairpin structures were formed by denaturing the double-stranded DNA in 1x SSC for 10 min in a boiling waterbath, and leaving them in the water while it cooled down to RT. The salt was removed by dialysing the hairpins for 1–4 h against 40 ml TE on Millipore VS filters. At each step, a 1/60 aliquot was saved for an analytical gel (Figure 4.3A). (13): 10 pmol hairpin and 100 pmol annealed adapter were ligated in 10-15 μ l with 1x T4 Ligase buffer and 200 U T4 ligase overnight at 16°C. The ligated product was separated from non-ligated adapter by agarose gel purification. After PCR, the 2x-product is the product of the final elongation, whereas the 1x-product is formed from the 2x-product by one cycle of denaturation.

palindromic repeat of these two bases. Ligation of two *BsaI*-cleaved PCR products results in a perfect head-to-head inverted repeat of the flanking DNA. When this molecule is denatured and re-annealed, a hairpin is created, which is identical in sequence and structure to the hypothetical hairpin left behind after *hobo* excision. *SfiI* cleavage (before denaturation) creates the sticky end for ligation of the adapter (Table 2.9 templates “-h”).

4.3.3 Plasmids for generating artificial hairpins directly

For later experiments (Section 4.4.8), plasmids, which allow the direct release of hairpinned and non-hairpinned control templates, were created (pKL96-pKL103, Table 2.7). The hairpin plasmids contain head-to-head ligated flanking DNA, which is released by *SfiI* digest. Thus, steps 4-9 of the previous protocol (Figure 4.2) are circumvented and large quantities of artificial hairpin templates can be produced. The non-hairpin plasmids contain one or more copies of the transposon end plus flanking DNA, which can be released by *SfiI/BsaI* digest (see Section 2.4.2). Digestion with *SfiI* creates a 3 bp 5' overhang, digestion with *BsaI* a 4 bp 3' overhang. Both overhangs are non-palindromic. This reduces the frequency of ligations of the isolated fragments to any molecule other than the adapter.

4.3.4 Frayed-duplex templates

For optimisation of the PCR conditions in FDPCR and as positive controls, the annealed adapter was pre-ligated to the artificial hairpin to form a frayed duplexed, hairpinned template (Table 2.9 templates “-hf”). For negative controls, the annealed adapter was ligated to a double-stranded restriction fragment, comprising the same flanking DNA as the hairpin plus some adjacent transposon DNA, and which is not hairpinned at the end (Table 2.9, templates “-df”).

4.4 Results and optimisation of FDPCR and SPPCR on control template

4.4.1 Analysis of FDPCR products

Initial experiments were carried out with preformed hairpins, ligated to frayed duplex adapter (templates “-hf”). After FDPCR, two major products were obtained, of 1x and 2x the size of the template (Figure 4.3A, lanes 14-17; Figure 4.3C, lane 1). Depending on the conditions, the 1x-product was c. 2-100 times more abundant than the 2x-product. Complete polymerisation of the ligated hairpin creates the double-stranded 2x-product.

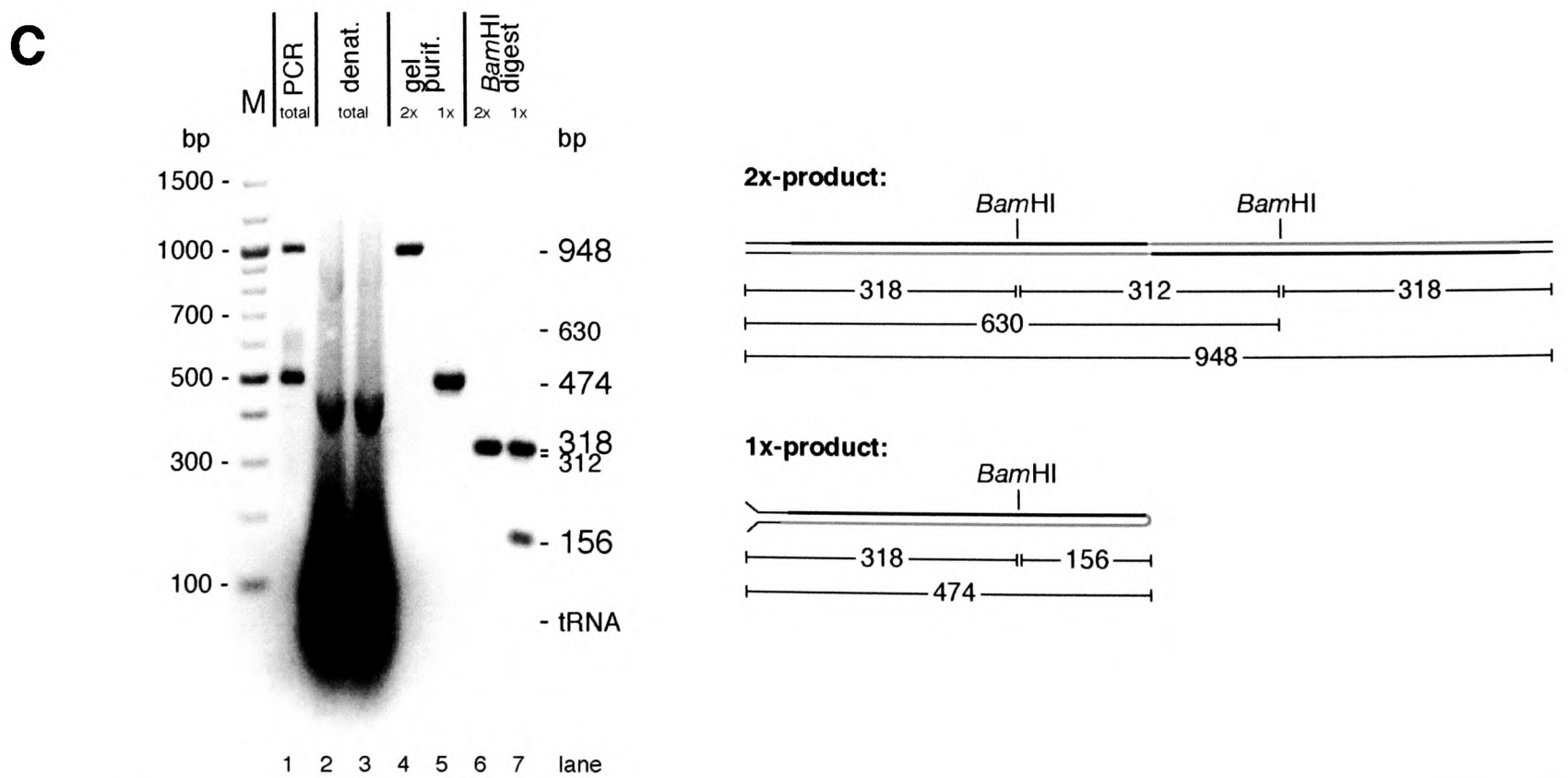
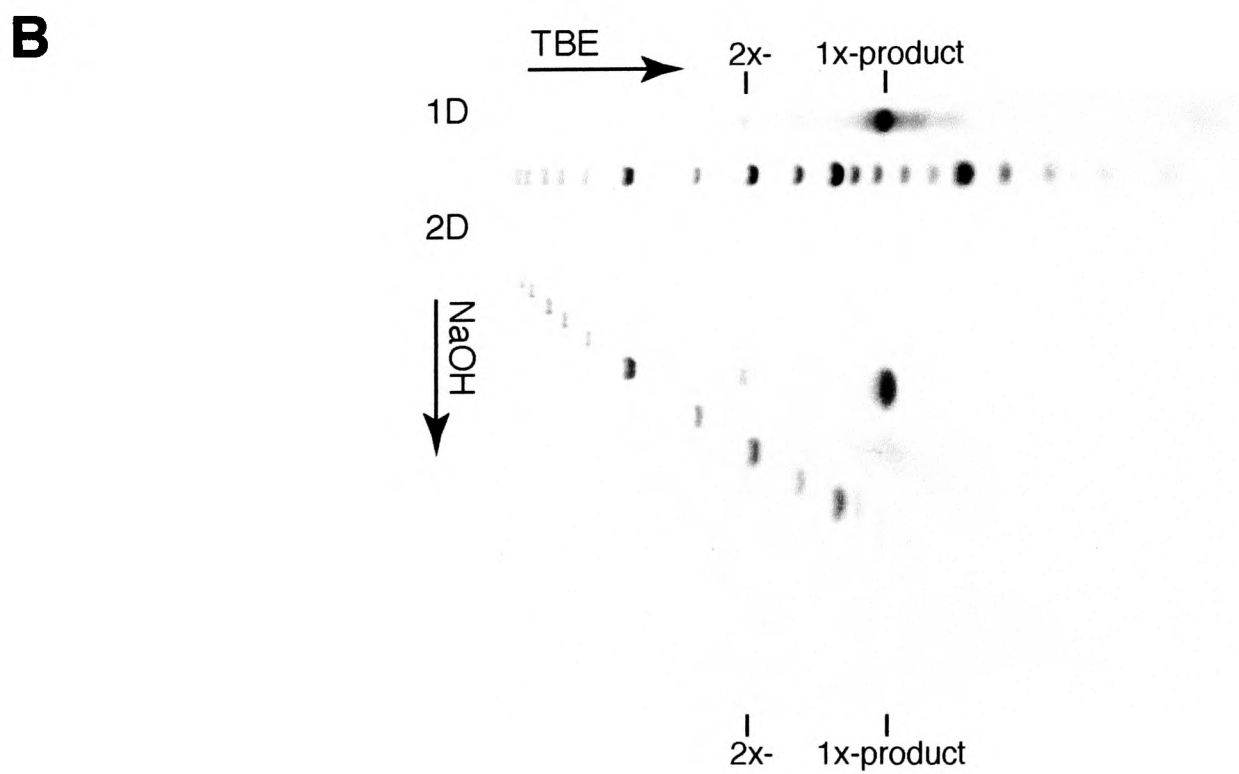
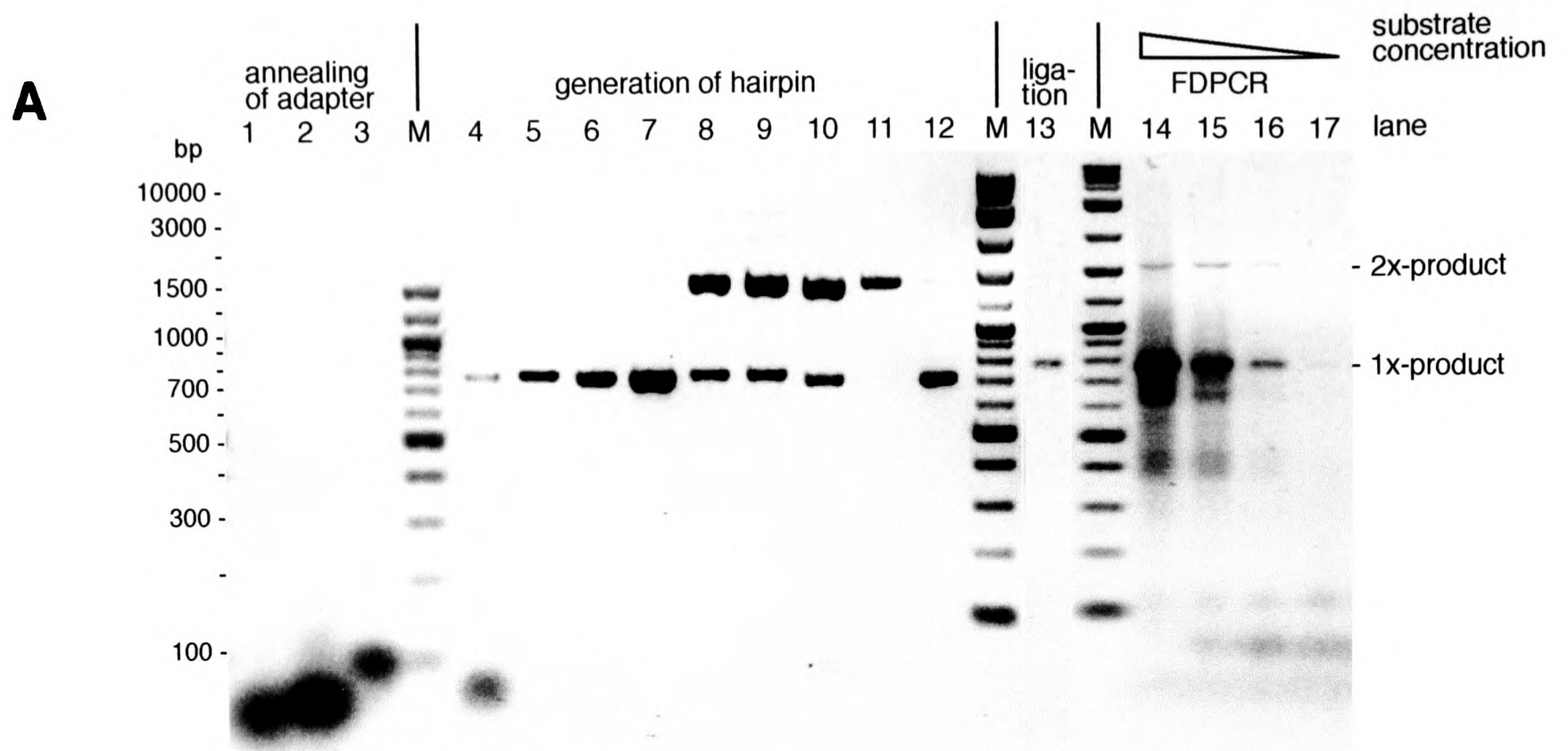


Fig. 4.3

Figure 4.3: Generation of assay components and analysis of FDPCR products

Reverse contrast images of ethidium bromide stained agarose gels.

- (A) Lanes 1-13: Generation of adapter KL135/KL136, hairpin 42S-h and adapter-ligated template 42S-hf. Lane numbers coincide with the numbers in parenthesis in Figure 4.2. Lane 1, free adapter oligo KL135; lane 2, free adapter oligo KL136; lane 3, annealed adapter; lane 4, product of PCR reaction which amplified the pKL42 flanking DNA with primers KL102 and KL70; lane 5, spin-column-purified PCR product; lane 6, *Bsal* digested PCR product; lane 7, spin-column-purified digestion product; lane 8, products of head-to-head ligation at *Bsal* overhang; lane 9, spin-column-purified ligation product; lane 10, *SfiI* digested ligation product; lane 11, gel-purified digestion product; lane 12, denatured and annealed hairpin 42S-h; lane 13, adapter-ligated hairpin 42S-hf. Lanes 14-17: Typical PCR reaction from template 42S-hf.
- (B) Two-dimensional gel of FDPCR products (template 42S-hf, primers KL87, KL88), showing that the 1x-band is hairpinned. 1 kb DNA ladder and a PCR reaction were loaded onto adjacent but one lanes of a 1.5% agarose, 1x TBE gel and run for 12 h at 30 V, RT. The upper panel shows the EtBr-stained gel after this first dimension. For the second dimension (lower panel), the gel was equilibrated in 50 mM NaOH, and run for 22 h at 25 V, 4°C in NaOH with buffer circulation. Although the PCR products are loaded in a different lane to the marker, the result is still clear. 2x-product and the background smear are linear; 1x-product is hairpinned, as it has the same low mobility as 2x-product if denatured.
- (C) Analytical denaturation and restriction digest of FDPCR products of template 10B-hf and primers KL87/KL77, confirming the expected identity of the products. M, 100 bp ladder; lane 1, FDPCR reaction; lanes 2-3, FDPCR reactions, boiled 10 min in 20 vol TEN (TE with 100 mM NaCl), cooled to RT quickly, EtOH-precipitated with tRNA and glycogen, resuspended in TE; the slightly higher mobility of the 1x- and 2x-band are probably due to entanglement during the quick cooling process; lane 4, top (2x) band of FDPCR reaction, gel purified; lane 5, bottom (1x) band of FDPCR reaction, gel purified; lane 6, *Bam*HI digest of purified top (2x) band; lane 7, *Bam*HI digest of purified bottom (1x) band. 1.5% agarose, 1x TBE gel, run for 2 h at 140 V, RT. Diagrams of expected PCR products from template 10B-hf: thick black lines, top strand of flanking DNA; bold grey lines, bottom strand of flanking DNA; thin black lines, frayed duplex adapter sequence; numbers, basepairs.

During the repeated denaturing and annealing cycles in the PCR, any product that is not used as template for the next polymerisation, will be converted into hairpinned 1x-product (Figure 4.1C and Figure 4.2).

To verify the identity of the products, PCR reactions were analysed by two-dimensional gel-electrophoresis, an example of which is shown in Figure 4.3B. As can be concluded from its position on the arc, the 2x-product is indeed linear. The shorter 1x-product behaves like a hairpin, in that it runs below the 2x-product in the native first dimension and at the same position as the 2x-product in the denaturing second dimension.

Furthermore, the two products were analysed by boiling and by restriction digest (Figure 4.3C). In accordance with their assumed identity, both products are converted to 1x-length when denatured (Figure 4.3C, lanes 2,3). Restriction digests of gel-purified products give rise to the expected fragments (Figure 4.3C, lanes 6,7 and adjacent diagrams).

4.4.2 Optimisation of SPPCR

Single-primer PCR was performed directly on preformed hairpins 42D-h and 42S-h and as nested PCR on the products of FDPCR reactions from template 42S-hf. Primers KL111 to KL115 initially yielded a large variety of highly amplified products, even in the absence of any background genomic DNA (not shown). After thorough optimisation (see Table 4.3, column 2 for result), the major products of KL111, KL113 and KL114 were of the right size for 1x-hairpins. These reactions were analysed on two-dimensional gels, similar to Figure 4.3B. Only for KL111 were some secondary structures visible, amounting to c. 5% of the total “1x” band. Southern blots of the 2D gels, hybridised with probe made from a 664 bp restriction fragment from within the flanking DNA (hairpin) sequence, confirmed a low abundance of hairpin among the KL111 products (not shown).

To counteract displacement of the primer with the second strand of the hairpin, a primer with a higher stability was designed (KL80, Table 2.5). By replacing all adenines and cytosines with modified nucleotides, the melting temperature was increased by 29°C over the natural sequence. After optimisation (result in Table 4.3, column 3), a single band of 1x-product size was obtained without background from hairpin 42S-h (Figure 4.4A). However, with the “high-melt” primer KL80, the sensitivity was a lot lower than FDPCR or nested FDPCR with primers KL108/KL109 (see Section 4.4.3, Table 4.3).

4.4.3 Optimisation of PCR for FDPCR

With two of the three pre-ligated control templates for FDPCR, 10B-hf and 42S-hf, PCR reactions were immediately successful and gave rise to the expected 1x- and 2x-products. However, initially, the sensitivity at low template concentrations was very low and the background of other signals was quite high. As high sensitivity is crucial for the success of this assay, a wide variety of parameters were optimised (Table 4.1; Table 4.2). (The third template, 42D-hf, did not give rise to any amplified products of the correct size.)

For both 10B-hf and 42S-hf templates, the efficiency of extension of the PCR product through the stem of the hairpin was found to be improved by conditions that are known to weaken basepairing, namely low magnesium concentration or high glycine betaine concentration. However, additional components to further weaken the basepairing, such as ethanol or DMSO, did not have a positive effect. The strand-displacing DNA polymerase *Tth* proved to work well. A significant improvement was achieved by using touchdown PCR, with high annealing temperatures during early cycles for high specificity of priming, and a slow, progressive decrease to lower annealing temperatures during later cycles for improved efficiency. A high concentration of amplification primers is beneficial, especially in the presence of excess adapter molecules. Diluting the template even in high concentrations of *E. coli* or mouse fibroblast DNA had very little effect, which shows that the primer binding and amplification is very specific.

The sensitivity was thereby improved to allow detection of template 10B-hf and 42S-hf to 10^{-16} M or 10^{-15} M, respectively (Figure 4.4B,C; Table 4.3, columns 4,5). The latter translates to being able to detect a concentration of 1 hairpin per 1.6 Gb of DNA or 1.8 hairpins per mouse genome (2.9 Gb).

4.4.4 Nested PCR

Due to the strong secondary structure and resulting low proportion of template molecules that get amplified each cycle, a higher number of cycles is needed for good amplification of hairpin template than of linear template. However, above 45 cycles with the same primer pair, the amplification of background surpasses that of specific product. Further amplification with a primer pair that binds inside the original one can improve both specificity and yield (nested PCR).

In the standard FDPCR assay, the amplification primers anneal to all of the frayed ends except for the last three bases before the start of the double-stranded stem. The nested

Table 4.1: Optimisation of PCR conditions for template 10B-hf with primers KL87/KL88

parameter	value or specification	signal strength of products	1x & 2x products to background	final
polymerase	<i>Tfl</i> , 0.01 U/ μ l	20%	80%	✓
	<i>Tth</i> , 0.025 U/ μ l	100%	100%	
MgCl ₂	1.5 mM	100%	100%	✓
	2.5 mM	50%	60%	
	3.5 mM	10%	30%	
<i>MasterAmp</i> PCR enhancer (glycine betaine)	0x	0%	-	
	2x	90%	100%	
	3x	90%	100%	
	4x	100%	100%	✓
dNTPs	200 μ M	100%	100%	✓
T4 DNA ligase buffer	0	100%	100%	✓
	1x	10%	80%	
primer concentration for rxn without extra adapter	0.25 μ M	50%	100%	
	0.5 μ M	100%	100%	✓
primer concentration for rxn with excess adapter	2.5 μ M	20%	100%	✓
	12.5 μ M	100%	50%	
competitor DNA	0	100%	100%	
<i>E. coli</i> , genomic	1 ng/ μ l	50%	10%	
SCGR _{II} , genomic	1.25 ng/ μ l	60%	40%	
SCGR _{II} , genomic, <i>Bgl</i> _I cut	1 ng/ μ l	50%	30%	✓
	10 ng/ μ l	45%	20%	
free adapter (to substrate)	1 : 1	80%	60%	
reaction volume	20 μ l	100%	50%	
	50 μ l	100%	100%	✓
	100 μ l	100%	100%	
denaturation temperature	92°C	100%	100%	
	93°C	100%	100%	
	94°C	100%	100%	✓
	95°C	100%	100%	
initial denaturation time	5'	100%	100%	✓
denaturation time	45'' (B) / 1' (S)	100%	100%	✓
annealing temperature	46°C	30%	10%	
	48°C	50%	20%	
	50°C	70%	30%	
	52°C	100%	50%	
	54°C	90%	50%	
with touch-down	50 - 44°C	50%	50%	
	52 - 46°C	60%	50%	
	54 - 49°C	70%	70%	
	56 - 50°C	80%	80%	
	58 - 52°C	90%	100%	
	60 - 54°C	100%	100%	
	60 - 52°C	100%	100%	✓
annealing time	45'' (B) / 1' (S)	100%	100%	✓

elongation temperature	68°C	100%	100%	✓
	69°C	80%	70%	
	70°C	60%	40%	
	71°C	50%	40%	
	72°C	40%	40%	
elongation time	2'	100%	100%	✓
final elongation time	10'	90%	100%	
	20'	100%	100%	✓
number of cycles	30	60%	100%	
	45	100%	100%	✓
	65	100%	20%	
thermocycler	Biometra TGradient (B)	100%	100%	
	Stratagene RoboCycler Gradient 96 (S)	100%	100%	✓
sensitivity: minimum substrate concentration for clear signal	w/ genomic DNA		initial: 10 ⁻¹⁰ M	final: 10 ⁻¹⁶ M
hairpin structure of products, as verified by 2D gels				✓

Optimisation of FDP-PCR conditions for template 10B-hf.

Optimisation of individual parameters was performed at otherwise constant conditions. Of each parameter, the highest signal strength or signal-to-noise ratio was noted as 100%. Percentages of other specifications or values were estimated in comparison to this optimal value. All signals were estimated by visual inspection of bands on EtBr-stained agarose gels. For touchdown PCR, the annealing temperature was reduced by 1°C every five cycles in the RoboCycler.

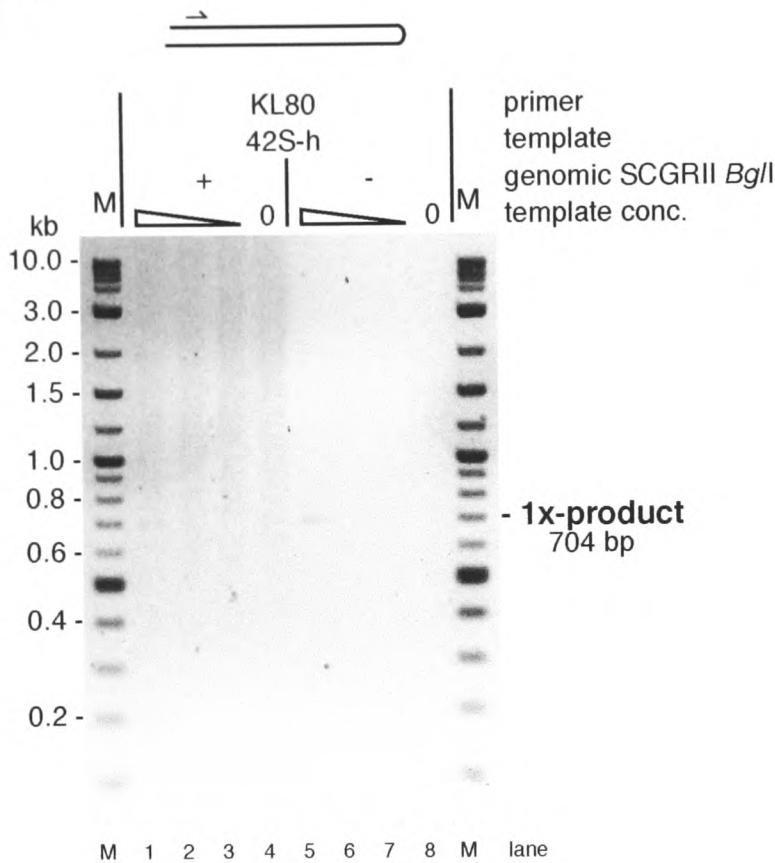
Table 4.2: Optimisation of PCR conditions for template 42S-hf with primers KL87/KL88

parameter	value or specification	signal strength of products	1x & 2x products to background	final
polymerase	<i>Tth</i> , 0.025 U/μl	100%	100%	✓
MgCl ₂	0.5 mM	1%	-	
	1.5 mM	100%	100%	✓
MgSO ₄	1.5 mM	1%	-	
MnSO ₄	0.5 mM	50%	50%	
	1.5 mM	1%	-	
	2.5 mM	100%	100%	
<i>MasterAmp</i> PCR enhancer (glycine betaine)	2x	100%	50%	
	4x	50%	100%	✓
DMSO	0	100%	100%	✓
	1%	90%	90%	
	5%	10%	40%	
	10%	0%	-	
ethanol	0	90%	100%	✓
	0.5%	100%	90%	
	1%	100%	90%	
	5%	70%	60%	
	10%	0%	-	

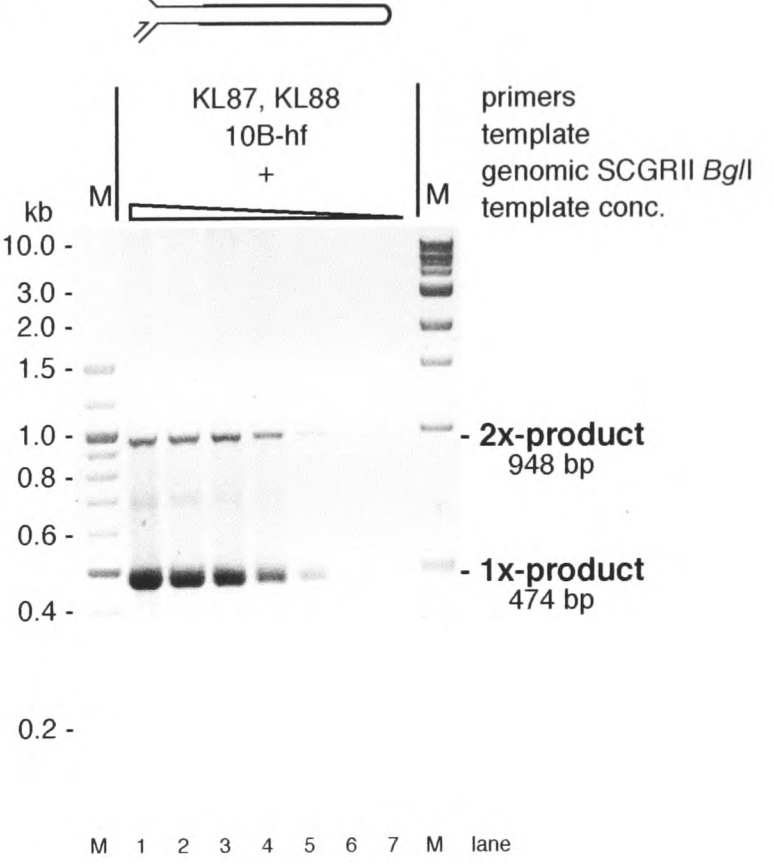
reagents	0	90%	90%	✓
BSA	4 ng/ μ l	90%	90%	
formamide	2%	100%	50%	
glycerol	2%	90%	90%	
Triton X-100	0.2%	30%	80%	
dNTPs	200 μ M	100%	100%	✓
primer concentration	0.5 μ M	100%	100%	✓
competitor DNA	0	100%	100%	
SCGRH genomic, <i>Bgl</i> I cut	20 pg/ μ l	100%	80%	
	200 pg/ μ l	100%	50%	
	2 ng/ μ l	100%	30%	
reaction volume	50 μ l	100%	100%	✓
denaturation temperature	94°C	100%	100%	✓
initial denaturation time	5'	100%	100%	✓
denaturation time	45"	100%	100%	✓
annealing temperature	48 - 39°C	10%	10%	
(with touch-down)	51 - 42°C	20%	10%	
	55 - 46°C	40%	50%	
	58 - 49°C	60%	50%	
	60 - 52°C	70%	100%	(✓)
	62 - 53°C	100%	100%	
	65 - 57°C	100%	100%	✓
	69 - 60°C	20%	80%	
	72 - 63°C	1%	-	
annealing time	45"	100%	100%	✓
elongation temperature	66°C	90%	50%	
	67°C	95%	70%	
	68°C	100%	100%	✓
	69°C	100%	100%	
	71°C	90%	90%	
	72°C	20%	80%	
	73°C	5%	80%	
	74°C	1%	80%	
elongation time	1' 40"	30%	50%	
	3'	100%	100%	✓
final elongation time	20'	100%	100%	✓
number of cycles	45	100%	100%	✓
thermocycler	Biometra TGradient	100%	100%	✓
sensitivity: minimum substrate concentration for clear signal	w/ genomic S <i>Bgl</i> I	10 ⁻¹² M		10 ⁻¹⁵ M
	w/ S10.1 <i>Bgl</i> I			10 ⁻¹⁴ M
hairpin structure of products, as verified by 2D gels				✓

Optimisation of FDPCR conditions for template 42S-hf. See Table 4.1 for description. For touchdown PCR, the annealing temperature was reduced by 0.2°C every cycle in the TGradient cyclor.

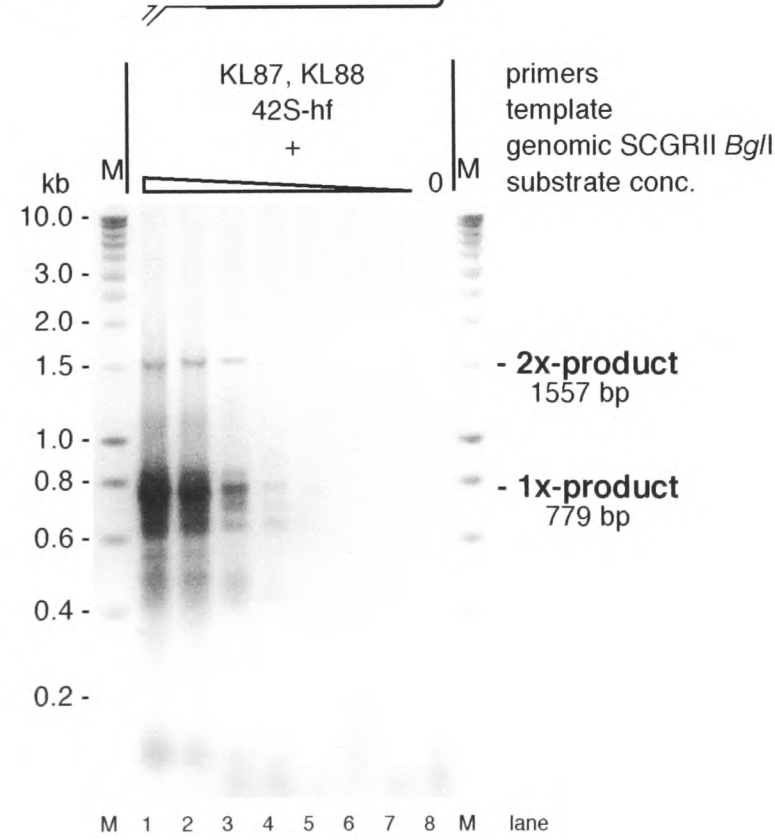
A



B



C



D

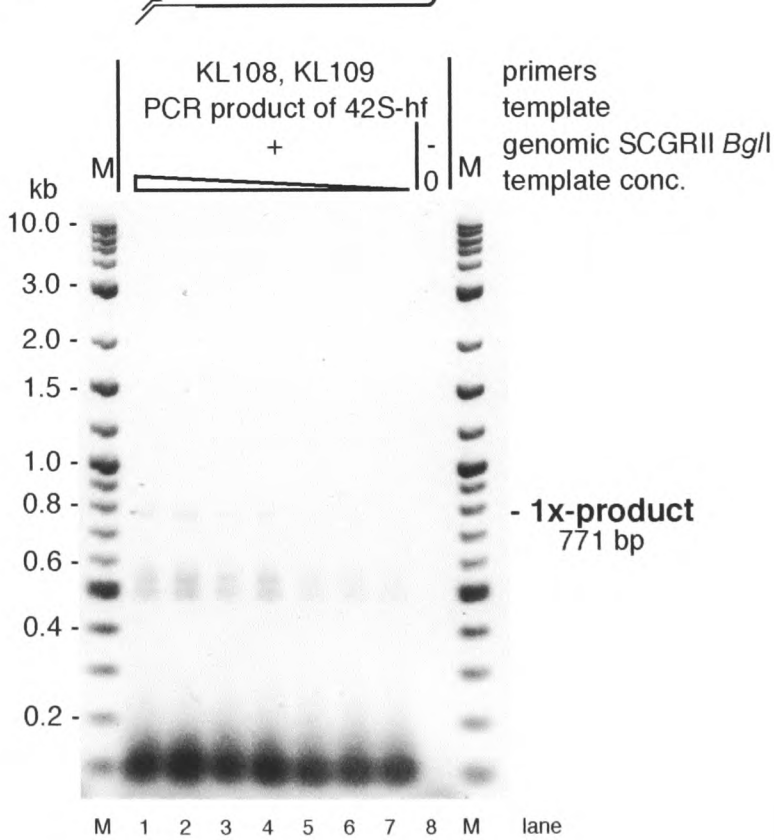


Fig. 4.4

Figure 4.4: Optimisation of FDPCR reactions

12.5 μ l of 50 μ l reactions loaded onto 1.5% agarose, 1x TBE gels. Gels were run for 2 h at 140 V, RT, stained 1 h in 1x TBE with EtBr, destained 2 h in 1x TBE.

- (A) SPPCR with high-melt primer KL80 from hairpinned template 42S-h. Genomic DNA: 100 ng per 50 μ l reaction (where indicated). Template concentrations: lanes 1,5, 10^{-9} M; lanes 2,6, 10^{-11} M; lanes 3,7, 10^{-13} M. M, 1 kb- and 100 bp ladder (NEB).
- (B) FDPCR with primers KL87 and KL88 from hairpinned template 10B-hf. Genomic DNA: 50 ng per 50 μ l reaction. Template concentrations: lane 1, 10^{-11} M; lane 2, 10^{-12} M; lane 3, 10^{-13} M; lane 4, 10^{-14} M; lane 5, 10^{-15} M; lane 6, 10^{-16} M; lane 7, 10^{-17} M. M, 100 bp ladder (NEB), 1 kb ladder (NEB).
- (C) FDPCR with primers KL87 and KL88 from hairpinned template 42S-hf. Genomic DNA: 100 ng per 50 μ l reaction. Template concentrations: lane 1, 10^{-11} M; lane 2, 10^{-12} M; lane 3, 10^{-13} M; lane 4, 10^{-14} M; lane 5, 10^{-15} M; lane 6, 10^{-16} M; lane 7, 10^{-17} M. M, 1 μ l HyperLadder™ I (Bioline).
- (D) Nested PCR with primers KL108 and KL109. Template: aliquots of FDPCR reaction from 42S-hf, panel (C), lane 5. Amount of PCR reaction (out of 50 μ l) used as template per reaction: lanes 1-2, 5 μ l; lanes 3-4, 2 μ l; lanes 5-6, 0.5 μ l; lane 7, 0.1 μ l (c. 10^{-13} M). M, 1 kb- and 100 bp ladder (NEB).

Table 4.3: Optimised SPPCR and FDPCR conditions

template	42S-h	42S-h	10B-hf	42S-hf	42S-hf(product)	96S-hf	
transposon	<i>hobo</i>	<i>hobo</i>	<i>hobo</i>	<i>hobo</i>	<i>hobo</i>	<i>Himar1</i>	
primers	KL111	KL80	KL87, KL88	KL87, KL88	KL108, KL109	KL87, KL88	
PCR method	SPPCR	SPPCR	FDPCR	FDPCR	nested FDPCR	FDPCR	
1x product	451 bp	704 bp	474 bp	779 bp	771 bp	121 bp	
<i>Tth</i> pol.	0.025 U/ μ l	0.025 U/ μ l	0.025 U/ μ l	0.025 U/ μ l	0.025 U/ μ l	0.025 U/ μ l	
dNTPs	200 μ M	200 μ M	200 μ M	200 μ M	200 μ M	200 μ M	
metal	1.5 mM MgCl ₂	1.5 mM MgSO ₄	1.5 mM MgCl ₂	1.5 mM MgCl ₂	1.5 mM MgSO ₄	3.5 mM MgCl ₂	
PCR enhancer	4x	2x	4x	4x	4x	4x	
primer conc.	2.5 μ M	0.5 μ M	0.5 / 2.5 μ M	0.5 μ M	0.5 μ M	0.5 μ M	
initial denat.	5' 94°C	5' 94°C	5' 94°C	5' 94°C	5' 94°C	15' 99.9°C	
denaturation	45'' 94°C	45'' 94°C	45'' 94°C	45'' 94°C	45'' 94°C	1' 99.9°C	
annealing	45'' 73-64°C	45'' 85-75°C	45'' 60-52°C	45'' 65-47°C	45'' 60-52°C	45'' 65-47°C	
elongation	2' 68°C	3' 68°C	2' 68°C	3' 68°C	3' 68°C	1' 68°C	
final elongat.	20' 68°C	20' 68°C	20' 68°C	20' 68°C	20' 68°C	20' 68°C	
no. of cycles	45	50	45	45	45	45	
sensitivity	(10 ⁻¹⁵ M) ^a	10 ⁻⁹ M	10 ⁻¹⁶ M	10 ⁻¹⁵ M	10 ⁻¹³ M	10 ⁻¹³ M ^b	
column	1	2	3	4	5	6	7

All reactions were in 50 μ l and 1x *Tth* buffer with ammonium sulphate. Sensitivity: lowest concentration of template that is detected; for comparison: a sensitivity of 10⁻¹⁶ M is equivalent to 30 hairpin template molecules per reaction. h, hairpin; hf, hairpin and frayed duplex adapter; ^a, <5% of 1x-product is hairpinned; ^b, in a background of 1,000x excess specific competitor 99S-df. Hairpinned templates from Table 2.9, which are not listed here, did not give rise to significant amounts of the expected products.

primer pair KL108/KL109 was therefore designed to partially overlap with the frayed ends, but to have a unique 3' end of 6-7 bp of which 3-4 bp is in the hairpin stem (Table 2.5). After optimisation, nested amplification of 42S-hf PCR products achieved a 10- to 100-fold amplification with low background (Table 4.3, column 6; Figure 4.4D). As only an aliquot of the first reaction can be taken into the second, nested reaction, this represents only a minor improvement.

4.4.5 Optimisation of ligation for FDPCR

Conditions for ligation of annealed adapter to low concentrations of preformed hairpin (1 fmol/10 μ l) in high concentrations of non-specific DNA background (up to 100 ng/10 μ l) were optimised for subsequent successful amplification by FDPCR. For both templates 10B-h and 42S-h, a relatively high ligase concentration of 4 U/ μ l and a relatively low excess of 10:1 adapter:hairpin was found to give the best results. The presence of

excess adapter required a higher amplification primer concentration of 2.5 μ M. Control ligation reactions, missing either the hairpin or the adapter, did not give rise to a detectable PCR product (not shown).

4.4.6 Size-selection of the substrate

The purpose of the FDPCR assay is the identification of hairpin structures, which are formed at a known position in the flanking DNA immediately adjacent to the transposon end. Pre-selection of template molecules for the size of the hairpinned template can therefore be used to increase the efficiency of detection. After transposon excision, followed by restriction digestion and ligation of the frayed duplex adapter, the putative hairpin template molecule exists within a large excess of other, unwanted, molecules of highly divergent size. Most notably, restriction fragments that include unreacted junctions, i.e. the same flanking DNA sequence without a hairpin, are significantly larger than the hairpinned template.

Sepharose 4B-CL spin columns and FPLC with a MonoQ anion-exchange column were found not to achieve a sufficient separation of closely spaced restriction fragments (not shown). Thus, agarose electrophoresis was used to evaluate size-selection of mixtures containing 3 μ g *Bgl*II digested genomic fibroblast DNA and decreasing amounts of adapter-ligated hairpinned template 10B-hf (474 bp). All DNA between 420 and 500 bp was excised and eluted from the gel and used as template for PCR reactions. FDPCR reactions detected down to 10^{-12} M or 10^{-15} M hairpinned template with 0.5 μ M or 2.5 μ M primers, respectively (not shown). Without size-selection, no hairpin was detected in the mixtures, even at 10^{-10} M and with high primer concentrations. Sizing the DNA before amplification is therefore clearly advantageous.

4.4.7 Effect of background DNA

During optimisation, all steps of the FDPCR protocol (ligation, FDPCR, nested PCR) were performed in parallel in a high background of non-hairpinned competitor DNA of various origins. For non-specific competition, genomic DNA from *E. coli* and from the mouse fibroblast cell line SCGR11 (see Table 2.4) was used. For specific competition, I used genomic DNA from the SCGR11-daughter cell line S10.1, which contains genomic integrations of the *hobo* reporter plasmid pKL42 (see Section 5.6.1; Table 2.4). Furthermore, a specific competitor substrate for hairpin 42S-hf was created by excising the same flanking DNA plus 203 bp of transposon end from pKL42 and ligating this double-stranded fragment to the frayed duplex adapter (42BB-df, Table 2.9; Section 4.3.3).

4.4.7.1 Effect of competitor DNA on SPPCR

When added to nested single-primer PCR (with primer KL111), even (*Bgl*I digested) DNA of non-transformed SCGR11 cells greatly increases the number of non-specific products and the level of background smear. Apparently, the primer can hybridise to two inversely oriented binding sites on several of the large DNA fragments. This adverse effect is abolished by digesting the SCGR11 or pKL42-transformed S10.1 DNA with *Rsa*I and *Sau*3AI. These four-cutters cleave the genomic DNA and pKL42 frequently, but do not cut between the left end of *hobo* and the *Sfi*I site in the flanking DNA where the adapter anneals.

4.4.7.2 Effect of competitor DNA on FDPCR

For the FDPCR reaction, any DNA that does not contain the adapter, such as genomic DNA from *E. coli*, SCGR11 fibroblasts, or (pKL42-transformed) S10.1 cells, whether digested or not, has very little effect (less than one order of magnitude; not shown). This clearly indicates that no significant mispriming of the amplification primers KL87 and KL88 occurs. The sequence of these primers had been specifically designed to have a very low probability of hybridising to any other sequence than the frayed ends of the adapter (see Section 4.3.1). A large excess of unligated adapter, on the other hand, decreases the sensitivity of the frayed duplex PCR greatly. This can be counteracted by increasing the amount of amplification primers (not shown).

Specific competition of adapter ligated to non-hairpinned flanking DNA and transposon sequence (template 42BB-df, Table 2.9) gives rise to a strong background smear throughout the entire length of the gel, which completely obscures any specific products (not shown). This effect is observed even at only equimolar concentrations of competitor and hairpinned template. Fortunately, this problem only arises if non-hairpinned flanking sequence cannot be removed before the FDPCR, e.g. by size selection. For a more detailed discussion, see Section 4.5.2.

4.4.7.3 Effect of competitor DNA on ligation

Ligation reactions of 10 nM adapter to 2 nM hairpin 10B-h were carried out in the presence of up to 100 ng *Bgl*I digested genomic SCGR11 DNA and used as templates for FDPCR. Only a 10-fold decrease in product strength and a doubling of the background was observed compared to reactions without any competitor DNA. As ligated, non-hairpinned template destroys the efficiency of FDPCR (Section 4.4.7.2), ligation to other DNA

fragments than the hairpin cannot have been efficient. This validates the choice of a restriction enzyme that creates an overhang of variable sequence (see Section 4.2.2).

4.4.8 FDPCR of short, pre-ligated templates

Many aspects of transposition reactions can be reconstituted *in vitro* with short linear DNA substrates containing the transposon end. If a frayed duplex is attached to the transposition substrate, the products could be analysed directly by FDPCR. This eliminates the need for DNA-purification, digestion and ligation and precludes the presence of any non-specific competitor DNA. Therefore, these reactions should be more specific than the full FDPCR protocol described above. Another big advantage is that it is easier to engineer templates with restriction sites at the most suitable positions.

As transposition substrates formed from annealed oligonucleotides or by PCR do not perform as well as plasmid restriction fragments in *in vitro* reactions (personal observation), plasmids were constructed from which transposition substrates can be released by restriction digest (Section 4.3.3). An additional advantage is that these plasmids allow the quick production of large amounts of substrate.

At this point, because of continuing technical difficulties with *hobo*, we became interested in whether the *Himar1 mariner* transposon creates hairpin intermediates in the flanking DNA during excision. Therefore, plasmids were created with transposition substrates for both *hobo* and *Himar1* transposase. For efficient PCR amplification of the putative hairpins, the substrates were designed to have only c. 50 bp of flanking DNA (in addition to the adapter), and c. 50 bp transposon end sequence, including the ITR-left. As positive controls, hairpinned templates with only the 50 bp flanking DNA were created (Table 2.9; Figure 4.5A).

4.4.8.1 Optimisation of FDPCR on short templates

FDPCR from the hairpinned *Himar1* template 96S-hf resulted in the familiar pattern of 1x- and 2x-products (Figure 4.5B, lane 2). Unexpectedly, the non-hairpinned control template (and putative transposition substrate) 99SB-df gave rise to a very similar pattern of products. The relative size of these (c. 100 and 200 bp) suggests that they are hairpin structures, equivalent to the 1x- and 2x-products from hairpinned templates. Such products could arise by template switching on partially denatured templates. The flanking DNA in these substrates contains a palindromic GC-rich stretch (CGGCCGCC) 20 bp upstream of the transposon end. If DNA polymerisation stops at this site and continues on the other strand

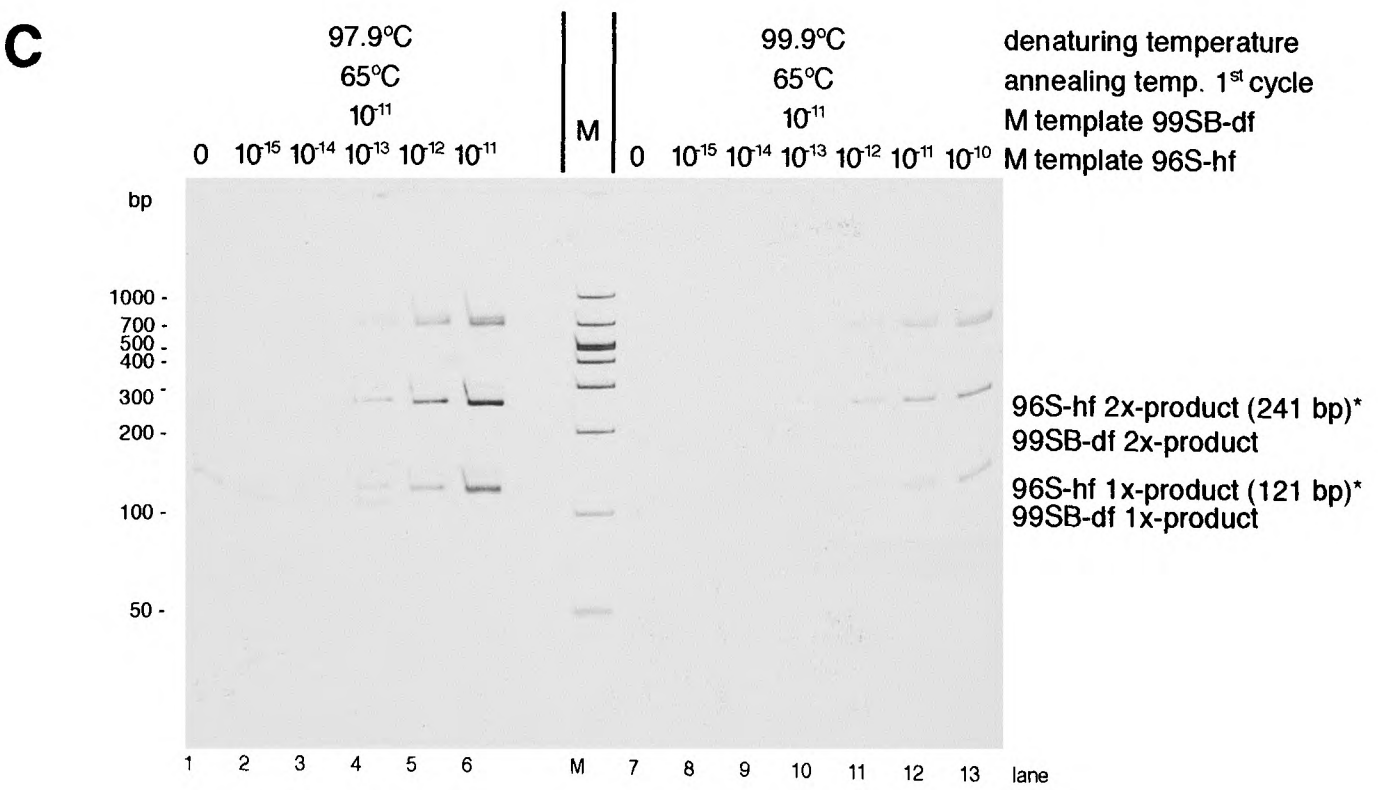
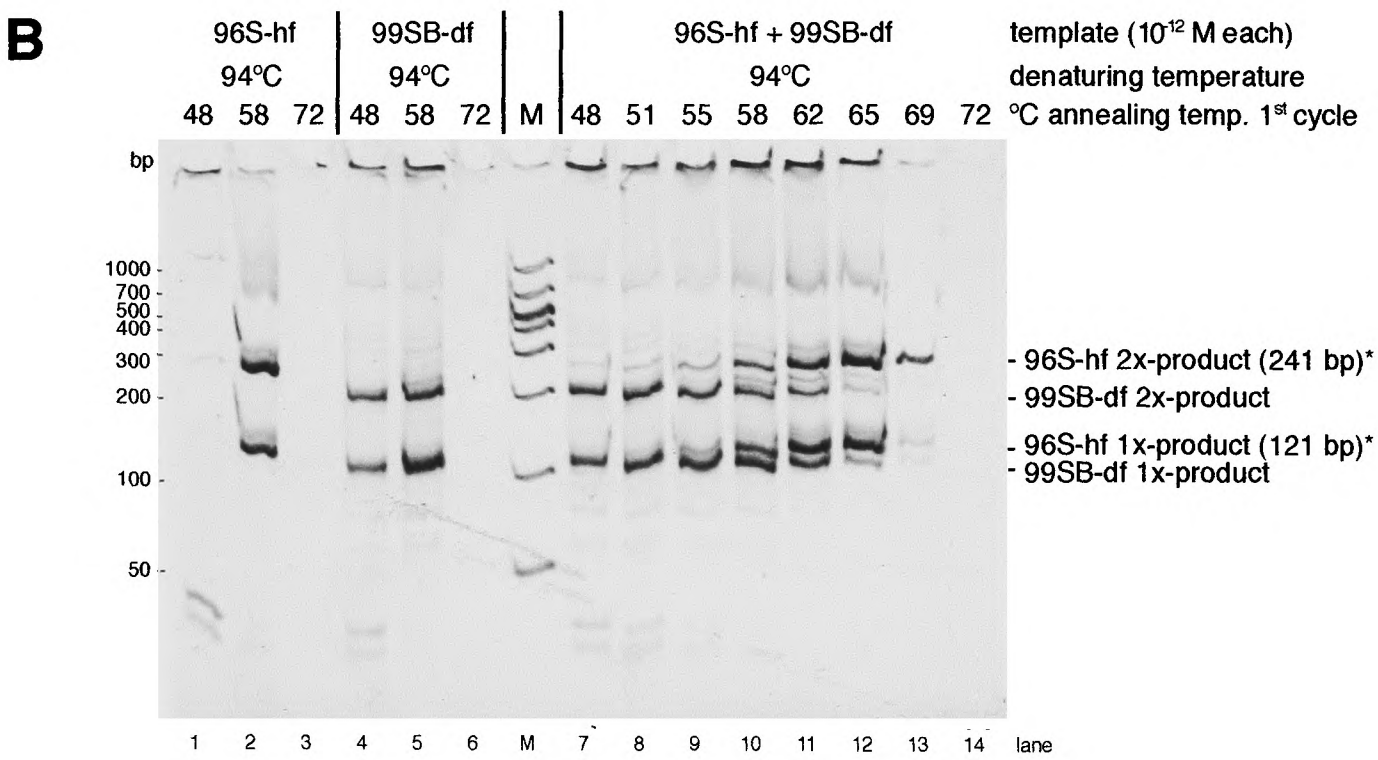
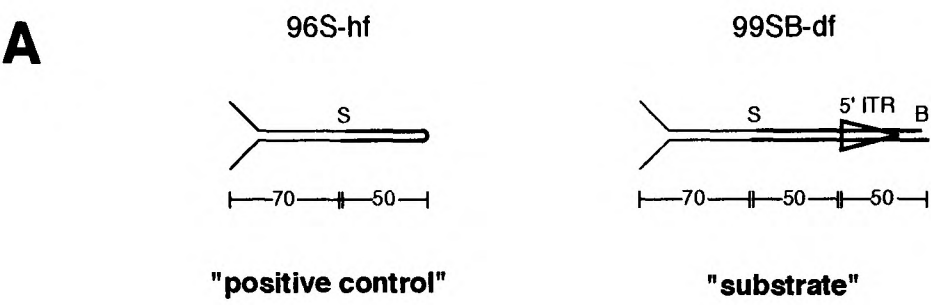


Fig. 4.5

Figure 4.5: FDPCR of short, pre-ligated templates

- (A) Schematic of templates, to scale in bp. Thin lines, frayed duplex adapter KL135/KL136, thick lines, restriction fragments from pKL96 or pKL99; triangle, *Himar1* ITR-left; S, overhang produced by *SfiI* cleavage; B, overhang produced by *BsaI* cleavage.
- (B) FDPCR reactions with pure or equal amounts of hairpinned and non-hairpinned competitor template (10^{-12} M). Annealing temperatures decrease by 0.2°C in every cycle. PCR parameters: 5 min initial denaturing at 94°C; 45 cycles of 45 s denaturing at 94°C, 45 s annealing, 1 min elongation at 68°C; 20 min final elongation at 68°C. Reaction buffer with 1.5 mM MgCl₂. M, PCR molecular weight marker 50-1000 bp (Amersham Life Science). Reverse-contrast image of EtBr-stained 12% native PAGE, run 30 min at 300 V, RT in 1x TBE. Asterisks mark the desired hairpinned 96S-hf products; the shorter bands are unwanted products from the non-hairpinned substrate 99SB-df, created by template-switching.
- (C) FDPCR reactions with constant competitor and a titration of hairpin template concentration. The annealing temperature decreases by 0.2°C in every cycle. PCR parameters: 15 min initial denaturing; 45 cycles of 1 min denaturing, 45 s annealing, 1 min elongation at 68°C; 20 min final elongation at 68°C. Reaction buffer with 1.5 mM MgCl₂. Marker, electrophoresis and labelling as in (B).

back to the frayed duplex, hairpins of the observed size will be created. Not all templates may be denatured to exactly the same basepair in each cycle. A PCR product, which has amplified one strand of a relatively completely denatured molecule can act as a primer in the next cycle and bridge the stem of an incompletely denatured molecule.

