Development of methods for combinatorial approaches to *cis*-regulatory module interactions

By

Maxim B. Joseph

Molecular organisation and assembly in cells

University of Warwick

A thesis submitted to the University of Warwick

for the degree of Doctor of Philosophy

March 2012

This book is dedicated to my brilliant and beautiful wife without whom I would be nothing. She always comforts and consoles, never complains or interferes, asks nothing and endures all. She also writes my dedications

Albert Malvino

Table of Contents

List of figures

List of tables

Acknowledgements

I would like to foremost acknowledge both my supervisors, Prof. James A. Covington and Prof. Georgy Koentges, for their tireless and diligent effort to support, encourage and inspire me during this project. Without their mentorship I would not have been able to successfully complete this project.

My advisory committee, consisting of Dr. Til Bretschneider and Dr. Michael Chappell, is thanked for their time and constructive feedback that has materially contributed to this project.

I would like to thank the many people from the different labs that I have worked with during this project for their help and assistance. From the biology lab: Dr. Kate Jordan, Dr. Danuta Jeziorska, Polly Downton and Xintao Zhang. From the engineering lab: Dr. Simon Leigh and Chris Purssell. Each of these people provided experience and expertise where I lacked and their support (and, at times, medical assistance) has been invaluable during this project. Without the assistance of each of these people, this project would have failed to get off the ground and crashed several times along the way. In addition to these people I would like to thank all the members of the MOAC and Systems Biology students, academics and support staff for providing a warm and nurturing environment from which to develop. Foremost of these people I would like to thank Prof. Alison Rodger. I cannot overemphasise the benefit of her mentorship and advice has been during this project.

I would like to thank my various collaborators in groups. From the Department of Chemistry: Dr. Mike Snowden, Dr. Eleni Bitzou and Xi Cheng. Dr. Sacha Ott from Systems Biology. Dr. Rich Boden from the School of Biomedical and Biosciences, University of Plymouth. From the Department of Statistics: Dr. Sach Mukherjee and Dr John Reid. Prof. Mario Nicodemi from the Department of Physics. The fruitful discussions I have had with each of these individuals have enhanced the various portions of the project with which they have been involved.

This work has received technical assistance from many members of technical staff from both the School of Engineering and the School of Life Sciences. In particular I would like to acknowledge the help and support I have received from Frank Courtney.

Finally, I would like to thank the people at Warwick and my family whose patience and support has sustained me throughout this project. Though too numerous to thank individually I would like to thank the following people principally: My mother, father and brother, Raquel and the members of flat 97.

Declaration

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy. The work described by the author is entirely original and my own unless otherwise stated.

Abstract

The complexity and size of the higher animal genome and relative scarcity of DNA-binding factors with which to regulate it imply a complex and pleiotropic regulatory system. *Cis*-regulatory modules (CRMs) are vitally important regulators of gene expression in higher animal cells, integrating external and internal information to determine an appropriate response in terms of gene expression by means of direct and indirect interactions with the transcriptional machinery. The interaction space available within systems of multiple CRMs, each containing several sites where one or more factors could be bound is huge. Current methods of investigation involve the removal of individual sites or factors and measuring the resulting effect on gene expression. The effects of investigations of this type may be masked by the functional redundancy present in some of these regulatory systems as a result of their evolutionary development. The investigation of CRM function is limited by a lack of technology to generate and analyse combinatorial mutation libraries of CRMs, where putative transcription factor binding sites are mutated in various combinations to achieve a holistic view of how the factors binding to those sites cooperate to bring about CRM function. The principle work of this thesis is the generation of such a library.

This thesis presents the development of microstereolithography as a method for making microfluidic devices, both directly and indirectly. A microfluidic device was fabricated that was used to generate oligonucleotide mixtures necessary to synthesise combinatorial mutants of a CRM sequence from the muscle regulatory factor MyoD. In addition, this thesis presents the development of the optimisation algorithms and assembly processes necessary for successful sequence assembly. Furthermore, it was found that the CRM, in combination with other CRMs, is able to synergistically regulate gene expression in a position and orientation independent manner in three separate contexts. Finally, by testing a small portion of the available combinatorial mutant library it was shown that mutation of individual binding sites within of the CRM is not sufficient to show a significant change in the level of reporter gene expression.

List of author's publications

Arasaradnam, R. P., Quraishi, N., Kyrou, I., Nwokolo, C. U., Joseph, M., Kumar, S., Bardhan, K. D., Covington J. A. (2011). Insights into "fermentonomics": evaluation of volatile organic compounds (VOCs) in human disease using an electronic "e-nose." *Journal of Medical Engineering & Technology*, ***35***(2), 87-91.

Fisher, R. D., Mbogoro, M. M., Snowden, M. E., Joseph, M. B., Covington, J. A., Unwin, P. R., Walton, R. I. (2011). Dissolution Kinetics of Polycrystalline Calcium Sulfate-Based Materials: Influence of Chemical Modification. *ACS applied materials & interfaces,* **3**, 3528-3537.

Cheng, X., Joseph M. B., Covington J. A., Dafforn T. A., Hicks M. A. and Rodger A. (2012). Continuous-channel flow linear dichroism. *Analytical methods*, **4**, 3169-3173

Sansuk S., Bitziou E., Joseph M. B., Covington, J. A., Boutelle M. B., Unwin P. R. and Macpherson, J. V. (2012) Ultrasensitive Detection of Dopamine Using a Carbon Nanotube Network Microfluidic Flow Electrode. *Analytical Chemistry*, *in press.*

List of abbreviations

3C	Chromatin conformation capture
3DP	3D printing
4C	Circularised chromatin conformation capture
5C	Carbon copy chromatin conformation capture
ABS	Acrylonitrile butadiene styrene
ALM	Additive layer manufacture
AP1	Activator protein 1
ASCII	American standard code for information interchange
BiFa	Binding factor
BMP-2	Bone morphogenic protein-2
BMP-4	Bone morphogenic