In these substrates, the GC-rich site is not part of the transposase recognition sequence. It could therefore be exchanged by re-cloning or site-directed mutagenesis of the substrate plasmid. However, sequences with strong basepairing cannot always be avoided, and incomplete denaturation of non-hairpinned templates is a general problem that could confound the sensitivity of the assay. Conditions were therefore sought to minimise this problem.

The positive signals from 96S-hf and the false-positive signals from 99SB-df are clearly distinguishable by PAGE. The PCR reactions were therefore optimised for the maximum purity of products from the hairpinned template, testing similar parameters as in Table 4.1 and Table 4.2. The results of this optimisation are shown in Table 4.3, column 7, and Figure 4.5B,C. A minor improvement was observed at increased Mg^{2+} concentration. Although it decreases the overall yield of products, the most dramatic increase in sensitivity was achieved by increasing the temperature and duration of the denaturation steps. This presumably denatures double-stranded molecules more effectively, thus preventing template-switching. The optimised conditions allow the detection of small concentrations (0.1 pM) of hairpin within a 1000-fold excess of non-hairpinned specific competitor.

4.5 Discussion

FDPCR and SPPCR were developed as novel methods to detect hairpin structures of known sequence. FDPCR was established as a very sensitive, specific and reliable assay. Ligation of an adapter to a non-palindromic overhang, created by an outside-cutting restriction enzyme, is very efficient and extremely specific. It is much more efficient and specific than the previous method of LMPCR, which relies on blunt-ended ligation. Primer hybridisation to the flared structure of the adapter is equally very efficient. Amplification of the hairpin itself is not as efficient as in a regular PCR reaction with double-stranded template, but due to its high specificity, this can be compensated by an increased number of PCR cycles. It is a lot more efficient and specific than SPPCR without flared adapters.

4.5.1 Efficiency of the PCR reaction

The efficiency of amplification in each PCR cycle can be estimated from the ratio of 2x- to 1x-product. In each cycle, virtually all 2x-product is probably converted to 1x in the denaturation phase (see Figure 4.3C, lane 2-3). Since the last cycle of polymerisation creates 2x-product, the proportion of 2x-product is equivalent to the proportion of DNA molecules present, which were successfully amplified in the last cycle. Gel purification, which includes an incubation step at 55°C, does not convert significant amounts of product into the other form ($1x \rightleftharpoons 2x$; Figure 4.3C, lane 4-5), therefore no conversion will occur during the final annealing and elongation steps.

This estimate suggests that amplification efficiency increases with decreasing template size. The 2x:1x ratio successively increases in FDPCR from templates 42S-hf (779 bp, Figure 4.4C: c. 1:10), 10B-hf (474 bp, Figure 4.4B: c. 1:3) and 96S-hf (121 bp, Figure 4.5B,C: c. 2:1).

4.5.2 Influence of competitor DNA on the specificity of FDPCR

The complete FDPCR assay was optimised in the presence of non-specific competitor DNA. Even large amounts were found to have very little effect on any reaction step, thus hairpins can be detected within a large excess of unrelated DNA. However, the presence of non-hairpinned competitor of the same sequence as the hairpin is problematic. This specific competitor can eliminate the ability of the FDPCR assay to detect any hairpinned DNA. Some more improvement might be achieved by further optimising the PCR parameters, such as denaturation time and temperature, in the presence of specific competitor. As this might not be sufficient for longer templates, any unreacted substrate molecules that cannot contain a hairpin should be removed prior to the PCR.

Size selection is suitable for the original purpose of the FDPCR assay, the detection of hairpins formed after *hobo* excision from free or genomically integrated plasmids (see Sections 5.2.1, 5.6.3.2; Figure 5.6). Digestion of unreacted pKL10 or pKL42 plasmid substrate with *Bgl*I or *Sfi*I results in ligatable restriction fragments of 4018 bp or 6215 bp (see Figure 2.2). These fragments are easily distinguishable from the 411 bp or 708 bp hairpins and can be removed by pre-sizing, as in Section 4.4.6.

For a more sophisticated method to remove unreacted substrate, magnetic beads (such as DynabeadsTM Streptavidin) can be used. With biotinylated primers that bind to the transposon end, such as KL118 and KL119 (Table 2.7), uncleaved transposon ends could be subtracted from the DNA mixture. Short linear transposition substrates could be ligated

to a biotinylated adapter at the transposon end before the transposition reaction, such that unreacted substrates can be pulled out directly. For enrichment of hairpin-containing flanking DNA, positive selection with a biotinylated primer that binds to the flanking sequence, such as KL120, could be used. Preliminary experiments of this type were not successful (not shown). With some further adjustments, however, pre-selection with magnetic beads should be a powerful method to increase the specificity and sensitivity of the FDPCR assay.

4.5.3 Applications of the assay

The FDPCR assay was developed to screen for potential hairpin structures formed during excision of *hobo* or *Himar1*. It can be used for the detection of any other DNA hairpins, provided a rare restriction site is present close by, or small dedicated substrates can be used from the start.

The assay is very sensitive and does not require radioactively labelled material. With pre-ligated substrate molecules, it is therefore a lot quicker and cheaper than the traditional methods to detect hairpin structures, namely sequencing gels and 2-dimensional gels, which require radioactive substrate or Southern blotting for the small amounts of hairpin produced. Once the system is set up for a particular substrate, it is very easy to test a large variety of conditions and components. It is therefore particularly suitable for screening for mutants and testing activities that might affect hairpin formation or -resolution. In the PCR cyclers used, up to 96 mutants can be tested in parallel. Although the PCR is not strictly quantitative, the sensitivity was found to be very reproducible (not shown), and 10-fold reductions in hairpin template concentration are clearly visible as a reduction in product (Figure 4.4; Figure 4.5).

It would be interesting as a control to try the FDPCR assay on *scid* thymocytes undergoing V(D)J recombination to see if the same results are obtained as with LMPCR, and hairpinned coding ends are found. A screen of mutagenised wt or *scid* thymocytes followed by FDPCR could help to complete the understanding of NHEJ at the coding ends (see Section 1.2.3.2). Similarly, by mutagenising either the transposase expression plasmid or the embryos for *hobo in vivo* excision experiments, it might be possible to increase the longevity of hairpins enough to be able to detect them with the FDPCR assay.

CHAPTER 5 - *HOBO* TRANSPOSITION *IN VITRO* AND *IN VIVO*

5.1 Introduction

In this work, several methods for reconstituting *hobo* transposition were attempted both *in vitro* and *in vivo*. The focus in both cases was first to detect activity and then more specifically the detection of hairpin intermediates.

In vitro excision reactions with purified transposase were initially assessed by ethidium bromide stained agarose electrophoresis. To achieve a higher sensitivity and to examine hairpin intermediates, these reactions were also analysed by FDPCR and Southern blotting. The DNA-binding capacity of the transposase was assayed in band shift experiments. *In vitro* insertions of short labelled transposon ends into supercoiled target plasmid were tested. Finally, *in vitro* plasmid-to-plasmid transposition was assayed genetically by transformation into *E. coli*. Both 6xHis-*hobo* transposase, purified from *E. coli*, and wt *hobo* transposase gained from an *in vitro* expression system, were used.

In vivo transposition and mating out assays were performed in *E. coli*. *In vivo* excision in *E. coli* was assayed by electrophoresis. *scid* mutant mouse fibroblasts, which were stably transfected with *hobo* transposon DNA, were assayed for transposase-mediated excision events. DNA extracted from these cells was examined for the presence of DNA hairpin intermediates by two-dimensional gel electrophoresis and by FDPCR and Southern blotting.

5.2 *hobo in vitro* experiments with purified 6xHis-transposase

5.2.1 Analysis of *hobo in vitro* reactions by gel electrophoresis and FDPCR

In vitro plasmid substrate reactions were carried out with the mini-*hobo*-plasmid pKL10 (see Figure 2.2A; Section 2.2.2.3; Table 2.7). This plasmid was incubated with purified 6xHis-*hobo* transposase in a variety of conditions (Section 2.5.1.1). One third of the volume of the *in vitro* reactions was analysed by agarose gel-electrophoresis and SYBR

Gold staining (Figure 5.1A). In all lanes, there was a band at the approximate size of free transposon. However, compared to untreated pKL10, no changes were detectable, i.e. there was no detectable excision or transposition activity.

For increased sensitivity, the remaining two-thirds of the *in vitro* reactions were analysed by FDPCR for the presence of hairpin in the right flanking DNA (Section 2.4.3; Figure 5.1B). In 18 out of 24 reactions, a faint product of just below 500 bp was obtained. This would correspond well to the expected 474 bp 1x-hairpinned product (see Table 2.9). The same (474 bp) product was obtained in the positive ligation control, containing digested pKL10 and pre-formed hairpin 10B-h (reaction 25), but not in negative controls without transposase or pre-formed hairpin (reaction 26). The negative PCR control, with genomic SCGR11 DNA as the only DNA (reaction 28), was also blank at this position.

To increase the signal, the putative 1x 474 bp band from four amplified transposition reactions, and one positive and negative control each (denoted below the gel in Figure 5.1B), were excised and eluted from a preparative gel. This DNA was subjected to another round of amplification with the same primers (Figure 5.2A), resulting in more clearly detectable bands of a size just below 500 bp (lanes 1-6). PCR controls with pre-ligated hairpin template 10B-hf in a high background of genomic DNA gave rise to very clear 1x- and 2x-product bands (lanes 9-10). The negative controls with only untransformed SCGR11 DNA present were blank (lanes 7-8).

In order to establish whether the PCR products are of the correct sequence to be hairpins of the right flanking DNA, a Southern blot was performed from this gel (Figure 5.2B). Clear signals were obtained, but the centre of the signals was located below the position of the expected hairpin products. It was therefore concluded that the PCR products obtained could not have come from hairpins in the pKL10 DNA after excision of the right *hobo* ITR.

5.2.2 Gel retardation

Band shift experiments were carried out to detect specific binding of the purified 6xHis-*hobo* transposase to radioactively labelled *hobo* ends, made from annealed, PAGE-purified oligonucleotides (KL49/KL50 and KL51/KL52; Table 2.10). Systematic testing of numerous binding conditions, buffer components and concentrations has produced several shifts, but none was specific to both the protein and the DNA. The eukaryotic DNA-binding protein HMGB1, which was tested as a possible cofactor, was found to bind to the *hobo* ends alone, but not in complex with the transposase (not shown).

Figure 5.1: *hobo* *in vitro* excision and FDPCR

In vitro excision reactions of a mini-*hobo* from pKL10 with 6xHis-*hobo* transposase, purified from *E. coli*. See Section 2.5.1.1 and text for details.

- (A) Direct gel-electrophoresis of reactions, SYBR Gold stained 1% agarose, 1x TBE gel. M, 100 bp and 1 kb ladders (NEB); P, pKL10, unreacted.
- (B) FDPCR of reactions, ligation- and PCR controls for amplification of hairpin 10B-h. SYBR Gold stained 1.5% agarose, 1x TBE gel. M, 100 bp and 1 kb ladder (NEB); p, positive control; n, negative control.

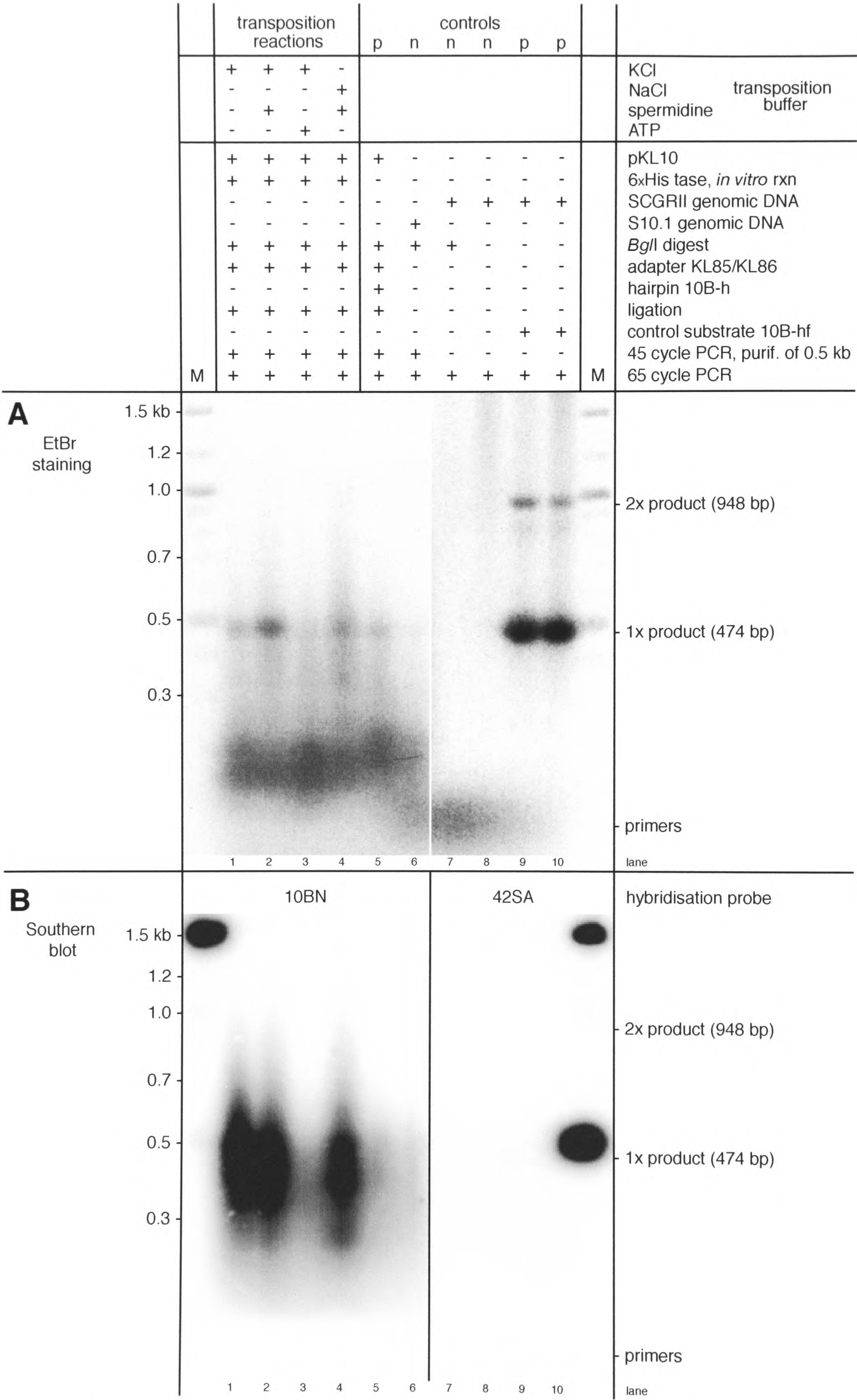


Fig. 5.2

Figure 5.2: Second PCR after *hobo* *in vitro* excision

- (A)** PCR amplification with primers KL87/KL88 of excised 480 bp FDPCR products from Figure 5.1B. 1.5% agarose gel, ethidium bromide stained.
- (B)** Southern blot of the gel in (A). M, 100 bp and 1 kb-ladder, NEB; p, positive control; n, negative control; 10BN, 276 bp *Bgl*I/*Nhe*I fragment of pKL10, flanking the *hobo* ITR-right; 42SA, 665 bp *Sfi*I/*Apa*I fragment of pKL42, flanking the *hobo* ITR-left; primers, KL87 and KL88, for amplification of frayed duplex.

5.3 *In vitro* experiments with *in vitro* translated transposase

The transposition activity of *hobo* transposase expressed *in vitro* in rabbit reticulocyte lysate (Section 3.6) was tested in two systems. (1) *In vitro* insertion of labelled transposon ends (102SX*; Table 2.10) into supercoiled plasmid (see Section 6.3.3 for a description of the system as it was applied to *HimarI* transposition). (2) *In vitro* plasmid-to-plasmid transposition of a *Kan^R*-marked *hobo* mini-transposon. Here, the donor-plasmid had a temperature sensitive origin of replication and a gene conveying streptomycin-sensitivity (*rpsL*) in the backbone (pKL93, Table 2.7); the acceptor was tetrameric pBluescript (*Amp^R*). The reactions were transformed into *E. coli* DH5 α (*Str^R*) and selected on Amp Kan Str at 42°C. Both types of experiments were performed in the presence or absence of *D. melanogaster* embryonic nuclear extract.

In both systems, the same high level of recombination was obtained with reticulocyte extracts whether or not the *hobo* transposase was expressed. This can be explained by nucleases and ligases of the extract with higher activities than the transposase. It confounded the analysis of any specific transposition activity, and these experiments were therefore not carried any further.

5.4 Mating out assay in *E. coli*

An *in vivo* mating out assay was performed to determine if *hobo* transposase was active in *E. coli*, and to see whether native and His-tagged transposase behave in the same way. If successful, an established *in vivo* transposition system would be very useful for many purposes such as screening for transposase mutants.

5.4.1 Components of the different mating out assays

The mating out assay detects transposition of a mini-transposon with a drug marker onto a conjugative plasmid. After conjugation to a different strain, successful transposition events are detected by selecting for the recipient strain, the conjugative plasmid and the mini-transposon (Figure 5.3).

The components of the mating out assay are (1) a streptomycin sensitive donor strain, which contains: (2) a pBluescript-based vector with a *Kan^R*-marked mini-transposon, (3) a pACYC based T7 transposase expression vector, and (4) a conjugative plasmid to act as a

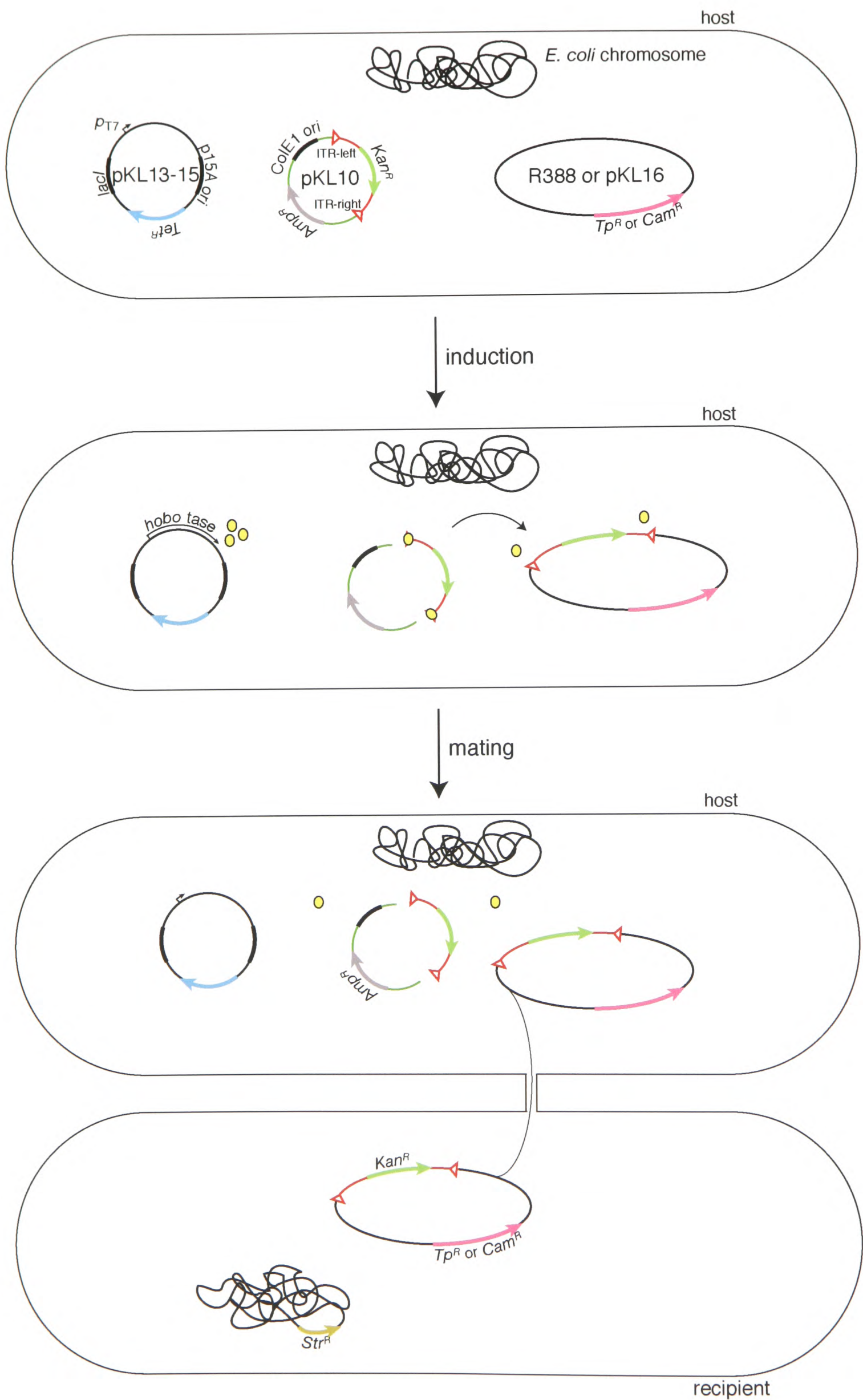


Fig. 5.3

Figure 5.3: Mating out assay

(Schematic representation.) Rationale: After induction, the transposase (tase) excises the Kan-resistant mini-transposon from pKL10 and inserts it into the conjugative plasmid (R388 or pKL16 = pOX38-*Cam^R*). This plasmid is then mated out to a Str-resistant recipient for detection. (Not to scale.)

target; and (5) a *Str^R* recipient strain to assay for *Kan^R* hops onto the conjugative plasmid.

5.4.1.1 Donor and recipient strains

As the transposase was expressed from a T7-promoter, the donor bacteria must contain the gene for the T7 polymerase (lambda DE3 lysogen). Three strains were used in parallel: BL21(DE3), an *E. coli* B strain, which is protease deficient; MM294(DE3), a *recA⁺* *E. coli* K12 strain; and HMS174(DE3), a K12 strain, which is deficient for *recA*. The recipient was AB1157, which has a chromosomal streptomycin resistance.

5.4.1.2 Construction of transposon and expression plasmids

Mini-*hobo* transposons were created with 300 bp transposon ends flanking a *Kan^R* marker in a high-copy *Amp^R*-marked backbone (see Section 2.2.2.3; Table 2.7; Figure 2.2A). pKL10 contains a left and a right *hobo* end, whereas pKL11 contains two left ends.

Since the existing expression plasmids, pKL7 to pKL9, contain the same origin of replication and selection marker as the mini-transposon-plasmids, they are not suitable for creating stable double-transformants. The expression cassettes were therefore cloned into pACYC184 using annealed oligonucleotide adapters between non-compatible overhangs to generate pKL13 (native transposase in pACYC) and pKL15 (10xHis-transposase in pACYC) (Table 2.7).

5.4.1.3 Selection of conjugative plasmids

As target DNA, the conjugative plasmids R388 and pOX38 were used. R388 encodes a trimethoprim resistance. pOX38, which is an F plasmid derivative and forms more stable pili, has no resistance marker. It was therefore marked with a chloramphenicol resistance by transposition of a mini-Tn10-Cam^R from pNK2884 (Table 2.8, Section 2.6.1). The marked pOX38 was recovered by conjugation into strain AB1157 and selection on streptomycin and chloramphenicol.

Since the mini-Tn10 carrying the Cam-resistance could in principle insert anywhere on pOX38, mating efficiencies of eight independent insertions were determined to ensure that no mating-specific genes had been destroyed. To this end, the marked pOX38 plasmids were mated into four different strains, BL21(DE3), ER2566, HMS174(DE3) and MM294(DE3). Of the eight Cam resistant pOX38 plasmids, the one with the highest mating efficiency was named pKL16 and used for all following experiments (Table 2.7).

5.4.2 *In vivo* transposition and mating out

Bacteria containing expression-, donor-, and acceptor-plasmids were grown to $OD_{600} = 0.1$ and induced for at least 4 h with IPTG. Transposition events onto the conjugative acceptor plasmid were detected by mating it out to a streptomycin-resistant recipient strain (AB1157). 10^8 host cells were mated to 10^9 recipients, as determined by previous OD_{600} -to-colony number correlations. The mating bacteria were disrupted by vortexing after 60 min and dilutions plated on selective media (modified after Morisato *et al.*, 1983; Roberts *et al.*, 1985).

To determine the mating efficiency, dilutions of the mating mixture were plated on Str-Tp plates (for R388) or Str-Cam plates (for pKL16). Here, colonies arise from all recipients that received a conjugative plasmid, independent of transposition events. Their number divided by the total number of donor cells in the plated dilution gives the mating efficiency.

Colonies on Str-Kan-Tp- or Str-Kan-Cam-plates arise from recipients of a conjugative plasmid onto which the mini-transposon has transposed or has otherwise recombined. Control cultures without transposase-expression vector (i.e. pKL10 alone) indicate the proportion of background recombination.

A single recombination event between donor and target plasmids would give Amp^R transconjugants. Transconjugants, which are Kan^R and Amp^S , are either transpositions or the products of a double crossover between donor and target. The Str^R - Kan^R - Tp^R -colonies from the mating of cultures with *hobo* mini-transposon and transposase were therefore restreaked on ampicillin and, if sensitive, stored as glycerol stocks.

The set-up and results of the *in vivo* transposition and mating out experiments are described in Table 5.1. The mating efficiency was found to be 100% if both strains were streaked out together on a plate for mating, rather than being shaken together in liquid culture (Table 5.1, experiments #2-3). In liquid culture, pKL16 was found to mate better than R388, which is consistent with the pOX38-derivative forming a more stable pilus (experiments #1,4,5). Control-cultures with Tn10 transposase expression and mini-transposon plasmids gave rise to up to several million transposition events, showing that the mating-out assay was working well. However, from 17 independent *hobo* transposition matings, a total of only 39 Str^R - Kan^R - Amp^S -colonies were recovered. All of these were from experiments with R388 as the target, and six events were from the no-transposase control experiments.

Table 5.1: Mating out assay

Experiment #		1	2	3	4	5	6	
STRAINS	hosts	BL21(DE3) HMS174(DE3) MM294(DE3)	HMS174(DE3)	HMS174(DE3)	HMS174(DE3) MM294(DE3)	HMS174(DE3) MM294(DE3)	HMS174(DE3)	
	acceptor	AB1157	AB1157	AB1157	AB1157	AB1157	AB1157	
PLASMIDS	<i>hobo</i> donor tase	pKL10	pKL10	pKL10	pKL10	pKL10	pKL56	
		pKL13 pKL15	pKL13	pKL13	pKL13 pKL15	pKL13 pKL15		
	<i>Tn10</i> donor tase	pNK3287	pNK3287	pNK3287	pNK81	n.d.	n.d.	
		pRC62	pRC62	pRC62				
	target	R388	R388	R388	pKL16	pKL16	pKL16	
CONDITIONS	growth	culture	flask (shaker)	flask (shaker)	flask (shaker)	test tube (slow roller)	test tube (slow roller)	flask (shaker)
		T	37°C	37°C	37°C	37°C	37°C	37°C
		to OD ₆₀₀	0.1	0.1	0.1	0.125	0.1	0.1
	induction	culture	flask (shaker)	flask (shaker)	flask (shaker)	plate (shelf)	flask (shaker)	flask (shaker)
		time, T	5 or 12h 30°C	4h 30°C	12h 30°C	12h 37°C	4h 30°C	4h 30°C
	mating	culture	test tube (slow roller)	plate (shelf)	plate (shelf)	test tube (slow roller)	test tube (slow roller)	plate (shelf)
time, T		3h 37°C	3h 37°C	3h 37°C	1h 37°C	1h 37°C	3h 37°C	
RESULTS (sum)	mating efficiency		c. 1%	100%	100%	c. 50%	c. 50%	n.d.
	<i>Str^R Kan^R Amp^S</i> recombinants	<i>hobo</i> + tase	1	36	2	0	0	0
		<i>hobo</i>	0	6	0	0	n.d.	n.d.
		<i>Tn10</i> + tase	c. 2000	>10 ⁶	>10 ⁶	142	n.d.	n.d.

Plasmids: see Table 2.7, Figure 2.2A, Table 2.8. Growth and mating: flask = in 250 ml Erlenmeyer flask, shaking at c. 100 rpm; roller = in 25 ml test tube, rolling at c. 3 rpm. Induction with 1 mM IPTG. Results: mating efficiency calculated from 10⁸ host cells.

5.4.3 Analysis of recombinant plasmids

For a molecular analysis of the recombination products, the original R388 plasmid and seven insertions were isolated. Direct sequencing of three insertions with primers binding within the transposon ends (KL17, KL18, KL20 and KL21) did not give readable sequence despite several trials of automated sequencing by two sequencing facilities and manual sequencing with two protocols.

Comparative restriction mapping revealed that all seven analysed insertions contain the full 1.5 kb-*Mlu*I-fragment of the Kan-resistance gene within the mini-transposon, but

they do not contain the pKL10-flanking DNA (Figure 5.4A). This suggested recombination close to the transposon ends and possibly transposition. A detailed analysis of the insertions was therefore continued.

Amplification of the insertion junctions was attempted by arbitrary PCR with one primer specific to the transposon sequence and the second primer with 10 bp of degenerate sequence to bind to the unknown target sequence; followed by a second round with nested primers (see Figure 5.4B for a more detailed description). With a first set of primers, only internal transposon sequence was amplified. With a second set of primers, carefully designed to avoid internal binding of the degenerate primers (see Appendix A.4), no specific products were detected.

Subcloning of the insertions was achieved after partial *Hpa*II digest of the recombinant plasmids. Sequencing and DNA alignments established that none of them were *bona fide hobo* insertions.

5.4.4 Detection of *in vivo* excisions by minipreps

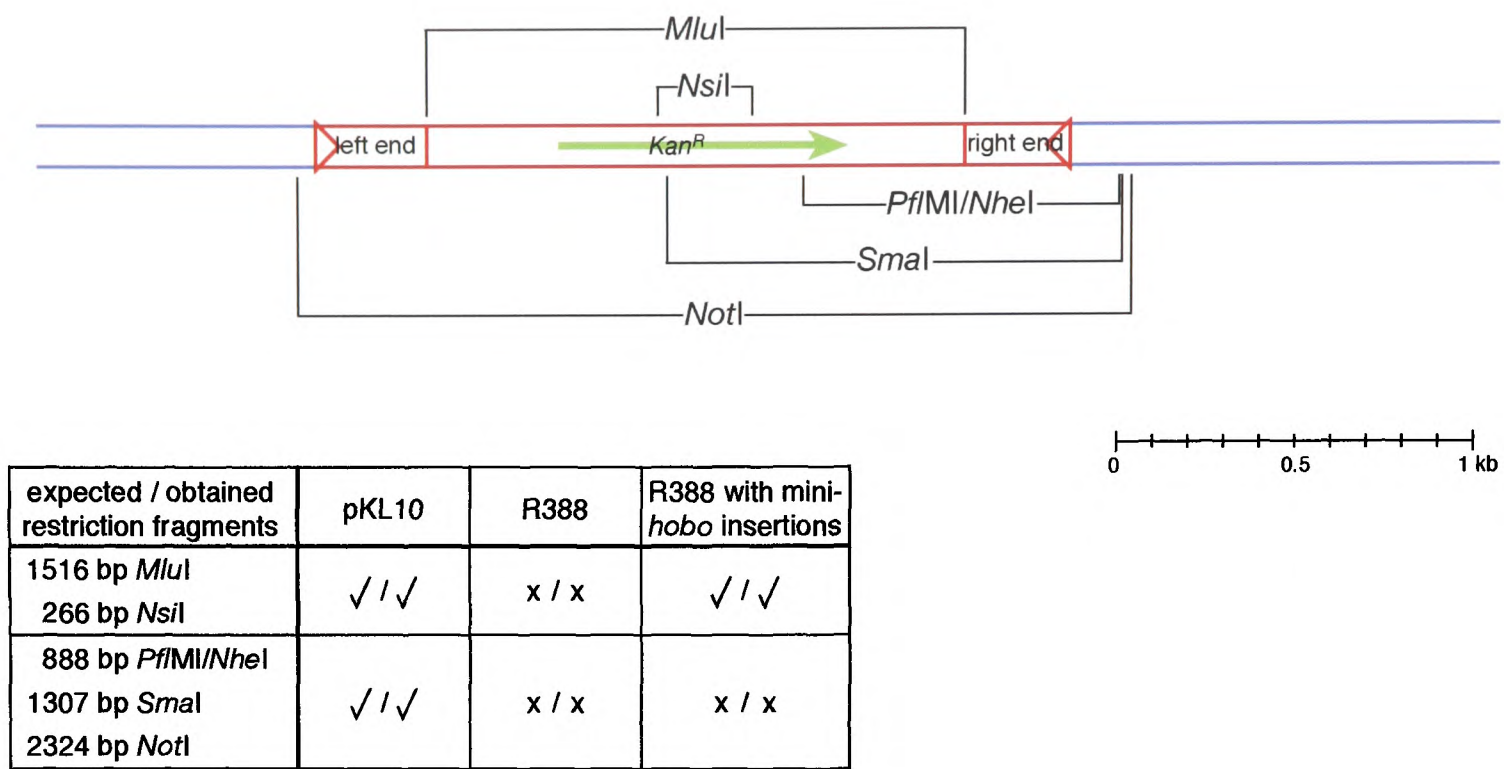
The experiments described above were designed to detect the complete transposition reaction. It is possible, however, that partial transposition reactions were taking place. An experiment was therefore carried out to detect excision events.

MM294(DE3) and BL21(DE3) were transformed with both pKL10 (mini-*hobo*) and pKL13 (native transposase), grown to $OD_{600} = 0.2-0.5$ in LB-Amp-Tet and induced with 1 mM IPTG. Aliquots of the culture were taken after 30 min to 4 h and overnight induction and the plasmid DNA was isolated by boiling preps (modified from Clewell and Helinski, 1970). The boiling preps were run alongside undigested pKL10 and pKL13 on an agarose gel. The sample prepared from the MM294(DE3) double transformant after overnight induction contained an extra band, running at c. 1.8 kb. This is too short for being the linear mini-transposon, which has a length of 2.1 kb. Restriction digests with enzymes which cut within the pKL10 mini-transposon did not alter the size of the 1.8 kb band. On a Southern blot, the extra band did not hybridise with a probe made from the *Kan^R* gene from within the mini-transposon. Thus, no excised mini-transposon was detected in this experiment.

5.5 Toxicity of *hobo* transposase

Low expression (Chapter 3) and undetectable activity (Sections 5.2, 5.4) of the *hobo* transposase could be a result of toxicity of the overexpressed protein to *E. coli*. This

A



B

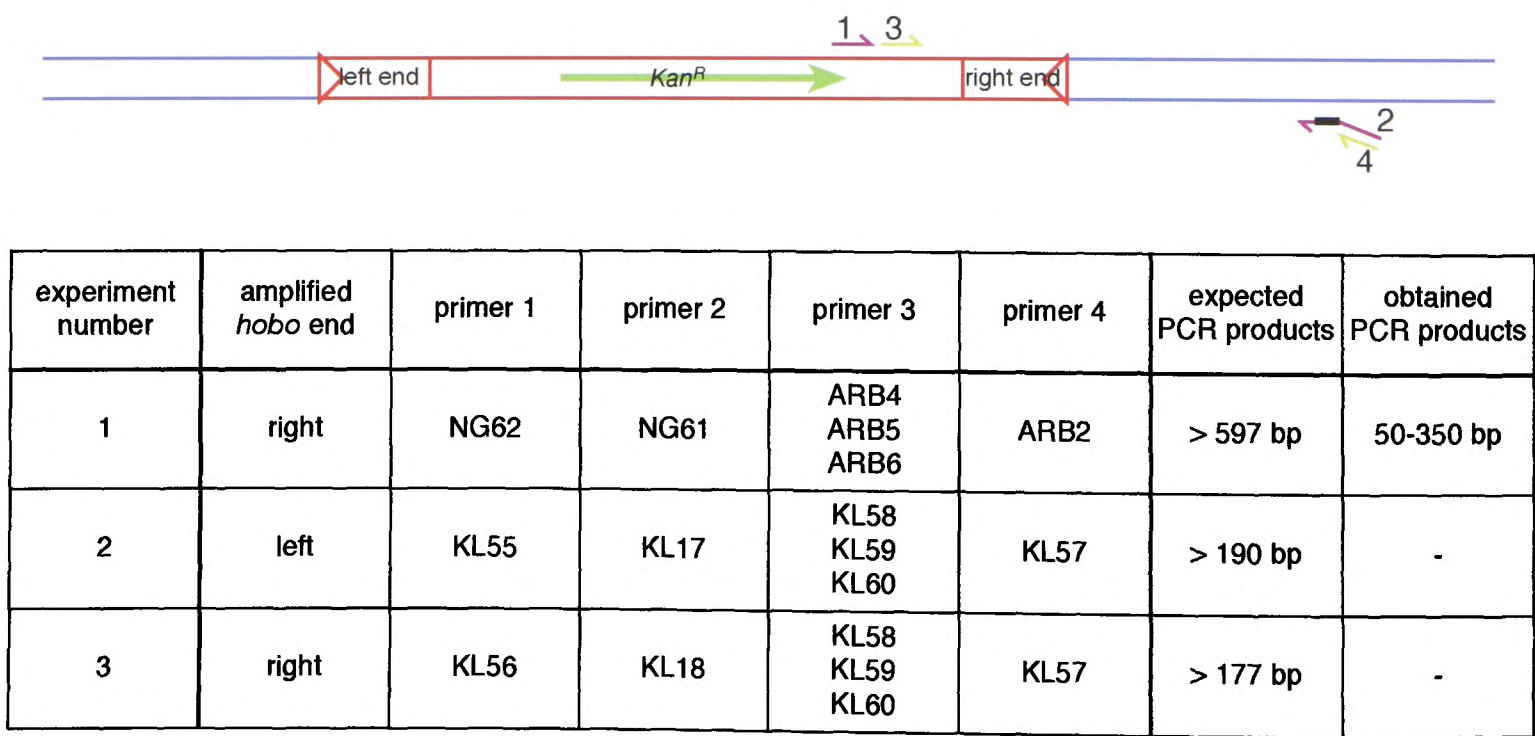


Fig. 5.4

Figure 5.4: Analysis of insertions into R388

- (A)** Comparative restriction mapping of the recombinant plasmids obtained from the mating out assay. Shown is the map of the *hobo* mini-transposon from pKL10, to scale. Native R388 and seven products of the mating-out-assay were digested with the selected restriction enzymes and run in adjacent lanes on agarose gels. The *Mlu*I- and *Nsi*I-fragments of stated length were present in the recombinant plasmids, but not in R388. This indicates that the inner part of the mini-transposon had recombined (or transposed) into the target plasmid. On the other hand, no extra *Not*I-, *Sma*I- or *Pfi*MI/*Nhe*I-fragments of the stated length were observed in the recombinant plasmids. This indicates that the flanking DNA on pKL10 has not been recombined into the target plasmid.
- (B)** Arbitrary PCR to analyse and subclone junctions of three insertions in R388. Violet half-arrows (1 & 2), primers for first round of PCR; yellow half-arrows (3 & 4), primers for second round of PCR; black bar in primer 2, arbitrary sequence. The first round of PCR reactions is with one primer specific to the transposon, facing outwards (primer 1) and a set of arbitrary primers (primer 2), some of which will bind to the unknown sequence of the target plasmid. The specificity is increased by a second, nested PCR, with primer 3 binding inside the transposon, closer to the transposon end, and primer 4 binding to the non-arbitrary 5'-sequence generated by primer 2. From each first-round reaction, one aliquot was amplified as described with primers 3 & 4, and a second aliquot with primer 4 only as a control: PCR-products that are formed without the transposon-specific primer 3, will not contain the transposon-target junction. For primer sequence and description, see Table 2.5, Table 2.6. Primer lengths and positions are not to scale.

toxicity could manifest itself in various forms, such as a reduced growth rate, an altered morphology or death of the expressing cells, and there would be a selection pressure in favour of loss, mutation or rearrangement of the expression plasmid. The toxic effects could be caused by properties of the overexpressed protein that are unrelated to the transposase function, or they could be due to the DNA-recombination activity of the transposase. In the latter case, this would indicate at least partial functionality of the protein. Hence, the toxicity of the *hobo* transposase on *E. coli* was assessed.

5.5.1 Toxicity of components of the mating-out assay

During the mating-out experiments, it was observed that the cell densities had declined during the matings. Poor cell viability could be one explanation for the low frequencies of transposition. In order to test this hypothesis, pKL16(pOX38-*Cam^R*)-containing HMS174(DE3) and MM294(DE3) cells were transformed with different plasmids, and the viability of the transformants compared under several conditions (Table 5.2). In general, the *recA*⁺ strain MM294 was healthier than the *recA*-deficient strain HMS174. Although there are slight differences between the reaction of the different strains to the plasmids, antibiotics and IPTG, a general pattern emerged: The presence of three antibiotics at once reduced the survival of all transformants. The backbone plasmid pBluescript had no effect, while pACYC184 had a very slight negative effect on cell survival. Double transformants of both plasmids also had an only mildly reduced viability. *Tn10* and *hobo* mini-transposons alone were not harmful. Conversely, *Tn10* and *hobo* transposase expression plasmids reduced cell survival, and cells with both mini-transposon and transposase of the same transposon were strongly compromised. IPTG was slightly detrimental to all cells, very bad for transposase transformants and lethal for transposase/mini-transposon double transformants. In summary, antibiotics and IPTG are harmful to all tested cells, but especially to transformants of both the transposase and the mini-transposon. In these experiments, *hobo* had a more severe effect than *Tn10*. This suggests that induced transposase interacts with the transposon ends to cause damage to the host cells, possibly by recombination.

This was confirmed in a further control experiment. Liquid cultures of *hobo* transposase/mini-transposon double transformants were found to have a severely reduced viability. If induced at 30°C, as in the mating-out experiments, a proportion survived, but cultures induced at 37°C did not give rise to a single colony, indicating complete lethality.

Another indication of the toxicity of *hobo* transposase was rapid plasmid loss. Within

Table 5.2: Influence of plasmids, antibiotics and inducer on *E. coli* viability

host	transformed plasmids	transposon	Tn ends	tase	anti-biotics (AB)	transformation efficiency on AB	size of transformants on AB	survival of restreaks on IPTG	survival of restreaks on AB + IPTG
BL21 (DE3) pLysS	pKL5	<i>hobo</i>	-	+	A	++			++
	pRC60	Tn10	-	+	A	++			--
HMS174(DE3)pKL16	pBSII-SK+	-	-	-	A,C	.			++
	pACYC184	-	-	-	T,C				-
	pBSII-SK+ & pACYC184	-	-	-	A,T,C	++	++	++	++
	pKL10	<i>hobo</i>	+	-	A,C				-
	pKL13	<i>hobo</i>	-	+	T,C				--
	pKL10 & pKL13	<i>hobo</i>	+	+	A,T,C	-	-	--	--
	pNK3287	Tn10	+	-	A,C				++
	pRC62	Tn10	-	+	T,C				--
	p1182 & pRC62	Tn10	+	+	A,T,C	--	-	--	--
	pNK81	Tn10	+	+	A,T,C	-	+	++	-
MM294(DE3) pKL16	pBSII-SK+	-	-	-	A,C				++
	pACYC184	-	-	-	T,C				++
	pBSII-SK+ & pACYC184	-	-	-	A,T,C	++	++	+	+
	pKL10	<i>hobo</i>	+	-	A,C				-
	pKL13	<i>hobo</i>	-	+	T,C				--
	pKL10 & pKL13	<i>hobo</i>	+	+	A,T,C	++	++	--	---
	pNK3287	Tn10	+	-	A,C				++
	pRC62	Tn10	-	+	T,C				--
	p1182 & pRC62	Tn10	+	+	A,T,C	++	++	-	--
	pNK81	Tn10	+	+	A,T,C	-	++	++	-
column 1	2	3	4	5	6	7	8	9	10

Different transposon-, transposase expression- and control plasmids (column 2) were transformed into the indicated *E. coli* host strains (column 2) and plated onto LB-agar plates with the appropriate antibiotics for the transformed and conjugative plasmids (column 6). Transformants ere counted and compared for size (columns 7-8). They were furthermore restreaked onto LB-IPTG plates without or with the same selection of antibiotics and assessed for a combination of size and number (survival, columns 9-10).

Columns 3-5: relevant features of transformed plasmids; +, present; -, absent. Columns 7-10: ++, large colonies, good survival; +, medium sized colonies, fair survival; -, small colonies, reduced survival; - -, tiny colonies, very few survivors; - - - , no survivors. AB, antibiotics: A, ampicillin; T, tetracycline; C, chloramphenicol.

p1182, mini-Tn10; pACYC184, cloning plasmid, backbone for pKL13; pBSII-SK+, cloning plasmid, backbone for pKL10; pKL5, CBP-*hobo* tase expression; pKL10, mini-*hobo*; pKL13, *hobo* tase expression; pNK81, Tn10; pNK3287, mini-Tn10; pRC60, IS10-tase expression; pRC62, IS10-tase expression. For detailed descriptions of strains and plasmids, see Table 2.1; Table 2.7; Table 2.8.

six hours of releasing antibiotic selection, the transposase expression plasmid (pKL13 or pKL15) was lost from double-transformants with pKL10. While pKL10 and pKL13 alone were still present after overnight induction, pKL15 was also lost quickly from single transformants. The fast plasmid loss argues for a strong selection pressure against the transposase-gene carrying plasmid, especially in combination with the transposon DNA.

These results show that native *hobo* transposase is toxic to *E. coli*, and that it might have a partial DNA recombination activity, which is detrimental to the host cells.

5.5.2 Morphology of expressing cells

The effect of transposase expression on the morphology of expressing bacteria was studied by light microscopy and compared with the effect of other recombinases (Figure 5.5). *E. coli* BL21(DE3)pLysS, untransformed or transformed with plasmids encoding 6xHis-*hobo* transposase (pKL8), 6xHis-Rec12 (pMW324) or IS10-transposase (pRC60), were compared. It has already been established that pMW324 and pRC60 drive very high expression levels of the respective proteins (Figure 3.2A; Figure 3.1B).

At the same time, the cells were checked for the formation of inclusion bodies. Rec12 is known to form inclusion bodies (M.C. Whitby, unpublished), whereas IS10 transposase does not (R. Chalmers, unpublished). M.C. Whitby observed that for Rec12, visual detection of inclusion bodies coincides with a smear near the boundary of stacking and separating gel on an SDS-polyacrylamide gel. After induction of 6xHis-transposase expression, a similar smear is formed, which could equally be due to inclusion bodies (Figure 3.2B).

Phase-bright, seemingly crystalline objects (inclusion bodies) were observed in pMW324 and pKL8 transformants, usually at the cell-poles or in between nucleoids (Figure 5.5). 6xHis-Rec12-inclusion bodies were seen in over 99% of pMW324 transformants. These were already large 40 minutes after induction and their size increased with time. 6xHis-*hobo* transposase inclusion bodies were smaller and visible only from 4 h post-induction onwards. Their appearance correlates with the observed smear on gels and blots. They were found in only c. 50% of the cells, variable between samples. Most inclusion bodies were in filamentous cells, at the place where the septum should have formed. Similar structures were never observed in untransformed controls or pRC60 transformants.

In transformants of all three plasmids, increased filamentation was observed. This is indicative of problems in cell division, which could arise through damaged DNA. Only untransformed cells had the small spherical morphology, which is typical for stationary

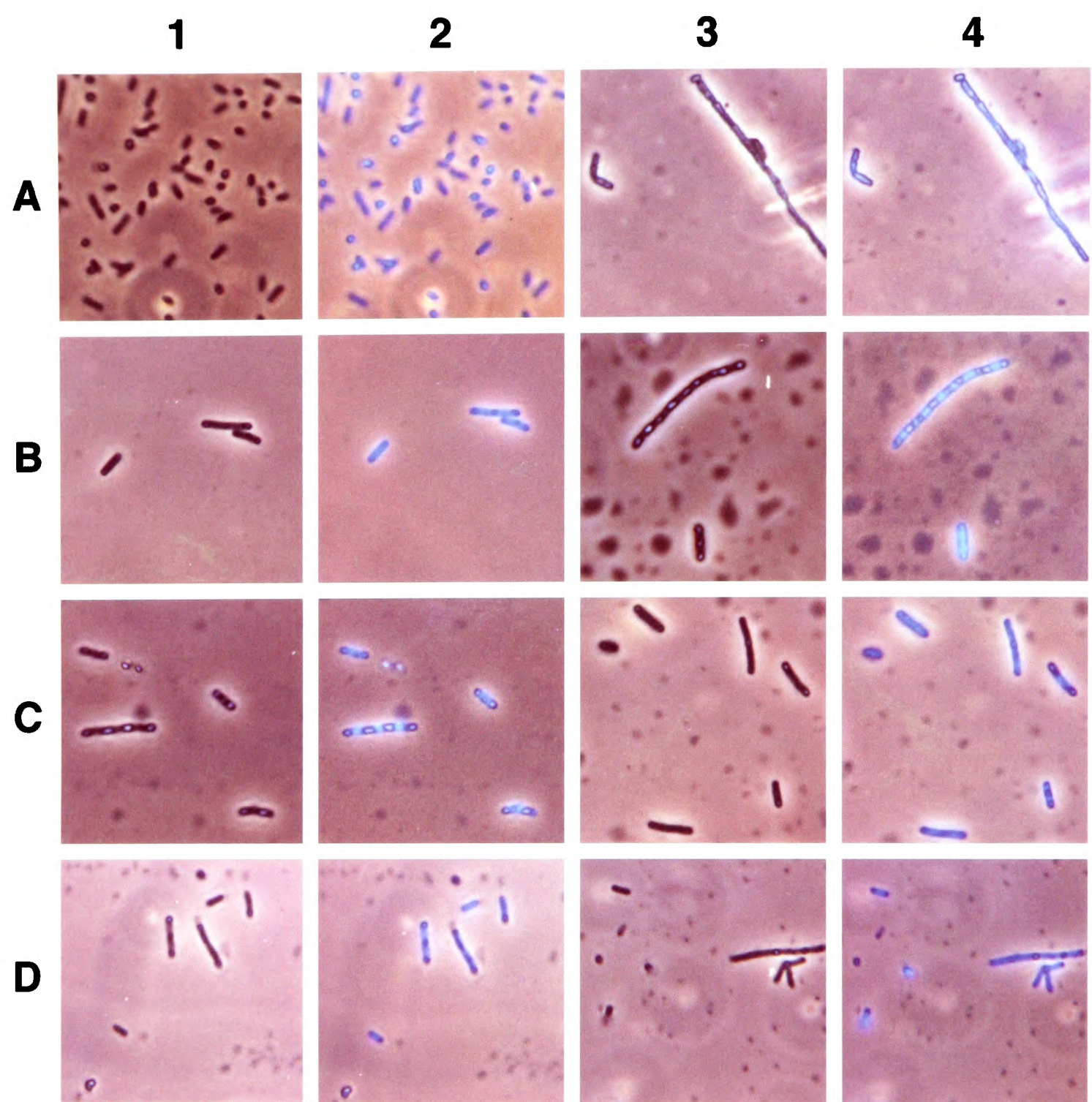


Fig. 5.5

Figure 5.5: Influence of protein expression on *E. coli* morphology

E. coli BL21(DE3)pLysS, transformed with the plasmids indicated in the table below, were grown at 37°C in LB-Amp-Cam to OD₆₀₀=0.3-0.4 and induced with 1 mM IPTG. Aliquots of the cultures were taken 40 min, 4 h, 6.5 h and 29 h after induction. The nucleoids were condensed with 200 mg/ml Cam and stained with DAPI.

Two adjacent pictures show the same sample, both viewed with phase contrast and at 1000-fold magnification. In each pair, the picture on the right is taken under UV light ($\lambda = 360$ nm), which causes fluorescence of DAPI stained DNA (emission at $\lambda = 420$ nm).

panel	transformed plasmid	induced protein	time post-induction	inclusion bodies	cell elongation	cell death
(A 1-2)	-	-	4 h	-	-	-
(A 3-4)	pRC60	IS 10 transposase	4 h	-	+++	++
(B 1-2)	pMW324	6xHis-Rec12	40 min	++	+	-
(B 3-4)	pMW324	6xHis-Rec12	4 h	+++	++	-
(C 1-2)	pMW324	6xHis-Rec12	4 h	+++	++	-
(C 3-4)	pKL8	6xHis- <i>hobo</i> transposase	4 h	+	+	-
(D 1-2)	pKL8	6xHis- <i>hobo</i> transposase	29 h	++	+	++
(D 3-4)	pKL8	6xHis- <i>hobo</i> transposase	29 h	++	+	++

phase in *E. coli*.

The number of dead cells, i.e. lysed, translucent and/or fragmented cells, was low in the untransformed control and in pMW324 transformants. The proportion of dead cells in pRC60 transformants was greatly increased from 4 h onwards, and hardly any cells survived overnight induction. The pKL8 transformants were still intact at earlier time points, but showed a high number of dead cells in the 29 h sample. This confirms that overexpressed 6xHis-*hobo* transposase is toxic to *E. coli*.

5.6 Excision of genomic *hobo* copies in mouse fibroblasts

A higher eukaryotic *in vivo* system would be advantageous for analysing the interactions between transposition and double strand break repair (DSBR). For cultured cells of the natural *hobo* host, *D. melanogaster*, there are no established methods for genetic modification. In contrast, for mouse cells, DSBR mutants are available and others can be produced as needed.

The main objective of these *in vivo* experiments is the reconstitution of *hobo* transposition activity and the detection of hairpin ends in the flanking DNA after excision (Figure 5.6). To increase the chance of detecting the otherwise unstable hairpin intermediate, we used fibroblasts from mice with the *scid* (severe combined immune deficiency) mutation, which have severely reduced levels of DNA-dependent protein kinase (DNA-PKcs), an essential component of the non-homologous end joining pathway, and involved in opening DNA hairpins (see Section 1.2.3.2.1).

To this end, we created *scid* and wt cell lines with a number of *hobo* transposons stably integrated into the chromosomes. The SCGR11 cell line is derived from *scid* mice; NIH 3T3 is their parental cell line, lacking the DNA-PKcs mutation (Table 2.4). Once the cell lines were established, transposase-expression vectors were introduced by transient transfection. One to three days after introduction of transposase, the genomic DNA of the transfected cells was tested for the presence of hairpin structures in the flanking DNA.

5.6.1 Stable transfectants

In order to generate stable *hobo* integrations, SCGR11 cells were co-transfected by calcium-phosphate treatment with the *hobo* mini-transposon plasmid pKL10 (Table 2.7; Figure 2.2A) and with the neomycin resistance plasmid pRC98 (Table 2.8) in ratios of 5:1 and 10:1. As the calcium-phosphate precipitate was formed with both DNA species, practically all cells that survive neomycin selection should also contain the mini-transposon plasmid.

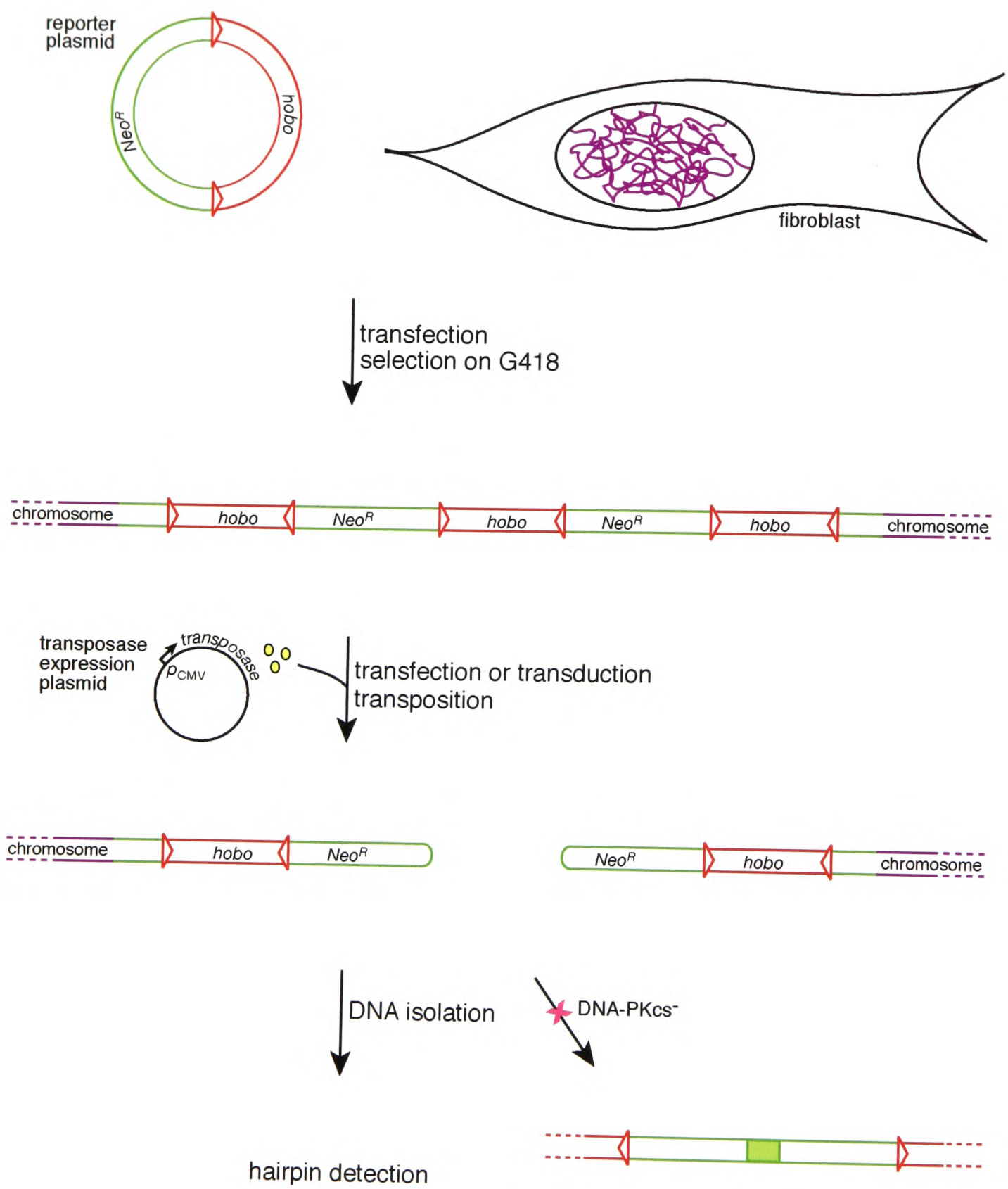


Fig. 5.6

Figure 5.6: *hobo* excision assay in mouse fibroblasts

Schematic of the *hobo* transposition assay in cell culture. Reporter plasmids containing the *hobo* transposon and a neomycin-resistance gene were transfected into fibroblasts. Growth in medium with G418, a functional homologue of neomycin, selected for those cells that have chromosomal integrations of the transfected plasmids (stable transfectants). In mammalian cells, supercoiled plasmids usually integrate as concatemers. To induce transposition of the integrated transposons, established cell lines were transfected or transduced with plasmid driving expression of *hobo* transposase from a retroviral promoter. The genomic DNA was harvested and tested for the presence of hairpin ends, which are expected to be formed at *hobo* excision. In cells deficient in non-homologous end joining (NHEJ), such as DNA-PKcs mutants, the repair of the hairpin ends is defective, which increases the chance of their detection.

Here, the assay is shown for reporter plasmid pKL42 (Figure 2.2B). With pKL10 (Figure 2.2A), transposon and *Neo^R*-selection marker are on two different plasmids, but the principle and procedure are the same. See text for more details.

Twelve clonal lines of transfected cells were established.

To assess the number of *hobo* reporter integrations, genomic DNA from these cell lines was digested and analysed by Southern blotting (see Section 2.6.2.1; not shown). Hybridisation-probes made from linearised pKL10 hybridise to both the reporter (pKL10) and the backbone of the selection plasmid (pRC98). The relative number of copies of each integrate can therefore be estimated from these blots. To determine the actual copy number, one blot was also hybridised with the cDNA of the single-copy gene for murine erythropoietin. The copy number of integrated plasmids varied between cell lines. They were estimated to contain 1-8 copies pRC98 and 1-3 copies pKL10 per cell (2.6.2.2.1).

A higher number of transposon copies per cell would be advantageous when attempting to detect *hobo* excision events, which may occur at a low frequency. To achieve this, (1) the ratio of pKL10 to pRC98 was increased, and (2) new plasmids were created with neomycin resistance marker and transposon on the same molecule (pKL41-48, see Table 2.7). All of these plasmids were constructed with the *Neo^R/Kan^R* cassette from pEGFP-N1 (Clontech). This aminoglycoside phosphotransferase gene has both a bacterial promoter and the SV40 early promoter and therefore provides kanamycin resistance in *E. coli* and neomycin or geneticin resistance in fibroblasts. In pKL47, the *Neo^R/Kan^R* cassette was inserted between *hobo* transposon ends, forming a mini-transposon. pKL42 has a full (wt) *hobo* transposon, and the *Neo^R/Kan^R* cassette in the backbone (Figure 2.2B). Theoretically, the presence of the transposase gene might make insertions of pKL42 unstable. However, as expression of *hobo* transposase from its natural promoter requires a factor specific to *Drosophila* embryonic germline cells (Calvi and Gelbart, 1994), it is very unlikely that transposase is going to be expressed in the transfected cells.

SCGR11 cells and the parental line NIH 3T3 were stably transfected with pKL10 / pRC98 (100:1), pKL42 and pKL47. From each transfection, c. 12 cell lines were established from single colonies and their genomic DNA tested for plasmid insertion by Southern blotting (see Section 2.6.2.1). After hybridisation with the transfected plasmid, the strongest, clearest signal was obtained from a line of pKL42-transfected SCGR11 cells, S10.1 (Figure 5.7A). This cell line was shown to contain c. 5 copies of the transposon plasmid (Section 2.6.2.2.2, Figure 5.7B). Line S10.1 was used for further experiments because of its relatively high number of insertions. Additionally, the presence of the intact transposons of pKL42 ensured that no sequences essential for transposition had been deleted. Finally, compared to transfected NIH 3T3 lines, the *scid* background should greatly facilitate the detection of hairpins.

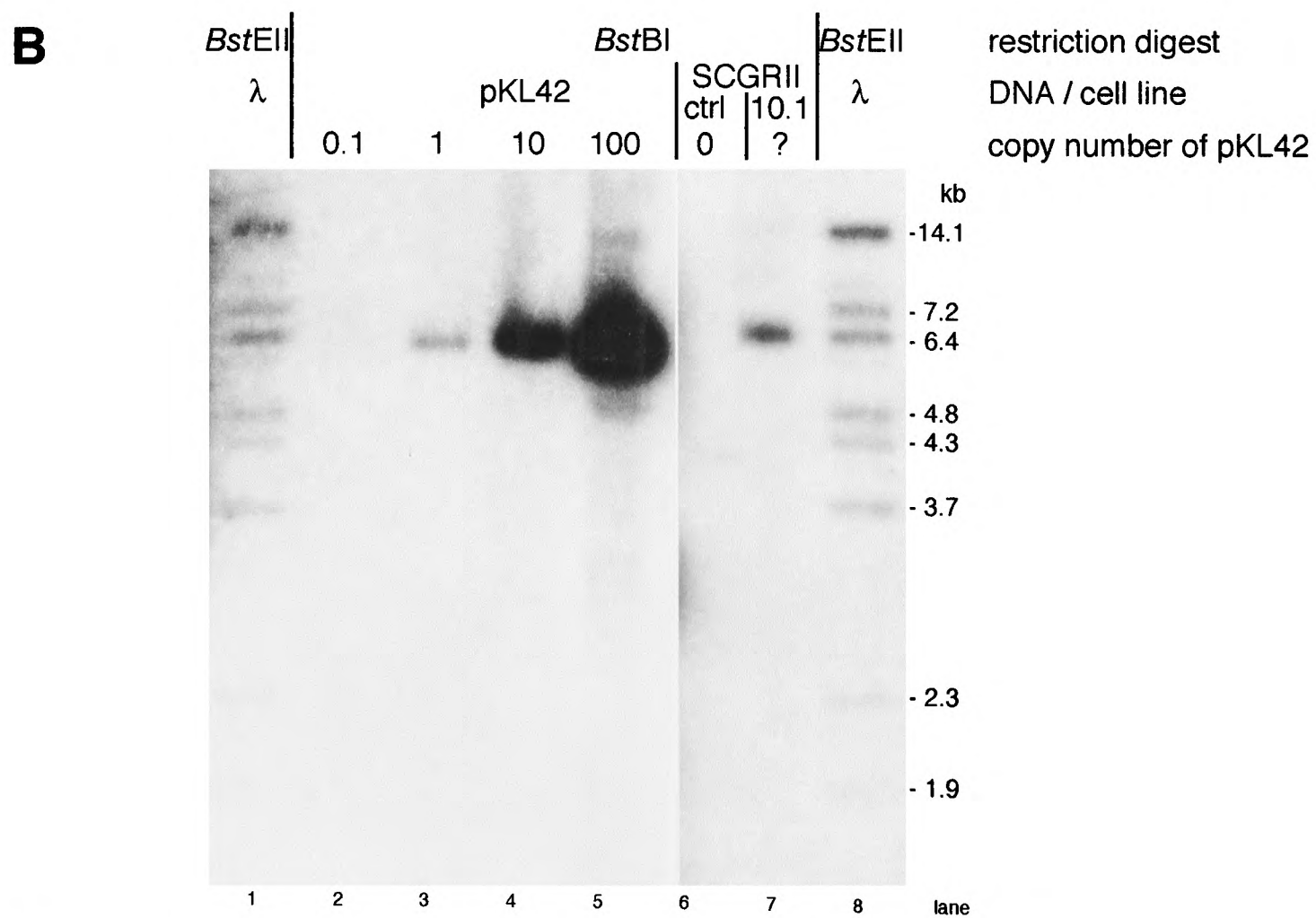
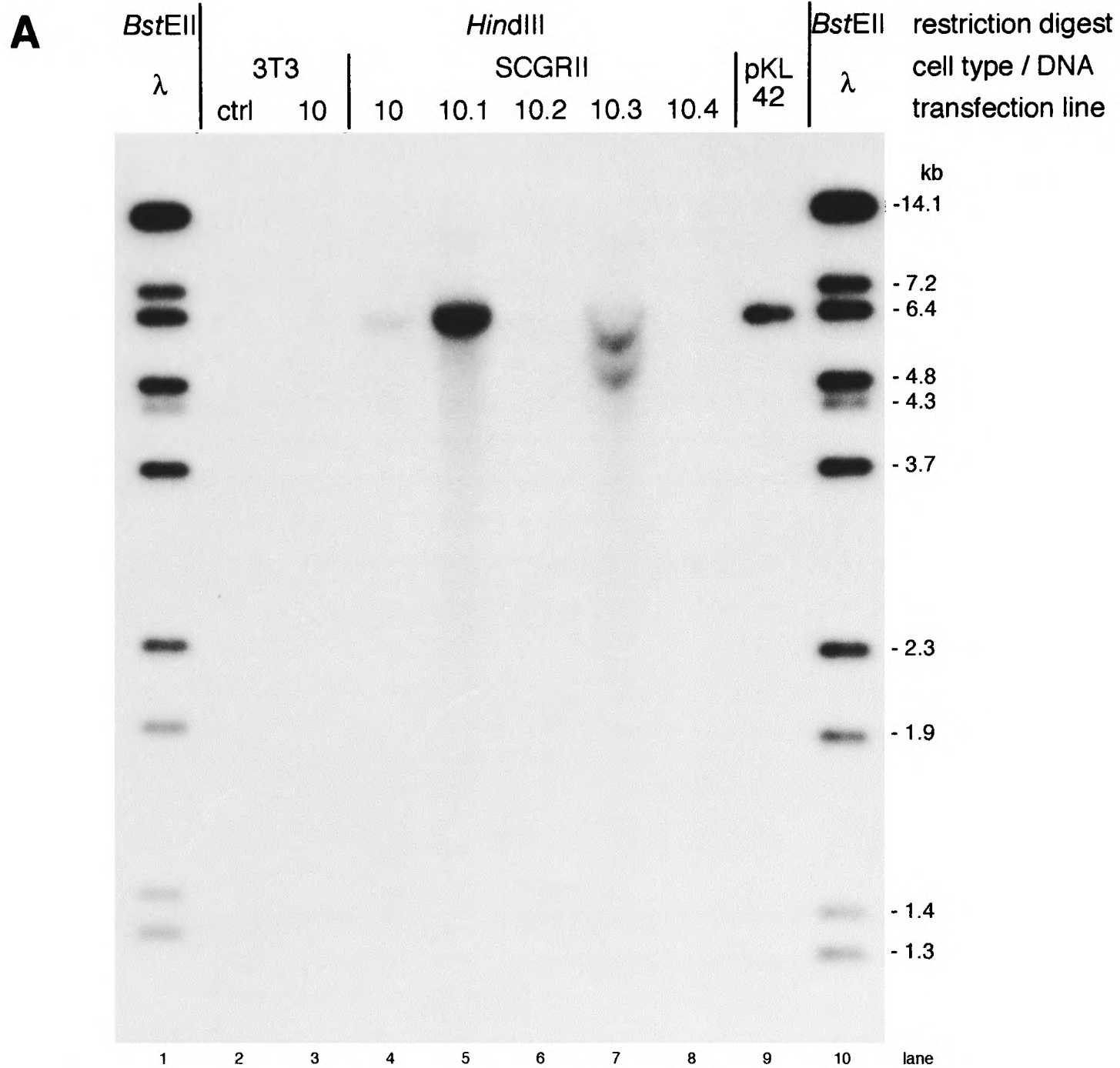


Fig. 5.7

Figure 5.7: Cell lines with stable *hobo* integrations

Southern blots of 1x TBE, 0.8% agarose gels.

- A)** Subset of established mouse fibroblast cell lines, transfected with pKL42. ctrl, non-transfected parental line. Per lane, 10 μ g genomic DNA or 0.1 ng plasmid pKL42 was analysed. Probed with *HindIII* digested pKL42 and lambda *BstEII* digest.
- B)** Determination of the copy number of pKL42 in *scid* cell line S10.1. ctrl, non-transfected parental line; 10.1, with stable pKL42 integrations. 20 μ g genomic DNA were loaded per lane. 40 pg pKL42 are equivalent to 1 copy per mouse genome (see Section 2.6.2.2.2). Probed with the 3.3 kb *Apal/EcoRV* fragment from the pKL42 backbone. Lanes 1-5 and 6-8 are from the same gel, separated by nine lanes. The question mark indicates the unknown copy number, to be determined from this gel.

5.6.2 Transfection and expression of *hobo* transposase constructs

To initiate transposition, transposase expression vectors were introduced into the newly created cell lines. For transposition to take place, (1) the transposase-coding plasmid must enter the cells efficiently, (2) the transposase protein must be expressed, and (3) the transposase must enter the nucleus. The performance of different methods of DNA-delivery concerning these three prerequisites was assessed with EGFP and *hobo* transposase-EGFP fusion proteins.

5.6.2.1 Chemical methods of transfection

Transfection was tested with pEGFP (pRC97) and its derivatives pKL35, pKL36, pKL37, pKL38 and pKL49 (Table 2.8; Table 2.7). In all six plasmids, expression of the protein is controlled by the constitutive CMV immediate early promoter. In pKL35 to pKL38, EGFP is fused to the C-terminus of the transposase, while, in pKL49, EGFP is fused to the N-terminus. pKL35, pKL37 and pKL49 are with wt transposase, but in pKL36 and pKL38, two arginine codons that are rare in *E. coli* had been exchanged (Section 3.4.1). The only difference between pKL35 and pKL36 versus pKL37, pKL38 and pKL49 is a shorter distance between *hobo* orf and CMV promoter in the latter three.

The transfection efficiency of the following methods was compared using pEGFP: calcium-phosphate transfection (Ausubel *et al.*, 1994), non-liposome micell forming Effectene (Qiagen), and three lipofection reagents, GenePORTER (Gene Therapy Systems), TransFast (Promega), and FuGENE (Roche). With GenePORTER and Effectene, 10-15% of the pEGFP-transfected cells were bright green after 48 h due to the expression of the EGFP-protein, compared to less than 1% with the other methods (not shown). Among SCGR11, NIH 3T3 and COS-7 cell-lines, the transfection efficiencies were comparable. COS-7 cells, which have very large nuclei, were used to facilitate the localisation of the protein. EGFP was observed uniformly in the cytoplasm and the nucleus.

No fluorescent signal was detected in cultures that had been transfected with the *hobo*-EGFP fusion-protein plasmids. Selection of transfectants with G418 for the Neo^R marker demonstrated that the absence of EGFP-signal in transposase-fusion transfected cells is not due to inefficient transfection. For increased sensitivity, an anti-GFP-antibody (Invitrogen) was used. Immunohistochemistry on transfected cells did not give rise to any signal, and Western blots of transfected and non-transfected cells, fractionated into nuclei and cytoplasm, showed no specific signal (not shown). In summary, *hobo*-EGFP

expression in chemically transfected cells is below the level of detection, as determined by fluorescence microscopy and immunodetection.

5.6.2.2 Retroviral transduction

Retroviruses have evolved a very high efficiency of infecting their mammalian or avian host cells. This capability can be used in retroviral transduction, which is several fold more efficient in delivering DNA into cells than any chemical method. As we had been unable to detect any transposase-fusion protein expression after lipofection, recombinant retroviruses for the transduction of the S10.1 reporter cell line were created.

The three-plasmid system with MoMuLV-based retroviral vectors (Soneoka *et al.*, 1995) was used, which employs simultaneous transfection of *gag-pol*, *env* and vector genome plasmids into packaging cells (Section 2.6.2.3). This system is simpler than generating stable cell lines that express all three components, and gives very high, helper-free titers of c. 10^7 infectious units per ml. The packaging cell line 293T (Table 2.4) stably expresses the SV40 large T antigen (T_{ag}), which drives replication of all three plasmids from their SV40 origin of replication (DuBridge and Calos, 1988). 293 and 293T cells support SV40 replication 3-4 times better than HeLa cells (Kelly *et al.*, 1989) and are highly transfectable.

Two alternative genomic vectors were used for *hobo* transposase and for the C- and N-terminal fusions of *hobo* transposase to EGFP (pKL64–pKL75, Table 2.8, Table 2.7). They differ by the position of the cytomegalovirus immediate early promoter (p_{CMV}) within the retroviral genome.

Using the three-plasmid system, the recombinant genomic plasmids and a control plasmid for pure EGFP, pGFP-i-lacZ, were transfected into 293T cells. After 48 h, the retrovirus-containing supernatant was used to transduce S10.1 cells (Section 2.6.2.3). Both 293T and S10.1 cells were scored for EGFP expression by fluorescent microscopy (Figure 5.8; Figure 5.9).

Judged by the presence of clear green signal, a transfection efficiency of c. 50% in 293T cells and a transduction efficiency of 90-95% in S10.1 cells was achieved. Expression in general was much stronger in 293T than in S10.1 cells. The brightest signal was always obtained from EGFP alone. The fusion proteins were just visible in S10.1 cells (Figure 5.9) but very clear in the packaging cells (Figure 5.8). It was found that N-terminal EGFP-tase fusions give rise to a brighter signal than C-terminal tase-EGFP. Expression from pLNCX-derivatives was found to be higher than from pRV583-derivatives. The pure

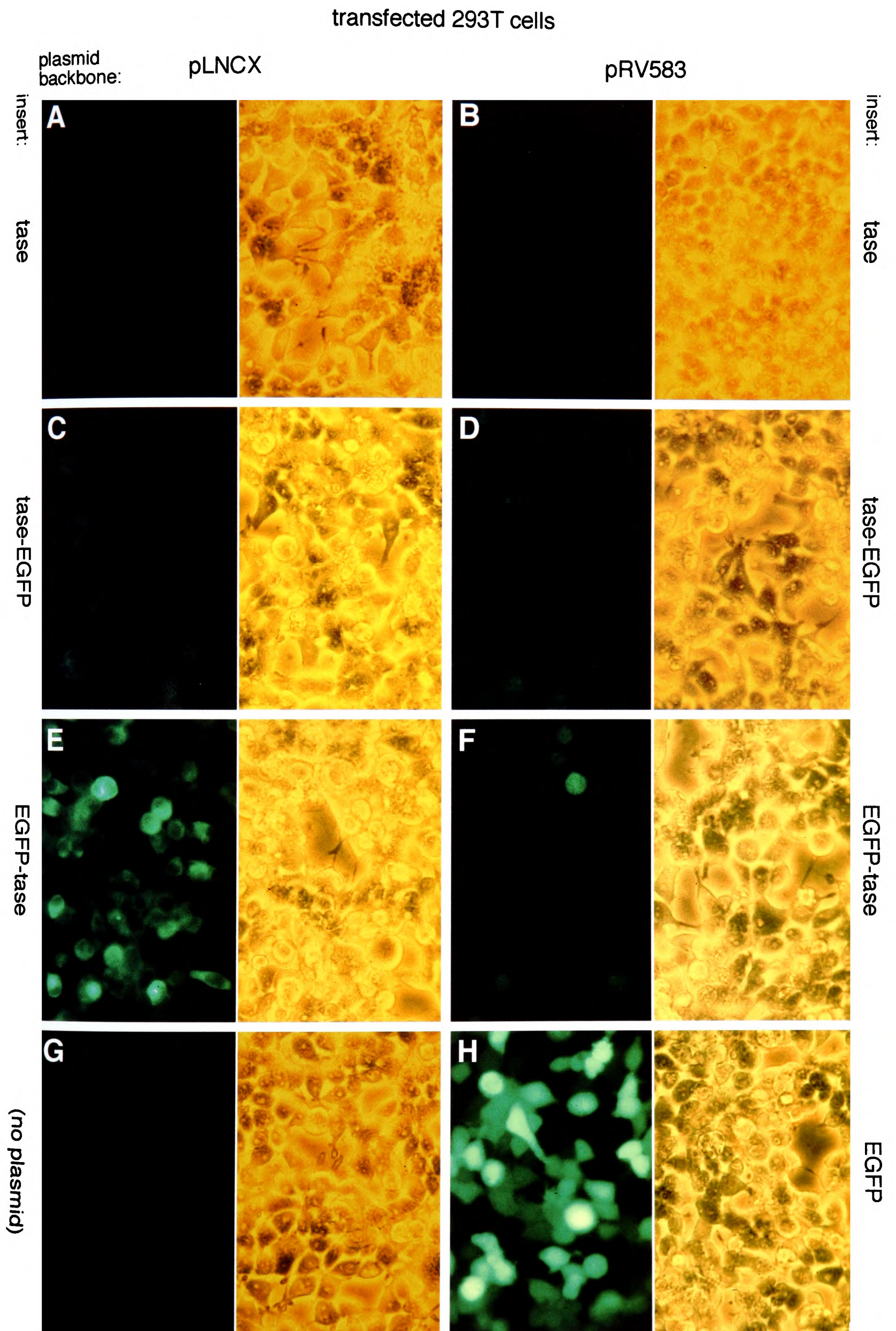


Fig. 5.8

Figure 5.8: Transfected 293T cells producing recombinant retrovirus

293T cells, 48 h after transfection with pHIT111, pHIT60 and the genomic plasmid specified below. In each panel, the picture on the left is taken under UV illumination, the picture on the right is taken under visible light. Both show the same selection of cells. Genomic plasmid:

- (A) pKL64: pLNCX-*hobo* tase
- (B) pKL70: pRV583-*hobo* tase
- (C) pKL66: pLNCX-*hobo* tase-EGFP
- (D) pKL72: pRV583-*hobo* tase-EGFP
- (E) pKL68: pLNCX-EGFP-*hobo* tase
- (F) pKL74: pRV583-EGFP-*hobo* tase
- (G) no plasmid
- (H) pGFP-i-lacZ: pRV583-EGFP-*lacZ*

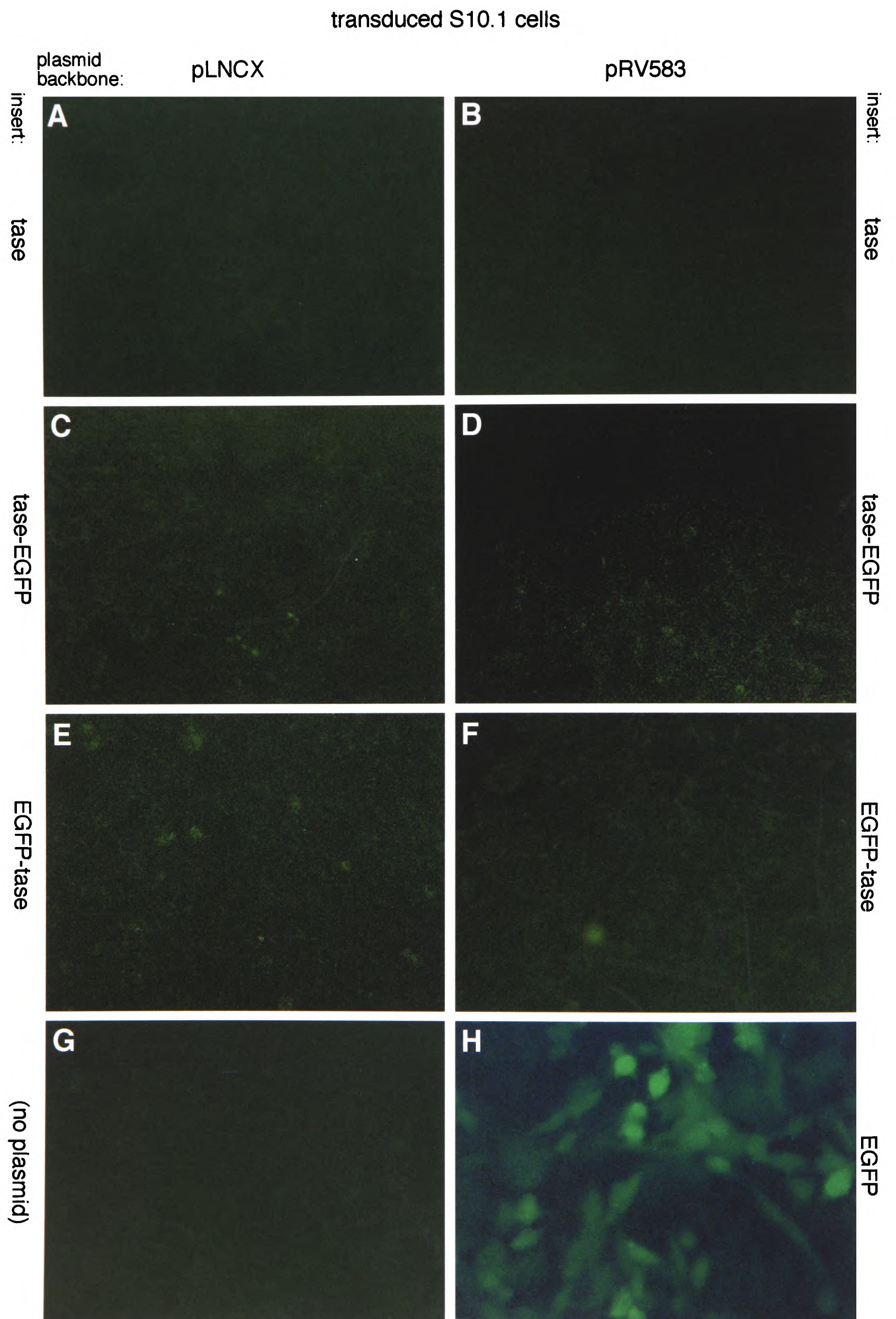


Fig. 5.9

Figure 5.9: Transduced S10.1 cells

S10.1 cells, 24 h after transduction with recombinant retroviruses. Pictures are taken under UV illumination. Retroviruses with genomes from:

- (A) pKL64: pLNCX-*hobo* tase
- (B) pKL70: pRV583-*hobo* tase
- (C) pKL66: pLNCX-*hobo* tase-EGFP
- (D) pKL72: pRV583-*hobo* tase-EGFP
- (E) pKL68: pLNCX-EGFP-*hobo* tase
- (F) pKL74: pRV583-EGFP-*hobo* tase
- (G) no plasmid
- (H) pGFP-i-lacZ: pRV583-EGFP-*lacZ*

EGFP protein is distributed evenly throughout the cell, whereas the fusion proteins show a fainter signal in the area of the nucleus, thus appearing to be mainly, but not entirely, cytoplasmic (Figure 5.8E). No influence of protein expression on cell survival was detected (Figure 5.8, yellow panels).

5.6.3 Analysis of S10.1 fibroblast DNA after expression of *hobo*-transposase

To initiate transposition of the integrated *hobo* transposon copies, cell line S10.1 was transiently transfected or retrovirally transduced with expression vectors for the native *hobo* transposase. The transfected cells were cultured for one to three days before being harvested. The total DNA was purified and analysed for evidence of transposition.

5.6.3.1 Analysis of excision by two-dimensional gel electrophoresis

The S10.1 *hobo* reporter cell line was transduced with retroviruses carrying the transposase expression cassette of plasmids pKL64 and pKL70 (Table 2.7; Section 2.6.2.3). After 24 h, the total DNA was harvested and analysed by two-dimensional agarose gel-electrophoresis with a denaturing second dimension, followed by a Southern blot (Figure 5.10). Before electrophoresis, the genomic DNA was digested with *Bst*BI, which cuts once in the centre of the flanking DNA of the pKL42 reporter construct integrated in the S10.1 cell line. The blots were hybridised with probe made from pKL42. The concatenated genomic pKL42 insertion gave rise to a very strong signal at 6.2 kb on the arc of linear lambda marker fragments, the position for non-recombined plasmid. However, no signal was detected at the expected position for hairpinned excision products (encircled position, Figure 5.10B). The experiment was repeated with a different set of restriction digests for substrate and probe preparation. This gave rise to equivalent results.

5.6.3.2 Hairpin detection with FDPCR

The S10.1 reporter cell line was chemically transfected with pKL39 (p_{CMV} -tase, Table 2.7) and harvested after 1, 2, or 3 days. Total DNA was purified by CsCl gradient centrifugation and analysed by FDPCR. Genomic DNA was digested with *Sfi*I and processed using various protocols, as follows: Adapter pairs were used in two different concentrations (1/10 or 20x excess of total *Sfi*I restriction ends). Ligations were performed with two different ligase- and ATP concentrations. Completed ligation reactions were or were not phenol extracted, and they were or were not size-selected by excision from a high-purity agarose gel. The PCR was performed as optimised for template 42S-hf, or with

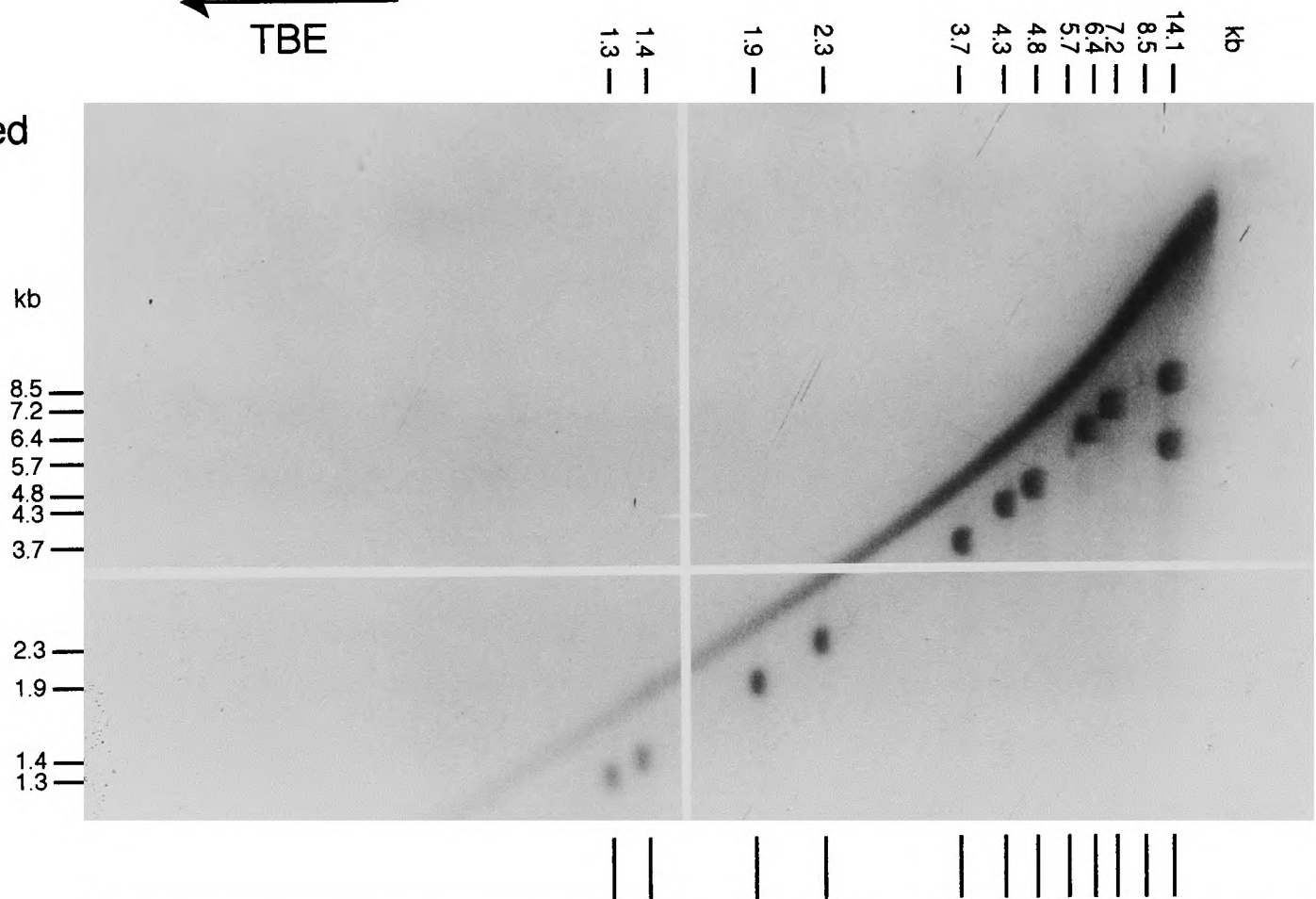
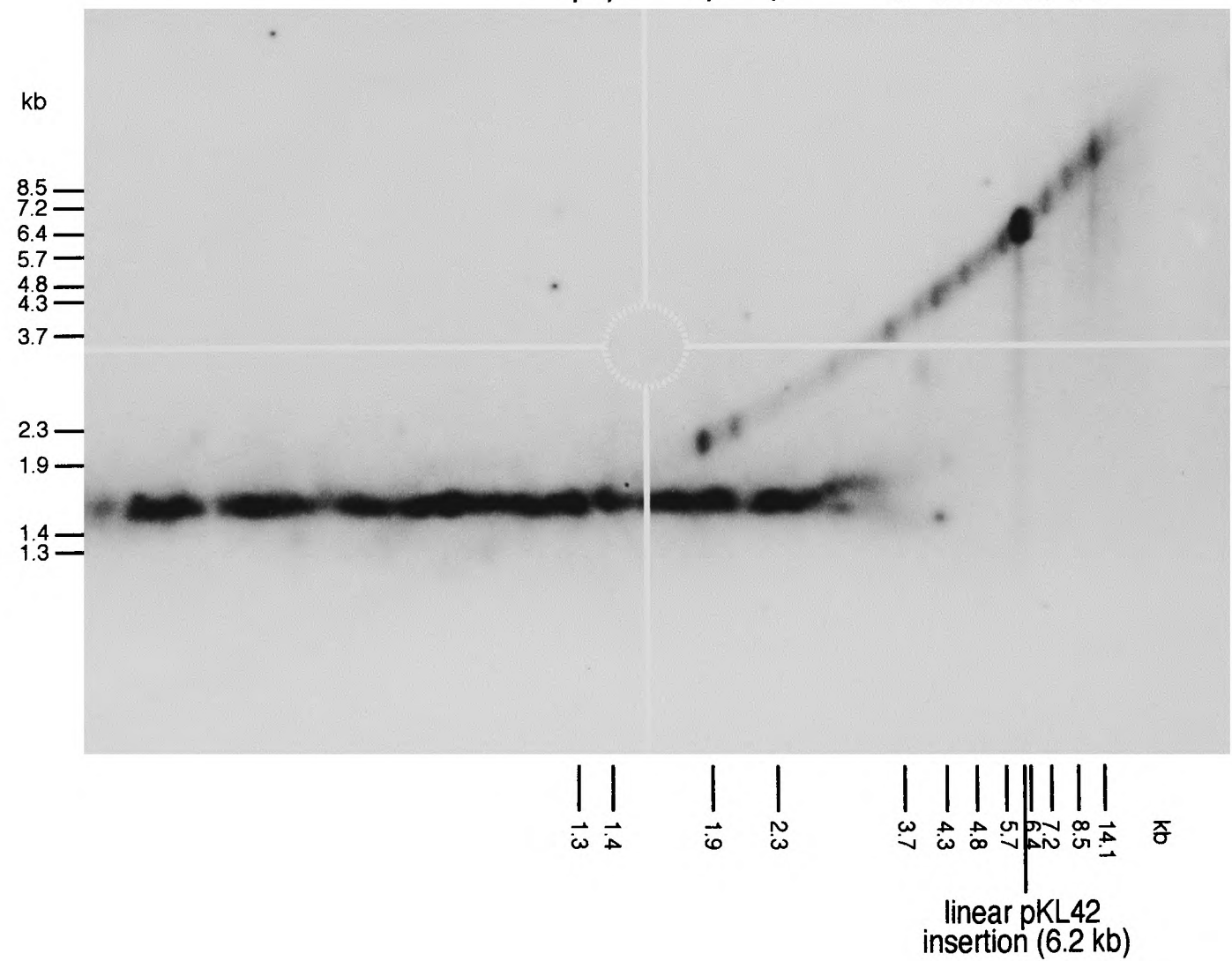
AEtBr-stained
test gelD2
NaOH
↓D1
←
TBE**B**Southern
blot

Fig. 5.10

Figure 5.10: 2D gels for the detection of hairpins from transduced S10.1 cells

Genomic DNA was isolated from S10.1 cells, which had been transduced with pKL64 and pKL70. For each of the two gels shown, 10 μ g of DNA from each transduction was digested with *Bst*BI and analysed by two-dimensional gel electrophoresis.

- (A) Digested genomic DNA and lambda *Bst*EII digest loaded onto adjacent lanes of a 1.1% agarose gel. First dimension run for 20 h in 1x TBE at RT, second dimension run for 24 h in 50 mM NaOH at 4°C. Reverse contrast image of an EtBr stained gel. The white cross marks the expected position of the putative hairpin.
- (B) Digested genomic DNA and lambda *Bst*EII digest loaded onto the same lane. The gel was run as in (A) and Southern blotted by alkaline transfer in denaturing solution. The blot was hybridised with probe made from *Dra*III-digested pKL42. Autoradiogram, exposed for 5 h. Excision of *hobo* elements would result in two fragments of flanking DNA after *Bst*BI digestion, both running at 1.6 kb in the first and at 3.2 kb in the second dimension (position of dashed white circle; see also Figure 2.2B).

higher concentrations of *Tth* polymerase, lower MgCl_2 , or with MnCl_2 instead of MgCl_2 . The expected 1x-hairpinned product band for pKL42 excisions is 779 bp (Table 2.9). While the positive and negative PCR controls behaved as expected, there were only faint bands of c. 800 and 950 bp detectable among the background smear in some of the experimental PCR reactions on agarose gels (Figure 5.11A). The 800 bp band would in theory be a good candidate for the sought-after hairpins. These bands were most prevalent in PCR reactions with MgCl_2 and the optimised concentration of polymerase from DNA that had undergone size-selection. A Southern blot of these gels, probed with a 665 bp (*SfiI/ApaI*) fragment from the left flanking DNA of pKL42 revealed strong signals in the same lanes as the visible 800 bp bands, especially in reactions from cells, which had expressed *hobo* transposase for 3 days (Figure 5.11B). However, similar to the *in vitro* excision experiments (Section 5.2.1), the signals were smears and mainly below 500 bp in length. This suggests that there had been no detectable level of excision and hairpin formation in the fibroblasts.

5.7 Discussion

Attempts were made to reconstitute the *hobo* transposition reaction both *in vitro* and *in vivo* in a variety of prokaryotic and eukaryotic systems. No specific transpositional activity or reaction intermediates were detected in any of the studied systems.

5.7.1 *In vitro* experiments

The activity of purified 6xHis-*hobo* transposase, expressed and purified from *E. coli*, was tested in plasmid-substrate reactions. These were scored by ethidium bromide stained gels and by FDPCR. Furthermore, the 6xHis-*hobo* transposase was assayed in gel shifts (Section 5.2). The three types of experiments examined a change of plasmid size by excision, the presence of hairpin ends, and the specific binding of protein to transposon ends, respectively, i.e. very different aspects of the reaction.

The FDPCR assay, in particular, is extremely sensitive (see Table 4.1). Only an extremely low abundance of hairpins would have been missed. Although a very low activity cannot be excluded, it appears that the purified 6xHis-*hobo* transposase was catalytically inactive in the conditions tested. For the band shift assay, c. 500 different reactions were analysed. It is therefore highly likely that either another protein, other than HMGB1, is required for activity, or that the purified transposase was inactive.

The inactivity of 6xHis-transposase, purified from *E. coli*, could have a variety of

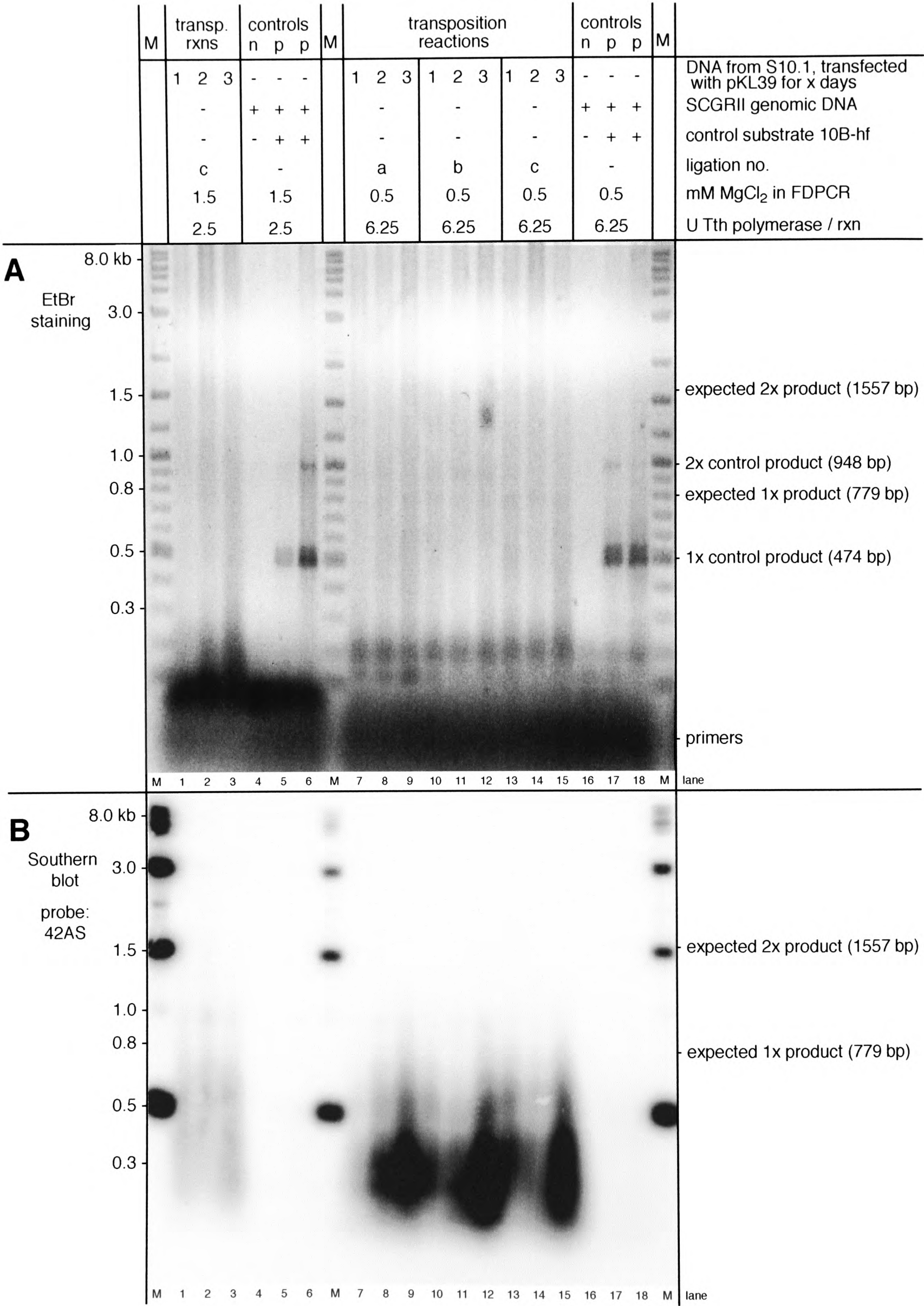


Fig. 5.11

Figure 5.11: FDPCR analysis of genomic DNA from mouse fibroblasts after *hobo* induction

- (A) Selected FDPCR reactions from *Sfi*I-digested genomic DNA, isolated from S10.1 fibroblasts, which had been transfected with the *hobo* transposase expression plasmid pKL39 1-3 days prior to DNA isolation. Note that S10.1 cells are stably transfected with pKL42 and that positive controls were carried out with artificial hairpin from the flanking DNA of pKL10. This leads to different sizes of the expected PCR products. Ligation conditions: a, 1x T4 ligase buffer, 1.8 mM ATP, 10 fmol adapter KL86/KL93, 2 μ g *Sfi*I-digested genomic DNA, 720 U T4 DNA ligase in 30 μ l; b, 1.8 mM ATP, 10 fmol adapter KL86/KL99, 2 μ g *Sfi*I-digested genomic DNA, 720 U T4 DNA ligase in 30 μ l; c, 1.0 mM ATP, 20 fmol adapter KL86/KL93, 4 μ g *Sfi*I-digested, phenol extracted genomic DNA, 800 U T4 DNA ligase in 20 μ l; all incubated overnight at 16°C. M, 1 kb and 100 bp ladder, NEB. 1.5% SeaKem Gold agarose, 1x TBE gel, run 2h 45 min at 140V, RT. Reverse contrast image of EtBr stained gel.
- (B) Southern blot of gel in (A), hybridised with probe made from the 665 bp *Ap*I-*Sfi*I fragment from pKL42. Note that this probe does not overlap with the substrate used in the controls.

causes. Most importantly, the 6xHis-tag could inactivate the transposase, the protein could be folded or modified incorrectly in the non-natural expression host, or the protein could lose its activity during purification. Although the protein was diluted before use, remnants of the elution buffer, such as Triton X-100, NaCl or imidazole, could inhibit the reaction, as do moderate concentrations of Triton X-100 for the Tn10 transposase (Chalmers and Kleckner, 1994). In addition, the co-purified truncations (see Chapter 3) might inhibit the activity of the full-length transposase through the formation of non-productive heterodimers. If the truncations still contain the dimerisation domain but lack the DNA-binding domain, synaptic complex formation would be inhibited. In Tn5, for example, transposition is regulated through the production of an inhibitor protein, which is an N-terminal truncation of transposase and forms non-productive heterodimers with the transposase (de la Cruz *et al.*, 1993; Braam *et al.*, 1999). (See Section 3.7.3 for suggested solutions to this problem.)

hobo transposase, produced *in vitro* in rabbit reticulocyte lysates, was hoped to have a higher transposition activity, as it was untagged, expressed in a higher eukaryotic system and contained very few truncations (Section 3.6). The expression reactions were used directly in transposition reactions without prior purification of the transposase. Here, the experiments failed due to a very high background nicking and ligation activity of the lysate itself (Section 5.3). However, purification of transposase from scaled-up *in vitro* expression reactions might have a good chance of producing an active protein.

5.7.2 *hobo* *in vivo* activity in *E. coli*

hobo transposition activity in bacteria was tested by a mating out assay. Analysis of the very few recombination events revealed that they were recombination artefacts, as determined by subcloning and sequencing (Section 5.4).

The mating out system itself worked extremely well, as shown by a very high transposition activity of the Tn10 controls. The inability to isolate *hobo* transposition events in a variety of bacterial strains indicates that *hobo* transposase is inactive in *E. coli*. Here, the inactivity cannot be caused by purification artefacts, as the reactions were performed *in vivo*, nor by a tag, as untagged transposase (expressed from pKL13) was inactive as well. Either the protein itself is inactive due to improper modification or folding, or the composition of the *E. coli* cytoplasm does not allow *hobo*-mediated transposition. This could be due to any number of intracellular conditions, such as the pH, viscosity or the concentration of ions or reducing agents, or due to protein components.

hobo transposition might require the presence of host proteins, such as DNA-binding or -bending proteins. Such requirements exist for *P* elements and Tn10 (see Section 1.1). Alternatively, DNA-modifying enzymes such as specific helicases could be necessary for activation of the transposon. In bacteria, these might be absent or very different from the *Drosophila* enzymes. On the other hand, some prokaryotic proteins might inhibit the activity of the transposon, e.g. by binding within its recognition sequence and obscuring the transposase binding sites, or by direct interaction with the protein.

5.7.3 Toxicity of *hobo* transposase to the host cells

In the experiments described, the expression of *hobo* transposase was found to be detrimental to the host cells. In *S. pombe*, one third of pKL6 (pESP-*hobo*) transformants lost the expression cassette for the GST-*hobo* transposase through recombination (see Section 3.5.1). In fission yeast, “these events typically occur at a low frequency” (Stratagene manual), but a frequency of 33% points toward a negative selection pressure against expression of the *hobo* transposase. In addition, a large number of dead cells was found in the growing cultures, even under repressing conditions (not shown).

In *E. coli*, T7 polymerase-expressed native and His-tagged *hobo* transposase reduces the viability of the host cells, as judged by low colony numbers (Table 5.2) and a high proportion of dead cells visible under the microscope (Figure 5.5). Similar to the situation in *S. pombe*, the transposase expression plasmid is lost extremely quickly if the antibiotic selection pressure is released (Section 5.4.4). Interestingly, the 10xHis-transposase expression plasmid is lost more quickly and has a stronger detrimental effect on survival than the plasmid encoding the native protein. It is conceivable that the tag mimics a modification of the *Drosophila*-expressed protein, which promotes correct folding.

The toxicity of the *hobo* transposase is especially interesting in the light of its low expression level. Highly expressed, potentially misfolded proteins trigger the stress response and inhibit vital functions by taking up a substantial part of the cytoplasm. But these effects should not be very pronounced from a protein that barely reaches 1% of total cell protein.

The toxicity is the only activity of the *hobo* transposase identified outside its natural host so far. It is not clear yet whether it is due to a partial transposition activity or due to an unrelated property. Both Tn5 and Tn10 transposase overexpression are lethal to *E. coli* and accompanied by similar filamentation phenotypes as observed for *hobo* transposase expression (Figure 5.5). In Tn5, the lethality and cell division defects are not due to the

transposition or DNA-binding activities of the transposase, but to a short N-terminal domain, which is thought to interact with an essential *E. coli* protein involved in cell division and/or chromosome segregation (Weinreich *et al.*, 1994b). A similar mechanism could be at work in *hobo* transposase expressing cells.

However, the observation that the presence of *hobo* mini-transposons increases the toxic effect of the transposase (Table 5.2) suggests that a transposition-related effect is also involved. Partial excisions and integrations of the mini-transposons would not have been detected in the experiments described, but they would be very detrimental to the host cells. The notion of DNA damage is supported by the observation that the *recA* mutant HMS174 strain is particularly sensitive to the presence of *hobo* transposase and mini-transposon.

The *hobo* transposase might also cleave the *E. coli* genome directly. In the published *E. coli* K12 genome sequence (4.64 Mb) (Rudd, 2000), the exact 12 bp *hobo* inverted repeat sequence is present once. Growing *E. coli* usually have two chromosomes on average (Skarstad *et al.*, 1983, 1985), so theoretically, synaptic complexes could form between the chromosomes, and DNA breaks could be inflicted by the transposase. Furthermore, ITR sequences with slight differences are probably also recognised by the transposase, and *E. coli* B strains (e.g. the employed expression strain BL21) have an additional 1.2 Mb of sequence and could contain more sequences that mimic the *hobo* inverted repeat.

The filamentous morphology of *hobo* transposase expressing bacteria supports the presence of DNA breaks in the bacterial chromosome. Severe DNA damage triggers the SOS response; which immediately inhibits cell division (via *recA*, *LexA* and *FtsZ*), but not cell growth, until the damage has been repaired. This could be tested by introducing the expression plasmid into SOS-reporter strains (e.g. Toman *et al.*, 1985).

The similar filamentous, lysed, fragmented and translucent morphology of *IS10* expressing cells further backs the notion that *hobo* transposase induces DNA damage by recombination. Several cryptic *IS10* ends have been detected in the *E. coli* chromosome. *In vitro*, recombination between similar cryptic ends in the *ColE1* origin of replication have been observed (Chalmers and Kleckner, 1996), so recombination of chromosomal sites might contribute to the lethal effect of overexpressed *IS10* transposase.

5.7.4 *hobo* transposase expression and activity in mouse fibroblasts

Cell lines with stably integrated *hobo* transposons were established and transiently transfected or transduced with transposase expression plasmids. Expression of EGFP-*hobo*

transposase fusions showed that the signal from fusion proteins was severely reduced compared to EGFP alone and that N-terminal tags gave a higher signal than C-terminal tags. This could be due to either higher expression of the entire protein, premature termination of translation before the C-terminal tag, or due to better folding of the EGFP at the N-terminal position. Expression of the EGFP-tagged transposases furthermore demonstrated localisation of the fusion protein; some protein was present in the nucleus, but the majority was located in the cytoplasm. To increase the concentration in the nucleus, a nuclear localisation sequence (NLS) could be added to the transposase gene. The NLS of SV40 large T antigen (PKKKRKV) (Kalderon *et al.*, 1984a, b) has been used successfully to target various proteins to mammalian nuclei (e.g. Dunn *et al.*, 1988; Lieber *et al.*, 1989). The effectiveness of an N- or C-terminal position of the NLS would have to be determined empirically with GFP-fusions.

On the other hand, a low concentration of *hobo* transposase might be advantageous for transposition activity, as it is for many other transposases such as *Ac* (Weil and Kunze, 2000), *Mos1* (Lohe and Hartl, 1996b), or *Himar1* (Lampe *et al.*, 1998). Detection by fluorescent microscopy requires at least a concentration of 100 nM or c. 4000 molecules of EGFP per cell. Compared with the optimum concentration of 8-10 nM for activity of *Himar1* transposase (Lampe *et al.*, 1998; Sections 6.4.3, 6.5.2), this indicates that a low level of fluorescent signal is probably beneficial for transposition.

DNA isolated from *hobo*-containing fibroblasts after the introduction of transposase contained no detectable excision events or hairpin structures. A correlation of transposase expression with cell necrosis or neoplastic transformations (indicating general mutations) was not observed. No transposon circles were conclusively identified by PCR (not shown). In summary, no activity of the *hobo* transposase was detected in mammalian cells.

5.7.5 Host specificity of *hobo*

Some eukaryotic transposons, such as *Himar1* have a very wide host range and transpose also in bacteria (Lampe *et al.*, 1999). Others, such as *P* elements, are only active in a very narrow range of closely related species (Chen, 1998; Robertson, 2002). The distribution and extent of horizontal transfer of elements among species is at least in part indicative of their ability to adapt to different hosts. The *hAT* family of transposons is widely distributed among eukaryotes, but individual members are only found in closely related species. There is good evidence for horizontal transfer of the *hAT* transposons *hobo* in the *D. melanogaster* species complex (Simmons, 1992) and *Tol2* in medaka fish *Oryzias*

The best chance to detect hairpins formed after *hobo* excision might still be the use of a carefully designed *in vitro* system. The transposase could be purified from insect cells as full-length protein with cleavable tags at both termini. Extracts from *Drosophila* embryos or other protein fractions could be added to provide necessary host factors. With the very high sensitivity of FDPCR on frayed linear transposon end substrates, even very low transposition activities would be detected.

CHAPTER 6 - *HIMAR1* TRANSPOSITION *IN VITRO*

6.1 Introduction

The reconstructed transposon *Himar1* from *Haematobia irritans* is the first *mariner*-like element (MLE) for which a functional *in vitro* system has been established (Lampe *et al.*, 1996, 1998). *Himar1* activity requires transposase as the only enzyme, and although the system has been used to transform a variety of prokaryotic and eukaryotic species (e.g. Akerley *et al.*, 1998; Zhang *et al.*, 1998; Rubin *et al.*, 1999; Pelicic *et al.*, 2000; Akerley and Lampe, 2002; Ashour and Hondalus, 2003; Ran Kim and Haeng Rhee, 2003), very little is known about the molecular mechanism of transposition. The same is true for *Mos1*, the second established MLE *in vitro* system (Tosi and Beverley, 2000). As MLEs are probably the most widespread eukaryotic mobile elements (Plasterk *et al.*, 1999), and as the available data suggests that there must be important differences to the transposition mechanism of other families of transposons, *Himar1* is a very worthwhile subject of study.

Himar1 transposition has so far mainly been analysed for transposition and insertion frequencies. In this study, the details of transposase-DNA interactions were analysed. Specific complexes could be identified which most probably are precursors of regular paired ends complexes. The protection pattern of these complexes were determined, which showed that transposase is bound directly only to one face of the ITRs. The high gel retardation, on the other hand, suggests that further molecules could be bound exclusively by protein-protein interactions. Insertion reactions performed with plasmid DNA substrates showed that after insertion, the supercoiling is retained in the target, even though these are not covalently closed. This indicates a very strong interaction of the transposase protein with the target DNA, a fact that was also visible by the relatively extreme methods necessary to remove the protein from strand-transfer complexes. Interactions with the target DNA were also analysed regarding target commitment. *Himar1* transposase was shown to exhibit target commitment only in the presence of moderate concentrations of crowding agents. Under these conditions, insertion takes place into the DNA encountered first, regardless of subsequent addition of another DNA species. This behaviour is reminiscent of the situation in V(D)J recombination, further strengthening the evolutionary link between these two systems. It was attempted to identify if catalysis proceeds *via* the

proposed hairpin mechanism. The data suggest that this is not the case, although clear evidence could not be obtained. These results could have been confounded by an exonuclease activity, which is thought to be intrinsic to *Himar1* and could possibly play a part in disassembly of the aforementioned stable strand-transfer complex.

6.2 Purification and activity of *Himar1* transposase

Initial experiments were carried out with aliquots of two different *Himar1* transposase preparations, kindly provided by David Lampe: (1) Wildtype, untagged *Himar1* transposase, recovered from inclusion bodies (Lampe *et al.*, 1996), and (2) the hyperactive C9 mutant Q131R/E137K (see Section 1.4.4), fused at the N-terminus to maltose binding protein (MBP) (Akerley and Lampe, 2002). The activity of both protein preparations was tested for insertions of short linear transposon end substrates and with plasmid substrates (Section 6.2.3). These comparisons confirmed the finding that the tagged C9 mutant is more active, but shows otherwise identical behaviour to the wildtype protein in all aspects tested (not shown). For all following experiments, the MBP-tagged C9-transposase was used.

6.2.1 Further purification of MBP-*Himar1* transposase

Himar1 transposase was initially purified as a fusion with maltose binding protein (MBP) using an amylose column and an early version of the published protocol (Akerley and Lampe, 2002) (Sections 2.3.4.1, 2.3.4.2; Figure 6.1A). The purification was extended by an additional FPLC purification step, which reduced non-specific nicking activity and increased the excision activity 3-fold (see Sections 6.2.2, 6.2.3).

The peak fractions from the amylose column were pooled, applied to a MonoQ anion exchange column and eluted with a linear gradient of NaCl (Section 2.3.4.3; Figure 6.1A,B). To enable elution from the column, the detergent n-decyl- β -D-maltopyranoside (nDMP) was added to all buffers. Transposase is recovered in two very broad peaks centred on fractions 13 and 47, eluting at 150 mM and 700 mM NaCl, respectively. Although the protein that eluted late in the gradient is more highly purified with respect to other proteins, most of the activity is associated with the small peak at fraction 11-13. The non-specific nicking activity was lowest with fraction 11. All subsequent experiments were therefore performed with MonoQ fraction 11, unless stated otherwise.

The 3-fold increase in specific activity is apparently a result of removing the largely inactive protein that elutes from the MonoQ column at 0.6–0.7 M NaCl, centred on fraction

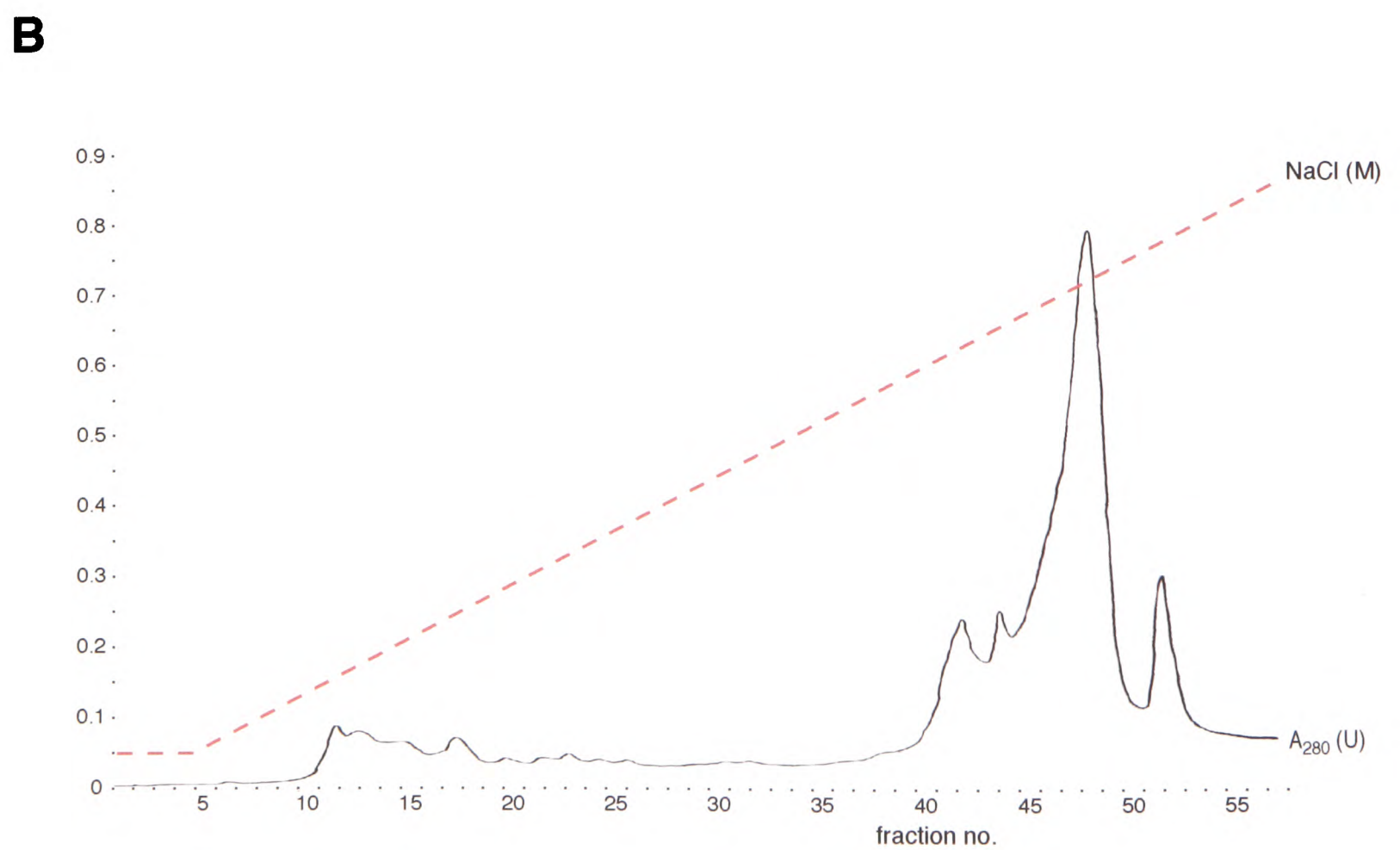
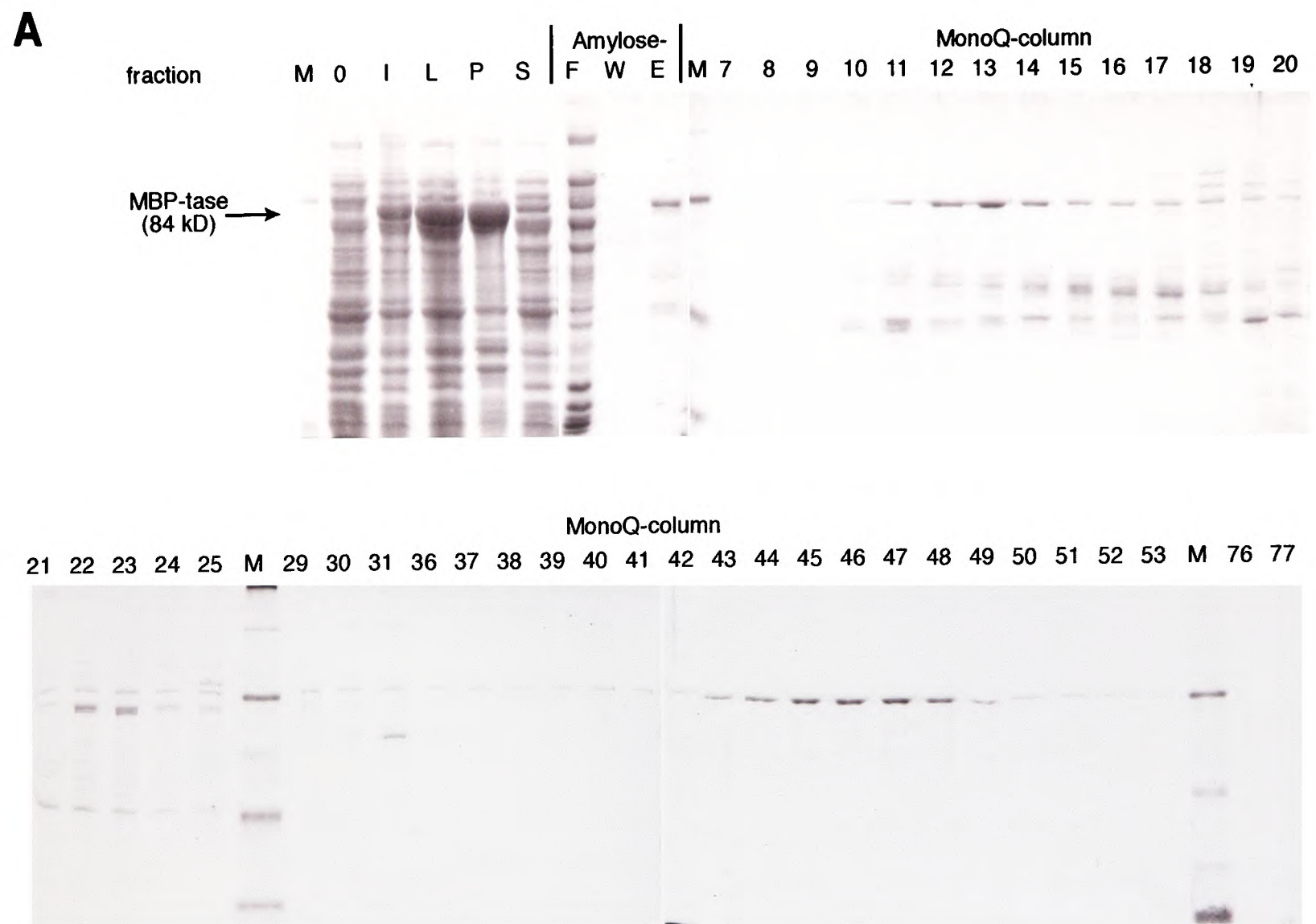


Fig. 6.1

Figure 6.1: Purification of MBP-*Himar1* C9 transposase

- (A) *Himar1* MBP-transposase fusion was purified by affinity chromatography on amylose resin as described in Sections 2.3.4.1 and 2.3.4.2. The pooled peak fractions were applied to a MonoQ column and eluted with a gradient of NaCl. The late peak is most highly purified with respect to other proteins but contains a significant amount of DNA that absorbs at 280 nm (see text). Coomassie stained 10% polyacrylamide SDS-Laemmli gels are shown. M, protein size marker; 0, uninduced culture, whole cells; I, cells induced for 2 h; L, lysate of induced cells after French press; P, pellet, insoluble fraction after 9,000x g spin of lysate; S, supernatant, soluble fraction; F, flowthrough; W, wash; E, eluate from amylose column; 7-77, MonoQ elution fractions, trace shown in (B).
- (B) Elution profile of the MonoQ column. Both the A_{280} of fractions and the molar concentration of the NaCl gradient in the elution buffer are shown on the same scale on the y-axis.

47. To our knowledge, it is unprecedented for a protein to elute from a DEAE column at so high a salt concentration. Fraction 47 has a much higher absorbance at 280 nm than would be expected from the amount of protein present (Figure 6.1A,B). Indeed, the position and pattern of the peaks in this region of the gradient is typical of DNA. Ethanol precipitation, ethidium bromide staining and DNaseI treatment of fraction 47 confirmed the presence of large amounts of DNA. It can therefore be concluded that much of the MBP-transposase recovered from the amylose column is tightly associated with DNA and is largely inactive.

6.2.2 Nicking activity

Both the amylose and MonoQ purified transposase are contaminated by an endonucleolytic activity that nicks supercoiled plasmids. This activity is independent of the presence of *Himar1* transposon ends (Figure 6.2A, lanes 1-5). This nicking activity might be due to an intrinsic activity of the transposase or due to a contaminating protein, co-purified with the transposase. One of the open questions is whether *Himar1* excision proceeds via hairpin intermediates. As the observed non-specific nicking activity would destroy any putative hairpin, separation of the transposase protein from any contaminating activities was attempted.

Compared to the original purification protocol using only one amylose column, a significant reduction in nicking activity was achieved by addition of the second purification column, either the MonoQ column or a second amylose column (Figure 6.2A,B). This suggests that at least some of the original nicking activity is due to a contaminating protein. Several additional attempts were undertaken to abolish the nicking activity, but all proved unsuccessful (not shown). MBP-transposase was expressed from two endonuclease deficient strains of *E. coli* (ER2566 and DH5 α , see Table 2.1). The amylose column was washed extensively with various combinations of detergents, denaturing reagents and high salt concentrations. All the reagents, such as a low concentration of sarcosyl, that removed the nicking activity, also inactivated the transposase. In ammonium sulphate precipitations, the transposase precipitated together with the nicking activity. Lastly, the amylose column peak fraction was further purified by gel filtration on a Superose 6 column. No resolution of the MBP-transposase and the nicking activity was obtained in any of these experiments. It therefore appears that the remaining nicking activity is either in a protein that is extremely tightly bound to the transposase, or that it is an integral part of the transposase itself.

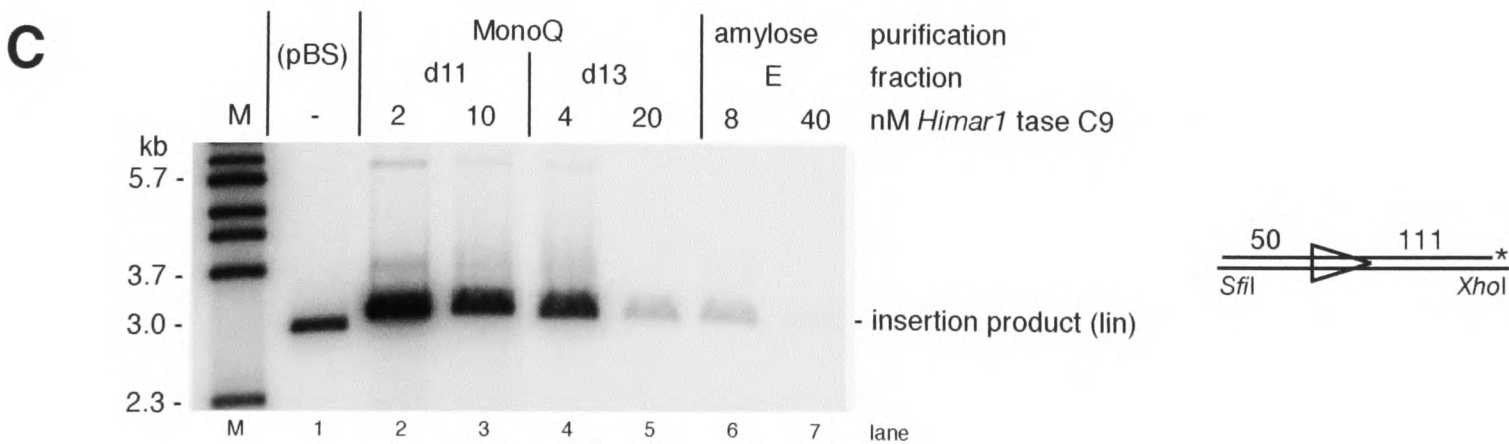
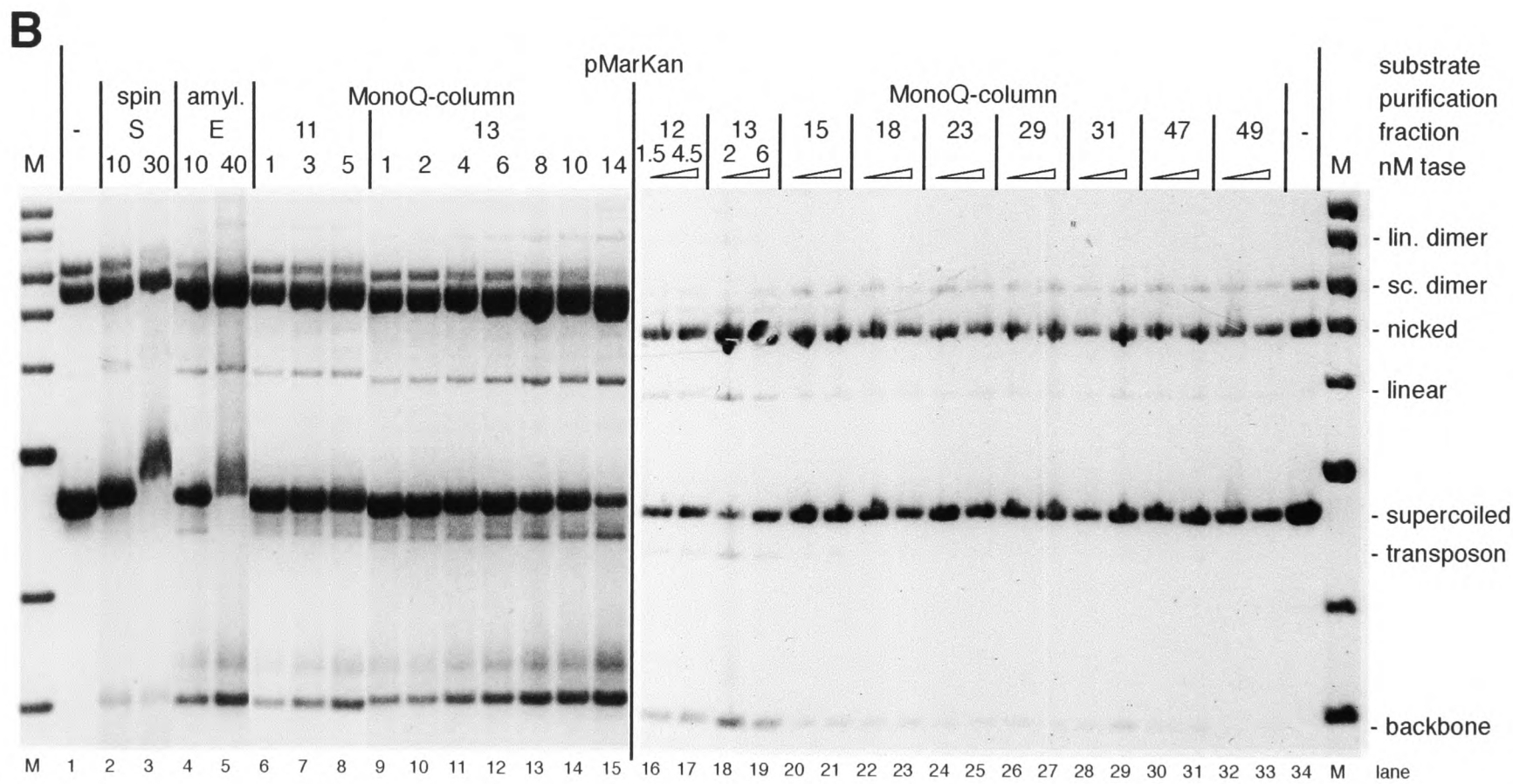
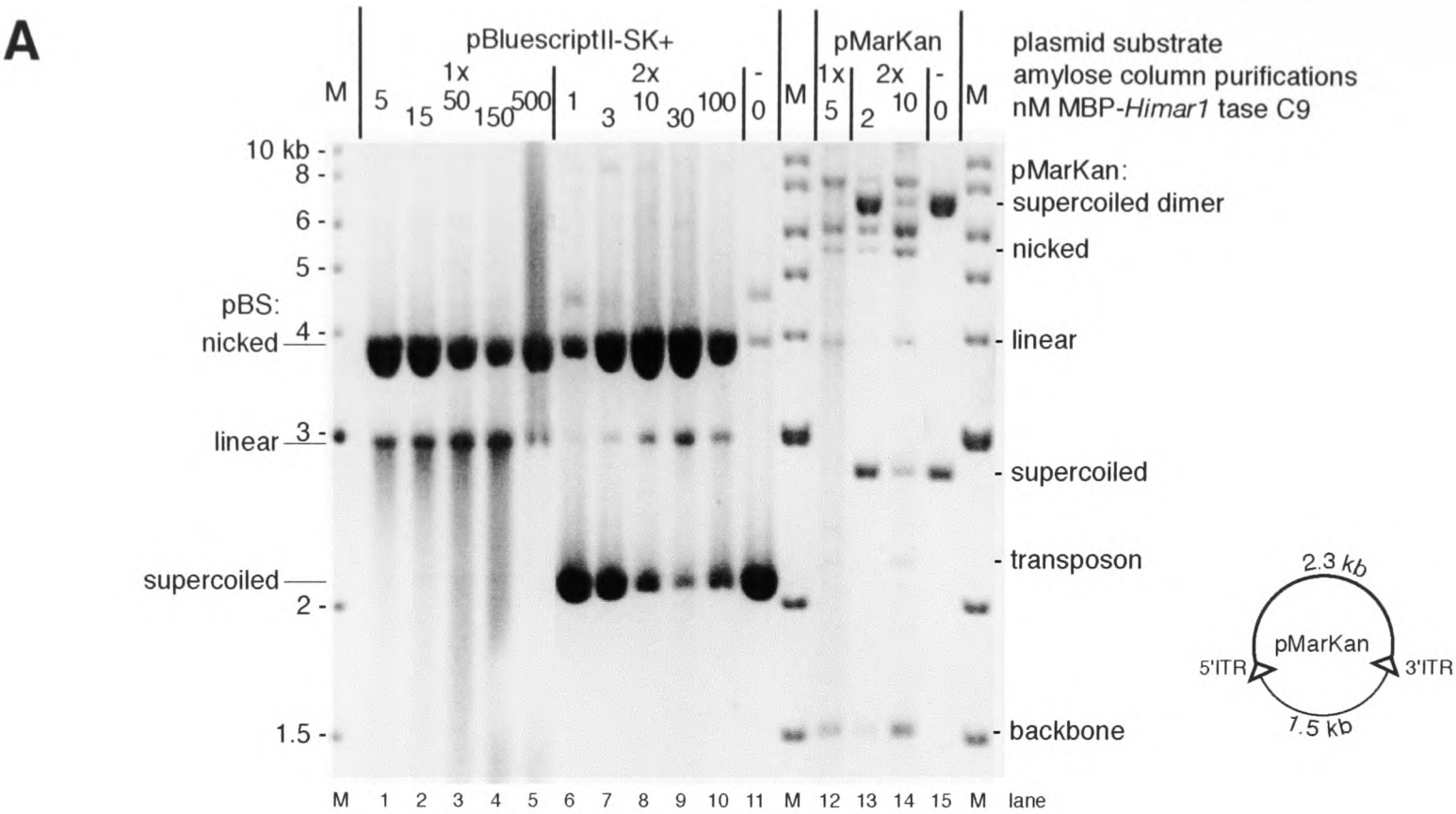


Fig. 6.2

Figure 6.2: Activities of *Himar1* transposase purifications

- (A) Nicking assay. Non-specific and specific plasmid-substrate reactions with MBP-*Himar1* transposase C9, purified with one or two passages through the amylose column (“1x” and “2x”, respectively). In plasmid substrate reactions, 250 ng of either pBluescriptII-SK+ (non-specific) or pMarKan (specific substrate) was incubated with the indicated transposase concentration and 10 mM MgCl₂ for 6 h at 30°C in 20 µl. Reverse contrast image of a 1.1% agarose TBE gel, stained with EtBr. Compared to the unreacted plasmid substrate (lane 11), incubation with 1x-purified MBP-transposase (lanes 1-5) transforms all of the supercoiled plasmid to nicked and linear product. In 2x-purified transposase (lanes 6-10), this nicking activity is reduced, as visible by a significant amount of supercoiled DNA. The same pattern is observed with pMarKan, when comparing the strength of the supercoiled band between reactions incubated without transposase (lane 15), 1x purified transposase (lane 12) and 2x-purified transposase (lanes 13-14). For a description of the bands formed at transposition from pMarKan, see Figure 6.3.
- (B) The total activity of purified *Himar1* transposase fractions was determined by measuring the amount of transposon excision from the supercoiled plasmid substrate, pMarKan. Each reaction (200 µl) contained 1 µg (396 fmol) of pMarKan and the indicated concentration of the respective transposase and 10 mM MgCl₂. Incubation was for 1.5 h at 30°C, followed by DNA-precipitation. Reverse contrast images of ethidium bromide stained 1.1% agarose, TBE gels are shown. Lane 1, 1 µg unreacted pMarKan; lanes 2-15, full reaction loaded; lanes 16-33, 80% of reaction loaded; lane 34, 0.8 µg unreacted pMarKan. Transposase concentrations were measured for the amylose-column eluate (lanes 4-5) and estimated for the other fractions, lanes 2-19, from Coomassie-stained SDS-PAGE. Compare lanes 4 and 14 for the 3-fold increase of transposition activity by MonoQ purification at equal transposase concentrations. Lanes 16-33: In even numbered lanes, 1 µl transposase fraction was used per reaction, in odd numbered lanes, 3 µl was used (see Figure 6.1A for estimation of concentrations). S, supernatant after ultracentrifugation of lysate; E, eluate from the amylose column. See Section 2.5.2.1 for details of the experimental setup.
- (C) Purified transposase fractions were tested for their ability to perform concerted insertion of a pair of transposon ends encoded on a linear radiolabelled DNA fragment into a supercoiled plasmid. Concerted insertion of transposon ends linearises the target plasmid. See Sections 2.5.2.2 and 6.3.3 for details of the reactions. The reactions (15 µl) contained 53 fmol (250 cps) of transposon end 99SX*, the indicated concentration of transposase and 63 fmol (125 ng) supercoiled pBluescriptII-SK+ as target. The transposon end fragment 99SX* was prepared by digesting pKL99 with *Sfi*I + *Xho*I (Table 2.10; Figure 2.1), and radioactively labelled by end filling (Section 2.2.3.1). Linear pBluescriptII-SK+ and a lambda *Bst*EII digestion were end labelled to act as size markers. The size of the transposon substrate is shown to the right of the gel. *, position of radioactive label on the transposon arm; the flanking DNA is unlabeled.

6.2.3 Activity tests for the purified transposase

The relative activities of the amylose and the MonoQ purified transposases were measured in a transposition assay (Figure 6.2B). The substrate is a supercoiled plasmid containing a mini-*mariner* construct in which a kanamycin marker is inserted between the inverted repeats (pMarKan, Table 2.8). After excision, the transposon is expected to insert into any DNA, which it encounters to produce a range of inter- and intra-molecular products that are difficult to quantify (Chalmers and Kleckner, 1996). However, excision of the transposon from the plasmid leaves behind the 1.5 kb plasmid backbone, which is an end product of the reaction and is a good measure of the total transposase activity. When 10 nM of each transposase was assayed, MonoQ fraction 13 has about 3-fold higher activity than the amylose-purified material (as estimated by eye) (Figure 6.2B, lanes 4,14).

The activity of the transposase preparations was also tested in an insertion assay (Figure 6.2C). The substrate was a linear DNA fragment encoding the terminal 61 bp from the 5'-end of *Himar1*, together with 50 bp of flanking DNA. The DNA fragment was labelled on the transposon end and supercoiled pBluescript was provided as target. Concerted insertion of a pair of transposon ends is expected to yield a radioactively labelled linear product 122 bp larger than the target provided. A product of the correct size is indeed detected (Figure 6.2C, lanes 2-7). The increase in activity by the anion exchange column was found to be even more pronounced for insertion reactions than for excision from a plasmid substrate.

6.2.4 Systematic variation of the *Himar1* *in vitro* reaction conditions

In order to facilitate analysis of delicate reaction intermediates and products, the reaction conditions were optimised for maximal excision and insertion activity and minimal non-specific nicking activity (data not shown). The *in vitro* activity of purified MBP-*Himar1* transposase was tested in different conditions with plasmid substrate reactions and with insertion reactions, as described in Section 6.2.3. Despite the different substrates (plasmid versus short ends) and purification protocols, the same optimal values were obtained for temperature, Mg^{2+} , NaCl, BSA and transposase concentration as had been published (Lampe *et al.*, 1998). In addition, the following parameters and buffer components were tested:

Exchanging Mg^{2+} for Mn^{2+} allowed a comparable level of excision, but significantly reduced inter- and intramolecular insertion. Glycerol ($\geq 30\%$), DMSO ($\geq 30\%$) and Ethidium bromide (≥ 0.2 mg/ml) were found to inhibit all aspects of the reaction, both in

the presence of Mg^{2+} or Mn^{2+} as the divalent cation. 10% DMSO only had very mild effects. Ca^{2+} concentrations above 10 mM (≥ 20 mM, in addition to 10 mM Mg^{2+}) and potassium glutamate (≥ 20 mM) are also inhibitory. Slightly higher reaction efficiencies are obtained if HEPES-buffer (pH 7.9) is used compared to Tris-buffer (pH 8.0). tRNA (50 μ g/ml) was found to be advantageous for Tris-buffered reactions, but it does not have a significant effect on the HEPES buffered ones. The most significant enhancement (5-10-fold) was achieved by the addition of 20 mM NH_4Cl to the reactions. It was also noted that, as expected, excision is more efficient in larger reaction volumes and insertion is enhanced in smaller volumes (not shown).

6.3 Catalytic activities of the *Himar1* transposase

6.3.1 Effect of the conformation of the substrate DNA on excision and insertion

One of the defining differences between prokaryotes and eukaryotes is the structure of the genome. Most eubacteria have circular and negatively supercoiled genomic DNA. Eukaryotic chromosomes, on the other hand, are linear and associated with dynamic chromatin structures. For DNA transposons, these differences have important mechanistic implications. For example, excision of the prokaryotic transposon *Tn10* is driven by supercoiling (Chalmers *et al.*, 1998). The closely related transposon *Tn5*, on the other hand, can excise from linear DNA (Reznikoff, 2002). We were therefore interested whether the conformation of substrate DNA had any effect on *Himar1* transposition.

Transposition reactions were set up with different topological forms of the mini-*mariner* plasmid pMarKan (Figure 6.3). The plasmid was either left in its supercoiled conformation (Figure 6.3B, lanes 1-2) or was linearised by cutting either within the donor backbone (lanes 3-5), within the transposon (lanes 6-8) or within both segments (lanes 9-11). Excision, as measured by accumulation of the excised backbone, was equally strong for all four substrates (Figure 6.3B, lanes 2,4,7,10, band “backbone”). *Himar1* excision is therefore neither dependent on nor inhibited by supercoiling. This is in contrast to *Tn10*, which requires supercoiling for excision in the absence of IHF (Chalmers *et al.*, 1998). Surprisingly, this is also different from *Mos1*, where excision is visible from linear, but not from supercoiled donor DNA (Y. Bigot, personal communication).

Insertion is more difficult to quantify in this assay, as it produces a number of large products, dependent on the structure and size of the chosen target. However, incubation of

A



B

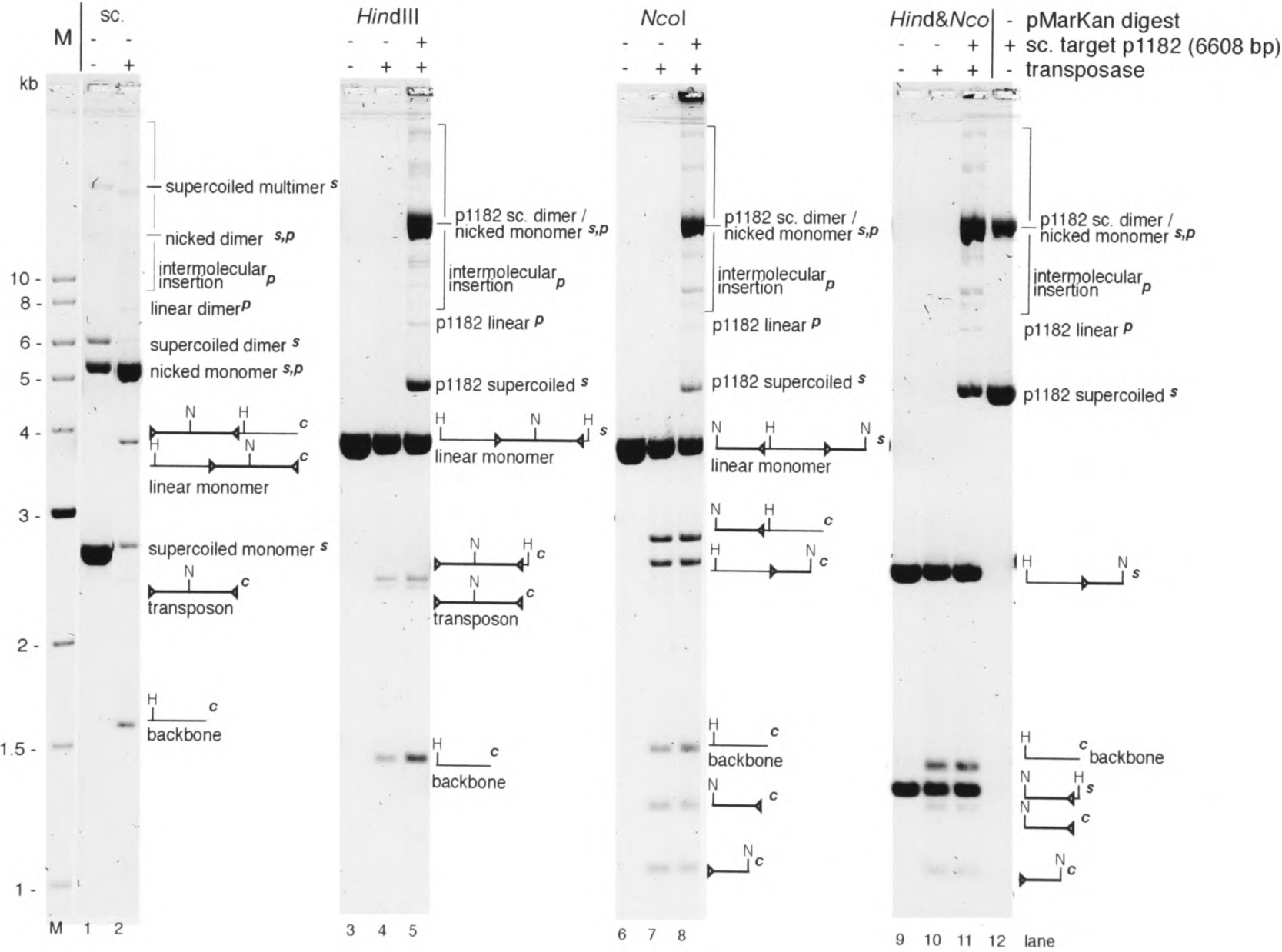


Fig. 6.3

Figure 6.3: *Himar1* transposition with supercoiled and linear plasmid substrates

Himar1 transposition reactions with plasmid substrate pMarKan: A first nick relaxes the supercoiled substrate to an open circle, which has a lower mobility (“nicked monomer”). A double strand break (at one of the transposon ends or due to excessive random nicking) linearises the substrate (to 3.8 kb, “linear”). A second double strand break for full excision releases the transposon (2.3 kb) and the plasmid backbone (1.5 kb). For complete transposition, the transposon inserts into another substrate molecule. If this target molecule is the supercoiled plasmid, the insertion product is a relaxed circular plasmid product of 6.1 kb (= 3.8+2.3 kb) length, which runs between nicked monomer and nicked dimer. Insertions into plasmid multimers have an even lower mobility. For reasons that are not entirely understood, pMarKan was found to spontaneously multimerise in *recA* mutant bacteria, making it impossible to get pure monomeric plasmid substrate directly from the cells. This slightly complicates the interpretation of the results.

- (A) Mini-*mariner* plasmid pMarKan, graphic representation. Thick lines, transposon; thin lines, backbone; triangles, inverted terminal repeats; N, *NcoI*, H, *HindIII* site.
- (B) Comparison of transposition reactions with linear or supercoiled DNA substrate. 1 μ g linearised or supercoiled pMarKan was incubated in reaction buffer with 10 mM $MgCl_2$ and 8 nM MBP-*Himar1* transposase C9, MonoQ fraction 13 for 1.5 h at 30°C. In lanes 5,8,11,12, 1 μ g supercoiled plasmid p1182 was included, which does not have any *Himar1* sequences. Fluorimager scan of an EtBr stained 1.1% agarose gel. sc., supercoiled (uncut); M, 1 kb ladder. Graphic representation as in (A). Indices: **s**, substrate DNA; **c**, cleavage products (specific cleavage at ITRs); **p**, other products of the reaction.

the supercoiled substrate plasmid with transposase results in a number of bands of low mobility, which are identifiable as insertion products (Figure 6.3B, lanes 1-2). With supercoiled substrate, insertion is thus weak, but detectable. When the only DNA present was linear, on the other hand, no insertion was detectable (lanes 4,7,10). Only if another supercoiled target plasmid (p1182, lane 12, Table 2.8) was added to the reactions, did insertions take place at detectable levels (lanes 5,8,11). Insertion is therefore significantly more efficient if the target DNA is supercoiled. This had not been observed directly before and fits the finding that *Himar1* prefers bent genomic DNA targets to straight ones (Lampe *et al.*, 1998; see Discussion 6.6.3.3).

Separating the transposon ends onto two linear molecules does not have a significant effect on either excision or insertion (lanes 7-8,10-11). This confirms that no supercoiling is required in the transposon sequence, as is also seen in insertion reactions of small labelled transposon end fragments (e.g. Figure 6.2C). In such assays with radioactively labelled ends, some insertion into linearised target DNA was observed (not shown) but, as here, insertion is more efficient if the target is supercoiled.

6.3.2 Analysis of excision products and intermediates

We were interested to discover whether MLEs create a DNA hairpin intermediate during excision. In order to detect hairpins formed during *Himar1* transposition, *in vitro* transposition reactions were performed with the plasmid substrate pMarKan in the absence or presence of genomic bacterial DNA to absorb any non-specific nicking activity. Reactions incubated for 40, 65 and 90 min in buffer containing Mg^{2+} were analysed by two-dimensional agarose gel electrophoresis, with the first dimension running in TBE and the second dimension in 50 mM NaOH (either directly or after elution of candidate bands from the first dimension). The gels were Southern blotted and hybridised with probe made from linearised pMarKan (not shown). Among the many signals obtained, none was in the correct position for either the donor-backbone or the released transposon fragment to contain hairpins on either one or two ends.

As an alternative method, radiolabelled *Himar1* ends were incubated with transposase as described, and analysed on sequencing gels (Figure 6.4A,B). Substrates were designed such that any hairpins formed during the reaction would appear above the position of the unreacted substrate. No band was detected at the correct position for a hairpin on either the flanking DNA (Figure 6.4A) or the transposon end (Figure 6.4B), at eleven time points

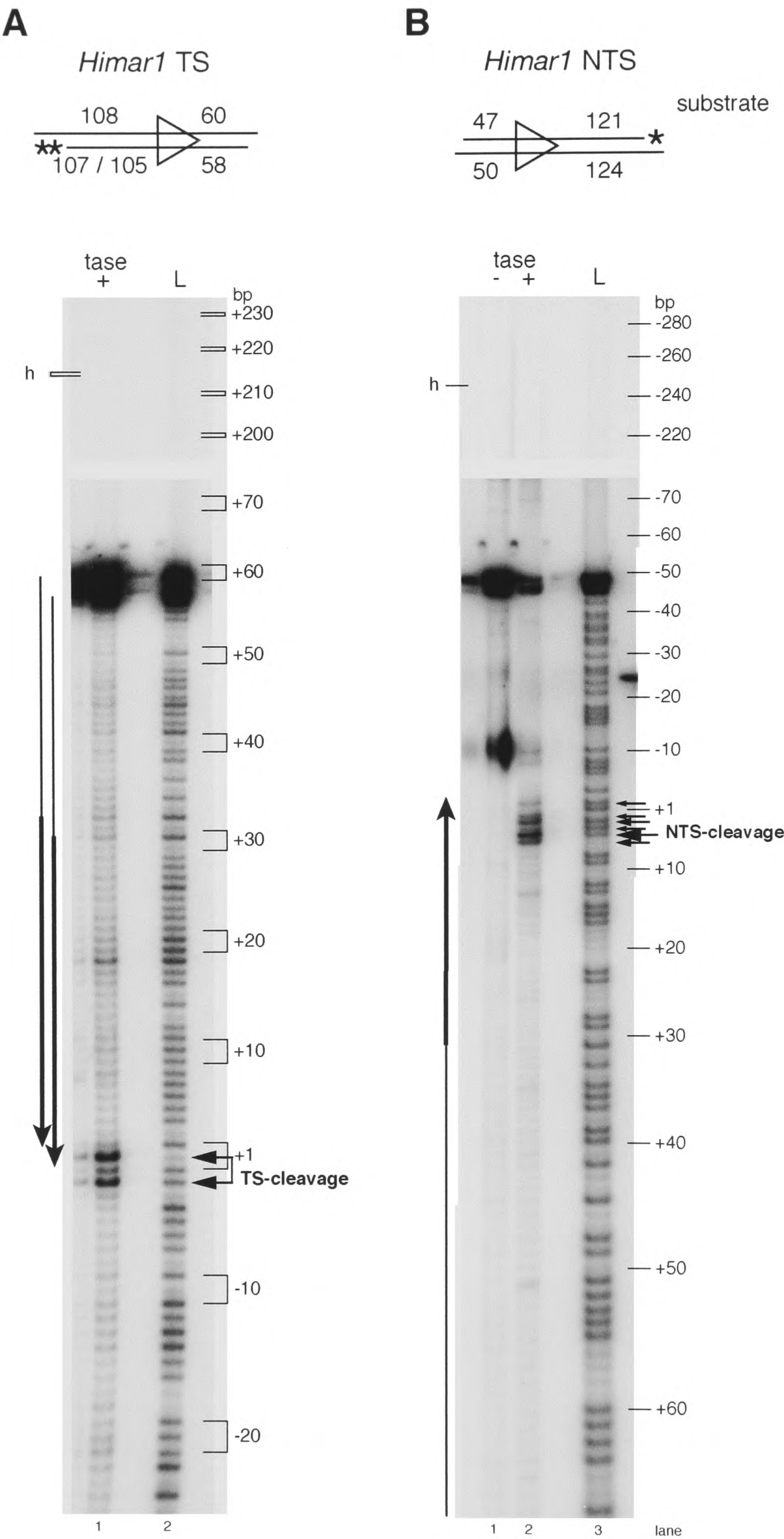


Fig. 6.4

Figure 6.4: *Himar1* excision

In vitro transposition reactions with 1 nM MBP-*Himar1* transposase (MonoQ fractions), 200 cps radiolabelled transposon end substrate per lane and 10 mM MgCl₂ are shown. Arrows, transposon end; bold arrow, ITR; L, Maxam-Gilbert G&A-ladder; h, expected position of putative hairpin. To determine the position of the larger size markers, a logarithmic curve was fitted through the coordinates of the visible markers and extended towards the top of the gel.

- (A) Substrate 97BD*, *Himar1* ITR-left, transferred strand labelled, incubated with MonoQ fraction 11 for 19 h at 30°C. Phosphoimager scan of 6% sequencing gel. Two α -³²P-dCTPs (asterisks) are incorporated at position 2 and 4 on the bottom strand of the flanking DNA, resulting in double bands on the sequencing gel. Size of the putative hairpin: 213/215 bp. The transferred strand is cleaved precisely at bp -1.
- (B) Substrate 97SX*, *Himar1* ITR-left, non-transferred strand labelled, incubated with MonoQ fraction 11 for 1.5 h at 30°C. Phosphoimager scan of 8% sequencing gel. Size of the putative hairpin: 216 bp. The non-transferred strand is cleaved imprecisely at bp -1 to +13. NB: The black dot around bp -10 in lane 1 is a contamination on the gel.

from two minutes to 19 h of incubation in MgCl_2 (not shown). The expected positions of the putative hairpins are indicated on the gels as “h”.

It is known that Mn^{2+} can lead to an increase of hairpin intermediates in related systems (Bolland and Kleckner, 1995). Experiments using Mn^{2+} instead of Mg^{2+} as the only divalent metal ion were carried out and analysed on sequencing gels as above. Here, no hairpin was detected either (not shown).

Excision was observed to occur precisely at bp-1 on the transferred strand, as described (Lampe *et al.*, 1996)(Figure 6.4A, lane 1, indicated “TS-cleavage”). Cleavage of the non-transferred strand was even more imprecise than published (Lampe *et al.*, 1997; Lampe *et al.*, 2001), varying from bp-1 to +13, with the strongest signals at bp +5, +6 and +3, i.e. the main cleavage product was recessed by 5 bp (Figure 6.4B, lane 2, indicated “NTS-cleavage”).

6.3.3 The *in vitro* insertion product retains a supercoiled conformation

Properties of the *Himar1* insertion reaction were analysed with the transposon ends encoded on radioactively labelled linear DNA fragments and unlabeled supercoiled plasmid as target (Figure 6.5A; see also Sections 2.5.2.2, 6.2.3; Table 2.10). Concerted insertion of a pair of transposon ends close to each other on opposite strands of the DNA gives a linear product slightly larger than the target plasmid. Insertion of just one transposon end results in a nicked circular product.

The formation of insertion products was confirmed to be specific to the presence of *Himar1* transposase and the *Himar1* inverted repeats in the DNA substrate: Labelled DNA fragments that carry the *hobo* inverted repeats are not inserted (Figure 6.5B).

When the reactions are analysed on a TAE-agarose gel, the majority of reaction products does not show the expected mobility of linear or nicked DNA (Figure 6.5B). Instead, the majority of the signal is located just above the position for supercoiled target plasmid, far below linear (denoted “linB” for linear, bottom). A small percentage of the product runs just above the linear plasmid (“linT” for linear, top).

The cause for the altered mobility of insertion products was investigated by different treatments of the insertion reactions prior to electrophoresis (Figure 6.6A). Addition of SDS (0.15%) and EDTA (15 mM) has no effect (lanes 1,5). Proteinase K digestion achieves partial conversion to the form with linear mobility (lanes 3,7), whereas both phenol-chloroform extraction (lanes 4,8) and heating the reactions for ten minutes at 75°C

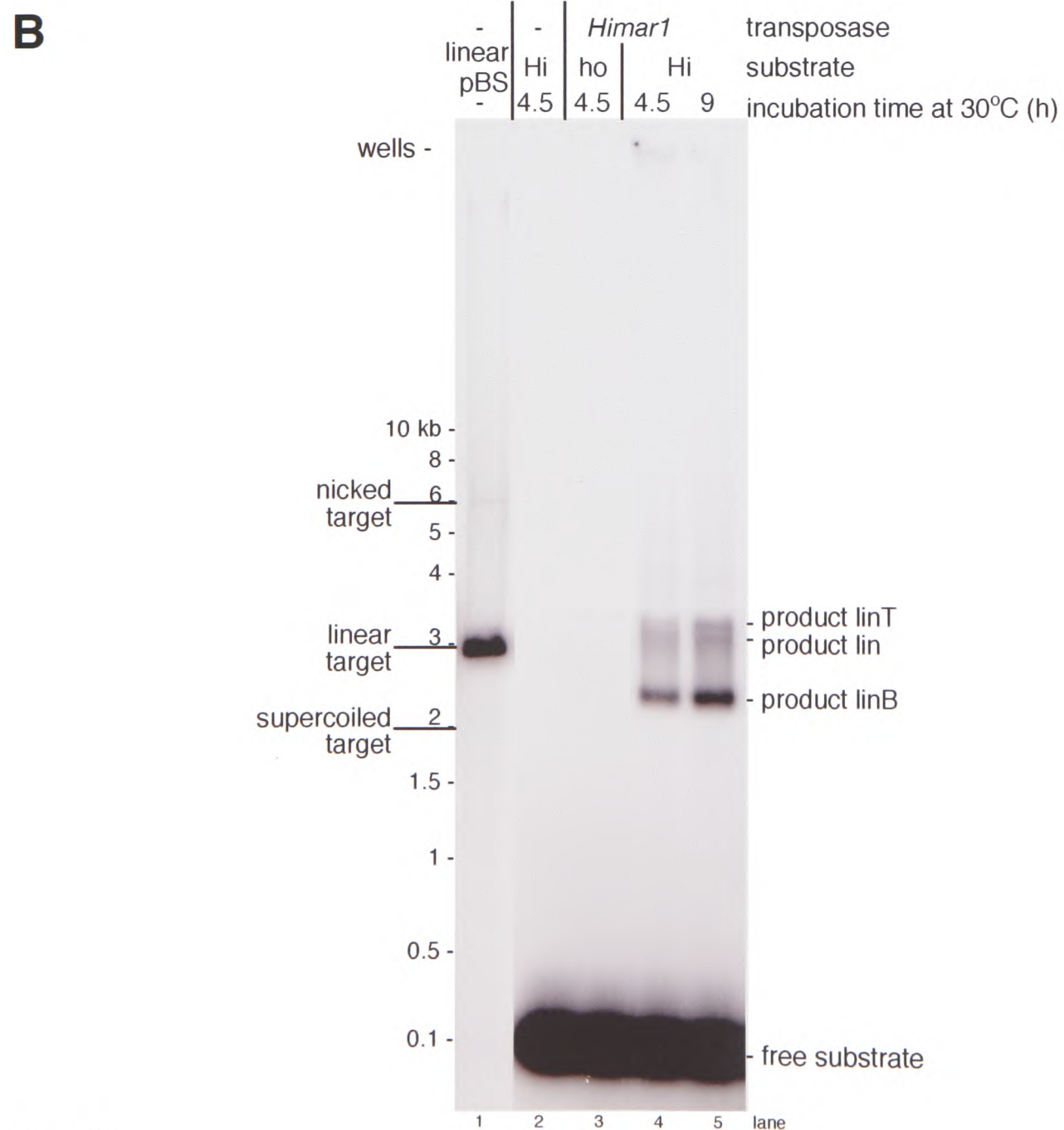
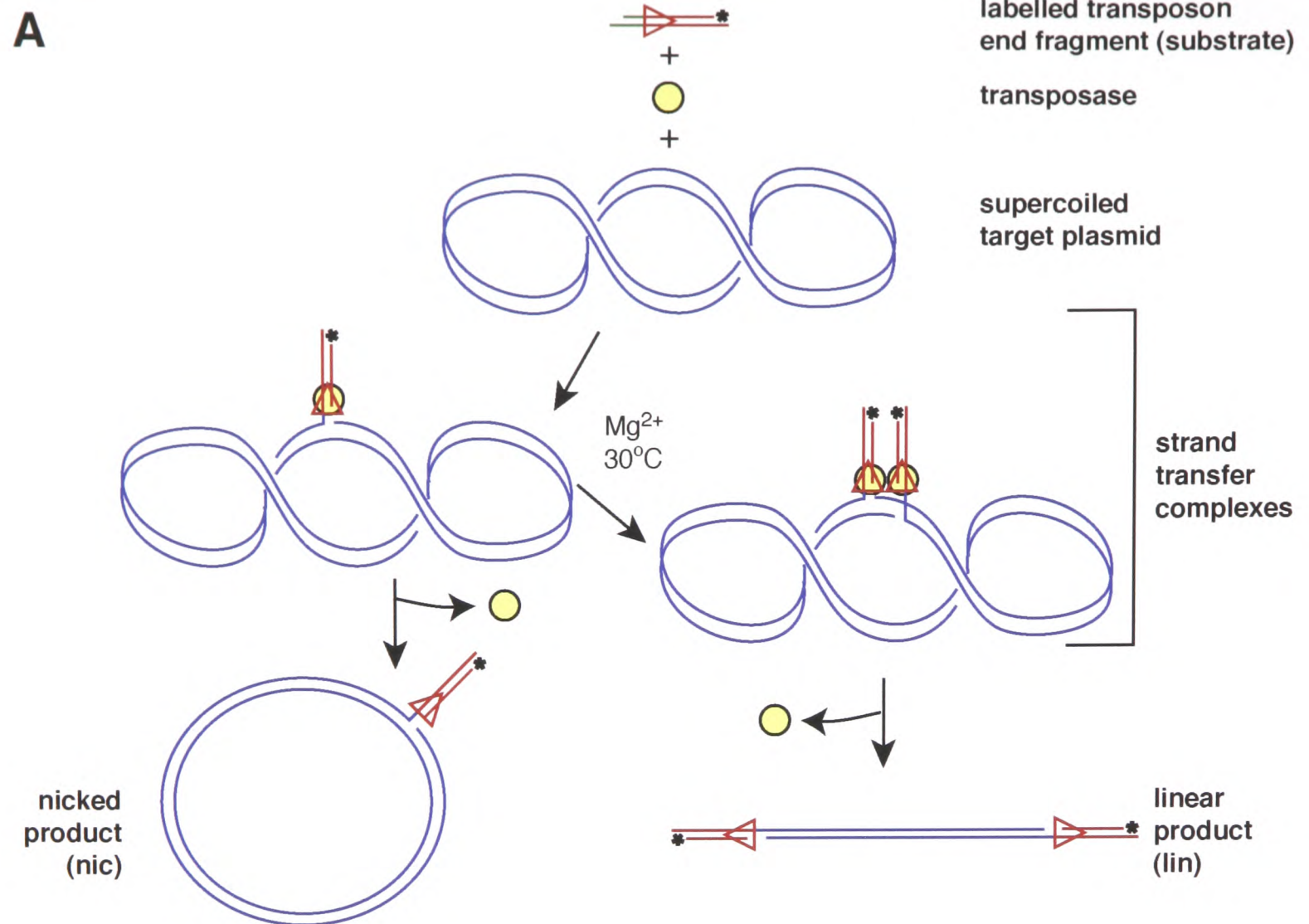
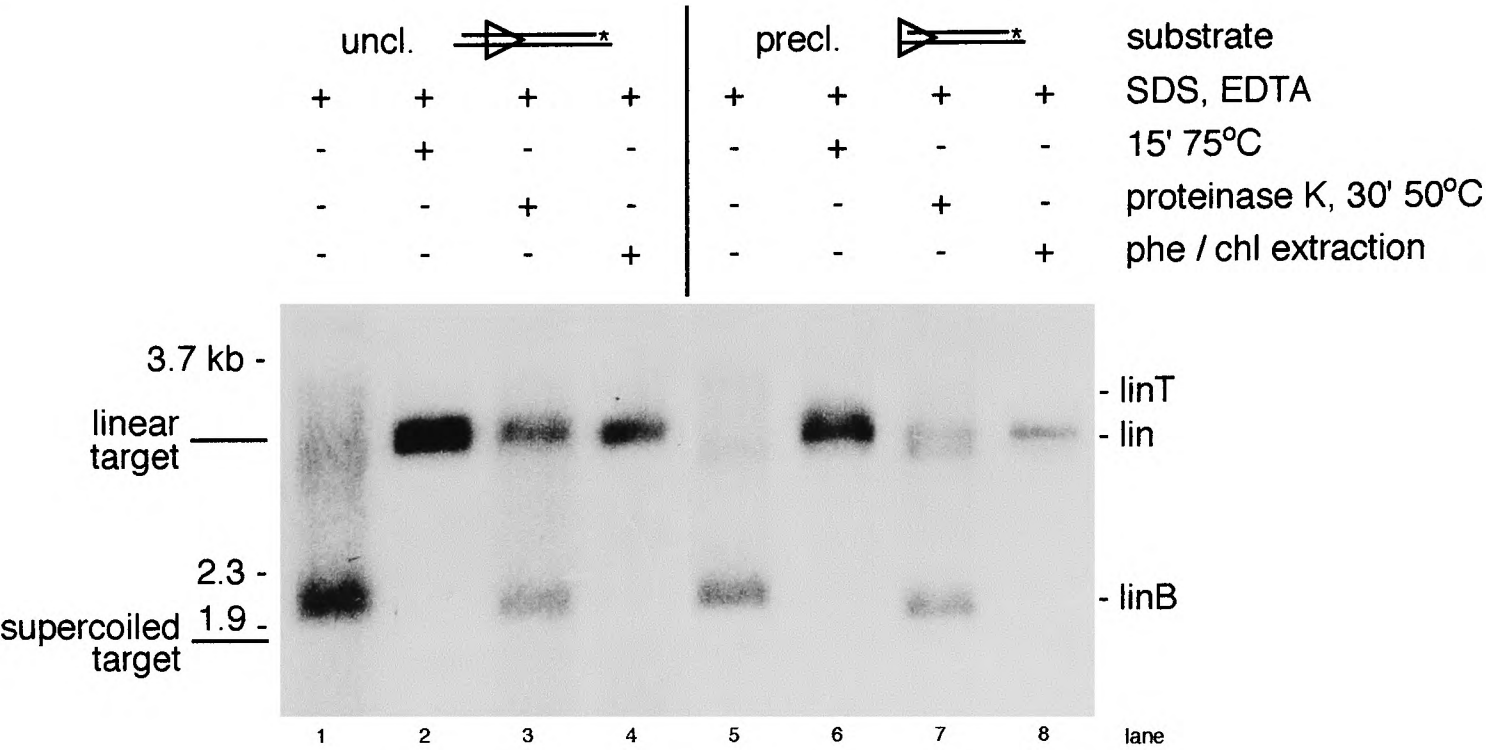


Fig. 6.5

Figure 6.5: Intermolecular insertion reactions

- (A)** Insertion of free transposon end fragments into supercoiled target plasmid; schematic of model. At insertion, the transferred strand of each transposon end is joined to one strand of the target plasmid by direct transesterification, which also creates a nick in the target strand. After dissociation of the insertion complex, the product of a one-ended insertion is therefore a nicked, relaxed circle; the product of a two-ended insertion has a double strand break and is linear. (Not to scale.)
- (B)** Reactions. Hi, *Himar1* substrate 99SX*; ho, *hobo* substrate 102SX*. Lane 1, labelled, linear pBluescript; lanes 2–5, insertion reactions with 10 mM MgCl₂, 10 nM MBP-*Himar1* C9 transposase. All lanes were loaded directly in gel loading buffer on a 0.7% agarose, 1x TAE gel and electrophoresed for 11 h at 65 V, 4°C. First, the gel was stained in EtBr and photographed for determining the positions of the marker and target plasmid standards running in adjacent lanes (not shown). Subsequently, the gel was dried and analysed by autoradiography.

A



B

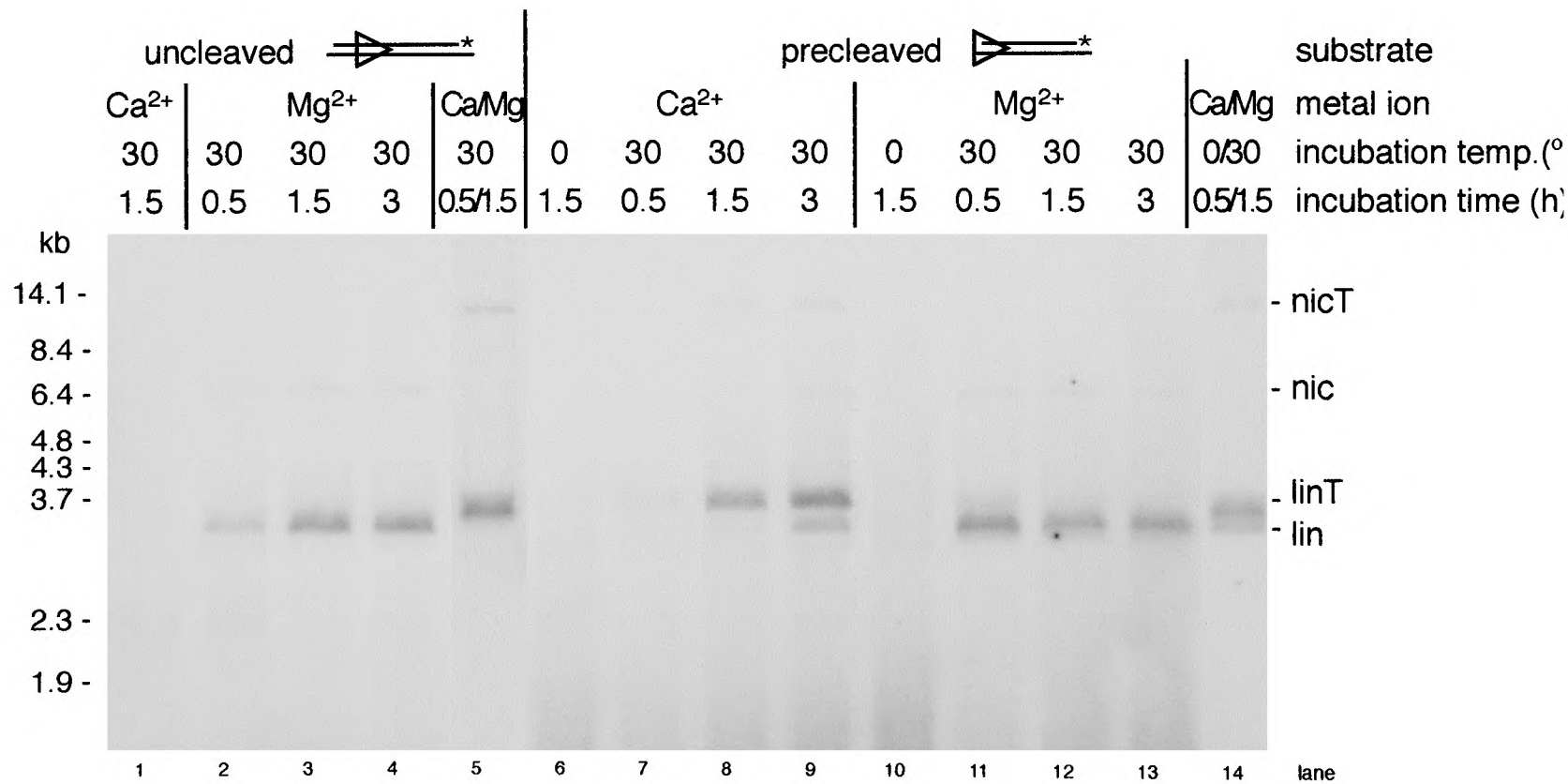


Fig. 6.6

Figure 6.6: Topology and stability of insertion products

Insertion reactions with 1 nM (15 fmol) MBP-*Himar1* C9 transposase into 250 ng pBluescript per lane. Uncleaved ends, 97SX*; precleaved ends, 105BX*, both 500 cps, 120 fmol per lane.

- (A)** Effect of treatment of completed insertion reactions on mobility. One 67.5 μ l insertion reaction for each substrate was set up in 10 mM MgCl_2 . After an incubation of 1.5 h at 30°C, 15 μ l aliquots were treated as follows: Lanes 1,5: addition of 1.5 μ l 10x stop mix to a final concentration of 0.15% SDS, 1.5 mM EDTA; lanes 2,6, addition of 1.5 μ l 10x stop mix, 15 min incubation at 75°C; lanes 3,7, addition of 1.5 μ l 10x stop mix and 1 μ g proteinase K, 30 min incubation 50°C; lanes 4,8, addition of TE to a total volume of 100 μ l, phenol/chloroform/IAA extraction, chloroform extraction, EtOH-precipitation with 10 ng glycogen, wash in 70% EtOH, pellet dissolved in 15 μ l TE, addition of 1.5 μ l 10x stop mix. All reactions were mixed with 1.7 μ l 8x gel loading buffer. Phosphoimager scan of a 0.7% agarose, 1xTBE gel, run for 14 h at 65 V, 4°C.
- (B)** Effects of metal ions on the generation and mobility of insertion products. Reactions performed as in (A), but with 2 mM CaCl_2 and/or 10 mM MgCl_2 , as indicated and with 4 nM (60 fmol) transposase. All reactions were stopped by incubating for 15 min at 75°C in the presence of SDS and EDTA. lin, position of linear insertion product; linT, top linear band; nic, nicked insertion product; nicT, top nicked band (see Figure 6.14). Lanes 5 and 14 are from a different gel, but aligned to the same marker. Here, the reactions were first incubated for 30 min in 2 mM CaCl_2 at 0°C (lane 5) or 30°C (lane 14), then MgCl_2 to a final concentration of 10 mM was added and incubation continued at 30°C for 90 min.

in the presence of SDS/EDTA (lanes 2,6) result in all reaction products having the mobility of linear insertion products. No difference was observed if uncleaved ends (lanes 1-4), blunt precleaved ends (not shown) or precleaved ends with a 2 bp 5' recess (lanes 5-8), were used.

The effect of these treatments suggest that the altered mobility of the insertion products linB and linT is caused by the protein remaining tightly attached to the DNA after insertion, resulting in an extremely stable strand transfer complex. From their position in the gel, the DNA in the linB product must have a supercoiled conformation, whereas the DNA in the linT product must be linear (see Figure 6.14). The linear topology of the heated product was confirmed by two dimensional gel electrophoresis in which the second dimension contained ethidium bromide (as in Chalmers *et al.*, 1998) (Figure 6.7). As the heating procedure causes less loss of DNA than the phenol/chloroform extraction, it was used for all further experiments.

6.3.4 Calcium supports strand transfer in *Himar1* transposition and stabilises the STC

Himar1 transposase contains a DDE motif, the catalytic triad that has been studied in great detail in other transposases and recombinases. The DDE residues serve to coordinate two divalent metal ions required for catalysis (Kennedy *et al.*, 2000; Beese and Steitz, 1991; Joyce and Steitz, 1995; Lovell *et al.*, 2002; Davies *et al.*, 2000). Divalent metal ions probably also have a structural role in the active site as they have as many points of coordination with the phosphate oxygens of the DNA as with the DDE residues (Davies *et al.*, 2000; Kennedy *et al.*, 2000). Indeed, the presence of divalent metal ions has been shown to promote the assembly and/or stability of transposon nucleoprotein complexes (Savilahti *et al.*, 1995; Sakai and Kleckner, 1997; Hiom and Gellert, 1998).

Despite this overall functional similarity, transposases differ in their specific requirements for divalent metal ions. Ca^{2+} does not support any of the chemical steps in Tn10 transposition (Sakai and Kleckner, 1997). However, it will support strand transfer in phage Mu and V(D)J transposition when the substrates are prenicked and precleaved, respectively (Savilahti *et al.*, 1995; Hiom and Gellert, 1997). I therefore set out to determine whether Ca^{2+} supported any of the steps of *Himar1* transposition.

Himar1 nicks the transferred strand precisely at the transposon end, but cleavage of the opposite strand is less precise (Lampe *et al.*, 1996, 1997; Figure 1.9; Figure 6.4A,B). In order to be able to assess the insertion capacity of *Himar1* after cleavage of the flanking

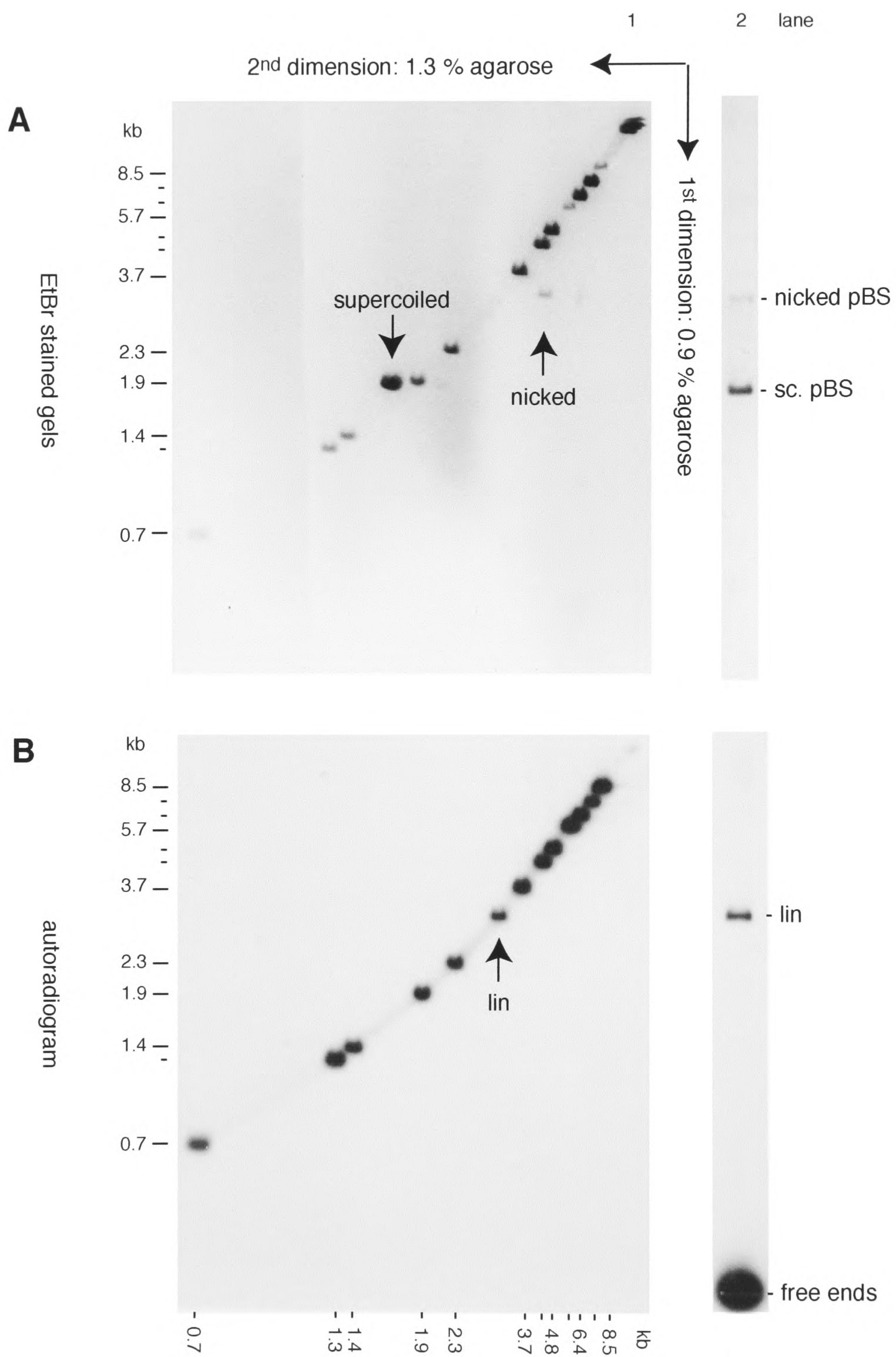


Fig. 6.7

Figure 6.7: Two-dimensional gel to investigate the conformation of the main insertion product

Insertion reaction of 1000 cps substrate 99SX* into 500 ng supercoiled unlabelled pBluescript with 10 nM MBP-*Himar1* C9 transposase and 10 mM MgCl₂, incubated for 7.5 h at 30°C. The reaction was stopped by adding SDS to 0.1%, EDTA to 10 mM, proteinase K to 0.6 mg/ml and incubating for 10 min at 75°C. It was divided into two aliquots and loaded onto a 0.9% agarose, 1x TAE gel. Aliquot 1 was loaded leftmost, together with 500 ng pBluescript, 1.5 µg unlabelled and 1000 cps radioactively labelled lambda *Bst*EII digest ("lane 1"). Aliquot 2 was loaded further to the right ("lane 2"). Electrophoresis in the first dimension was performed in 1x TAE for 14.5 h at 65 V, 4°C. The gel was cut 3.5 cm from the left edge. Lane 2 was EtBr-stained. Lane 1 was turned by 90° and embedded in a fresh 1.3% agarose, 1x TAE gel with EtBr and electrophoresed in the second dimension for 6 h at 150 V, 4°C with EtBr in the running buffer. Both gels were dried onto Hybond-N+ and analysed by autoradiography.

- (A) Reverse contrast image of EtBr-stained gels. Visible are the unlabelled target plasmid (750 ng pBluescript) in supercoiled and nicked conformation in both lanes and the unlabelled marker (linear *Bst*EII restriction fragments of phage lambda DNA) in the 2D gel.
- (B) Autoradiogram of the gels in (A). Visible are free labelled substrate ("free ends"), insertion product ("lin") and labelled marker. Note that the mobility of the insertion product coincides exactly with the arc of linear lambda fragments.

DNA, a *Bsr*DI site was engineered adjacent to the transposon end in the flanking DNA (pKL105; see Table 2.7). This restriction enzyme cleaves DNA with a 2 bp stagger immediately to one side of the recognition site and therefore mimics cleavage at one of the preferred positions, 2 bp inside the transposon end.

*Himar*1 transposase can perform insertion of both precleaved and uncleaved ends in the presence of Mg^{2+} (Figure 6.6A and Figure 6.6B, lanes 2-4,11-13). If Ca^{2+} is used instead of Mg^{2+} , only precleaved, but not uncleaved ends are inserted (Figure 6.6B, lanes 1,7-9). This indicates that Ca^{2+} supports insertion, but not cleavage, i.e. only transesterification is possible, but not hydrolysis. In the absence of either metal ion, no insertion is observed (not shown). Equally, no insertion can be detected at 0°C with either end and with either metal ion (Figure 6.6B, lanes 6,10). This is important, because staged target commitment assays (Section 6.4.1) with precleaved ends are incubated on ice with Ca^{2+} .

Figure 6.6B allows another conclusion: Strand transfer complexes formed in the presence of Mg^{2+} as the only divalent metal ion are completely disrupted and transformed to linear conformation by heating in SDS/EDTA (lanes 2-4,11-13). If, however, Ca^{2+} had been present during complex formation either exclusively or in addition to Mg^{2+} , the supercoiling of the target is released, but most of the transposase remains bound to the insertion product (band “linT”, lanes 5,7-9,14). The same appears to be true for single-insertion products if we assume that band “nicT” is nicked product with attached transposase (see also Figure 6.14). This is the first demonstration of the stabilising effect of Ca^{2+} on the *Himar*1 strand transfer complex.

6.3.5 Insertion of uncleaved ends results in more stable strand transfer complexes

Figure 6.6B shows that with uncleaved substrate and preincubation in Ca^{2+} (lane 5), 100% of the product runs as product linT after heat-treatment with SDS/EDTA. With precleaved substrate (lane 14), only c. 70% is present as linT, the rest is converted into linear product. This shows that post-insertion binding of transposase is more stable if the original substrate had been uncleaved.

6.4 Target commitment and the influence of macromolecular crowding

Integration of a transposon into a target site occurs in at least two stages. The initial contact between the transposon and the target site is by non-covalent interactions. In a subsequent step, the non-covalent interactions are converted into a covalent attachment during the strand transfer step.

The capacity of a transposase or recombinase to engage in productive non-covalent target interactions can be assessed in staged reactions in which different target molecules are presented sequentially (e.g. Sakai and Kleckner, 1997). Target commitment evaluates the preferential insertion of transposon ends into the target DNA that is encountered first. In addition, target commitment gives structural clues into whether the target interaction site is accessible prior to cleavage (Neiditch *et al.*, 2001). The use of uncleaved and precleaved transposon ends enables two steps in the reaction process to be examined for target commitment: before or after transposon cleavage, respectively.

6.4.1 *Himar1* exhibits target commitment before and after donor cleavage

Target commitment assays were performed as outlined in Figure 6.8A. Complexes of labelled transposon ends and *Himar1* transposase were presented with target plasmids of distinguishable size in a sequential manner under conditions that allow DNA binding but not insertion. After an incubation period, magnesium is added, which initiates catalysis. If insertion predominantly occurs into the target DNA that was encountered first, target commitment is said to have occurred. No target commitment is present if insertion occurs at equal levels into both targets. The target plasmids were chosen to be as isogenic as possible, to prevent biased insertion into one of them. To this end, the multiple cloning site of pBluescript was removed to render the smaller target, pKL104 (Table 2.7). Polyethylene glycol was added to some of the reactions (see below, Section 6.4.2).

For each condition, at least two reactions were set up: A control with both target plasmids added at the same time (Figure 6.8B, lanes 4,7,10), and an experimental reaction with either the large target added first (lanes 5,8,11), or the small target added first (lanes 6,9,12). Comparison of the insertion signal strengths into the individual targets of the sequential reactions with those of the control determined the level of target commitment (Figure 6.8B, Appendix A.2).

A

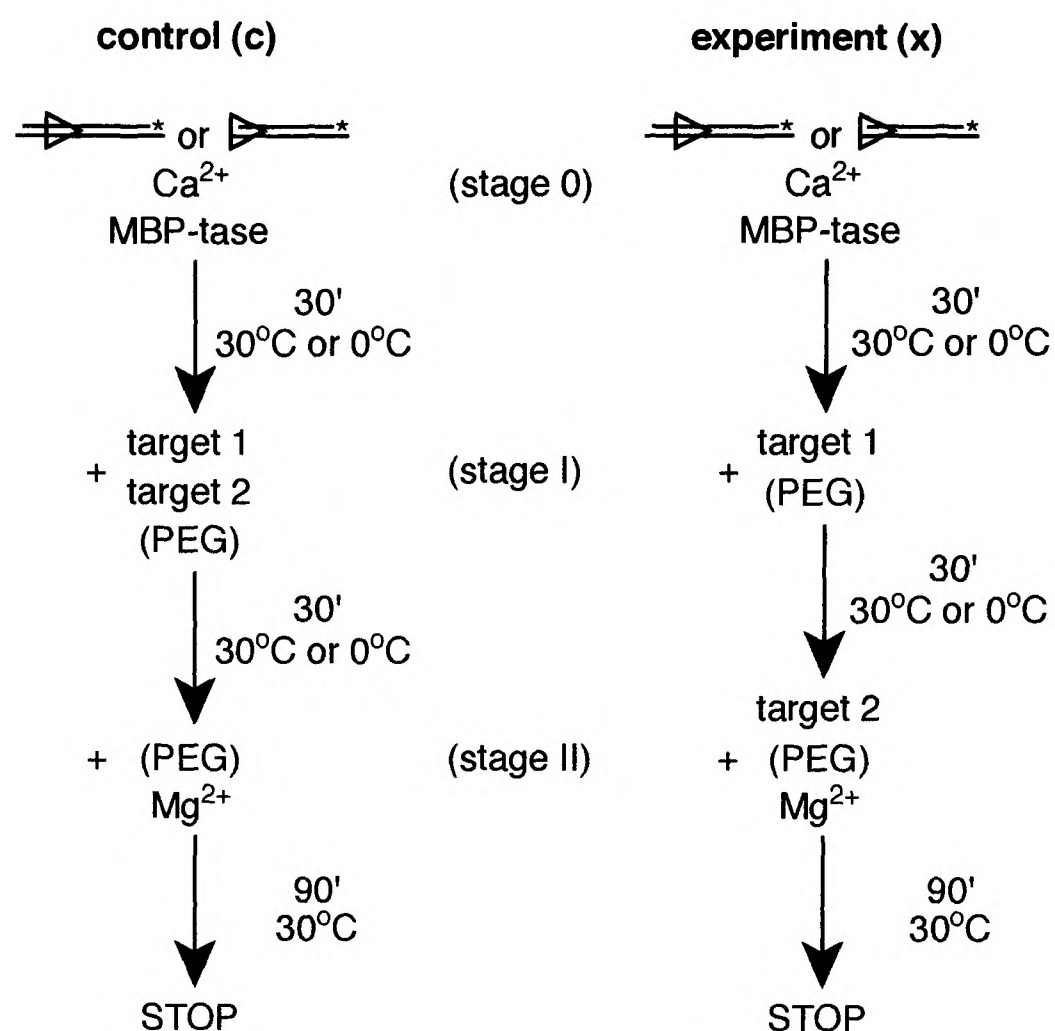
**B**

Figure 6.8: Target commitment of *Himar1* transposase

- (A) Schematic of experimental set-up. The capacity of *Himar1* transposon complexes to become committed to a specific individual target DNA molecule was assayed in a three-stage protocol. First, complexes of transposase and labelled transposon ends were formed in the presence of 2 mM Ca^{2+} , which stabilises the complexes (see Section 6.3.4). Second, supercoiled plasmid DNA (target 1) was added in molar excess and the complexes were incubated for half an hour under conditions that do not support strand transfer (at 30°C for uncleaved substrate, on ice for precleaved substrate). During this incubation, interactions between target 1 and transposon complexes should be established. Third, target 2, a second supercoiled plasmid of a distinguishable size and in equimolar amount to target 1 was added to the reactions, together with Mg^{2+} to 10 mM. The reactions were then incubated for 1.5 hours at the optimal temperature of 30°C to allow strand transfer. The insertion products were analysed by agarose gel-electrophoresis and autoradiography. The target plasmids were pBluescriptII-SK+ (L) and its derivative pKL104 (S), which lacks the 448 bp polylinker.
- (B) Typical experiment, analysed by 0.7% agarose electrophoresis and phosphoimaging. Done as in (A) with 60 fmol uncleaved substrate 97SX* or 60 fmol precleaved 105BX* and 1 nM (15 fmol) MBP-tase. Concentrations of PEG8000 were either 0% (-) or 10% (+). Stage I and II as in (A). M, lambda *Bst*EI digest, 3.7 and 2.3 kb bands; L, (insertion into) the larger target pBluescriptII-SK+ (3.0 kb); S, (insertion into) the smaller target pKL104 = pBluescriptII-SK+ without MCS (2.5 kb); %L, percentage of total counts inserted in large plasmid (linT L + lin L); %S, percentage of total counts inserted in small plasmid (linT S + lin S); % target commitment, as calculated with equation (5) in Appendix A.1. The percentages of insertion into each target and the level of target commitment were calculated from the phosphoimager units listed in Appendix A.2. The data for the gels shown here are identified in columns 7 (c, controls) and 8 (x, experiments) under the heading “lane in Fig. 6.8B”. There is a small but reproducible bias of insertion into the small target, which increases with increasing PEG concentration. This is taken into account in the calculation of target commitment.

For example, a target commitment of 4% (Figure 6.8B(i), lane 5) means that 4% of the total inserted substrate has committed itself to target 1 during stage I. During stage II, these 4% insert into target 1; the remaining 96% insert into target 1 and target 2 at the same ratio as if both targets had been presented simultaneously. For the reaction in Figure 6.8B(i), lane 5, this control ratio is 48 : 52 (target 1 : target 2), as determined by the reaction in lane 4. The target commitment value (in this case, 4%) is achieved by inserting the absolute values listed in Appendix A.2 into equation (5) derived in Appendix A.1. Target commitment is said to have taken place if the calculated percentage is positive.

The values obtained in a number of experiments are summarised in Appendix A.2 and Figure 6.9A. There is clear, positive target commitment with both uncleaved and precleaved substrates and in all tested conditions. With uncleaved ends, the level of target commitment averaged at 6, 33 or 68%, with precleaved ends it averaged at 23, 38 or 76%, dependent on presence or absence of PEG (Figure 6.9A). This means that *Himar1* synaptic complexes bind to target DNA both before and after excision. A slightly stronger level of target commitment was displayed by cleaved ends in all conditions (see also Section 6.4.2).

6.4.2 Macromolecular crowding enhances target commitment

Tn10 target interactions were detected in a relatively simple buffer containing glycerol, spermidine and divalent metal ion (Sakai and Kleckner, 1997), while in similar experiments on RAG1/2 transposition, the reactions were supplemented with polyethylene glycol (PEG, MW 8000) at stage II of the reaction (Neiditch *et al.*, 2001). As the effects of molecular crowding agents on target commitment had not been clearly established, it was investigated by including PEG8000 at different stages of the reaction (Figure 6.8).

It was found that in *Himar1*, the level of target commitment detected increases dramatically in the presence of polyethylene glycol. The earlier PEG is present, the larger the effect (Figure 6.9A; from left to right: without PEG, PEG at stage II, PEG at stages I and II) (see Discussion 6.6.3.1).

The significance of the difference in effects on target commitment between different treatments was investigated by analysis of variance (ANOVA). As the effects of two factors (crowding and end cleavage) and their interaction was of interest, a factorial design was used (Appendix A.1), which is more accurate and efficient than one-factor-at-a-time analyses (Montgomery, 1997; Samuels, 1991).

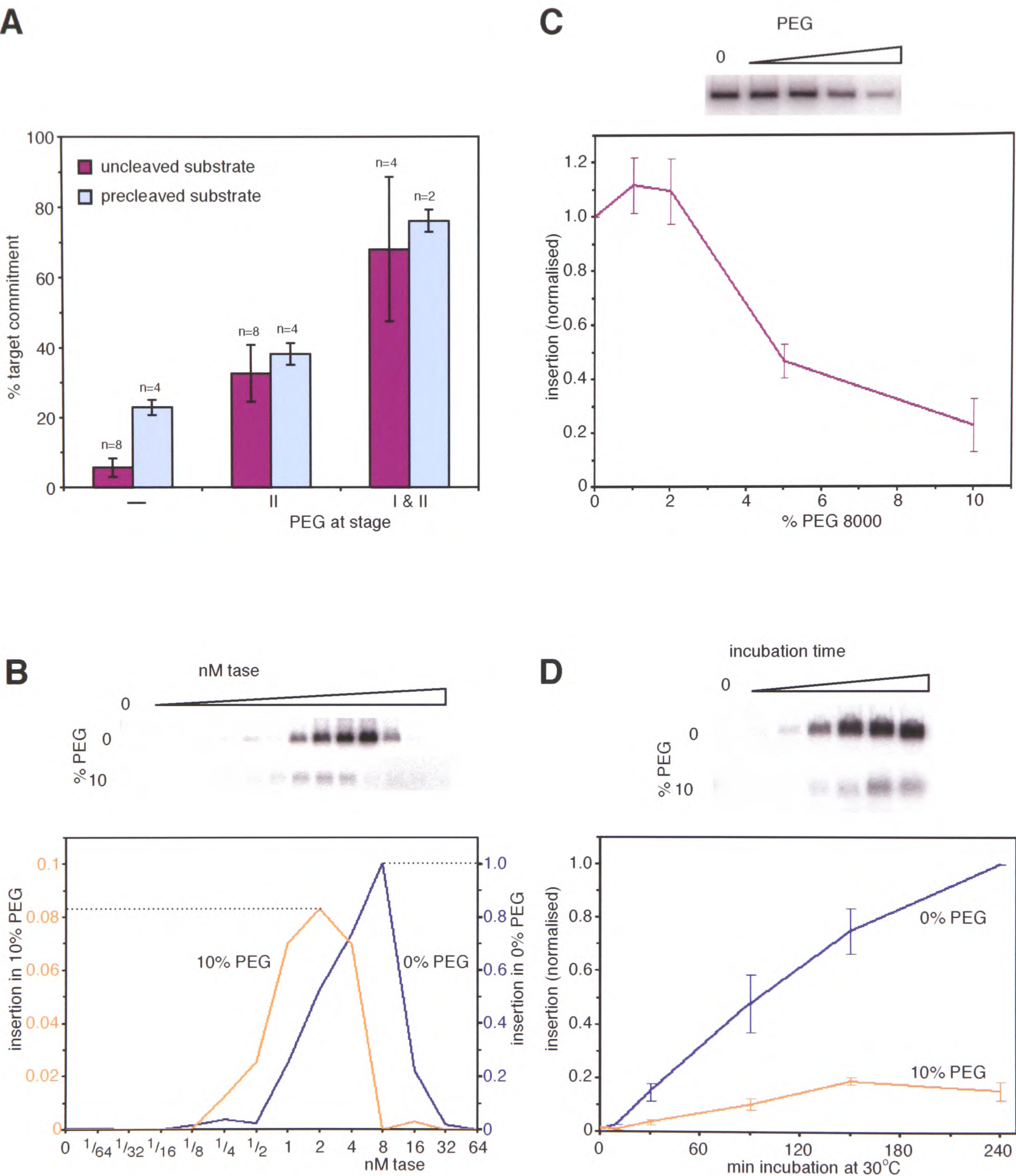


Fig. 6.9

Figure 6.9: Influence of polyethylene glycol on *Himar1* insertion properties

- (A) Plot of mean percentages of target commitment from Appendix A.2. If present, the concentration of PEG8000 is 10%. Uncleaved substrate: 97SX*; cleaved substrate: 141/142* and 105BX*. Error bars, standard error. Pairs of bars from left to right: no PEG, as in Figure 6.8B, lanes 5-6; PEG at stage II, as in Figure 6.8B, lanes 8-9; PEG at stages I & II, as in Figure 6.8B, lanes 11-12. The presence of PEG in stage II increases the level of target commitment. This effect is more pronounced, when PEG is present during both stages of the experiment (PEG at stages I & II, right most columns).
- (B) Titration of transposase concentration in the presence or absence of PEG. 500 cps (39 fmol) 97SX*, 125 ng pBS and PEG, if indicated, were mixed in buffer containing 10 mM MgCl₂ with dilutions of MBP-*Himar1* C9 transposase and incubated for 3 h at 30°C. All reactions (A-C) were stopped and analysed as described in Section 2.5.1.1. Addition of PEG shifts the optimal transposase concentration for insertion to 2 nM, down from 8 nM without PEG. The peak insertion level is also affected and is reduced to 8.2% of the PEG-free optimum.
- (C) Titration of PEG concentration. 500 cps (25 fmol) 99SX* ends, 125 ng pBS and 1 nM (15 fmol) MBP-*Himar1* C9 transposase were allowed to form complexes in transposition buffer with 2 mM CaCl₂ for 30 min at 30°C. For insertion, 10 mM MgCl₂ and the indicated concentration of PEG8000 were added, and the incubation continued for 1.5 h at 30°C. Error bars, standard error; n=2.
- (D) Kinetics of insertion in the presence or absence of PEG. Complexes of 500 cps (60 fmol) 97SX* ends and 1 nM (15 fmol) MBP-*Himar1* C9 transposase were formed in transposition buffer with 2 mM CaCl₂ for 30 min at 30°C. On ice, 10 mM MgCl₂, 125 ng pBS and, if indicated, PEG8000 to a final concentration of 10% were mixed in. The insertion reactions were started by placing the tubes into a 30°C waterbath; they were stopped by taking aliquots into 10x STOP mix buffer (final 1.5% SDS, 1.5 mM EDTA) and placing them into dry ice. Error bars, standard error; n=2. The level of insertion is decreased six-fold if PEG is present. The reaction kinetics remains largely unaffected.

ANOVA and F tests show that the increase of target commitment with PEG concentration is highly significant: With 99.998% confidence, the real means of values for the three conditions are different from each other (Table 6.1, row 2).

Conversely, the increase of target commitment observed upon end cleavage (Figure 6.9A) is at best marginally significant (83% confidence) (Table 6.1, row 3). There is no statistical evidence for interaction between the influences of PEG concentration and end cleavage (Table 6.1, row 4), i.e. there is no statistical evidence that the effect of the PEG concentration on target commitment depends on the cleavage state of the DNA substrate or *vice versa*.

A similar statistical analysis of target commitment itself was confounded by the large number of different conditions tested and therefore small number of samples in identical conditions.

Table 6.1: Analysis of variance in a factorial design for influences on *Himar1* target commitment

source of variation	sum of squares (SS)	degrees of freedom (df)	mean square (MS)	F_0	P -value
PEG	14126.743	2	7063.37	18.08	0.00002
end cleavage	765.214	1	765.21	1.96	0.17447
interaction	193.957	2	96.98	0.25	0.78218
error	9377.423	24	390.73		
total	24463.336	29			

SS: sum of squares; df , degrees of freedom; MS, mean square; F_0 , test statistic for the hypothesis of no differences in means; P -value, significance level.

For definitional formulas, see Appendix A.2 (from Montgomery, 1997). P -values are from the HyperStat Online Textbook: http://davidmlane.com/hyperstat/F_table.html.

6.4.3 Effects of PEG on *Himar1* transposition *in vitro*

Because of the strong influence of PEG on target commitment, the effect of PEG on other properties of *Himar1* transposition was investigated. Insertion reactions with labelled linear transposon ends and supercoiled pBluescript as target were set up and stopped as described in Section 6.3.3.

The transposition reaction was titrated with transposase at 0% and 10% PEG8000 (Figure 6.9B). In the absence of PEG, the optimum transposase concentration is 8 nM (Figure 6.9B, blue curve), which agrees closely with the published value of 10 nM (Lampe *et al.*, 1998). At 10% PEG, the optimum transposase concentration is 2 nM (Figure 6.9B,

orange curve). The reaction is extremely sensitive to an excess of transposase and is almost completely abolished by transposase concentrations only 4-fold above the optimum. This holds true both in the presence or absence of PEG. Thus, at 10% PEG, the optimum transposase concentration and the amount of transposase required to inhibit the reaction are both reduced to the same extent. This is consistent with the molecular crowding activity of PEG, which excludes large molecules from a significant proportion of the volume of the solution.

Next, the *HimarI* insertion reaction was titrated with 0–10% PEG (Figure 6.9C). The transposase concentration in this experiment was limited to 1 nM to preclude aggregation at higher PEG concentrations. Under these conditions, up to 2% PEG has a slight beneficial effect on the reaction. However, at 10% PEG the reaction is inhibited and only 20% of the insertion activity remains. This is similar to *Tn10* transposition, which is also severely inhibited by PEG (R. Chalmers, unpublished).

Finally, to determine whether inhibition has a kinetic effect on the reaction, a time course was performed at 0 and 10% PEG (Figure 6.9D). As before, 10% PEG reduced transposition to about 20%. With PEG, the reaction reaches completion earlier.

In summary, PEG influences the insertion reaction quantitatively, but not qualitatively. Insertion reactions retain their characteristics. Regarding target commitment, we interpret these results as PEG preserving real non-covalent interactions long enough to be converted to covalent insertions. The high level of target commitment in the presence of PEG is therefore considered to be an accurate reflection of the behaviour of the paired ends complex.

6.5 Identification and analysis of nucleoprotein complexes

Transposition, like many other aspects of nucleic acid metabolism, takes place in the context of a high order protein-DNA complex. These complexes are often large and have a complicated structure, which reflects the presence of sophisticated regulatory mechanisms. One of the earliest steps of the transposition reaction is when the ends of the element are brought together in a synaptic complex. All of the subsequent steps of the reaction take place within this complex, which generally becomes more stable at later stages, because the reaction is energetically downhill and involves no high-energy cofactors. Thus, the initial synaptic complex is often the least stable intermediate and the most difficult to study.

For the qualitative and quantitative analysis of DNA-protein interactions, the gel shift (or gel retardation) assay is one of the most effective and simple methods (Revzin, 1989; Carey, 1991). The technique is very sensitive because the mobility of a complex is influenced by its size, shape and charge.

6.5.1 Specific *Himar1* transposase-DNA complexes can be formed *in vitro*

Gel retardation experiments were carried out to analyse complexes formed between *Himar1*-transposase and the transposon end. When a DNA fragment encoding the *Himar1* left end is incubated with transposase, two major complexes are detected (Figure 6.10). To test whether these complexes are specific, complex assembly was challenged by mixing unlabeled competitor DNA with the labelled transposon end before the addition of transposase. Both complexes are resistant to a very large excess of poly(dIdC) or pBluescript DNA (Figure 6.10A, lane 2-7). However, both complexes are very sensitive to challenge with unlabeled transposon end fragments (Figure 6.10A, lane 8-12). It can therefore be concluded that complexes 1 and 2 are both specific for the interaction between transposase and the transposon end. This conclusion is further supported by hydroxyl radical footprints presented in Section 6.5.7, in which transposase was observed to contact DNA specifically at the terminal inverted repeats. Assembly of these complexes is extremely reproducible, and they are detected with a variety of transposon ends of different lengths (see Section 6.5.3).

6.5.2 Transposase titration reveals additional complexes

To investigate the relationship between complex 1 and 2, the complex formation reaction was titrated with an increasing amount of transposase (Figure 6.10B). This was performed with two different DNA fragments with 67 bp or 124 bp transposon arms and different lengths of flanking DNA. Complex 1 and 2 are both present at the lowest transposase concentration tested. As the transposase concentration is increased, complex 2 begins to dominate the reaction. At 16 nM transposase, only complex 2 is present, the free DNA has disappeared and a significant amount of material has aggregated in the wells.

The highest proportion of ends is present in discrete complexes at transposase concentrations between 8 and 16 nM. Interestingly, this is the same optimal concentration as for insertion reactions (Figure 6.9B). Just above this concentration, almost all material is aggregated in the wells, below this concentration, most of the DNA is unbound (Figure 6.10B,C).

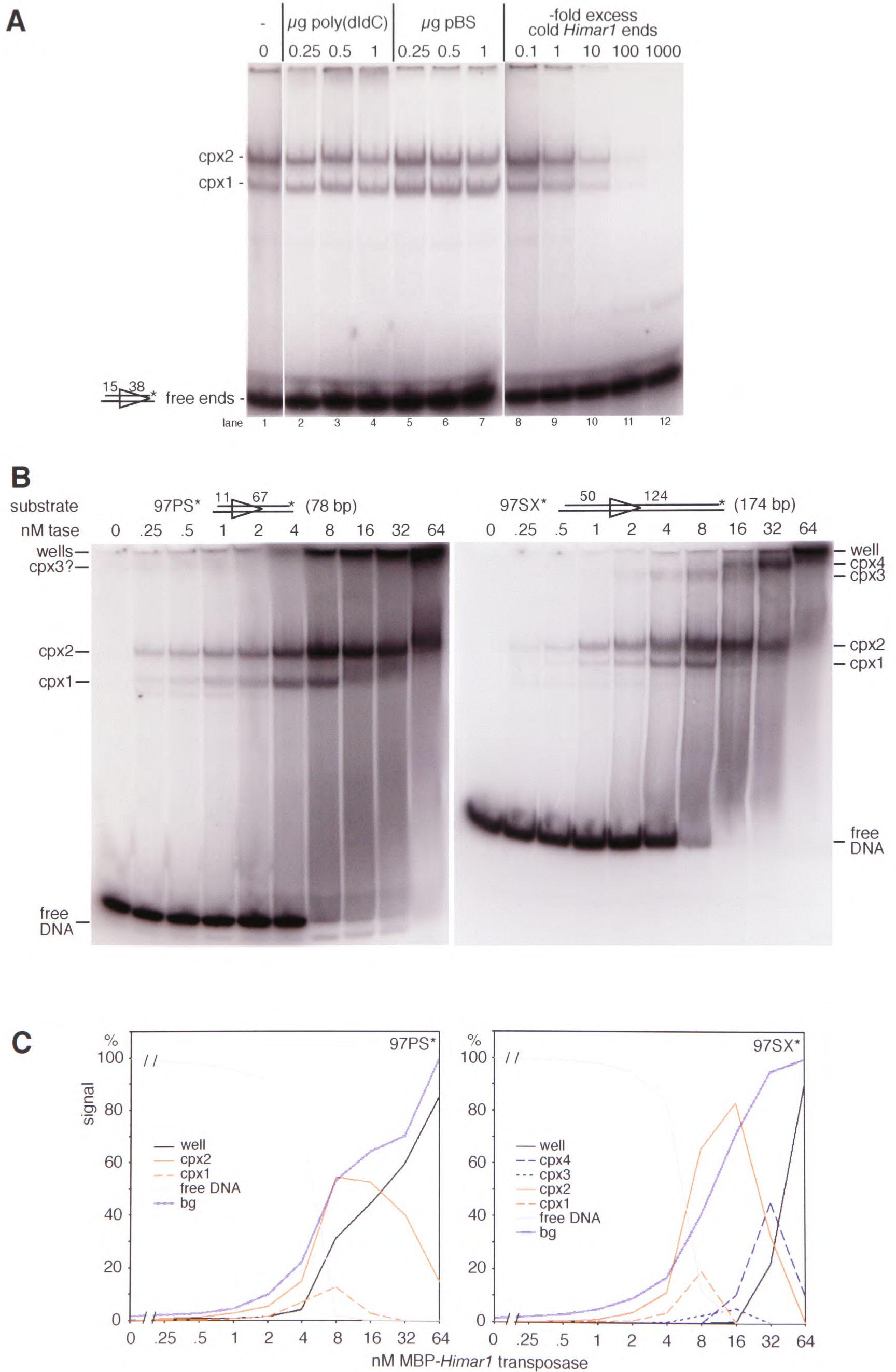


Fig. 6.10

Figure 6.10: Specific complexes between transposase and the left end of *Himar1*

Binding reactions of labelled *Himar1* transposon ends with MBP-*Himar1*-transposase C9 in reaction buffer with 2 mM Ca^{2+} for 1.5 h at 30°C. The transposon end fragment 143/144* was created by annealing oligonucleotides KL143 and KL144; 97PS* and 97SX* were created by restriction digest from plasmid pKL97 (see Table 2.10).

- (A) A linear DNA fragment encoding the left end of *Himar1* was used to form complexes with purified transposase. 15 μl reactions were set up with 10 fmol (350 cps) transposon end 143/144* (Table 2.10; Figure 2.1) and 15 fmol (1 nM) transposase. Assembly of the complexes was challenged with unlabeled competitor DNA, which was present in the reaction mixture before the addition of transposase. An autoradiogram of a gel retardation experiment is shown. *, position of radioactive label on the transposon arm. Lanes 2-7, competition with the synthetic double-stranded alternating copolymer poly(dIdC) or with supercoiled pBluescriptII-SK+ plasmid DNA. 1 μg competitor DNA corresponds to a c. 300-fold nucleotide excess over the labelled transposon end substrate. Lanes 8-12, specific competition with unlabelled transposon end substrate 143/144. Phosphoimager scans of native 5% PA gels, run for 120 min at 300 V, 4°C in 1xTAE.
- (B) Transposase titration of the complex assembly reaction. Binding reactions of substrate 97PS* (78 bp, 29 fmol, 500 cps/lane) or 97SX* (174 bp, 50 fmol, 500 cps/lane) with increasing concentrations of MBP-*Himar1* transposase C9, MonoQ fraction d11. In these 15 μl reactions, 1 nM transposase corresponds to 15 fmol. Phosphoimager scan of a native 5% acrylamide, 1x TAE gel, run for 2.5 h at 300 V, 4°C.
- (C) Quantification of the bands in (B), measured with ImageGauge software (Fuji). The background (bg) was measured in an area of constant size between cpx2 and cpx3. The bg curve is normalised against its maximum value (at 64 nM tase). Values of free DNA, complexes and wells are plotted after subtraction of the bg value of the appropriate lane. Shown are the percentages of the individual bands per lane.

Transposase titration of the DNA fragment with the 124 bp transposon arm reveals two new bands, complexes 3 and 4 (Figure 6.10B, right). At higher transposase concentrations, complex 4 increases at the expense of complex 3. Close inspection of the titration of the 67 bp transposon arm (Figure 6.10B, left) reveals faint bands that may represent complexes 3 and 4. Presumably, these complexes require the longer transposon arm or flanking DNA, either for assembly or to impart stability during electrophoresis. One obvious possibility is that complexes 1+2 represent transposase bound to a single transposon end, while complexes 3+4 represent a synaptic complex with two transposon ends. This possibility will be addressed in Section 6.5.6.

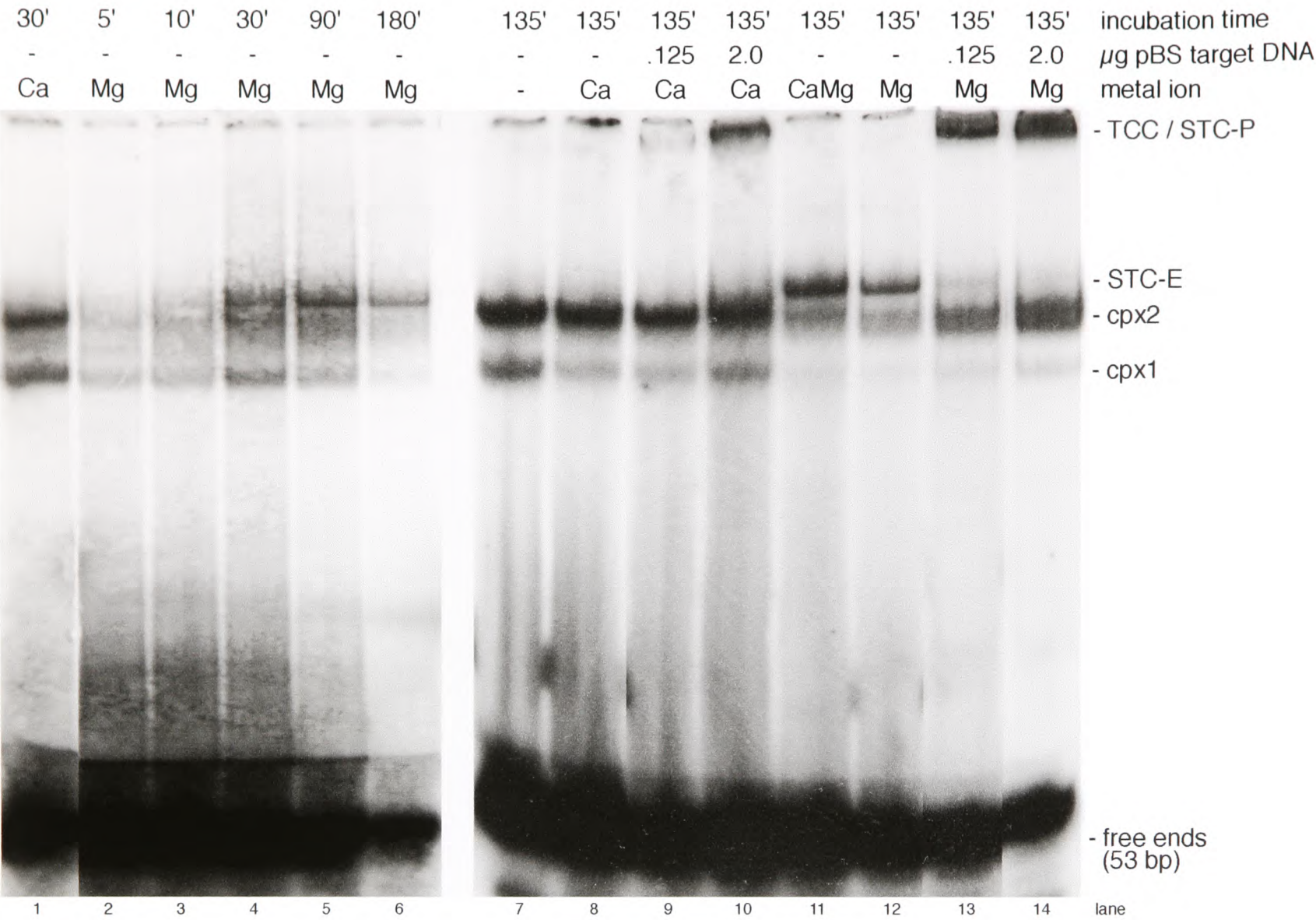
6.5.3 Formation of strand-transfer and target-capture complexes

The next question addressed was whether the *Himar1* complexes are catalytically active. A time course experiment was carried out in the presence of magnesium to follow the formation of reaction products (Figure 6.11A, lanes 1-6). At early time points, complexes 1 and 2 are present (lanes 1-2). After 30 min, a new band (STC-E) is apparent above complex 2. This band prevails throughout several hours, while complex 1 and 2 decrease in strength.

The observed Mg-specific band apparently represents a strand transfer complex after insertion into a free substrate molecule ("STC-E", for strand-transfer complex into transposon end substrate). This interpretation is supported by the following observations: If a supercoiled target plasmid is added to the reactions, the majority of the STC-E band is replaced by a band of very low mobility near the top of the gel (STC-P, lanes 13-14), the expected position for plasmid DNA. This latter reaction differs from the non-denatured insertion reactions of Figure 6.5B only by the kind of gel used, and in both cases, the predominant species is the strand transfer complex with plasmid as target ("linB" or "STC-P" in Figure 6.5B and Figure 6.11A, respectively). The same dependency of the mobility of the strand transfer complexes on target size is observed when comparing gel shift experiments using DNA substrates of different sizes (Figure 6.11B, lanes 6,8,10, band STC-E).

It is also apparent that insertion favours the supercoiled target molecule. This can be seen by comparing the strength of STC-P and STC-E in Figure 6.11A, lane 13. This reaction contains only 1.6 times more supercoiled plasmid than linear ends (64 and 40 fmol, respectively), but the abundance of insertion products into the plasmid is at least 20-fold

A



B

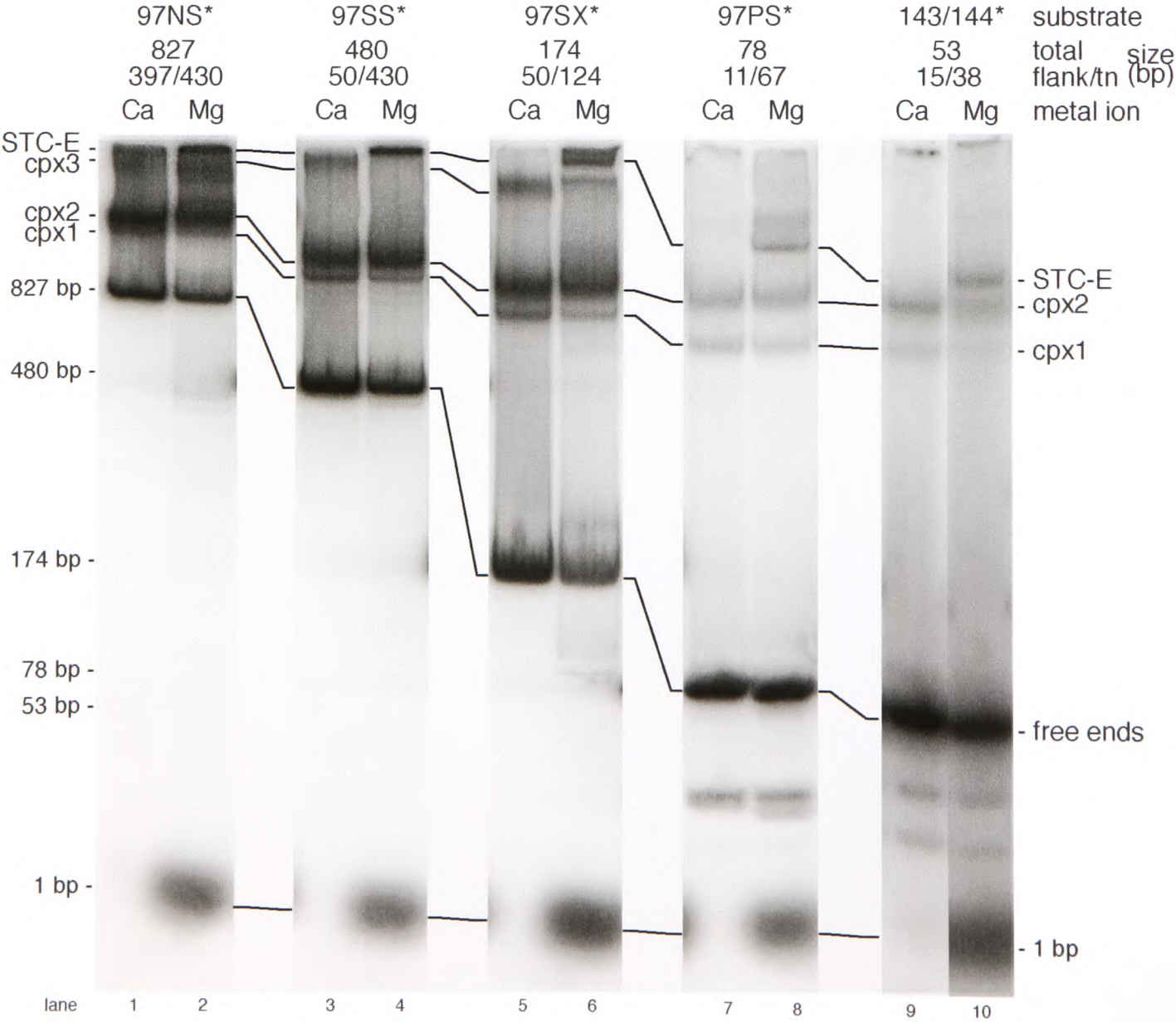


Fig. 6.11

Figure 6.11: Effect of metal ions and substrate sizes on *Himar1* complexes

- (A) Binding and catalysis of substrate 143/144* (40 fmol, 500 cps/lane) with 1 nM (15 fmol) MBP-*Himar1* transposase with 2 mM CaCl₂ and/or 10 mM MgCl₂. If indicated, 125 ng (64 fmol) or 2 µg (1024 fmol) supercoiled pBluescript plasmid were included in the reactions. For lanes 1-6, aliquots of substrate and buffer were kept on ice. Reactions were started at intervals by adding freshly diluted transposase and transferring the aliquot to 30°C; all aliquots were stopped simultaneously by putting them on ice and quickly loading them on the gel. For lanes 7-14, the reactions were set up directly and incubated at 30°C for 2 h 15 min before PAGE. TCC, target capture complex with plasmid target (lanes 9-10); STC-P, strand transfer complex with plasmid target (lanes 13-14); STC-E, strand transfer complex with transposon end target (lanes 3-6, 11-14). Autoradiograms of native 5% PA, 1x TAE gels, run for 2-3 h at 300 V, 4°C.
- (B) Complexes were assembled in 15 µl reactions with 60 fmol (4 nM) MBP-*Himar1* transposase, using transposon end fragments with a range of different lengths (see Table 2.10). In all of these, the transposon end is labelled on the NTS. Incubation time at 30°C: lanes 1-8, 1.5 h; lanes 9-10, 2 h 15 min. Amounts of DNA fragments: 97NS* (827 bp), 50 fmol; 97SS* (480 bp), 33 fmol; 97SX* (174 bp), 22 fmol; 97PS* (78 bp), 50 fmol; 143/144* (53 bp), 42 fmol. Note that all lanes contained 500 cps label but the actual amount of DNA (given above) is variable because of differences in labelling and crush and soak purification. Lanes 1-8 came from the same gel, lanes 9-10 from a second gel. An autoradiogram of a gel retardation experiment is shown (1xTBE, 5% PA gel, run for 2.5 h at 250 V, 4°C).

higher. This is in agreement with our previous observations (Section 6.3.1), which also showed preferential insertion into supercoiled target.

In reactions, which had been incubated with target plasmid for over two hours, a complex of low mobility is observed also in the absence of magnesium (Figure 6.11A, lanes 9-10). As Ca^{2+} alone does not support transposition of uncleaved ends (Figure 6.6B, lane 1), this complex cannot be a strand transfer complex. Instead, it appears to be a non-covalent target capture complex (TCC): It runs at the same position in the gel as the STC-P, and it is more abundant at higher concentrations of the target DNA. Upon addition of magnesium, this TCC is probably directly converted to the STC. The presence of a relatively stable target capture complex is in good accordance with the observation of *Himar1* target commitment before donor cleavage (see Section 6.4.1, Discussion 6.6.3.2).

6.5.4 *Himar1* transposase displays exonucleolytic activity in the presence of magnesium

When magnesium is included in the reaction, an accumulation of labelled nucleotides is observed at the bottom of the gel. These signals appear at the same position, independent of the length of the DNA substrates (Figure 6.11B). In all of the substrates used, the labelled nucleotide is positioned 2-4 nucleotides from the fragment end. This means that the activity cannot be a phosphatase, but is an exonuclease, which sequentially cuts off nucleotides from free DNA ends. As single nucleotides rather than a ladder of digested fragments are observed and as the label is on the 3' end of the strand, the activity must be a 3'→5' exonuclease. When different fractions of the MonoQ purification were incubated with non-*Himar1* labelled DNA and analysed on a sequencing gel, only the fraction with the highest transposition activity (fraction 11, see Section 6.2.1) displayed significant exonucleolytic activity (not shown).

6.5.5 The mobility of complexes is predominantly determined by the protein component

Complexes were formed with various DNA substrates, ranging from 53 to 827 bp of total fragment length (Figure 6.11B). When comparing the positions in the gel of complexes 1 and 2 formed in the presence of calcium, it is evident that their distance from the wells is very similar throughout all of the substrates used.

If the mobility were mainly due to the molecular weight of the DNA component, the *ratio* of the distance travelled by the complex versus that travelled by the free DNA would

remain constant, independent of the DNA length. This is clearly not the case in Figure 6.11B: With the longest DNA fragment, complex 1 and 2 run at about half the distance between free DNA and wells, whereas with the shortest fragment, the complexes have covered less than one third of this distance.

Thus, the mobility of the complexes is not dominated by the molecular weight of the DNA fragment used, but by the protein component, which is taken to be constant in each species of complex. The molecular weight of one monomer of transposase is 83 kD, and therefore seems insufficient to account for the change in mobility on its own. More than one transposase monomer bound per DNA end would be one explanation for these results.

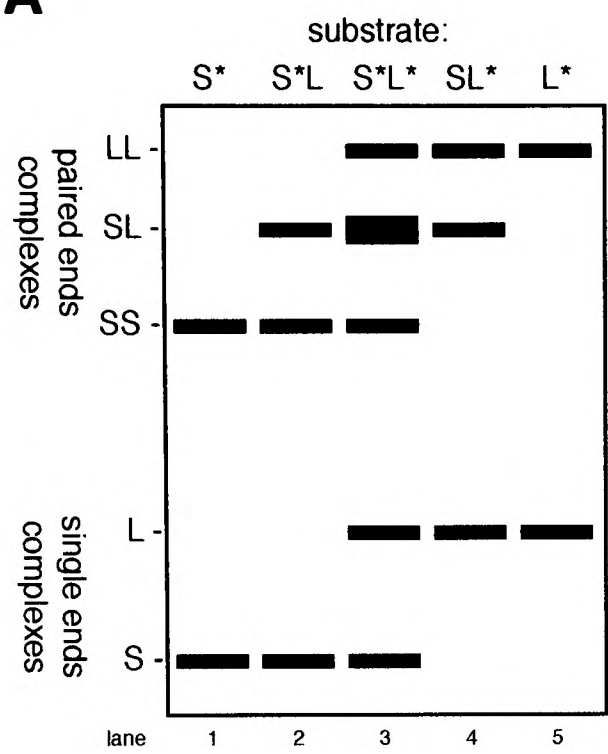
6.5.6 Complexes 1 and 2 contain a single transposon end

All transposons studied so far form synaptic complexes, which are a prerequisite for excision and insertion. This pairing ensures concerted excision and insertion of both transposon ends. Forming complexes from mixtures of long and short DNA fragments is an established method to determine whether a complex contains one or a pair of transposon arms (e.g. Sakai *et al.*, 1995). If a complex contains two DNA molecules, using mixed substrates results in the formation of the original short (SS) and long (LL) complexes, and of an additional complex of intermediate mobility, which contains one long and one short substrate molecule (SL) (Figure 6.12A). In such reactions, if only one of the two substrate species is labelled, two species are observed, one characteristic of the labelled fragment alone (e.g. S*S*) and a second with the intermediate mobility (e.g. S*L). If only one DNA molecule is present per complex, addition of an unlabelled fragment of different length will not generate any additional complexes.

Complexes were assembled in the presence of Ca²⁺ with transposon end fragments that allow the maximum separation between complexes formed with either the short (S) or the long (L) substrate (Figure 6.11B; Figure 6.12C). With 174 bp and 480 bp substrates, the short complexes 1 and 2 are clearly separated from the long complexes 1 and 2. However, no complexes of intermediate size are detectable in reactions with mixed substrates, regardless of which substrate is labelled (Figure 6.12C, lanes 2-4). It can therefore be concluded that complexes 1 and 2 contain a single transposon end.

Candidate for true PECs are complexes 3 and 4. Due to their instability at longer electrophoresis time, the results were inconclusive, even after glutaraldehyde crosslinking (not shown).

A



B

	S*	L
S*	S*S*	S*L
L	S*L	LL

C

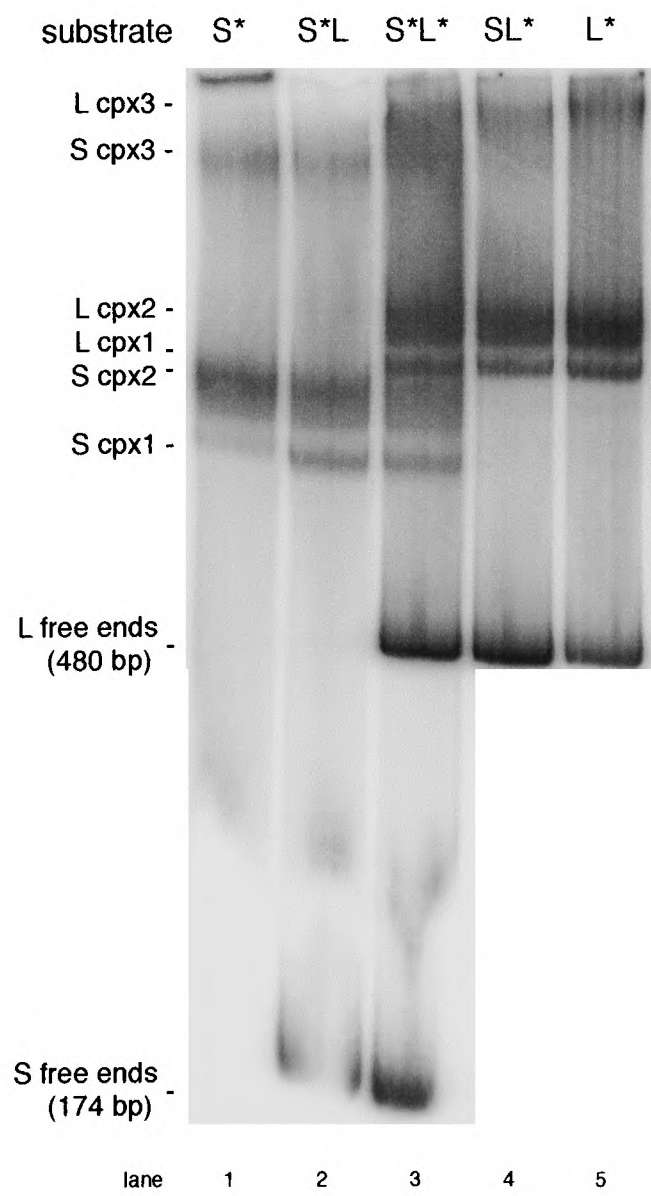


Fig. 6.12

Figure 6.12: Identification of complex stoichiometry

- (A) Schematic of assay (see text Section 6.5.6).
- (B) Matrix of paired ends complexes formed with labelled short and unlabelled long ends. If equimolar amounts of both ends are employed, stochastically twice as many heterogeneous complexes as homogeneous ones are expected. If only one of the two ends is labelled in the mixed complexes, the signal strength of both complexes is expected to be identical. Following the same reasoning, mixed complexes of two labelled substrate species (S^*L^*) have twice the signal strength as the homogeneous complexes (S^*S^* and L^*L^*). If only one DNA molecule is present per complexes, it is impossible to generate any other sizes than the original complexes.
- (C) Complexes were assembled with either short (31 fmol) or long (47 fmol) radioactively labelled transposon end fragments and an equal amount of unlabelled fragments, if shown. Assembly reactions with 8 nM (120 fmol) transposase were incubated for 1.5 h in 1x reaction buffer with 2 mM CaCl_2 . S, short substrate 97SX (174 bp); L, long substrate 97SS (480 bp); *, position of radioactive label on the transposon arm. An autoradiogram of a 5% native TAE-PA gel is shown after 4 h electrophoresis at 300 V, 4°C. No complexes of intermediate mobility are formed from complex 1 and 2; these are therefore single ends complexes.

6.5.7 The two single end complexes show identical hydroxyl radical protection patterns

To determine whether the different mobility of complexes 1 and 2 is due to different numbers of transposase monomers binding to the transposon end, the complexes were footprinted with hydroxyl radicals (Schickor and Heumann, 1994; Crellin and Chalmers, 2001.)

Complexes were assembled in a standard reaction, treated with hydroxyl radicals and separated in a gel retardation experiment. The free (uncomplexed) DNA, complex 1 and complex 2 were recovered from the gel and the footprints were displayed on a sequencing gel and quantified (Figure 6.13).

Complex 1 and complex 2 have identical protection patterns from bp+1 to bp+23 on the NTS. There are two strong patches of hydroxyl radical protection centred on bp10 and bp20 of the transposon end. These patches are in phase with the helical repeat of DNA indicating that transposase interacts with one face of the helix. Similar 10 bp phasing of transposase contacts has been observed in other transposon systems, such as *Tn10*, phage Mu and RAG1/2 (Crellin and Chalmers, 2001; Zou *et al.*, 1991; Swanson and Desiderio, 1998).

This pattern is of significantly higher resolution than the published DNaseI footprints, which were derived from in-bulk-experiments instead of from isolated complexes. DNaseI footprinting gave a continuous block of protection from bp+3 to +31 on the NTS and bp+2 to +29 on the TS (Lampe *et al.*, 1996; Lampe *et al.*, 2001).

The hydroxyl-radical footprints show that the interaction between transposase and the transposon ends is the same for complexes 1 and 2. Although the complexes have different mobilities in the gel, it appears that the transposase protects only the ITR sequences in the terminal 23 bp of the transposon, as no additional binding sites were detected. Polymerisation of transposase monomers on the transposon end is therefore not responsible for the differences in mobility. The difference in mobility between complexes 1 and 2 may therefore be due to recruitment of additional transposase monomers by protein-protein interaction.

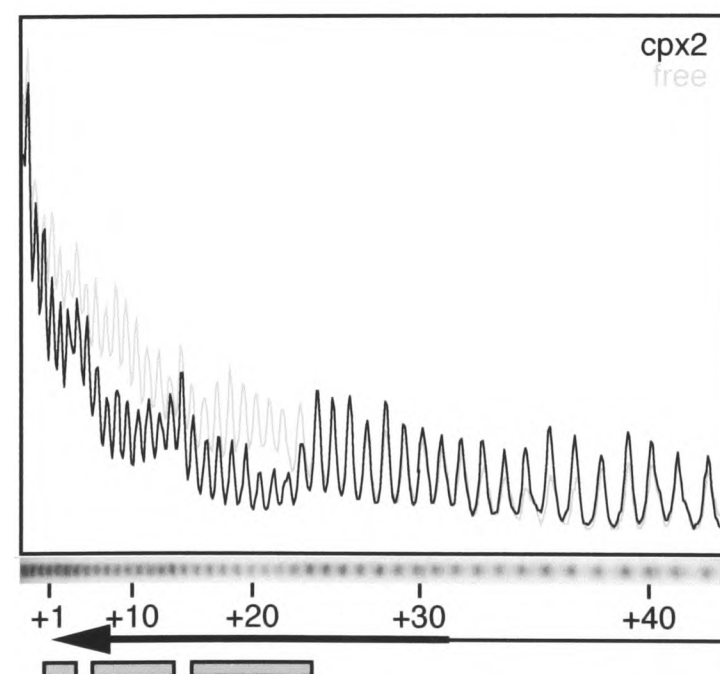
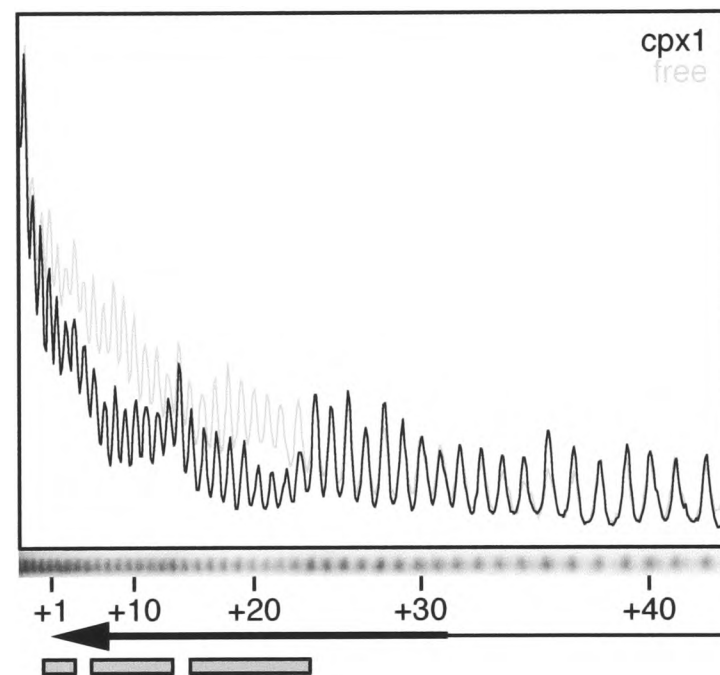
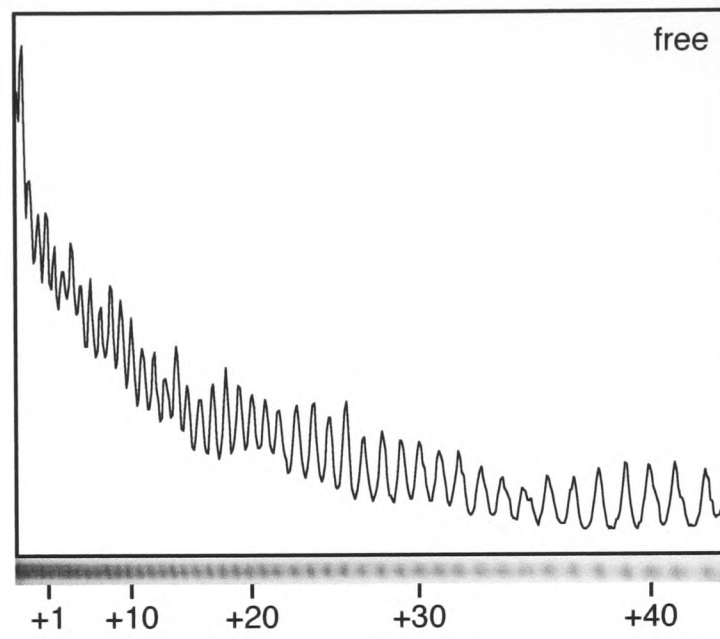
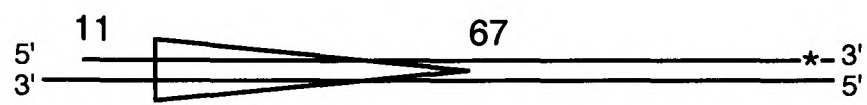


Fig. 6.13

Figure 6.13: Footprints of *Himar1* transposase-transposon end complexes

Complexes were assembled in 9 x 20 μ l with 720 fmol (4 nM) transposase and 333 fmol of transposon end 97PS* with the label on the non-transferred strand. The reaction mixture was treated with hydroxyl radicals to generate a footprint. Complex 1, 2 and the free DNA were purified in a gel retardation experiment and the footprint was displayed on a sequencing gel. Traces were generated using ImageGauge software. In the second and third panel, the trace from the bound complex (black) is shown with the aligned trace of the free DNA (light grey) in the background. The region of the sequencing gel containing the footprint of the complexes is shown aligned below the traces. The arrowhead is the position of the transposon end; the thick line indicates the inverted repeats. The number of bp inside the transposon end are indicated. Shaded boxes indicate regions of hydroxyl radical protection.

6.6 Discussion

The *Himar1* transposition reaction, as derived from the results of this chapter, is summarised in Figure 6.14. The major complexes and processes are discussed in this section.

6.6.1 Single end complexes

The most abundant and stable nucleoprotein complexes 1 and 2 have been shown to contain only one DNA molecule each (Section 6.5.6). These single ends complexes (SECs) could be precursors and/or disintegration products of paired ends complexes. It was also observed that with increasing transposase concentrations, the strength of complex 2 increases at the expense of complex 1 (Section 6.5.2), which suggests that complex 2 is formed from complex 1 by binding of additional transposase molecules. A likely model is therefore that the unstable *Himar1* PEC is composed of the more stable SECs, complex 1 and 2, and that it falls apart into the same complexes if subjected to electrophoresis (see also Section 6.6.1.3).

6.6.1.1 Composition of single ends complexes

Considering that complexes 1 and 2 both contain only one DNA molecule per complex (Sections 6.5.6), it still remains to be determined why they show distinct migratory behaviour during electrophoresis.

To test the influence of the shape of the DNA on migration through the gel, DNA substrates of equal length, but with the transposase binding site (ITR) at different positions can be used (“circular permutation”). If the transposase induces a bend in the DNA, retardation should be greatest when the ITR is positioned in the centre of the DNA fragment (Orchard and May, 1993; Tavakoli *et al.*, 1997).

The number of protein molecules per synaptic complex varies among transposons (see Section 1.1.3). Thus, the SECs complex 1 and 2, which are either precursors to or disintegration products of PECs (see above), can potentially harbour one or several protein monomers.

The footprint analysis showed identical protection patterns for both complexes within the transposon sequence (Section 6.5.7). It is therefore unlikely that the complexes differ by the number of transposase monomers bound directly to the end sequence. Instead, the mobility difference might be due to protein-protein interactions, where a dimer of

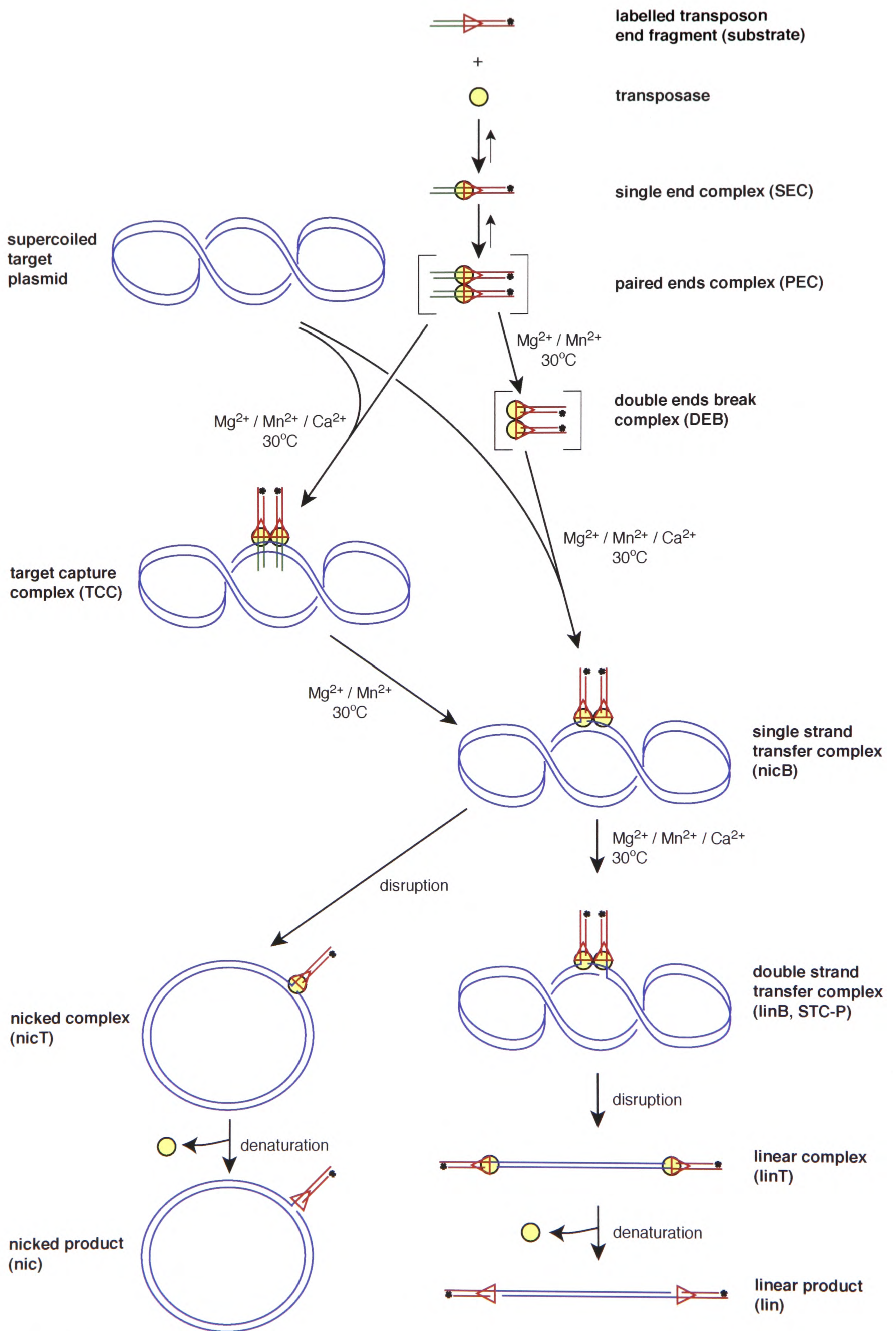


Fig. 6.14

Figure 6.14: Summary of complexes formed during *Himar1* insertion

Shown is the insertion of short tlinear transposon ends into supercoiled target plasmid, catalysed by *Himar1* transposase.

The first step of *Himar1* transposition is specific binding of *Himar1* transposase to the transposon ends (ITRs), forming single end complexes (SECs). A synaptic complex (PEC) is probably formed by the entry of naked DNA or a second SEC. Both uncleaved and cleaved end complexes can bind to target DNA to form a target capture complex (TCC). Transposon end cleavage requires the presence of Mg^{2+} or Mn^{2+} ions, whereas for insertion, either Mg^{2+} , Mn^{2+} or Ca^{2+} is sufficient. After insertion, the transposase remains bound in the stable strand transfer complex (STC).

After insertion of one transposon end into supercoiled target plasmid, the DNA product is a nicked circle with a transposon end flap (nic); after insertion of two transposon ends, the insertion product is linear (lin). However, the transposase remains bound in the STC unless forcibly removed, preserving the supercoiled conformation and high electrophoretic mobility of the target plasmid (nicB / linB). Disruption of the STC by heating in SDS and EDTA initially releases the supercoils from the donor DNA. Transposase is still bound at this stage (nicT / linT). Further denaturation or deproteinisation finally removes the protein (nic / lin).

Red, transposon ends; green, flanking donor DNA; blue, target plasmid; yellow, transposase; lin, insertion product with linear DNA conformation (double-ended insertion); nic, insertion product with nicked DNA conformation (single-ended insertion); linB, nicB, nucleoprotein complexes with higher mobility than free DNA (bottom band in an agarose gel); linT, nicT, nucleoprotein complexes with lower mobility than free DNA (top band in an agarose gel). Not to scale. Brackets [] signify complexes which have not yet been detected and whose existence remains unclear. The number of transposase monomers per complex and the relative orientation of transposon ends is not known (see Figure 6.15 for a model).

transposase occupies only one binding site on the DNA. It is also possible that additional monomers bind to the flanking DNA, where the protection pattern was not determined.

The similar mobility of complexes formed with DNA substrates of very different lengths suggests that the protein component is the major determinant of mobility. To establish the number of transposon molecules per complex, their molecular weights could be estimated by a Ferguson analysis (Ferguson, 1964; Orchard and May, 1993). In *Mos1*, which forms a pair of complexes analogous to *Himar1* complexes 1 and 2 (Augé-Gouillou *et al.*, 2001b), the result of such an analysis was consistent with a monomer of transposase for the lower complex, whereas the larger complex appeared to contain either one or two monomers (Augé-Gouillou *et al.*, 2001a).

To determine the number of subunits directly, band shifts could be performed with a mixture of proteins of different molecular weights, such as with and without the MBP-fusion. Following the same reasoning as detecting paired ends complexes with long and short DNA molecules (Section 6.5.6), complexes of intermediate size would indicate the presence of at least two transposase monomers (Bhasin *et al.*, 2000).

Alternatively, the number of transposase monomers bound in the SECs and PECs could be determined by crosslinking complexes with one of a large number of available crosslinking reagents, followed by SDS-PAGE analysis. With increasing concentration of the crosslinking reagent, there should be a stepwise increase in the molecular weight of the crosslinked complex, forming a ladder. Ideally, the number of steps would reflect the number of protein molecules (Bolt *et al.*, 2001).

Another approach would be to determine the molar ratio of protein to DNA in each complex in double-label experiments (Carey, 1991). The protein and DNA in the complexes could be labelled with different isotopes to allow quantification (Fox, 1976).

6.6.1.2 Specificity of transposon DNA-binding

Assembly of the *Himar1* nucleoprotein complexes is insensitive to the presence of 1-2 μ g competitor DNA, which corresponds to a 300-fold excess (in nucleotides) over the labelled transposon end (Section 6.5.1). This is in marked contrast to the behaviour of many bacterial IS elements where complex assembly is extremely sensitive to the presence of competitor DNA. *Tn10* is probably a typical example where genetic and *in vitro* experiments have shown that the transposase is strongly *cis*-acting. There, transposition is severely inhibited by competitor DNA, presumably because a strong non-specific DNA binding activity causes the transposase to interact strongly with the first DNA it encounters

(Kleckner, 1989; Kwon *et al.*, 1995; R. Chalmers, unpublished). In bacteria, this is usually the transposon DNA from which the transposase is expressed. Such *cis*-action in bacterial IS elements is probably an adaptation to minimize the rate of transposition; with a *cis*-acting transposase the rate of transposition will increase linearly with the transposon copy number and not exponentially as would be expected for a *trans* acting transposase (Kleckner, 1989). In contrast, eukaryotic transposases must be *trans*-acting because transcription and translation are physically separate, and most transposase molecules would otherwise get absorbed by genomic DNA before reaching their recognition sites. This is probably one factor contributing to the large copy number of full length transposons and miniature inverted repeat transposable elements (MITEs) in eukaryotic genomes.

Even greater precision of DNA binding is required of transposase when discriminating between cognate transposon ends and those of related transposons. In an alignment of MLE ITRs, the most conserved positions are bp+5 and +15 (Lampe *et al.*, 2001). Interestingly, exactly these positions were not protected in my hydroxyl radical footprints; instead, the protection pattern is the strongest where the sequence variability among different MLEs is the highest (Section 6.5.7). This means that the transposase has the ability to probe the DNA sequence precisely at those bases that convey specificity.

6.6.1.3 Influence of DNA sequences beyond the ITR

I found that DNA substrate length was an important factor governing complex stability. In band shifts as well as insertion assays, DNA substrates with longer transposon arms showed higher levels of complex formation (Figure 6.11B) and higher insertional activity (not shown). Complex 3 is only detectable using substrates with transposon arm lengths of at least 124 bp, but not of 67 bp or less (Figure 6.10B). The extra sequences are probably involved in stabilising protein-DNA interactions (see Figure 6.15A,B for a model). In the substrates used, *Himar1*-specific sequences only extend to bp+50, therefore any interactions with the DNA beyond bp+67 must be of a sequence non-specific nature. This is in accordance with *in vitro* plasmid-to-plasmid transposition and insertion experiments in both *Himar1* and *Mos1*, where the specific sequences required extend for only a few basepairs beyond the ITR (Lampe *et al.*, 1998; Tosi and Beverley, 2000).

In contrast to the *in vitro* findings, *Mos1 in vivo* excisions from chromosomal locations indicate the importance of several internal stretches of sequence (Lohe and Hartl, 1996a, 2002; Lozovsky *et al.*, 2002). Although these effects may be due to unrelated mechanisms, such as reorganisation of the chromatin, these sequences could be involved in

higher-order structures. An example of these is found in phage Mu, where the binding of the internal enhancer sequence is essential for LER formation (Watson and Chaconas, 1996). The existence and function of any such internal sequences in *Himar1* could offer a future line of investigation.

6.6.1.4 The molecular basis for overproduction inhibition

For *Mos1* transposition *in vivo* and *Himar1* insertion *in vitro*, transposition appears to be repressed when the transposase is present at high concentrations (Lohe and Hartl, 1996b; Lampe *et al.*, 1998). It has been suggested that this is mediated via concentration dependent aggregation and is an adaptation to limit the rate of transposition as the copy number of the transposon increases (Hartl *et al.*, 1997b). In agreement with the literature, I observed a sharp activity peak for *Himar1* insertion centring on a transposase concentration of 8 nM (Figure 6.9B, blue curve). More importantly, I found that nucleoprotein complex formation is also concentration dependent, and the highest concentration of distinct complexes is found in sharp peaks centring around 8 nM transposase (Figure 6.10B,C). Above this concentration, most of the DNA-bound transposase remains in the wells (Figure 6.10B,C), indicating that transposase inactivity at high concentrations is indeed coupled to aggregation.

Apart from the transposase concentration activity profile, there are additional parallels between *Himar1* insertion activity and complex formation. For instance, I observed that short DNA ends (143/144*) are inserted with a c. 6-fold reduced efficiency into supercoiled plasmid than longer ones (97SX*) (not shown). This is consistent with the reduced stability of complexes with shorter substrates described above (Section 6.6.1.3). Further evidence supporting the relationship between the complexes and the active species is the inhibitory effect of Ca²⁺ ions (≥ 10 mM) on both processes (not shown). As efficient insertion is dependent on the efficient formation of single- and paired ends complexes, these parallels confirm that the observed complexes are indeed early reaction intermediates.

6.6.2 The synaptic complex

A synaptic complex, which consists of paired transposon ends and the required number of transposase monomers, could not be identified conclusively.

It appears that the synaptic complex is unstable during electrophoresis and either dissociates into complexes 1 and 2 or disappears into a smear. A candidate for this PEC would be complex 3. In this case, altering the reaction and electrophoresis conditions

should eventually lead to the detection of a stable PEC. This could, for example, be achieved by removing the MBP-tag from the transposase, using a higher ratio of acrylamide to bisacrylamide or by adding metal ions or crowding reagents to the buffer. For example, detection of the recently published *Mos1* PEC requires the addition of 20% DMSO and a low concentration of magnesium ions (Dawson and Finnegan, 2003).

Alternatively, it is possible that the transposon ends only synapse stably in conjunction with target DNA. This is unusual for transposons, but has been observed for Tn7, which, as *Himar1*, can bind to target DNA before donor cleavage (Craig, 2002a) (see also Section 6.6.3.2). The high stability of the SECs is in agreement with this possibility. In addition, DMSO has a known effect on the structural requirements for transposition complexes (Mizuuchi and Mizuuchi, 1989), indicating that the *Mos1* PECs might not be representative of the *in vivo* situation. The data presented here conforms to both scenarios.

6.6.3 The target capture complex

6.6.3.1 Influence of macromolecular crowding on the TCC

Cellular interiors are occupied by a high concentration of macromolecules. In the eukaryotic cytoplasm, these typically take up 20-30% of the available volume; in some organelles such as the nucleus and in bacteria, the level of crowding is even higher, in the range of 30-45% (Fulton, 1982; Zimmerman and Trach, 1991; Bohrmann *et al.*, 1993). All intracellular systems have evolved to deal with and to make use of the resulting effects of volume exclusion and non-specific interactions. Although largely ignored in experimental biochemistry, the addition of crowding agents has already been proven essential for the *in vitro* activity of DNA replication, transcription, cell division and many other processes (Minton, 2001). The effect of crowding can be positive or negative. Solubility and diffusion rates are decreased, whereas protein folding and binding are generally increased (Ellis, 2001). As an example, binding of HU to *E. coli* DNA is increased by over 10-fold in the presence of 12% PEG (Murphy and Zimmerman, 1994).

I observed that *Himar1* target commitment was barely detectable in dilute buffer but very pronounced (c. 70%) in the presence of 10% polyethylene glycol (Section 6.4.2). This is a good example of the general observation that intracellular crowding preserves interactions between macromolecules. With 10% PEG and 0.0025%w/v BSA, the experimental reaction conditions still exhibited a lesser degree of crowding than is present inside the nucleus. Thus, *in vivo*, *Himar1* transposase probably inserts exclusively into the target it encounters first.

PEG furthermore decreased overall *Himar1* insertion efficiency (Section 6.4.3). This could be a result of PEG-mediated DNA aggregation, which would make large proportions of the DNA inaccessible to the transposase. DNA aggregation and condensation is promoted by volume exclusion (crowding), multivalent cations, high DNA concentrations and supercoiling, i.e. conditions present in both the nucleus and the *in vitro* reactions (Sikorav *et al.*, 1994; Pelta *et al.*, 1996; Bloomfield, 1996). *In vivo*, the problem of DNA aggregation is circumvented by DNA binding proteins such as histones in eukaryotes and histone-like proteins in prokaryotes (Minsky *et al.*, 1997).

6.6.3.2 Target commitment

The experiments presented here demonstrate that *Himar1* transposase can commit to target DNA before and after donor cleavage (Section 6.4.1). This is similar to the transposition of V(D)J and phage Mu, which recruit target DNA to the transpososome at various steps along the reaction pathway, before and after cleavage (Neiditch *et al.*, 2001; Naigamwalla and Chaconas, 1997). Tn7 and Tn10, on the other hand, capture target DNA at a defined point, before or after donor cleavage, respectively (Bainton *et al.*, 1991, 1993; Sakai and Kleckner, 1997).

6.6.3.2.1 Genomic distribution and target commitment

The timing of target commitment has been proposed to influence the genomic distribution of different transposons by affecting the distance over which the elements can move (Sakai and Kleckner, 1997). If this is true, the early target commitment of *Himar1* should lead to a preference for target sites nearby the donor site. However, there is no evidence for *mariner* transposons exhibiting such a preference. An analysis of the distribution of the honeybee *mariner* transposon TnM1a revealed that the 435 elements per haploid genome were dispersed throughout the genome and not tightly clustered (Ebert *et al.*, 1995). If the distribution of this element is representative of other MLEs, this would indicate that timing of target capture has no effect on the distance between the sites of excision and insertion.

6.6.3.2.2 Structural implications of target commitment for the TCC

The ability to detect target commitment with uncleaved donor DNA shows that the target interaction site is distinct from the site occupied by the donor DNA. If, on the other hand, the target DNA had occupied the same binding site within the transposase as the donor, then target commitment would have only been visible after cleavage. Similarly, the

early target commitment of *Himar1* transposase shows that, in contrast to *Tn10*, the *Himar1* target binding site is not obstructed by the transposon DNA in the TCC (Sewitz *et al.*, 2003) (see Figure 6.15C for a model). The slightly stronger amount of target commitment observed with precleaved versus uncleaved ends (Figure 6.9A; Section 6.4.2) might be due to a slight electrostatic repulsion of the target DNA by the flanking sequence.

The extended time frame for binding target DNA, before and after donor cleavage, can be envisioned to be beneficial for the completion of the *mariner* transposition reaction. The flexible target commitment might be one factor contributing to the evolutionary success of MLEs.

6.6.3.3 Implications of target DNA conformation

We observed that *Himar1* insertion is considerably more efficient into supercoiled than into linear target DNA (Sections 6.3.1, 6.5.3). Furthermore, target capture complexes were only ever observed with supercoiled plasmid target, but never with short linear target (Figure 6.11A, lanes 9-10; Figure 6.11B, lanes 5,7,9).

Equally, in the *Mos1* *in vitro* system, relaxation of supercoiled target plasmid reduces the transposition frequency by 16-fold (Tosi and Beverley, 2000). As demonstrated by increased sensitivity to DNaseI, preferred *Himar1* target sites are in regions with a high degree of DNA bending (Lampe *et al.*, 1998; Brukner *et al.*, 1995). A computational study showed preferred insertion of *Himar1* and other DDE transposons into regions with a bendable structure and a propensity to form an A-DNA-like double helix (“low A-philicity”)(Vigdal *et al.*, 2002).

One structural distinction of A-DNA is its wide and shallow minor groove that allows proteins easier access to the bases for binding (Fuller *et al.*, 1965; Ivanov and Minchenkova, 1994). DNA bending affects the width of the major and minor grooves (Brukner *et al.*, 1995), as does DNA supercoiling (Kopka *et al.*, 1985; Störl *et al.*, 1993). As negatively supercoiled DNA is underwound, it furthermore facilitates helix opening (Alberts *et al.*, 1994), which is intuitively helpful for the strand transfer step.

Preferred target sites for *Himar1* insertion therefore appear to be those stretches of DNA that either have a distorted structure or can easily be bent to adopt one. This requirement for distorted target appears to be a general principle for DNA transposition: Phage Mu displays a similar preference for supercoiling, high deformability and low A-philicity in its target sites as *Himar1* (Naigamwalla and Chaconas, 1997; Haapa-Paananen *et al.*, 2002). Studies on *Tn10* target interaction (Pribil and Haniford, 2000), HIV

integration (Muller and Varmus, 1994), RAG1/2 transposition (Lee *et al.*, 2002) and Tn7 insertion (Kuduvalli *et al.*, 2001) all show the same preferences for bent or bendable DNA. Furthermore, Tn10 transposase mutants with relaxed target site specificity have an increased ability to bend target DNA (Pribil and Haniford, 2003).

DNA in the prokaryotic nucleoid is negatively supercoiled. In the eukaryotic nucleus, most of the DNA in chromatin is also underwound, either toroidally around nucleosomes, or in plectonemic domains (Sinden, 1994). The natural state of the genomic DNA therefore provides a substrate that is close to the ideal state for the insertion reaction to take place efficiently.

6.6.4 The strand transfer complex

6.6.4.1 High stability of strand transfer complexes

The chemical steps of transposition involve no high-energy cofactors and the reaction is therefore energetically downhill. One consequence for transposons in general is that the intermediates therefore become more stable as the reaction progresses (Mizuuchi, 1992b; Kleckner *et al.*, 1996). This makes the spontaneous dissociation of complexes after completion of the reaction less likely. Dissociation even requires the input of extra energy in some cases.

In Tn10, the strand transfer complex is more stable than the double end break complex, which is more stable than the PEC (Sakai *et al.*, 1995; R. Chalmers, unpublished). Tn5 transposase remains bound to the transposition product *in vitro*, unless treated with phenol or detergent (Goryshin and Reznikoff, 1998). It has been proposed that DnaA and/or a yet unidentified protease are required to release the Tn5 transposase and allow repair of the 9 bp single stranded gaps of target DNA on either site of the transposon (Yin and Reznikoff, 1987; Reznikoff, 2002). In Mu transposition, the strand transfer complex is also extremely stable, resisting high temperature and requiring a protease for disassembly *in vivo* (Surette *et al.*, 1987; Levchenko *et al.*, 1995). Mu uses the unfolding chaperone and ATPase ClpX to destabilise the transposition complex at the junction between the recombination and replication stages of replicative transposition (Levchenko *et al.*, 1995; Kruklitis *et al.*, 1996; Jones *et al.*, 1998; Burton *et al.*, 2001). This is not only the case for prokaryotic complexes, as V(D)J post-cleavage complexes are extremely stable (Agrawal and Schatz, 1997), and after transposition, the RAG proteins appear to remain tightly associated with strand transfer complexes (Agrawal *et al.*, 1998). So far it has not been reported how these complexes are resolved.

Upon insertion into target, the *Himar1* transpososome becomes extremely stable and is resistant to SDS and partially resistant to proteinase K treatment (Section 6.3.3). From this data it is likely that after strand transfer in *Himar1*, host enzymes are employed to disassemble the complex. So far, no host proteins responsible for disassembling the STC and for repairing the insertion site have been identified. Due to its high stability, the assembled strand transfer complex would be an excellent substrate to load an affinity column and to pull down interacting proteins from cell extracts (as in Brymora *et al.*, 2001). For an alternative model for the STC disassembly mechanism, see Section 6.6.4.3.

6.6.4.2 The prevalence of supercoiling after strand transfer

The electrophoretic mobility of the *Himar1* strand transfer complex before deproteinisation shows that, despite the linear structure of the DNA, the target molecule retains its supercoiled conformation (Section 6.3.3). As the target molecule contains two nicks adjacent to the site of the insertion, the transposon ends must remain tightly associated with each other, and the transposase contacts must restrain unwinding of the target DNA flanking the insertion site.

Recent experiments from our lab indicate that the *Himar1* transposon DNA also keeps its supercoiled state after insertion (N. Buisine, unpublished). Therefore, both the target and the transposon DNA remain topologically constrained, although the molecules are not covalently closed at this point. This is similar to *Tn10*, where the supercoils of both the transposon and the target are retained after insertion (R. Chalmers, unpublished; Sakai and Kleckner, 1997).

The question arises whether the retention of the supercoils is just a result of the high stability of the strand transfer complex, which is a thermodynamic necessity, or whether there are some selectable advantages in it. As an example, self-insertion could be limited if the transposon DNA maintains its compact conformation even after excision, thus limiting the chances of it encountering the active site before an external target does.

6.6.4.3 Exonuclease activity of *Himar1* transposase

Our *Himar1* transposase preparation displayed a strong Mg^{2+} -dependent 3'→5' exonuclease activity (Section 6.5.4). In the *Mos1 in vitro* system, an equivalent exonucleolytic activity has been observed, which is abolished together with other catalytic activity in DDD-to-DDE active site mutants (C. Augé-Gouillou, personal communication). This suggests that the *Himar1* exonucleolytic activity observed here is also an intrinsic feature of the transposase and not due to a contamination.

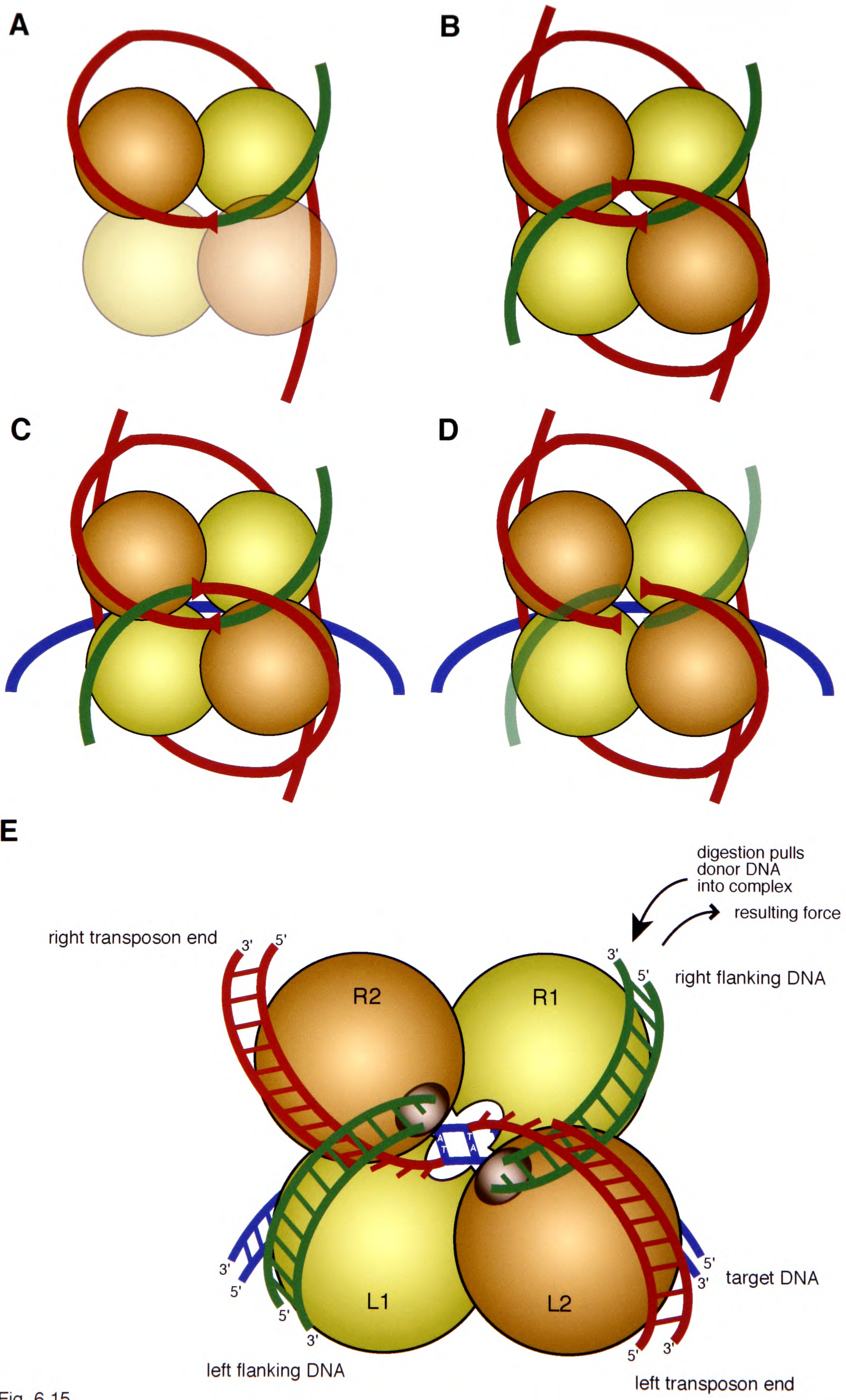


Fig. 6.15

Figure 6.15: Proposed structures of MLE nucleoprotein complexes

Constructed from the data presented here and from published data on *Himar1* and *Mos1* (see text for details and references). Red, transposon ends; green, flanking donor DNA; blue, target DNA; spheres, monomers or oligomers of transposase. Not to scale.

- (A) Single end complex (SEC). It is not yet known whether the SEC contains all proteins of the PEC (transparent spheres). Shown is specific binding of the transposon end to one molecule of transposase and non-specific contacts of flanking and subterminal transposon sequences. The shape of the DNA wrapped around the complex is hypothetical.
- (B) Putative paired ends complex (PEC). Drawn for catalysis *in trans*, as has been observed for all transposases so far.
- (C) Uncleaved target capture complex (uTCC). The target DNA probably has a sharper bend than shown here.
- (D) Cleaved target capture complex (cTCC). The optional presence of donor DNA is hypothetical.
- (E) Strand transfer complex (STC). Hypothetical structure and function. R1, binding of right flanking DNA, transposition of left transposon end; R2, specific binding of right transposon end, exonucleolytic digestion of left flanking DNA; L1, binding of left flanking DNA, transposition of right transposon end; L2, specific binding of left transposon end, exonucleolytic digestion of right flanking DNA.

Excision of MLEs and other eukaryotic transposons frequently causes deletions of variable size in the donor DNA (Bryan *et al.*, 1990). In the light of our data, it seems conceivable that the transposase itself catalyses these deletions. This requires the donor DNA to remain associated with the transposase after excision. As target capture is possible even with uncleaved ends (Sections 6.4.1, 6.5.3) the TCC has the ability to accommodate donor DNA (see Section 6.6.3.2.2). The same probably holds true for the STC. This could be shown in insertion experiments using transposon end fragments with labelled flanking DNA. Indirect evidence for the retention of the donor DNA in the STC is provided by the observation that protein binding in the STC is more stable if uncleaved rather than donor-free ends had been inserted (Section 6.3.5). Thus, a model can be constructed in which the strand transfer complex additionally contains cleaved donor DNA, and a subset of the bound transposase monomers has exonucleolytic activity (Figure 6.15E).

If true, this model could provide the basis for two speculative mechanisms that, for the first time, give the deletions a biological function. First, digestion of the donor DNA could be instrumental in the dissociation of the otherwise extremely stable strand transfer complex. The elasticity, twist and/or writhe of the DNA acts as a force, which increases in strength as the DNA is shortened, and disrupts the complex once a threshold is reached. Second, it is conceivable that the large deletions cause an increased frequency of homologous recombination versus non-homologous end joining (NHEJ) and micro-homology repair. The long-term survival of non-replicative transposons depends on the restoration of the empty donor site from the transposon-containing sister-chromatid by homologous recombination. By transposing predominantly during S-phase to anaphase (Wilson *et al.*, 2003), MLEs appear to make use of the fact that the host DNA is in duplicate and homologous recombination is increased (West, 2003). It is not yet understood what influences the choice of repair pathways (Jackson, 2002), but even a slight shift towards homologous recombination would convey a significant selective advantage.

6.6.5 Conclusion

In this study, all steps of the *Himar1* transposition process, from binding of the transposase to its recognition sites and excision of the transposon to capture of target DNA and insertion were studied in detail with a variety of methods. These results provide a foundation for the understanding of the *mariner* transposition process.

APPENDIX

A.1 Calculating the fraction of committed synaptic complexes

Target commitment was calculated from measurements of insertion products, quantified through phosphoimaging with ImageGauge software (Fuji). Measurements are given in arbitrary units, which were obtained by the following formula: 1 unit = [(phosphoimager units of measured band) – (phosphoimager units of background)] / area [pixel²]. Values obtained from insertion into target 1 are denoted “T1”, those into target 2, “T2”. Index “x” indicates the experimental reaction, where target 1 is added at stage I and target 2 at stage II. Index “c” indicates the control reaction, in which both target plasmids are added simultaneously at stage I. (Modified after Sakai and Kleckner, 1997.)

T1_x is the total of insertions into target 1. This divides into targets captured and committed during stage I (“T1_I”) and those captured during stage II (“T1_{II}”):

$$T1_x = T1_I + T1_{II} \quad (1).$$

T2_x is the total of insertions into target 2, which, due to the time of addition, was all selected during stage II:

$$T2_x = T2_{II} \quad (2).$$

T1_c and T2_c are the units of target 1 or target 2 captured and inserted into in the control reaction, where both targets had been added simultaneously. Due to the large excess of target molecules to complexes and otherwise equal conditions, it follows that:

$$\frac{T1_c}{T2_c} = \frac{T1_{II}}{T2_{II}} \quad (3).$$

Or, from (2):

$$\frac{T1_c}{T2_c} = \frac{T1_{II}}{T2_x} \quad (3a),$$

$$T1_{II} = T2_x \times \frac{T1_c}{T2_c} \quad (3b).$$

Substituting (3b) into (1) gives:

$$T1_I = T1_x - T2_x \times \frac{T1_c}{T2_c} \quad (4).$$

T1_I is a measure of the complexes that have committed to target 1 during stage I. For a percentage of target commitment (tc), this is put in relation to the total insertion in the reaction:

$$tc(\%) = 100 \times \frac{T1_x - T2_x \times \frac{T1_c}{T2_c}}{T1_x + T2_x} \quad (5).$$

A.2 Target commitment measurements and results of calculations

substrate		target 1	target 2	PEG (%)		lane in		insertion (units)				% target commit- ment	
		stage I	stage II	I	II	c	x	control		experiment			
								T1 _c	T2 _c	T1 _x	T2 _x		
uncleaved	97SX	L	S	-	-	i,4	i,5	103.2	113.3	49.4	49.5	4.4	
		S	L	-	-	i,4	i,6	113.3	103.2	60.0	36.2	21.1	
		L	S	-	10	i,7	i,8	7.9	15.5	12.9	9.8	35.0	
		S	L	-	10	i,7	i,9	15.5	7.9	31.2	3.6	69.7	
		L	S	10	10	i,10	i,11	9.8	25.2	13.1	0.0	100.0	
		S	L	10	10	i,10	i,11	25.2	9.8	67.8	0.0	100.0	
	97SX	L	S	-	-			32.5	33.1	32.2	25.6	12.3	
		S	L	-	-			33.1	32.5	31.8	30.2	1.7	
		L	S	-	10			16.0	18.4	15.6	3.8	63.4	
		S	L	-	10			18.4	16.0	29.3	12.7	35.1	
	97SX	L	S	-	-			29.0	29.3	28.2	26.1	4.4	
		S	L	-	-			29.3	29.0	34.9	32.2	3.5	
		L	S	-	10			16.1	19.0	10.7	9.5	13.2	
		S	L	-	10			19.0	16.1	30.7	18.5	18.1	
	97SX	L	S	-	-			22.9	25.4	19.4	23.6	-4.4	
		L	S	-	10			13.2	16.7	12.4	13.0	8.5	
		L	S	10	10			10.0	14.5	16.0	5.2	58.2	
		L	S	-	-			26.6	28.6	26.5	27.0	2.7	
		L	S	-	10			12.6	17.8	17.3	15.7	18.7	
		L	S	10	10			4.0	6.2	0.8	0.9	13.9	
precleaved	105BX	L	S	-	-	ii,4	ii,5	93.8	181.5	33.7	33.1	24.9	
		S	L	-	-	ii,4	ii,6	181.5	93.8	153.5	50.8	27.1	
		L	S	-	10	ii,7	ii,8	56.0	125.9	56.2	43.2	37.2	
		S	L	-	10	ii,7	ii,9	125.9	56.0	139.6	29.4	43.6	
		L	S	10	10	ii,10	ii,11	47.4	108.0	85.0	19.8	72.9	
		S	L	10	10	ii,10	ii,12	108.0	47.4	218.9	14.8	79.2	
	141/142	L	S	-	-			9.0	12.9	8.0	6.7	22.6	
		L	S	-	10			11.5	19.5	15.1	11.8	30.1	
		L	S	-	-			6.8	11.6	8.9	9.8	17.2	
		L	S	-	10			15.1	26.7	12.2	7.1	42.2	
	column	1	2	3	4	5	6	7	8	9	10	11	12

L, larger target, pBluescriptII-SK+; S, smaller target, pKL104, added at stage I or II as indicated in columns 2-3. T1, T2: phosphorimager units of insertion products (lin + linT) into target 1 or target 2, respectively. Index c (T1/2_c), insertion into target 1/2 in control reaction (columns 8-9); index x (T1/2_x), insertion in experimental reaction (columns 10-11); % target commitment, level of target commitment (column 12). Horizontal lines separate data from different gels.

target commitment (%) = 100 x (T1_x - T2_x x T1_c / T2_c) / (T1_x + T2_x) (equation (5), Appendix A.1).

A.3 ANOVA definitional formulas

source of variation	sum of squares (SS)	degrees of freedom (df)	mean square (MS)	F_0
A treatments (between)	$SS_A = \sum_{i=1}^a \frac{y_{i.}^2}{n_{i.}} - \frac{y_{..}^2}{n_{..}}$	$df_A = a - 1$	$MS_A = \frac{SS_A}{df_A}$	$F_0 = \frac{MS_A}{MS_E}$
B treatments (between)	$SS_B = \sum_{j=1}^b \frac{y_{.j}^2}{n_{.j}} - \frac{y_{..}^2}{n_{..}}$	$df_B = b - 1$	$MS_B = \frac{SS_B}{df_B}$	$F_0 = \frac{MS_B}{MS_E}$
interaction	$SS_{AB} = \sum_{i=1}^a \sum_{j=1}^b \frac{y_{ij.}^2}{n_{ij.}} - \frac{y_{..}^2}{n_{..}} - SS_A - SS_B$	$df_{AB} = (a - 1)(b - 1)$	$MS_{AB} = \frac{SS_{AB}}{df_{AB}}$	$F_0 = \frac{MS_{AB}}{MS_E}$
error (within treatments)	$SS_E = SS_T - SS_A - SS_B - SS_{AB}$	$df_E = \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)$	$MS_E = \frac{SS_E}{df_E}$	
total	$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} y_{ijk}^2 - \frac{y_{..}^2}{n_{..}}$	$df_T = n_{..} - 1$		

Analysis of variance (ANOVA) with unbalanced proportional data in a two-factor factorial design (Modified from Montgomery, 1997).

SS: sum of squares; df, degrees of freedom; MS, mean square; F_0 , test statistic; $y_{i.}$, total of all observations under the i^{th} level of factor A; $y_{.j}$, total of all observations under the j^{th} level of factor B; $y_{ij.}$, total of all observations in the ij^{th} cell; $y_{...}$, grand total of all the observations. Mathematically:

$$y_{i.} = \sum_{j=1}^b \sum_{k=1}^{n_{ij}} y_{ijk} ; y_{.j} = \sum_{i=1}^a \sum_{k=1}^{n_{ij}} y_{ijk} ; y_{ij.} = \sum_{k=1}^{n_{ij}} y_{ijk} ; y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} y_{ijk} .$$

A.4 DnaCalc for designing arbitrary primers

Every base is given a value from 0 to 4: A=0, C=1, G=2, T=3. Sequences are displayed as quaternary numbers. For a triplet: (value of 3rd base) + (value of 2nd base x 4) + (value of 1st base x 16); for a quadruplet: (value of 4th base) + (value of 3rd base x 4) + (value of 2nd base x 16) + (value of 1st base x 64); e.g.: ACTG = 2 + 3x4 + 1x16 + 0x64 = 30. These quaternary numbers were calculated for all triplets and quadruplets in the sequence of interest. [Here: the sequences between binding site for primer 1 (KL55/KL56) and the transposon end; sum of left and right *hobo* end = 400 bp.] The frequency of all these quaternary numbers in the analysed sequence was calculated. Those numbers with a frequency of zero were converted back to sequence. From these, 5-nucleotide-sequences were designed for the specific 3' ends of the arbitrary primers. Calculations were done in Microsoft Excel. (With advice from Sean D. Colloms.)

REFERENCES

- Abad, P., Cerutti, M., Pauron, D., Quiles, C., Palin, B., Devauchelle, G. and Dalmasso, A. (1993) Expression and biochemical characterization of the DNA binding activity of TcA, the putative transposase of *Caenorhabditis elegans* transposable element Tc1. *Biochem Biophys Res Commun* **192**, 1445-52.
- Adams, M.D., Celniker, S.E., Holt, R.A. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-95.
- Agrawal, A. and Schatz, D.G. (1997) RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell* **89**, 43-53.
- Agrawal, A., Eastman, Q.M. and Schatz, D.G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744-51.
- Aguilar, L.K. and Belmont, J.W. (1991) V gamma 3 T cell receptor rearrangement and expression in the adult thymus. *J Immunol* **146**, 1348-52.
- Akamatsu, Y., Tsurushita, N., Nagawa, F., Matsuoka, M., Okazaki, K., Imai, M. and Sakano, H. (1994) Essential residues in V(D)J recombination signals. *J Immunol* **153**, 4520-9.
- Akamatsu, Y., Monroe, R., Dudley, D.D., Elkin, S.K., Gartner, F., Talukder, S.R., Takahama, Y., Alt, F.W., Bassing, C.H. and Oettinger, M.A. (2003) Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. *Proc Natl Acad Sci U S A* **100**, 1209-14.
- Akerley, B.J., Rubin, E.J., Camilli, A., Lampe, D.J., Robertson, H.M. and Mekalanos, J.J. (1998) Systematic identification of essential genes by *in vitro* mariner mutagenesis. *Proc Natl Acad Sci U S A* **95**, 8927-32.
- Akerley, B.J. and Lampe, D.J. (2002) Analysis of gene function in bacterial pathogens by GAMBIT. In Clark, V.L. and Bavoil, P.M. (eds.), *Bacterial pathogenesis, Part C: Identification, regulation, and function of virulence factors*. Academic Press, San Diego, Vol. 358, pp. 100-8.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994) *Molecular biology of the cell*. Garland Publishing, New York.
- Amyes, S.G. and Smith, J.T. (1976) The purification and properties of the trimethoprim-resistant dihydrofolate reductase mediated by the R-factor, R388. *Eur J Biochem* **61**, 597-603.
- Aravind, L. and Koonin, E.V. (1999) Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J Mol Biol* **287**, 1023-40.
- Arbuckle, J.L., Fauss, L.A., Simpson, R., Ptaszek, L.M. and Rodgers, K.K. (2001) Identification of two topologically independent domains in RAG1 and their role in macromolecular interactions relevant to V(D)J recombination. *J Biol Chem* **276**, 37093-101.
- Arkhipova, I. and Meselson, M. (2000) Transposable elements in sexual and ancient asexual taxa. *Proc Natl Acad Sci U S A* **97**, 14473-7.
- Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J. and Kourilsky, P. (1999) A direct estimate of the human alphabeta T cell receptor diversity. *Science* **286**, 958-61.
- Ashour, J. and Hondalus, M.K. (2003) Phenotypic mutants of the intracellular actinomycete *Rhodococcus equi* created by *in vivo* Himar1 transposon mutagenesis. *J Bacteriol* **185**, 2644-52.
- Atkinson, P.W., Warren, W.D. and O'Brochta, D.A. (1993) The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc Natl Acad Sci U S A* **90**, 9693-7.

- Augé-Gouillou, C., Bigot, Y., Pollet, N., Hamelin, M.H., Meunier-Rotival, M. and Periquet, G. (1995) Human and other mammalian genomes contain transposons of the *mariner* family. *FEBS Lett* **368**, 541-6.
- Augé-Gouillou, C., Notareschi-Leroy, H., Abad, P., Periquet, G. and Bigot, Y. (2000) Phylogenetic analysis of the functional domains of *mariner*-like element (MLE) transposases. *Mol Gen Genet* **264**, 506-13.
- Augé-Gouillou, C., Hamelin, M.H., Demattei, M.V., Periquet, G. and Bigot, Y. (2001a) The ITR binding domain of the *mariner Mos-1* transposase. *Mol Genet Genomics* **265**, 58-65.
- Augé-Gouillou, C., Hamelin, M.H., Demattei, M.V., Periquet, M. and Bigot, Y. (2001b) The wild-type conformation of the *Mos-1* inverted terminal repeats is suboptimal for transposition in bacteria. *Mol Genet Genomics* **265**, 51-7.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds.). (1994) *Current protocols in molecular biology*. John Wiley and Sons, Inc.
- Avila, P. and de la Cruz, F. (1988) Physical and genetic map of the IncW plasmid R388. *Plasmid* **20**, 155-7.
- Bachmann, B.J. (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev* **36**, 525-57.
- Bainton, R., Gamas, P. and Craig, N.L. (1991) Tn7 transposition *in vitro* proceeds through an excised transposon intermediate generated by staggered breaks in DNA. *Cell* **65**, 805-16.
- Bainton, R.J., Kubo, K.M., Feng, J.N. and Craig, N.L. (1993) Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified *in vitro* system. *Cell* **72**, 931-43.
- Baran, G., Echt, C., Bureau, T. and Wessler, S. (1992) Molecular analysis of the maize wx-B3 allele indicates that precise excision of the transposable Ac element is rare. *Genetics* **130**, 377-84.
- Bassing, C.H., Swat, W. and Alt, F.W. (2002) The mechanism and regulation of chromosomal V(D)J recombination. *Cell* **109 Suppl**, S45-55.
- Basu, M., Hegde, M.V. and Modak, M.J. (1983) Synthesis of compositionally unique DNA by terminal deoxynucleotidyl transferase. *Biochem Biophys Res Commun* **111**, 1105-12.
- Bazin, C. and Higuët, D. (1996) Lack of correlation between dysgenic traits in the *hobo* system of hybrid dysgenesis in *Drosophila melanogaster*. *Genet Res* **67**, 219-26.
- Bazin, C., Denis, B., Capy, P., Bonnivard, E. and Higuët, D. (1999) Characterization of permissivity for *hobo*-mediated gonadal dysgenesis in *Drosophila melanogaster*. *Mol Gen Genet* **261**, 480-6.
- Beall, E.L., Admon, A. and Rio, D.C. (1994) A *Drosophila* protein homologous to the human p70 Ku autoimmune antigen interacts with the *P* transposable element inverted repeats. *Proc Natl Acad Sci U S A* **91**, 12681-5.
- Becker, D., Lutticke, R., Li, M. and Starlinger, P. (1992) Control of excision frequency of maize transposable element *Ds* in *Petunia* protoplasts. *Proc Natl Acad Sci U S A* **89**, 5552-6.
- Becker, H.A. and Kunze, R. (1996) Binding sites for maize nuclear proteins in the subterminal regions of the transposable element *Activator*. *Mol Gen Genet* **251**, 428-35.
- Becker, H.A. and Kunze, R. (1997) Maize *Activator* transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. *Mol Gen Genet* **254**, 219-30.
- Beese, L.S. and Steitz, T.A. (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *Embo J* **10**, 25-33.
- Bender, J. and Kleckner, N. (1992) IS10 transposase mutations that specifically alter target site recognition. *Embo J* **11**, 741-50.

- Bhasin, A., Goryshin, I.Y. and Reznikoff, W.S. (1999) Hairpin formation in Tn5 transposition. *J Biol Chem* **274**, 37021-9.
- Bhasin, A., Goryshin, I.Y., Steiniger-White, M., York, D. and Reznikoff, W.S. (2000) Characterization of a Tn5 pre-cleavage synaptic complex. *J Mol Biol* **302**, 49-63.
- Biery, M.C., Lopata, M. and Craig, N.L. (2000) A minimal system for Tn7 transposition: the transposon-encoded proteins TnsA and TnsB can execute DNA breakage and joining reactions that generate circularized Tn7 species. *J Mol Biol* **297**, 25-37.
- Blackman, R.K., Grimaila, R., Koehler, M.M. and Gelbart, W.M. (1987) Mobilization of *hobo* elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. *Cell* **49**, 497-505.
- Blackman, R.K., Koehler, M.M., Grimaila, R. and Gelbart, W.M. (1989) Identification of a fully-functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. *Embo J* **8**, 211-7.
- Bloomfield, V.A. (1996) DNA condensation. *Curr Opin Struct Biol* **6**, 334-41.
- Boehm, U., Heinlein, M., Behrens, U. and Kunze, R. (1995) One of three nuclear localization signals of maize *Activator* (*Ac*) transposase overlaps the DNA-binding domain. *Plant J* **7**, 441-51.
- Bogue, M.A., Wang, C., Zhu, C. and Roth, D.B. (1997) V(D)J recombination in Ku86-deficient mice: distinct effects on coding, signal, and hybrid joint formation. *Immunity* **7**, 37-47.
- Bohrmann, B., Haider, M. and Kellenberger, E. (1993) Concentration evaluation of chromatin in unstained resin-embedded sections by means of low-dose ratio-contrast imaging in STEM. *Ultramicroscopy* **49**, 235-51.
- Bolland, S. and Kleckner, N. (1995) The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. *Proc Natl Acad Sci U S A* **92**, 7814-8.
- Bolt, E.L., Lloyd, R.G. and Sharples, G.J. (2001) Genetic analysis of an archaeal Holliday junction resolvase in *Escherichia coli*. *J Mol Biol* **310**, 577-89.
- Bonnivard, E., Bazin, C., Denis, B. and Higuete, D. (2000) A scenario for the *hobo* transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem TPE repeats. *Genet Res* **75**, 13-23.
- Boussy, I.A. and Daniels, S.B. (1991) *hobo* transposable elements in *Drosophila melanogaster* and *D. simulans*. *Genet Res* **58**, 27-34.
- Braam, L.A., Goryshin, I.Y. and Reznikoff, W.S. (1999) A mechanism for Tn5 inhibition. carboxyl-terminal dimerization. *J Biol Chem* **274**, 86-92.
- Breitling, R. and Gerber, J.K. (2000) Origin of the paired domain. *Dev Genes Evol* **210**, 644-50.
- Breslauer, K.J., Frank, R., Blocker, H. and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A* **83**, 3746-50.
- Brukner, I., Sanchez, R., Suck, D. and Pongor, S. (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *Embo J* **14**, 1812-8.
- Bryan, G., Garza, D. and Hartl, D. (1990) Insertion and excision of the transposable element *mariner* in *Drosophila*. *Genetics* **125**, 103-14.
- Brymora, A., Cousin, M.A., Roufogalis, B.D. and Robinson, P.J. (2001) Enhanced protein recovery and reproducibility from pull-down assays and immunoprecipitations using spin columns. *Anal Biochem* **295**, 119-22.
- Bureau, T.E., Ronald, P.C. and Wessler, S.R. (1996) A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proc Natl Acad Sci U S A* **93**, 8524-9.

- Burton, B.M., Williams, T.L. and Baker, T.A. (2001) ClpX-mediated remodeling of Mu transpososomes: selective unfolding of subunits destabilizes the entire complex. *Mol Cell* **8**, 449-54.
- Cabaniols, J.P., Fazilleau, N., Casrouge, A., Kourilsky, P. and Kanellopoulos, J.M. (2001) Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. *J Exp Med* **194**, 1385-90.
- Callebaut, I. and Mornon, J.P. (1998) The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis. *Cell Mol Life Sci* **54**, 880-91.
- Calvi, B.R., Hong, T.J., Findley, S.D. and Gelbart, W.M. (1991) Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*. *Cell* **66**, 465-71.
- Calvi, B.R. and Gelbart, W.M. (1994) The basis for germline specificity of the *hobo* transposable element in *Drosophila melanogaster*. *Embo J* **13**, 1636-44.
- Capy, P., Langin, T., Bigot, Y., Brunet, F., Daboussi, M.J., Periquet, G., David, J.R. and Hartl, D.L. (1994) Horizontal transmission versus ancient origin: *mariner* in the witness box. *Genetica* **93**, 161-70.
- Capy, P., Vitalis, R., Langin, T., Higuier, D. and Bazin, C. (1996) Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? *J Mol Evol* **42**, 359-68.
- Carey, J. (1991) Gel Retardation. In Sauer, R.T. (ed.), *Protein - DNA Interactions*. Academic Press, Inc., San Diego, CA, Vol. 208, pp. 103-17.
- Chaconas, G. and Harshey, R.M. (2002) Transposition of phage Mu DNA. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 384-402.
- Chalmers, R., Guhathakurta, A., Benjamin, H. and Kleckner, N. (1998) IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* **93**, 897-908.
- Chalmers, R., Sewitz, S., Lipkow, K. and Crellin, P. (2000) Complete nucleotide sequence of Tn10. *J Bacteriol* **182**, 2970-2.
- Chalmers, R.M. and Kleckner, N. (1994) Tn10/IS10 transposase purification, activation, and *in vitro* reaction. *J Biol Chem* **269**, 8029-35.
- Chalmers, R.M. and Kleckner, N. (1996) IS10/Tn10 transposition efficiently accommodates diverse transposon end configurations. *Embo J* **15**, 5112-22.
- Chalmers, R.M. and Blot, M. (1999) IS elements and transposons. In Charlebois, R.L. (ed.), *Organization of the Prokaryotic Genome*. ASM Press, Washington, DC.
- Chandler, M. and Mahillon, J. (2002) Insertion sequences revisited. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 305-66.
- Chen, C.-L. (1998) *Distribution of P elements in Muscamorpha*. Ph.D. thesis, University of Illinois, Urbana-Champaign.
- Chen, L., Trujillo, K., Sung, P. and Tomkinson, A.E. (2000) Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J Biol Chem* **275**, 26196-205.
- Cheung, S., Arndt, K. and Lu, P. (1984) Correlation of *lac* operator DNA imino proton exchange kinetics with its function. *Proc Natl Acad Sci U S A* **81**, 3665-9.
- Ciubotaru, M., Ptaszek, L.M., Baker, G.A., Baker, S.N., Bright, F.V. and Schatz, D.G. (2003) RAG1-DNA binding in V(D)J recombination. Specificity and DNA-induced

conformational changes revealed by fluorescence and CD spectroscopy. *J Biol Chem* **278**, 5584-96.

- Claudianos, C., Brownlie, J., Russell, R., Oakeshott, J. and Whyard, S. (2002) maT-A clade of transposons intermediate between *mariner* and Tc1. *Mol Biol Evol* **19**, 2101-9.
- Clewell, D.B. and Helinski, D.R. (1970) Properties of a supercoiled deoxyribonucleic acid-protein relaxation complex and strand specificity of the relaxation event. *Biochemistry* **9**, 4428-40.
- Coen, E.S., Carpenter, R. and Martin, C. (1986) Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*. *Cell* **47**, 285-96.
- Collins, J., Forbes, E. and Anderson, P. (1989) The Tc3 family of transposable genetic elements in *Caenorhabditis elegans*. *Genetics* **121**, 47-55.
- Colot, V., Haedens, V. and Rossignol, J.L. (1998) Extensive, nonrandom diversity of excision footprints generated by *Ds*- like transposon *Ascot-1* suggests new parallels with V(D)J recombination. *Mol Cell Biol* **18**, 4337-46.
- Corish, P., Black, D.M., Featherston, D.W., Merriam, J. and Dover, G.A. (1996) Natural repressors of *P*-induced hybrid dysgenesis in *Drosophila melanogaster*: a model for repressor evolution. *Genet Res* **67**, 109-21.
- Craig, N.L. (2002a) Tn7. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 423-56.
- Craig, N.L. (2002b) Mobile DNA: An introduction. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 3-11.
- Craigie, R. (2002) Retroviral DNA integration. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 613-30.
- Crellin, P. and Chalmers, R. (2001) Protein-DNA contacts and conformational changes in the Tn10 transpososome during assembly and activation for cleavage. *Embo J* **20**, 3882-91.
- Cui, Z., Geurts, A.M., Liu, G., Kaufman, C.D. and Hackett, P.B. (2002) Structure-function analysis of the inverted terminal repeats of the *Sleeping Beauty* transposon. *J Mol Biol* **318**, 1221-35.
- Cuomo, C.A. and Oettinger, M.A. (1994) Analysis of regions of RAG-2 important for V(D)J recombination. *Nucleic Acids Res* **22**, 1810-4.
- Cuomo, C.A., Mundy, C.L. and Oettinger, M.A. (1996) DNA sequence and structure requirements for cleavage of V(D)J recombination signal sequences. *Mol Cell Biol* **16**, 5683-90.
- Daboussi, M.J., Langin, T. and Brygoo, Y. (1992) *Fot1*, a new family of fungal transposable elements. *Mol Gen Genet* **232**, 12-6.
- Davies, D.R., Braam, L.M., Reznikoff, W.S. and Rayment, I. (1999) The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-Å resolution. *J Biol Chem* **274**, 11904-13.
- Davies, D.R., Goryshin, I.Y., Reznikoff, W.S. and Rayment, I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**, 77-85.
- Dawson, A. and Finnegan, D.J. (2003) Excision of the *Drosophila mariner* transposon *Mos1*. Comparison with bacterial transposition and V(D)J recombination. *Mol Cell* **11**, 225-35.
- de la Cruz, N.B., Weinreich, M.D., Wiegand, T.W., Krebs, M.P. and Reznikoff, W.S. (1993) Characterization of the Tn5 transposase and inhibitor proteins: a model for the inhibition of transposition. *J Bacteriol* **175**, 6932-8.
- DeVault, J.D., Hughes, K.J., Leopold, R.A., Johnson, O.A. and Narang, S.K. (1996) Gene transfer into corn earworm (*Helicoverpa zea*) embryos. *Genome Res* **6**, 571-9.

- Difilippantonio, M.J., McMahan, C.J., Eastman, Q.M., Spanopoulou, E. and Schatz, D.G. (1996) RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* **87**, 253-62.
- Doak, T.G., Doerder, F.P., Jahn, C.L. and Herrick, G. (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proc Natl Acad Sci U S A* **91**, 942-6.
- Dowe, M.F., Jr., Roman, G.W. and Klein, A.S. (1990) Excision and transposition of two *Ds* transposons from the bronze mutable 4 derivative 6856 allele of *Zea mays* L. *Mol Gen Genet* **221**, 475-85.
- DuBridge, R.B. and Calos, M.P. (1988) Recombinant shuttle vectors for the study of mutation in mammalian cells. *Mutagenesis* **3**, 1-9.
- Dunn, J.J., Kripl, B., Bernstein, K.E., Westphal, H. and Studier, F.W. (1988) Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene* **68**, 259-66.
- Dynan, W.S. and Yoo, S. (1998) Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Res* **26**, 1551-9.
- Eastman, Q.M., Leu, T.M. and Schatz, D.G. (1996) Initiation of V(D)J recombination *in vitro* obeying the 12/23 rule. *Nature* **380**, 85-8.
- Ebert, P.R., Hileman, J.P.t. and Nguyen, H.T. (1995) Primary sequence, copy number, and distribution of *mariner* transposons in the honey bee. *Insect Mol Biol* **4**, 69-78.
- Eide, D. and Anderson, P. (1988) Insertion and excision of *Caenorhabditis elegans* transposable element *Tc1*. *Mol Cell Biol* **8**, 737-46.
- Elkin, S.K., Matthews, A.G. and Oettinger, M.A. (2003) The C-terminal portion of RAG2 protects against transposition *in vitro*. *Embo J* **22**, 1931-8.
- Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* **26**, 597-604.
- Elroy-Stein, O., Fuerst, T.R. and Moss, B. (1989) Cap-independent translation of mRNA conferred by encephalomyocarditis virus 5' sequence improves the performance of the vaccinia virus/bacteriophage T7 hybrid expression system. *Proc Natl Acad Sci U S A* **86**, 6126-30.
- Engels, W.R., Johnson-Schlitz, D.M., Eggleston, W.B. and Sved, J. (1990) High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell* **62**, 515-25.
- Engler, P., Klotz, E. and Storb, U. (1992) N region diversity of a transgenic substrate in fetal and adult lymphoid cells. *J Exp Med* **176**, 1399-404.
- Essers, L., Adolphs, R.H. and Kunze, R. (2000) A highly conserved domain of the maize *Activator* transposase is involved in dimerization. *Plant Cell* **12**, 211-24.
- Fayet, O., Ramond, P., Polard, P., Prere, M.F. and Chandler, M. (1990) Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol Microbiol* **4**, 1771-7.
- Feeney, A.J., Tang, A. and Ogwaro, K.M. (2000) B-cell repertoire formation: role of the recombination signal sequence in non-random V segment utilization. *Immunol Rev* **175**, 59-69.
- Feldmar, S. and Kunze, R. (1991) The ORFa protein, the putative transposase of maize transposable element *Ac*, has a basic DNA binding domain. *Embo J* **10**, 4003-10.
- Feng, J.A., Johnson, R.C. and Dickerson, R.E. (1994) *Hin* recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science* **263**, 348-55.
- Ferguson, K.A. (1964) Starch-gel electrophoresis - application to the classification of pituitary proteins and polypeptides. *Metabolism* **13**, 985-1002.

- Feschotte, C. and Wessler, S.R. (2002) Mariner-like transposases are widespread and diverse in flowering plants. *Proc Natl Acad Sci U S A* **99**, 280-5.
- Feschotte, C., Swamy, L. and Wessler, S.R. (2003) Genome-wide analysis of *mariner*-like transposable elements in rice reveals complex relationships with *Stowaway* miniature inverted repeat transposable elements (MITEs). *Genetics* **163**, 747-58.
- Fischer, S.E., Wienholds, E. and Plasterk, R.H. (2001) Regulated transposition of a fish transposon in the mouse germ line. *Proc Natl Acad Sci U S A* **98**, 6759-64.
- Foster, T.J., Davis, M.A., Roberts, D.E., Takeshita, K. and Kleckner, N. (1981) Genetic organization of transposon Tn10. *Cell* **23**, 201-13.
- Fox, B.W. (1976) *Techniques of sample preparation for liquid scintillation counting*. Elsevier, New York.
- Frank, K.M., Sekiguchi, J.M., Seidl, K.J., Swat, W., Rathbun, G.A., Cheng, H.L., Davidson, L., Kangaloo, L. and Alt, F.W. (1998) Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173-7.
- Franz, G., Loukeris, T.G., Dialektaki, G., Thompson, C.R. and Savakis, C. (1994) Mobile Minos elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. *Proc Natl Acad Sci U S A* **91**, 4746-50.
- Freemont, P.S., Hanson, I.M. and Trowsdale, J. (1991) A novel cysteine-rich sequence motif. *Cell* **64**, 483-4.
- Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. and Schatz, D.G. (2000) The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol* **18**, 495-527.
- Fugmann, S.D. and Schatz, D.G. (2001) Identification of basic residues in RAG2 critical for DNA binding by the RAG1-RAG2 complex. *Mol Cell* **8**, 899-910.
- Fujimoto, S. and Yamagishi, H. (1987) Isolation of an excision product of T-cell receptor alpha-chain gene rearrangements. *Nature* **327**, 242-3.
- Fuller, W., Wilkins, M.H.F., Wilson, H.R., Hamilton, L.D. and Arnott, S. (1965) The molecular configuration of deoxuribonucleic acid. IV. X-ray diffraction study of the A form. *J Mol Biol* **12**, 60-80.
- Fulton, A.B. (1982) How crowded is the cytoplasm? *Cell* **30**, 345-7.
- Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fujiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J., Bronson, R.T., Malynn, B.A., Bryans, M., Zhu, C., Chaudhuri, J., Davidson, L., Ferrini, R., Stamato, T., Orkin, S.H., Greenberg, M.E. and Alt, F.W. (1998) A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**, 891-902.
- Garcia-Fernández, J., Bayascas-Ramirez, J.R., Marfany, G., Munoz-Marmol, A.M., Casali, A., Baguña, J. and Saló, E. (1995) High copy number of highly similar *mariner*-like transposons in planarian (*Platyhelminthe*): evidence for a trans-phyla horizontal transfer. *Mol Biol Evol* **12**, 421-31.
- Garcia-Saez, I. and Plasterk, R.H. (2000) Purification of the *Caenorhabditis elegans* transposase Tc1A refolded during gel filtration chromatography. *Protein Expr Purif* **19**, 355-61.
- Gauss, G.H. and Lieber, M.R. (1996) Mechanistic constraints on diversity in human V(D)J recombination. *Mol Cell Biol* **16**, 258-69.
- Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G. and Wuthrich, K. (1994) Homeodomain-DNA recognition. *Cell* **78**, 211-23.
- Gellert, M. (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* **71**, 101-32.

- Gerstein, R.M. and Lieber, M.R. (1993) Extent to which homology can constrain coding exon junctional diversity in V(D)J recombination. *Nature* **363**, 625-7.
- Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C. and Mathis, D. (1993) Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* **261**, 1175-8.
- Glover, D.M. (ed.). (1985-1987) *DNA cloning: a practical approach*. IRL, Oxford.
- Gluzman, Y. (1981) SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**, 175-82.
- Godderz, L.J., Rahman, N.S., Risinger, G.M., Arbuckle, J.L. and Rodgers, K.K. (2003) Self-association and conformational properties of RAG1: implications for formation of the V(D)J recombinase. *Nucleic Acids Res* **31**, 2014-23.
- Gomez, S.P. and Handler, A.M. (1997) A *Drosophila melanogaster* *hobo-white*⁺ vector mediates low frequency gene transfer in *D. virilis* with full interspecific *white*⁺ complementation. *Insect Mol Biol* **6**, 165-71.
- Gomulski, L.M., Torti, C., Malacrida, A.R. and Gasperi, G. (1997) *Ccmar1*, a full-length *mariner* element from the Mediterranean fruit fly, *Ceratitidis capitata*. *Insect Mol Biol* **6**, 241-53.
- Gorbunova, V. and Levy, A.A. (1997a) Circularized *Ac/Ds* transposons: formation, structure and fate. *Genetics* **145**, 1161-9.
- Gorbunova, V. and Levy, A.A. (1997b) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res* **25**, 4650-7.
- Gorbunova, V. and Levy, A.A. (2000) Analysis of extrachromosomal *Ac/Ds* transposable elements. *Genetics* **155**, 349-59.
- Goryshin, I.Y. and Reznikoff, W.S. (1998) Tn5 *in vitro* transposition. *J Biol Chem* **273**, 7367-74.
- Grant, S.G., Jessee, J., Bloom, F.R. and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci U S A* **87**, 4645-9.
- Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K. and Lieber, M.R. (1998) DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Mol Cell* **2**, 477-84.
- Grawunder, U. and Harfst, E. (2001) How to make ends meet in V(D)J recombination. *Curr Opin Immunol* **13**, 186-94.
- Grindley, N.D.F. (2002) The movement of Tn3-like elements: Transposition and cointegrate resolution. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 272-302.
- Guyer, M.S., Reed, R.R., Steitz, J.W. and Low, K.B. (1980) Identification of a sex-factor affinity site in *E. coli* as $\gamma\delta$. *Cold Spring Harbor Symp. Quant. Biol.* **45**, 135-40.
- Haapa-Paananen, S., Rita, H. and Savilahti, H. (2002) DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection *in vitro*. *J Biol Chem* **277**, 2843-51.
- Han, J.O., Erskine, L.A., Purugganan, M.M., Stamato, T.D. and Roth, D.B. (1998) V(D)J recombination intermediates and non-standard products in XRCC4-deficient cells. *Nucleic Acids Res* **26**, 3769-75.
- Han, J.O., Steen, S.B. and Roth, D.B. (1999) Intermolecular V(D)J recombination is prohibited specifically at the joining step. *Mol Cell* **3**, 331-8.
- Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-80.
- Handler, A.M. and Gomez, S.P. (1995) The *hobo* transposable element has transposase-dependent and - independent excision activity in drosophilid species. *Mol Gen Genet* **247**, 399-408.

- Haniford, D.H. (2002) Transposon Tn10. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 457-83.
- Haren, L., Ton-Hoang, B. and Chandler, M. (1999) Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol* **53**, 245-81.
- Harrington, J.J. and Lieber, M.R. (1994) The characterization of a mammalian DNA structure-specific endonuclease. *Embo J* **13**, 1235-46.
- Hartl, D.L. (1996) The most unkindest cut of all. *Nat Genet* **12**, 227-9.
- Hartl, D.L., Lohe, A.R. and Lozovskaya, E.R. (1997a) Modern thoughts on an ancient mariner: function, evolution, regulation. *Annu Rev Genet* **31**, 337-58.
- Hartl, D.L., Lozovskaya, E.R., Nurminsky, D.I. and Lohe, A.R. (1997b) What restricts the activity of *mariner*-like transposable elements. *Trends Genet* **13**, 197-201.
- Hehl, R., Nacken, W.K., Krause, A., Saedler, H. and Sommer, H. (1991) Structural analysis of *Tam3*, a transposable element from *Antirrhinum majus*, reveals homologies to the *Ac* element from maize. *Plant Mol Biol* **16**, 369-71.
- Heinlein, M., Brattig, T. and Kunze, R. (1994) *In vivo* aggregation of maize *Activator* (*Ac*) transposase in nuclei of maize endosperm and *Petunia* protoplasts. *Plant J* **5**, 705-14.
- Henikoff, S. (1992) Detection of *Caenorhabditis* transposon homologs in diverse organisms. *New Biol* **4**, 382-8.
- Hesse, J.E., Lieber, M.R., Gellert, M. and Mizuuchi, K. (1987) Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* **49**, 775-83.
- Hesse, J.E., Lieber, M.R., Mizuuchi, K. and Gellert, M. (1989) V(D)J recombination: a functional definition of the joining signals. *Genes Dev* **3**, 1053-61.
- Hiom, K. and Gellert, M. (1997) A stable RAG1-RAG2-DNA complex that is active in V(D)J cleavage. *Cell* **88**, 65-72.
- Hiom, K. and Gellert, M. (1998) Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol Cell* **1**, 1011-9.
- Hiom, K., Melek, M. and Gellert, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463-70.
- Houba-Herlin, N., Becker, D., Post, A., Larondelle, Y. and Starlinger, P. (1990) Excision of a *Ds*-like maize transposable element (*Ac delta*) in a transient assay in *Petunia* is enhanced by a truncated coding region of the transposable element *Ac*. *Mol Gen Genet* **224**, 17-23.
- Hudson, A.D., Carpenter, R. and Coen, E.S. (1990) Phenotypic effects of short-range and aberrant transposition in *Antirrhinum majus*. *Plant Mol Biol* **14**, 835-44.
- Huet, F., Lu, J.T., Myrick, K.V., Baugh, L.R., Crosby, M.A. and Gelbart, W.M. (2002) A deletion-generator compound element allows deletion saturation analysis for genomewide phenotypic annotation. *Proc Natl Acad Sci U S A* **99**, 2.
- Huye, L.E., Purugganan, M.M., Jiang, M.M. and Roth, D.B. (2002) Mutational analysis of all conserved basic amino acids in RAG-1 reveals catalytic, step arrest, and joining-deficient mutants in the V(D)J recombinase. *Mol Cell Biol* **22**, 3460-73.
- Ivanov, V.I. and Minchenkova, L.E. (1994) [The A-form of DNA: in search of the biological role]. *Mol Biol (Mosk)* **28**, 1258-71.
- Ivics, Z., Izsvak, Z., Minter, A. and Hackett, P.B. (1996) Identification of functional domains and evolution of *Tc1*-like transposable elements. *Proc Natl Acad Sci U S A* **93**, 5008-13.
- Izsvak, Z., Ivics, Z. and Plasterk, R.H. (2000) *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol* **302**, 93-102.

- Izsvak, Z., Khare, D., Behlke, J., Heinemann, U., Plasterk, R.H. and Ivics, Z. (2002) Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. *J Biol Chem* **277**, 34581-8.
- Jackson, S.P. (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687-96.
- Jacobson, J.W., Medhora, M.M. and Hartl, D.L. (1986) Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc Natl Acad Sci U S A* **83**, 8684-8.
- Jainchill, J.L., Aaronson, S.A. and Todaro, G.J. (1969) Murine sarcoma and leukemia viruses: assay using clonal lines of contact-inhibited mouse cells. *J Virol* **4**, 549-53.
- Jarvik, T. and Lark, K.G. (1998) Characterization of *Soymar1*, a *mariner* element in soybean. *Genetics* **149**, 1569-74.
- Jarvis, P., Belzile, F., Page, T. and Dean, C. (1997) Increased *Ac* excision (*iae*): *Arabidopsis thaliana* mutations affecting *Ac* transposition. *Plant J* **11**, 907-19.
- Jones, J.D., Carland, F., Lim, E., Ralston, E. and Dooner, H.K. (1990) Preferential transposition of the maize element *Activator* to linked chromosomal locations in tobacco. *Plant Cell* **2**, 701-7.
- Jones, J.M., Welty, D.J. and Nakai, H. (1998) Versatile action of *Escherichia coli* ClpXP as protease or molecular chaperone for bacteriophage Mu transposition. *J Biol Chem* **273**, 459-65.
- Jones, J.M. and Gellert, M. (2001) Intermediates in V(D)J recombination: a stable RAG1/2 complex sequesters cleaved RSS ends. *Proc Natl Acad Sci U S A* **98**, 12926-31.
- Jones, J.M. and Gellert, M. (2002) Ordered assembly of the V(D)J synaptic complex ensures accurate recombination. *Embo J* **21**, 4162-71.
- Joyce, C.M. and Steitz, T.A. (1995) Polymerase structures and function: variations on a theme? *J Bacteriol* **177**, 6321-9.
- Junop, M.S., Modesti, M., Guarne, A., Ghirlando, R., Gellert, M. and Yang, W. (2000) Crystal structure of the XRCC4 DNA repair protein and implications for end joining. *Embo J* **19**, 5962-70.
- Kalderon, D., Richardson, W.D., Markham, A.F. and Smith, A.E. (1984a) Sequence requirements for nuclear location of simian virus 40 large-*T* antigen. *Nature* **311**, 33-8.
- Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984b) A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499-509.
- Kale, S.B., Landree, M.A. and Roth, D.B. (2001) Conditional RAG-1 mutants block the hairpin formation step of V(D)J recombination. *Mol Cell Biol* **21**, 459-66.
- Katinka, M.D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretailade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J. and Vivares, C.P. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450-3.
- Kaufman, P.D., Doll, R.F. and Rio, D.C. (1989) *Drosophila* P element transposase recognizes internal P element DNA sequences. *Cell* **59**, 359-71.
- Kawakami, K., Shima, A. and Kawakami, N. (2000) Identification of a functional transposase of the *Tol2* element, an *Ac*-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc Natl Acad Sci U S A* **97**, 11403-8.
- Kelly, J.J., Munholland, J.M. and Wildeman, A.G. (1989) Comeasurement of simian virus 40 early and late promoter activity in HeLa and 293 cells in the presence of *T* antigen. *J Virol* **63**, 383-91.
- Kempken, F. and Kück, U. (1996) *restless*, an active *Ac*-like transposon from the fungus *Tolypocladium inflatum*: structure, expression, and alternative RNA splicing. *Mol Cell Biol* **16**, 6563-72.

- Kempken, F. and Kück, U. (1998) Evidence for circular transposition derivatives from the fungal *hAT*-transposon *restless*. *Curr Genet* **34**, 200-3.
- Kempken, F. and Windhofer, F. (2001) The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. *Chromosoma* **110**, 1-9.
- Kennedy, A.K., Guhathakurta, A., Kleckner, N. and Haniford, D.B. (1998) *Tn10* transposition via a DNA hairpin intermediate. *Cell* **95**, 125-34.
- Kennedy, A.K. (1999) *Mechanistic aspects of Tn10 transposition*. PhD Thesis, University of Western Ontario, London, Ontario, Canada.
- Kennedy, A.K., Haniford, D.B. and Mizuuchi, K. (2000) Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. *Cell* **101**, 295-305.
- Khan, E., Mack, J.P., Katz, R.A., Kulkosky, J. and Skalka, A.M. (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res* **19**, 851-60.
- Kidwell, M.G. (1993) Lateral transfer in natural populations of eukaryotes. *Annu Rev Genet* **27**, 235-56.
- Kidwell, M.G. and Lisch, D.R. (2002) Transposable elements as sources of genomic variation. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 59-90.
- Kim, D.R., Dai, Y., Mundy, C.L., Yang, W. and Oettinger, M.A. (1999) Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev* **13**, 3070-80.
- Kirch, S.A., Sudarsanam, P. and Oettinger, M.A. (1996) Regions of RAG1 protein critical for V(D)J recombination. *Eur J Immunol* **26**, 886-91.
- Kitamura, K., Hashida, S.N., Mikami, T. and Kishima, Y. (2001) Position effect of the excision frequency of the *Antirrhinum* transposon *Tam3*: implications for the degree of position-dependent methylation in the ends of the element. *Plant Mol Biol* **47**, 475-90.
- Kleckner, N. (1989) Transposon *Tn10*. In Berg, D.E. and Howe, M.M. (eds.), *Mobile DNA*. American Society for Microbiology, Washington, DC.
- Kleckner, N., Bender, J. and Gottesman, S. (1991) Uses of transposons with emphasis on *Tn10*. *Methods Enzymol* **204**, 139-80.
- Kleckner, N., Chalmers, R.M., Kwon, D., Sakai, J. and Bolland, S. (1996) *Tn10* and *IS10* transposition and chromosome rearrangements: mechanism and regulation *in vivo* and *in vitro*. *Curr Top Microbiol Immunol* **204**, 49-82.
- Koga, A., Shimada, A., Shima, A., Sakaizumi, M., Tachida, H. and Hori, H. (2000) Evidence for recent invasion of the medaka fish genome by the *Tol2* transposable element. *Genetics* **155**, 273-81.
- Komori, T., Okada, A., Stewart, V. and Alt, F.W. (1993) Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* **261**, 1171-5.
- Kopka, M.L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985) Binding of an antitumor drug to DNA, Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G. *J Mol Biol* **183**, 553-63.
- Korswagen, H.C., Durbin, R.M., Smits, M.T. and Plasterk, R.H. (1996) Transposon *Tc1*-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. *Proc Natl Acad Sci U S A* **93**, 14680-5.
- Kredich, N.M. (1996) Biosynthesis of cysteine. In Neidhardt, F.C. (ed.), *Escherichia coli and Salmonella: cellular and molecular biology*. ASM Press, Washington, DC, Vol. 1, pp. 514-27.

- Krukltis, R., Welty, D.J. and Nakai, H. (1996) ClpX protein of *Escherichia coli* activates bacteriophage Mu transposase in the strand transfer complex for initiation of Mu DNA synthesis. *Embo J* **15**, 935-44.
- Kuduvalli, P.N., Rao, J.E. and Craig, N.L. (2001) Target DNA structure plays a critical role in Tn7 transposition. *Embo J* **20**, 924-32.
- Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P. and Skalka, A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol Cell Biol* **12**, 2331-8.
- Kumar, C.S., Muthukumaran, G., Frost, L.J., Noe, M., Ahn, Y.H., Mariano, T.M. and Pestka, S. (1989) Molecular characterization of the murine interferon gamma receptor cDNA. *J Biol Chem* **264**, 17939-46.
- Kunze, R. and Starlinger, P. (1989) The putative transposase of transposable element *Ac* from *Zea mays L.* interacts with subterminal sequences of *Ac*. *Embo J* **8**, 3177-85.
- Kunze, R., Behrens, U., Courage-Franzkowiak, U., Feldmar, S., Kuhn, S. and Lutticke, R. (1993) Dominant transposition-deficient mutants of maize *Activator* (*Ac*) transposase. *Proc Natl Acad Sci U S A* **90**, 7094-8.
- Kunze, R. (1996) The maize transposable element activator (*Ac*). *Curr Top Microbiol Immunol* **204**, 161-94.
- Kunze, R. and Weil, C.W. (2002) The *hAT* and CACTA superfamilies of plant transposons. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 565-610.
- Kwon, D., Chalmers, R.M. and Kleckner, N. (1995) Structural domains of *IS10* transposase and reconstitution of transposition activity from proteolytic fragments lacking an interdomain linker. *Proc Natl Acad Sci U S A* **92**, 8234-8.
- Kwon, J., Imbalzano, A.N., Matthews, A. and Oettinger, M.A. (1998) Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. *Mol Cell* **2**, 829-39.
- Ladevèze, V., Galindo, M.I., Pascual, L., Periquet, G. and Lemeunier, F. (1994) Invasion of the *hobo* transposable element studied by *in situ* hybridization on polytene chromosomes of *Drosophila melanogaster*. *Genetica* **93**, 91-100.
- Ladevèze, V., Galindo, I., Chaminade, N., Pascual, L., Periquet, G. and Lemeunier, F. (1998) Transmission pattern of *hobo* transposable element in transgenic lines of *Drosophila melanogaster*. *Genet Res* **71**, 97-107.
- Ladevèze, V., Aulard, S., Chaminade, N., Biemont, C., Periquet, G. and Lemeunier, F. (2001) Dynamics of the *hobo* transposable element in transgenic lines of *Drosophila melanogaster*. *Genet Res* **77**, 135-42.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.
- Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y. and Tonegawa, S. (1989) Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell* **59**, 859-70.
- Lampe, D.J., Churchill, M.E. and Robertson, H.M. (1996) A purified *mariner* transposase is sufficient to mediate transposition *in vitro*. *Embo J* **15**, 5470-9.
- Lampe, D.J., Churchill, M.E. and Robertson, H.M. (1997) A purified *mariner* transposase is sufficient to mediate transposition *in vitro* (erratum). *Embo J* **16**, 4152-3.
- Lampe, D.J., Grant, T.E. and Robertson, H.M. (1998) Factors affecting transposition of the *Himar1* *mariner* transposon *in vitro*. *Genetics* **149**, 179-87.

- Lampe, D.J., Akerley, B.J., Rubin, E.J., Mekalanos, J.J. and Robertson, H.M. (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc Natl Acad Sci U S A* **96**, 11428-33.
- Lampe, D.J., Walden, K.K. and Robertson, H.M. (2001) Loss of transposase-DNA interaction may underlie the divergence of *mariner* family transposable elements and the ability of more than one *mariner* to occupy the same genome. *Mol Biol Evol* **18**, 954-61.
- Lander, E.S., Linton, L.M., Birren, B. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- Landree, M.A., Wibbenmeyer, J.A. and Roth, D.B. (1999) Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev* **13**, 3059-69.
- Landree, M.A., Kale, S.B. and Roth, D.B. (2001) Functional organization of single and paired V(D)J cleavage complexes. *Mol Cell Biol* **21**, 4256-64.
- Langin, T., Capy, P. and Daboussi, M.J. (1995) The transposable element *impala*, a fungal member of the Tc1-mariner superfamily. *Mol Gen Genet* **246**, 19-28.
- Lee, G.S., Neiditch, M.B., Sinden, R.R. and Roth, D.B. (2002) Targeted transposition by the V(D)J recombinase. *Mol Cell Biol* **22**, 2068-77.
- Leroy, H., Castagnone-Sereno, P., Renault, S., Auge-Gouillou, C., Bigot, Y. and Abad, P. (2003) Characterization of *Mcmar1*, a *mariner*-like element with large inverted terminal repeats (ITRs) from the phytoparasitic nematode *Meloidogyne chitwoodi*. *Gene* **304**, 35-41.
- Levchenko, I., Luo, L. and Baker, T.A. (1995) Disassembly of the Mu transposase tetramer by the ClpX chaperone. *Genes Dev* **9**, 2399-408.
- Lewis, S., Gifford, A. and Baltimore, D. (1985) DNA elements are asymmetrically joined during the site-specific recombination of *kappa* immunoglobulin genes. *Science* **228**, 677-85.
- Lewis, S.M., Hesse, J.E., Mizuuchi, K. and Gellert, M. (1988) Novel strand exchanges in V(D)J recombination. *Cell* **55**, 1099-107.
- Lewis, S.M. and Hesse, J.E. (1991) Cutting and closing without recombination in V(D)J joining. *Embo J* **10**, 3631-9.
- Li, M.G. and Starlinger, P. (1990) Mutational analysis of the N terminus of the protein of maize transposable element *Ac*. *Proc Natl Acad Sci U S A* **87**, 6044-8.
- Li, X., Lobo, N., Bauser, C.A. and Fraser, M.J., Jr. (2001) The minimum internal and external sequence requirements for transposition of the eukaryotic transformation vector *piggyBac*. *Mol Genet Genomics* **266**, 190-8.
- Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E. and Alt, F.W. (1995) The *XRCC4* gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83**, 1079-89.
- Liao, L.W., Rosenzweig, B. and Hirsh, D. (1983) Analysis of a transposable element in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* **80**, 3585-9.
- Lidholm, D.A., Gudmundsson, G.H. and Boman, H.G. (1991) A highly repetitive, *mariner*-like element in the genome of *Hyalophora cecropia*. *J Biol Chem* **266**, 11518-21.
- Lieber, A., Kiessling, U. and Strauss, M. (1989) High level gene expression in mammalian cells by a nuclear T7-phase RNA polymerase. *Nucleic Acids Res* **17**, 8485-93.
- Lieber, M.R., Hesse, J.E., Mizuuchi, K. and Gellert, M. (1988) Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. *Proc Natl Acad Sci U S A* **85**, 8588-92.
- Lipkowitz, S., Stern, M.H. and Kirsch, I.R. (1990) Hybrid T cell receptor genes formed by interlocus recombination in normal and *ataxia-telangiectasis* lymphocytes. *J Exp Med* **172**, 409-18.

- Lister, C., Jackson, D. and Martin, C. (1993) Transposon-induced inversion in *Antirrhinum* modifies *nivea* gene expression to give a novel flower color pattern under the control of *cycloidearadialis*. *Plant Cell* **5**, 1541-53.
- Livak, F. and Schatz, D.G. (1996) T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol Cell Biol* **16**, 609-18.
- Lohe, A.R., Lidholm, D.A. and Hartl, D.L. (1995a) Genotypic effects, maternal effects and grand-maternal effects of immobilized derivatives of the transposable element *mariner*. *Genetics* **140**, 183-92.
- Lohe, A.R., Moriyama, E.N., Lidholm, D.A. and Hartl, D.L. (1995b) Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements. *Mol Biol Evol* **12**, 62-72.
- Lohe, A.R. and Hartl, D.L. (1996a) Reduced germline mobility of a *mariner* vector containing exogenous DNA: effect of size or site? *Genetics* **143**, 1299-306.
- Lohe, A.R. and Hartl, D.L. (1996b) Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. *Mol Biol Evol* **13**, 549-55.
- Lohe, A.R., Sullivan, D.T. and Hartl, D.L. (1996) Subunit interactions in the *mariner* transposase. *Genetics* **144**, 1087-95.
- Lohe, A.R., De Aguiar, D. and Hartl, D.L. (1997) Mutations in the *mariner* transposase: the D,D(35)E consensus sequence is nonfunctional. *Proc Natl Acad Sci U S A* **94**, 1293-7.
- Lohe, A.R., Timmons, C., Beerman, I., Lozovskaya, E.R. and Hartl, D.L. (2000) Self-inflicted wounds, template-directed gap repair and a recombination hotspot. Effects of the *mariner* transposase. *Genetics* **154**, 647-56.
- Lohe, A.R. and Hartl, D.L. (2002) Efficient mobilization of *mariner* *in vivo* requires multiple internal sequences. *Genetics* **160**, 519-26.
- Lovell, S., Goryshin, I.Y., Reznikoff, W.R. and Rayment, I. (2002) Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat Struct Biol* **9**, 278-81.
- Lozovskaya, E.R., Nurminsky, D.I., Hartl, D.L. and Sullivan, D.T. (1996) Germline transformation of *Drosophila virilis* mediated by the transposable element *hobo*. *Genetics* **142**, 173-7.
- Lozovsky, E.R., Nurminsky, D., Wimmer, E.A. and Hartl, D.L. (2002) Unexpected stability of *mariner* transgenes in *Drosophila*. *Genetics* **160**, 527-35.
- Lu, F. and Craig, N.L. (2000) Isolation and characterization of Tn7 transposase gain-of-function mutants: a model for transposase activation. *Embo J* **19**, 3446-57.
- Ma, Y., Pannicke, U., Schwarz, K. and Lieber, M.R. (2002) Hairpin opening and overhang processing by an *Artemis*/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781-94.
- Mansilla-Soto, J. and Cortes, P. (2003) VDJ recombination: *Artemis* and its *in vivo* role in hairpin opening. *J Exp Med* **197**, 543-7.
- Maruyama, K. and Hartl, D.L. (1991) Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *J Mol Evol* **33**, 514-24.
- Maurer, P., Rejasse, A., Capy, P., Langin, T. and Riba, G. (1997) Isolation of the transposable element *hupfer* from the entomopathogenic fungus *Beauveria bassiana* by insertion mutagenesis of the nitrate reductase structural gene. *Mol Gen Genet* **256**, 195-202.
- McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M. and Oettinger, M.A. (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83**, 387-95.
- McClintock, B. (1948) Mutable loci in maize. *Carnegie Inst. Wash. Year Book* **47**, 155-69.
- McClintock, B. (1949) Mutable loci in maize. *Carnegie Inst. Wash. Year Book* **48**, 142-54.

- McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **36**, 344-9.
- McCormack, W.T., Tjoelker, L.W., Carlson, L.M., Petryniak, B., Barth, C.F., Humphries, E.H. and Thompson, C.B. (1989) Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single nonrandom nucleotides to both coding segments. *Cell* **56**, 785-91.
- McGinnis, W., Shermoen, A.W. and Beckendorf, S.K. (1983) A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* **34**, 75-84.
- Melek, M., Gellert, M. and van Gent, D.C. (1998) Rejoining of DNA by the RAG1 and RAG2 proteins. *Science* **280**, 301-3.
- Melek, M. and Gellert, M. (2000) RAG1/2-mediated resolution of transposition intermediates: two pathways and possible consequences. *Cell* **101**, 625-33.
- Melek, M., Jones, J.M., O'Dea, M.H., Pais, G., Burke, T.R., Jr., Pommier, Y., Neamati, N. and Gellert, M. (2002) Effect of HIV integrase inhibitors on the RAG1/2 recombinase. *Proc Natl Acad Sci U S A* **99**, 134-7.
- Meselson, M. and Yuan, R. (1968) DNA restriction enzyme from *E. coli*. *Nature* **217**, 1110-4.
- Messier, T.L., O'Neill, J.P., Hou, S.M., Nicklas, J.A. and Finette, B.A. (2003) *In vivo* transposition mediated by V(D)J recombinase in human T lymphocytes. *Embo J* **22**, 1381-8.
- Michel, K. and Atkinson, P.W. (2003) Nuclear localization of the *Hermes* transposase depends on basic amino acid residues at the N-terminus of the protein. *J Cell Biochem* **89**, 778-90.
- Miller, A.D. and Rosman, G.J. (1989) Improved retroviral vectors for gene transfer and expression. *Biotechniques* **7**, 980-2, 4-6, 9-90.
- Minsky, A., Ghirlando, R. and Reich, Z. (1997) Nucleosomes: a solution to a crowded intracellular environment? *J Theor Biol* **188**, 379-85.
- Minton, A.P. (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem* **276**, 10577-80.
- Mizushima, S. and Nagata, S. (1990) pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res* **18**, 5322.
- Mizuuchi, K. (1992a) Polynucleotidyl transfer reactions in transpositional DNA recombination. *J Biol Chem* **267**, 21273-6.
- Mizuuchi, K. (1992b) Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu Rev Biochem* **61**, 1011-51.
- Mizuuchi, K. (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells* **2**, 1-12.
- Mizuuchi, K. and Baker, T.A. (2002) Chemical mechanisms for mobilizing DNA. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 12-23.
- Mizuuchi, M. and Mizuuchi, K. (1989) Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. *Cell* **58**, 399-408.
- Mizuuchi, M., Baker, T.A. and Mizuuchi, K. (1992) Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. *Cell* **70**, 303-11.
- Mo, X., Bailin, T. and Sadofsky, M.J. (2001) A C-terminal region of RAG1 contacts the coding DNA during V(D)J recombination. *Mol Cell Biol* **21**, 2038-47.

- Modesti, M., Hesse, J.E. and Gellert, M. (1999) DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *Embo J* **18**, 2008-18.
- Montgomery, D.C. (1997) *Design and analysis of experiments*. John Wiley & Sons, Inc., New York, NY.
- Moran, J.V. and Gilbert, N. (2002) Mammalian LINE-1 retrotransposons and related elements. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 836-69.
- Morgan, G.T. (1995) Identification in the human genome of mobile elements spread by DNA-mediated transposition. *J Mol Biol* **254**, 1-5.
- Morisato, D., Way, J.C., Kim, H.J. and Kleckner, N. (1983) Tn10 transposase acts preferentially on nearby transposon ends *in vivo*. *Cell* **32**, 799-807.
- Morisato, D. and Kleckner, N. (1987) Tn10 transposition and circle formation *in vitro*. *Cell* **51**, 101-11.
- Moshous, D., Li, L., Chasseval, R., Philippe, N., Jabado, N., Cowan, M.J., Fischer, A. and de Villartay, J.P. (2000) A new gene involved in DNA double-strand break repair and V(D)J recombination is located on human chromosome 10p. *Hum Mol Genet* **9**, 583-8.
- Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A. and de Villartay, J.P. (2001) *Artemis*, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**, 177-86.
- Moshous, D., Pannetier, C., Chasseval Rd, R., Deist Fl, F., Cavazzana-Calvo, M., Romana, S., Macintyre, E., Canioni, D., Brousse, N., Fischer, A., Casanova, J.L. and Villartay, J.P. (2003) Partial T and B lymphocyte immunodeficiency and predisposition to lymphoma in patients with hypomorphic mutations in *Artemis*. *J Clin Invest* **111**, 381-7.
- Moss, B., Elroy-Stein, O., Mizukami, W.A., Alexander, W.A. and Fuerst, T.R. (1990) New mammalian expression vectors. *Nature* **348**, 91-2.
- Muller, H.P. and Varmus, H.E. (1994) DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *Embo J* **13**, 4704-14.
- Mundy, C.L., Patenge, N., Matthews, A.G. and Oettinger, M.A. (2002) Assembly of the RAG1/RAG2 synaptic complex. *Mol Cell Biol* **22**, 69-77.
- Murphy, L.D. and Zimmerman, S.B. (1994) Macromolecular crowding effects on the interaction of DNA with *Escherichia coli* DNA-binding proteins: a model for bacterial nucleoid stabilization. *Biochim Biophys Acta* **1219**, 277-84.
- Naigamwalla, D.Z. and Chaconas, G. (1997) A new set of Mu DNA transposition intermediates: alternate pathways of target capture preceding strand transfer. *Embo J* **16**, 5227-34.
- Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28**, 292.
- NCBI. (2002) Mouse Genome Sequencing.
- Neiditch, M.B., Lee, G.S., Landree, M.A. and Roth, D.B. (2001) RAG transposase can capture and commit to target DNA before or after donor cleavage. *Mol Cell Biol* **21**, 4302-10.
- Neiditch, M.B., Lee, G.S., Huye, L.E., Brandt, V.L. and Roth, D.B. (2002) The V(D)J recombinase efficiently cleaves and transposes signal joints. *Mol Cell* **9**, 871-8.
- Newfeld, S.J. and Takaesu, N.T. (1999) Local transposition of a *hobo* element within the *decapentaplegic* locus of *Drosophila*. *Genetics* **151**, 177-87.
- Newfeld, S.J. and Takaesu, N.T. (2002) An analysis using the *hobo* genetic system reveals that combinatorial signaling by the Dpp and Wg pathways regulates *dpp* expression in leading edge cells of the dorsal ectoderm in *Drosophila melanogaster*. *Genetics* **161**, 685-92.

- Nick McElhinny, S.A., Snowden, C.M., McCarville, J. and Ramsden, D.A. (2000) Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* **20**, 2996-3003.
- O'Brochta, D.A., Warren, W.D., Saville, K.J. and Atkinson, P.W. (1994) Interplasmid transposition of *Drosophila hobo* elements in non-drosophilid insects. *Mol Gen Genet* **244**, 9-14.
- O'Brochta, D.A., Warren, W.D., Saville, K.J. and Atkinson, P.W. (1996) *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica*. *Genetics* **142**, 907-14.
- O'Hare, K. and Rubin, G.M. (1983) Structures of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* **34**, 25-35.
- Oettinger, M.A., Schatz, D.G., Gorka, C. and Baltimore, D. (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517-23.
- Okazaki, K., Davis, D.D. and Sakano, H. (1987) T cell receptor beta gene sequences in the circular DNA of thymocyte nuclei: direct evidence for intramolecular DNA deletion in V-D-J joining. *Cell* **49**, 477-85.
- Oosumi, T., Belknap, W.R. and Garlick, B. (1995) *mariner* transposons in humans. *Nature* **378**, 672.
- Orchard, K. and May, G.E. (1993) An EMSA-based method for determining the molecular weight of a protein-DNA complex. *Nucleic Acids Res* **21**, 3335-6.
- Patterson, J.T. (1943) Studies in the genetics of *Drosophila* III. The *Drosophilidae* of the Southwest. *Univ. Texas Publs* **4313**, 7-216.
- Paull, T.T. and Gellert, M. (2000) A mechanistic basis for Mre11-directed DNA joining at microhomologies. *Proc Natl Acad Sci U S A* **97**, 6409-14.
- Peacock, W.J., Dennis, E.S., Gerlach, W.L., Sachs, M.M. and Schwartz, D. (1984) Insertion and excision of *Ds* controlling elements in maize. *Cold Spring Harb Symp Quant Biol* **49**, 347-54.
- Pelacic, V., Morelle, S., Lampe, D. and Nassif, X. (2000) Mutagenesis of *Neisseria meningitidis* by *in vitro* transposition of *Himar1 mariner*. *J Bacteriol* **182**, 5391-8.
- Pelta, J., Livolant, F. and Sikorav, J.L. (1996) DNA aggregation induced by polyamines and cobalthexamine. *J Biol Chem* **271**, 5656-62.
- Peters, J.E. and Craig, N.L. (2001) Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev* **15**, 737-47.
- Pinkerton, A.C., O'Brochta, D.A. and Atkinson, P.W. (1996) Mobility of *hAT* transposable elements in the Old World bollworm, *Helicoverpa armigera*. *Insect Mol Biol* **5**, 223-7.
- Plasterk, R.H. (1991) The origin of footprints of the *Tc1* transposon of *Caenorhabditis elegans*. *Embo J* **10**, 1919-25.
- Plasterk, R.H. and Groenen, J.T. (1992) Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following *Tc1* excision. *Embo J* **11**, 287-90.
- Plasterk, R.H. (1996) The *Tc1/mariner* transposon family. *Curr Top Microbiol Immunol* **204**, 125-43.
- Plasterk, R.H., Izsvak, Z. and Ivics, Z. (1999) Resident aliens: the *Tc1/mariner* superfamily of transposable elements. *Trends Genet* **15**, 326-32.
- Plasterk, R.H.A. and van Luenen, H.G.A.M. (2002) The *Tc1/mariner* family of transposable elements. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 519-32.
- Pohlman, R.F., Fedoroff, N.V. and Messing, J. (1984a) The nucleotide sequence of the maize controlling element *Activator*. *Cell* **37**, 635-43.

- Pohlman, R.F., Fedoroff, N.V. and Messing, J. (1984b) Correction: nucleotide sequence of *Ac*. *Cell* **39**, 417.
- Pribil, P.A. and Haniford, D.B. (2000) Substrate recognition and induced DNA deformation by transposase at the target-capture stage of *Tn10* transposition. *J Mol Biol* **303**, 145-59.
- Pribil, P.A. and Haniford, D.B. (2003) Target DNA bending is an important specificity determinant in target site selection in *Tn10* transposition. *J Mol Biol* **330**, 247-59.
- Purugganan, M.M., Shah, S., Kearney, J.F. and Roth, D.B. (2001) Ku80 is required for addition of N nucleotides to V(D)J recombination junctions by terminal deoxynucleotidyl transferase. *Nucleic Acids Res* **29**, 1638-46.
- Qiu, J.X., Kale, S.B., Yarnell Schultz, H. and Roth, D.B. (2001) Separation-of-function mutants reveal critical roles for RAG2 in both the cleavage and joining steps of V(D)J recombination. *Mol Cell* **7**, 77-87.
- Radice, A.D. and Emmons, S.W. (1993) Extrachromosomal circular copies of the transposon *Tc1*. *Nucleic Acids Res* **21**, 2663-7.
- Ramsden, D.A. and Gellert, M. (1995) Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes Dev* **9**, 2409-20.
- Ramsden, D.A., McBlane, J.F., van Gent, D.C. and Gellert, M. (1996) Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *Embo J* **15**, 3197-206.
- Ramsden, D.A. and Gellert, M. (1998) Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *Embo J* **17**, 609-14.
- Ran Kim, Y. and Haeng Rhee, J. (2003) Flagellar basal body *flg* operon as a virulence determinant of *Vibrio vulnificus*. *Biochem Biophys Res Commun* **304**, 405-10.
- Rao, J.E., Miller, P.S. and Craig, N.L. (2000) Recognition of triple-helical DNA structures by transposon *Tn7*. *Proc Natl Acad Sci U S A* **97**, 3936-41.
- Rao, J.E. and Craig, N.L. (2001) Selective recognition of pyrimidine motif triplexes by a protein encoded by the bacterial transposon *Tn7*. *J Mol Biol* **307**, 1161-70.
- Revzin, A. (1989) Gel electrophoresis assays for DNA-protein interactions. *Biotechniques* **7**, 346-55.
- Reznikoff, W.S. (2002) *Tn5* transposition. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 403-22.
- Rice, P., Craigie, R. and Davies, D.R. (1996) Retroviral integrases and their cousins. *Curr Opin Struct Biol* **6**, 76-83.
- Robbins, P.R., Carpenter, R. and Coen, E.S. (1989) A chromosome rearrangement suggests that donor and recipient sites are associated during *Tam3* transposition in *Antirrhinum majus*. *Embo J* **8**, 5-13.
- Roberts, D., Hoopes, B.C., McClure, W.R. and Kleckner, N. (1985) *IS10* transposition is regulated by DNA adenine methylation. *Cell* **43**, 117-30.
- Robertson, H.M. (1993) The *mariner* transposable element is widespread in insects. *Nature* **362**, 241-5.
- Robertson, H.M. and MacLeod, E.G. (1993) Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol Biol* **2**, 125-39.
- Robertson, H.M. (1995) The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiol.* **41**, 99-105.
- Robertson, H.M. and Lampe, D.J. (1995a) Recent horizontal transfer of a *mariner* transposable element among and between *Diptera* and *Neuroptera*. *Mol Biol Evol* **12**, 850-62.

- Robertson, H.M. and Lampe, D.J. (1995b) Distribution of transposable elements in arthropods. *Annu Rev Entomol* **40**, 333-57.
- Robertson, H.M. and Asplund, M.L. (1996) *Bmmar1*: a basal lineage of the *mariner* family of transposable elements in the silkworm moth, *Bombyx mori*. *Insect Biochem Mol Biol* **26**, 945-54.
- Robertson, H.M., Zumpano, K.L., Lohe, A.R. and Hartl, D.L. (1996) Reconstructing the ancient *mariners* of humans. *Nat Genet* **12**, 360-1.
- Robertson, H.M. (1997) Multiple *mariner* transposons in flatworms and hydras are related to those of insects. *J Hered* **88**, 195-201.
- Robertson, H.M. (2002) Evolution of DNA transposons in eukaryotes. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 1093-110.
- Rodgers, K.K., Bu, Z., Fleming, K.G., Schatz, D.G., Engelman, D.M. and Coleman, J.E. (1996) A zinc-binding domain involved in the dimerization of RAG1. *J Mol Biol* **260**, 70-84.
- Rooney, S., Sekiguchi, J., Zhu, C., Cheng, H.L., Manis, J., Whitlow, S., DeVido, J., Foy, D., Chaudhuri, J., Lombard, D. and Alt, F.W. (2002) Leaky *Scid* phenotype associated with defective V(D)J coding end processing in *Artemis*-deficient mice. *Mol Cell* **10**, 1379-90.
- Rooney, S., Alt, F.W., Lombard, D., Whitlow, S., Eckersdorff, M., Fleming, J., Fugmann, S., Ferguson, D.O., Schatz, D.G. and Sekiguchi, J. (2003) Defective DNA repair and increased genomic instability in *Artemis*-deficient murine cells. *J Exp Med* **197**, 553-65.
- Ros, F. and Kunze, R. (2001) Regulation of *Activator/Dissociation* transposition by replication and DNA methylation. *Genetics* **157**, 1723-33.
- Rosenzweig, B., Liao, L.W. and Hirsh, D. (1983) Sequence of the *C. elegans* transposable element *Tc1*. *Nucleic Acids Res* **11**, 4201-9.
- Roth, D.B., Menetski, J.P., Nakajima, P.B., Bosma, M.J. and Gellert, M. (1992a) V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in *scid* mouse thymocytes. *Cell* **70**, 983-91.
- Roth, D.B., Nakajima, P.B., Menetski, J.P., Bosma, M.J. and Gellert, M. (1992b) V(D)J recombination in mouse thymocytes: double-strand breaks near T cell receptor delta rearrangement signals. *Cell* **69**, 41-53.
- Roth, D.B., Zhu, C. and Gellert, M. (1993) Characterization of broken DNA molecules associated with V(D)J recombination. *Proc Natl Acad Sci U S A* **90**, 10788-92.
- Roth, D.B. and Craig, N.L. (1998) VDJ recombination: a transposase goes to work. *Cell* **94**, 411-4.
- Rousseau, P., Normand, C., Loot, C., Turlan, C., Alazard, R., Duval-Valentin, G. and Chandler, M. (2002) Transposition of *IS911*. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 367-83.
- Rubin, E., Lithwick, G. and Levy, A.A. (2001) Structure and evolution of the *hAT* transposon superfamily. *Genetics* **158**, 949-57.
- Rubin, E.J., Akerley, B.J., Novik, V.N., Lampe, D.J., Husson, R.N. and Mekalanos, J.J. (1999) *In vivo* transposition of *mariner*-based elements in enteric bacteria and mycobacteria. *Proc Natl Acad Sci U S A* **96**, 1645-50.
- Rubin, G.M., Kidwell, M.G. and Bingham, P.M. (1982) The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* **29**, 987-94.
- Rudd, K.E. (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res* **28**, 60-4.
- Sadofsky, M.J., Hesse, J.E., McBlane, J.F. and Gellert, M. (1993) Expression and V(D)J recombination activity of mutated RAG-1 proteins [published erratum appears in *Nucleic Acids Res* 1994 Feb 11;22(3):550]. *Nucleic Acids Res* **21**, 5644-50.

- Sadofsky, M.J., Hesse, J.E. and Gellert, M. (1994) Definition of a core region of RAG-2 that is functional in V(D)J recombination. *Nucleic Acids Res* **22**, 1805-9.
- Saedler, H. and Nevers, P. (1985) Transposition in plants: a molecular model. *Embo J* **4**, 585-90.
- Sakai, J., Chalmers, R.M. and Kleckner, N. (1995) Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn10 transposition. *Embo J* **14**, 4374-83.
- Sakai, J. and Kleckner, N. (1997) The Tn10 synaptic complex can capture a target DNA only after transposon excision. *Cell* **89**, 205-14.
- Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S. (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280**, 288-94.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- Samuels, M.L. (1991) *Statistics for the life sciences*. Dellen Publishing Company, San Francisco.
- Santagata, S., Besmer, E., Villa, A., Bozzi, F., Allingham, J.S., Sobacchi, C., Haniford, D.B., Vezzoni, P., Nussenzweig, M.C., Pan, Z.Q. and Cortes, P. (1999) The RAG1/RAG2 complex constitutes a 3' flap endonuclease: implications for junctional diversity in V(D)J and transpositional recombination. *Mol Cell* **4**, 935-47.
- Sarnovsky, R.J., May, E.W. and Craig, N.L. (1996) The Tn7 transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *Embo J* **15**, 6348-61.
- Savilahti, H., Rice, P.A. and Mizuuchi, K. (1995) The phage Mu transpososome core: DNA requirements for assembly and function. *Embo J* **14**, 4893-903.
- Schatz, D.G. (1997) V(D)J recombination moves *in vitro*. *Semin Immunol* **9**, 149-59.
- Schickor, P. and Heumann, H. (1994) Hydroxyl radical footprinting. In Kneale, G.G. (ed.), *DNA-Protein Interactions: Principles and Protocols*. Humana Press, Totowa, NJ, Vol. 30, pp. 21-32.
- Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M. and Peng, A. (1993) Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. *Genes Dev* **7**, 2520-32.
- Schlissel, M.S. (1998) Structure of nonhairpin coding-end DNA breaks in cells undergoing V(D)J recombination. *Mol Cell Biol* **18**, 2029-37.
- Schlissel, M.S. (2002) Does *Artemis* end the hunt for the hairpin-opening activity in V(D)J recombination? *Cell* **109**, 1-4.
- Schultz, H.Y., Landree, M.A., Qiu, J.X., Kale, S.B. and Roth, D.B. (2001) Joining-deficient RAG1 mutants block V(D)J recombination *in vivo* and hairpin opening *in vitro*. *Mol Cell* **7**, 65-75.
- Scofield, S.R., English, J.J. and Jones, J.D. (1993) High level expression of the Activator transposase gene inhibits the excision of Dissociation in tobacco cotyledons. *Cell* **75**, 507-17.
- Scott, L., LaFoe, D. and Weil, C.F. (1996) Adjacent sequences influence DNA repair accompanying transposon excision in maize. *Genetics* **142**, 237-46.
- Sewitz, S., Crellin, P. and Chalmers, R. (2003) The positive and negative regulation of Tn10 transposition is mediated by structurally asymmetric transposon arms. *Nucleic Acids Res* **31**, (in press).
- Shao, H. and Tu, Z. (2001) Expanding the diversity of the IS630-Tc1-mariner superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics* **159**, 1103-15.

- Shoemaker, C.B. and Mitsock, L.D. (1986) Murine erythropoietin gene: cloning, expression, and human gene homology. *Mol Cell Biol* **6**, 849-58.
- Sikorav, J.L., Pelta, J. and Livolant, F. (1994) A liquid crystalline phase in spermidine-condensed DNA. *Biophys J* **67**, 1387-92.
- Silhavy, T., Berman, M.L. and Enquist, L.W. (1984) *Experiments with gene fusions*. Cold Spring Harbor Laboratory, New York.
- Silver, D.P., Spanopoulou, E., Mulligan, R.C. and Baltimore, D. (1993) Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. *Proc Natl Acad Sci U S A* **90**, 6100-4.
- Simmons, G.M. (1992) Horizontal transfer of *hobo* transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. *Mol Biol Evol* **9**, 1050-60.
- Simmons, M.J., Haley, K.J., Grimes, C.D., Raymond, J.D. and Fong, J.C. (2002a) Regulation of *P* element transposase activity in *Drosophila melanogaster* by *hobo* transgenes that contain *KP* elements. *Genetics* **161**, 205-15.
- Simmons, M.J., Haley, K.J., Grimes, C.D., Raymond, J.D. and Niemi, J.B. (2002b) A *hobo* transgene that encodes the *P* element transposase in *Drosophila melanogaster*. Autoregulation and cytotype control of transposase activity. *Genetics* **161**, 195-204.
- Sinden, R.R. (1994) *DNA structure and function*. Academic Press, San Diego.
- Skarstad, K., Steen, H.B. and Boye, E. (1983) Cell cycle parameters of slowly growing *Escherichia coli* B/r studied by flow cytometry. *J Bacteriol* **154**, 656-62.
- Skarstad, K., Steen, H.B. and Boye, E. (1985) *Escherichia coli* DNA distributions measured by flow cytometry and compared with theoretical computer simulations. *J Bacteriol* **163**, 661-8.
- Smit, A.F. and Riggs, A.D. (1996) *Tiggers* and DNA transposon fossils in the human genome. *Proc Natl Acad Sci U S A* **93**, 1443-8.
- Smith, C.J. and Parker, A.C. (1993) Identification of a circular intermediate in the transfer and transposition of Tn4555, a mobilizable transposon from *Bacteroides* spp. *J Bacteriol* **175**, 2682-91.
- Smith, D., Wohlgemuth, J., Calvi, B.R., Franklin, I. and Gelbart, W.M. (1993) *hobo* enhancer trapping mutagenesis in *Drosophila* reveals an insertion specificity different from *P* elements. *Genetics* **135**, 1063-76.
- Soneoka, Y., Cannon, P.M., Ramsdale, E.E., Griffiths, J.C., Romano, G., Kingsman, S.M. and Kingsman, A.J. (1995) A transient three-plasmid expression system for the production of high titer retroviral vectors. *Nucleic Acids Res* **23**, 628-33.
- Spanopoulou, E., Zaitseva, F., Wang, F.H., Santagata, S., Baltimore, D. and Panayotou, G. (1996) The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* **87**, 263-76.
- Steen, S.B., Han, J.O., Mundy, C., Oettinger, M.A. and Roth, D.B. (1999) Roles of the "dispensable" portions of RAG-1 and RAG-2 in V(D)J recombination. *Mol Cell Biol* **19**, 3010-7.
- Störl, K., Burckhardt, G., Lown, J.W. and Zimmer, C. (1993) Studies on the ability of minor groove binders to induce supercoiling in DNA. *FEBS Lett* **334**, 49-54.
- Streck, R.D., MacGaffey, J.E. and Beckendorf, S.K. (1986) The structure of *hobo* transposable elements and their insertion sites. *Embo J* **5**, 3615-23.
- Surette, M.G., Buch, S.J. and Chaconas, G. (1987) Transpososomes: stable protein-DNA complexes involved in the *in vitro* transposition of bacteriophage Mu DNA. *Cell* **49**, 253-62.

- Swanson, P.C. and Desiderio, S. (1998) V(D)J recombination signal recognition: distinct, overlapping DNA- protein contacts in complexes containing RAG1 with and without RAG2. *Immunity* **9**, 115-25.
- Swanson, P.C. and Desiderio, S. (1999) RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of a V(D)J initiation complex. *Mol Cell Biol* **19**, 3674-83.
- Swanson, P.C. (2001) The DDE motif in RAG-1 is contributed *in trans* to a single active site that catalyzes the nicking and transesterification steps of V(D)J recombination. *Mol Cell Biol* **21**, 449-58.
- Swanson, P.C. (2002) Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. *Mol Cell Biol* **22**, 1340-51.
- Tarchini, R., Biddle, P., Wineland, R., Tingey, S. and Rafalski, A. (2000) The complete sequence of 340 kb of DNA around the rice *Adh1-adh2* region reveals interrupted colinearity with maize chromosome 4. *Plant Cell* **12**, 381-91.
- Tavakoli, N.P., DeVost, J. and Derbyshire, K.M. (1997) Defining functional regions of the IS903 transposase. *J Mol Biol* **274**, 491-504.
- Teo, S.H. and Jackson, S.P. (2000) Lif1p targets the DNA ligase Lig4p to sites of DNA double-strand breaks. *Curr Biol* **10**, 165-8.
- Tevelev, A. and Schatz, D.G. (2000) Intermolecular V(D)J recombination. *J Biol Chem* **275**, 8341-8.
- Thai, T.H., Purugganan, M.M., Roth, D.B. and Kearney, J.F. (2002) Distinct and opposite diversifying activities of terminal transferase splice variants. *Nat Immunol* **3**, 457-62.
- Thompson, C.B. (1995) New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3**, 531-9.
- Thomson, A.M., Roberts, N.A., Sloane Stanley, J.A., Souyri, M. and Wood, W.G. (1999) Haemopoietic progenitor cell lines generated by the myeloproliferative leukaemia virus: a model system to analyse murine and human lineage- affiliated genes. *Br J Haematol* **107**, 33-48.
- Timsit, Y., Vilbois, E. and Moras, D. (1991) Base-pairing shift in the major groove of (CA)_n tracts by B-DNA crystal structures. *Nature* **354**, 167-70.
- Toman, Z., Dambly-Chaudière, C., Tenenbaum, L. and Radman, M. (1985) A system for detection of genetic and epigenetic alterations in *Escherichia coli* induced by DNA-damaging agents. *J Mol Biol* **186**, 97-105.
- Ton-Hoang, B., Polard, P. and Chandler, M. (1998) Efficient transposition of IS911 circles *in vitro*. *Embo J* **17**, 1169-81.
- Ton-Hoang, B., Polard, P., Haren, L., Turlan, C. and Chandler, M. (1999) IS911 transposon circles give rise to linear forms that can undergo integration *in vitro*. *Mol Microbiol* **32**, 617-27.
- Tonegawa, S. (1983) Somatic generation of antibody diversity. *Nature* **302**, 575-81.
- Tosi, L.R. and Beverley, S.M. (2000) *cis* and *trans* factors affecting *Mos1 mariner* evolution and transposition *in vitro*, and its potential for functional genomics. *Nucleic Acids Res* **28**, 784-90.
- Tsai, C.L. and Schatz, D.G. (2003) Regulation of RAG1/RAG2-mediated transposition by GTP and the C-terminal region of RAG2. *Embo J* **22**, 1922-30.
- Turcotte, K. and Bureau, T. (2002) Phylogenetic analysis reveals *Stowaway*-like elements may represent a fourth family of the IS630-Tc1-mariner superfamily. *Genome* **45**, 82-90.
- van Gent, D.C., Mizuuchi, K. and Gellert, M. (1996a) Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271**, 1592-4.

- van Gent, D.C., Ramsden, D.A. and Gellert, M. (1996b) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* **85**, 107-13.
- van Luenen, H.G., Colloms, S.D. and Plasterk, R.H. (1994) The mechanism of transposition of Tc3 in *C. elegans*. *Cell* **79**, 293-301.
- van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H. and Sixma, T.K. (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. *Embo J* **16**, 6044-54.
- Venter, J.C., Adams, M.D., Myers, E.W. *et al.* (2001) The sequence of the human genome. *Science* **291**, 1304-51.
- Vieira, C., Lepetit, D., Dumont, S. and Biemont, C. (1999) Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol* **16**, 1251-5.
- Vigdal, T.J., Kaufman, C.D., Izsvak, Z., Voytas, D.F. and Ivics, Z. (2002) Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/mariner transposable elements. *J Mol Biol* **323**, 441-52.
- Vink, C., Yeheskiely, E., van der Marel, G.A., van Boom, J.H. and Plasterk, R.H. (1991) Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. *Nucleic Acids Res* **19**, 6691-8.
- Vos, J.C. and Plasterk, R.H. (1994) Tc1 transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain. *Embo J* **13**, 6125-32.
- Vos, J.C., De Baere, I. and Plasterk, R.H. (1996) Transposase is the only nematode protein required for *in vitro* transposition of Tc1. *Genes Dev* **10**, 755-61.
- Voytas, D.F. and Boeke, J.D. (2002) Ty1 and Ty5 of *Saccharomyces cerevisiae*. In Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds.), *Mobile DNA II*. ASM Press, Washington, DC, pp. 631-62.
- Walker, J.R., Corpina, R.A. and Goldberg, J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**, 607-14.
- Wang, L. and Kunze, R. (1998) Transposase binding site methylation in the epigenetically inactivated Ac derivative *Ds-cy*. *Plant J* **13**, 577-82.
- Warren, W.D., Atkinson, P.W. and O'Brochta, D.A. (1994) The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family. *Genet Res* **64**, 87-97.
- Watson, M.A. and Chaconas, G. (1996) Three-site synapsis during Mu DNA transposition: a critical intermediate preceding engagement of the active site. *Cell* **85**, 435-45.
- Way, J.C., Davis, M.A., Morisato, D., Roberts, D.E. and Kleckner, N. (1984) New Tn10 derivatives for transposon mutagenesis and for construction of *lacZ* operon fusions by transposition. *Gene* **32**, 369-79.
- Weil, C.F. and Kunze, R. (2000) Transposition of maize *Ac/Ds* transposable elements in the yeast *Saccharomyces cerevisiae*. *Nat Genet* **26**, 187-90.
- Weinreich, M.D., Gasch, A. and Reznikoff, W.S. (1994a) Evidence that the *cis* preference of the Tn5 transposase is caused by nonproductive multimerization. *Genes Dev* **8**, 2363-74.
- Weinreich, M.D., Yigit, H. and Reznikoff, W.S. (1994b) Overexpression of the Tn5 transposase in *Escherichia coli* results in filamentation, aberrant nucleoid segregation, and cell death: analysis of *E. coli* and transposase suppressor mutations. *J Bacteriol* **176**, 5494-504.
- Welch, D.M. and Meselson, M. (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* **288**, 1211-5.
- Wessler, S.R., Bureau, T.E. and White, S.E. (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr Opin Genet Dev* **5**, 814-21.

- West, S.C. (2003) Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol* **4**, 435-45.
- Wiegand, T.W. and Reznikoff, W.S. (1992) Characterization of two hypertransposing Tn5 mutants. *J Bacteriol* **174**, 1229-39.
- Williams, G.S., Martinez, A., Montalbano, A., Tang, A., Mauhar, A., Ogwaro, K.M., Merz, D., Chevillard, C., Riblet, R. and Feeney, A.J. (2001) Unequal VH gene rearrangement frequency within the large VH7183 gene family is not due to recombination signal sequence variation, and mapping of the genes shows a bias of rearrangement based on chromosomal location. *J Immunol* **167**, 257-63.
- Wilson, R., Orsetti, J., Klocko, A.D., Aluvihare, C., Peckham, E., Atkinson, P.W., Lehane, M.J. and O'Brochta, D.A. (2003) Post-integration behavior of a *Mos1 mariner* gene vector in *Aedes aegypti*. *Insect Biochem Mol Biol* **33**, 853-63.
- Yamashita, S., Takano-Shimizu, T., Kitamura, K., Mikami, T. and Kishima, Y. (1999) Resistance to gap repair of the transposon *Tam3* in *Antirrhinum majus*: a role of the end regions. *Genetics* **153**, 1899-908.
- Yannopoulos, G., Stamatis, N., Monastirioti, M., Hatzopoulos, P. and Louis, C. (1987) *hobo* is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF. *Cell* **49**, 487-95.
- Yannopoulos, G., Zabalou, S. and Alahiotis, S.N. (1994a) Distribution of *P* and *hobo* mobile elements in environmentally manipulated long-term *Drosophila melanogaster* cage populations. *Hereditas* **121**, 87-102.
- Yannopoulos, G., Zabalou, S., Stamatis, N. and Tsamathis, G. (1994b) Differential regulation of *P* and *hobo* mobile elements by two laboratory strains of *Drosophila melanogaster*. *Genet Res* **63**, 129-37.
- Yin, J.C. and Reznikoff, W.S. (1987) *dnaA*, an essential host gene, and Tn5 transposition. *J Bacteriol* **169**, 4637-45.
- Zhang, L., Sankar, U., Lampe, D.J., Robertson, H.M. and Graham, F.L. (1998) The *Himar1 mariner* transposase cloned in a recombinant adenovirus vector is functional in mammalian cells. *Nucleic Acids Res* **26**, 3687-93.
- Zhang, L., Dawson, A. and Finnegan, D.J. (2001) DNA-binding activity and subunit interaction of the *mariner* transposase. *Nucleic Acids Res* **29**, 3566-75.
- Zhu, C. and Roth, D.B. (1995) Characterization of coding ends in thymocytes of *scid* mice: implications for the mechanism of V(D)J recombination. *Immunity* **2**, 101-12.
- Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* **222**, 599-620.
- Zou, A.H., Leung, P.C. and Harshey, R.M. (1991) Transposase contacts with Mu DNA ends. *J Biol Chem* **266**, 20476-82.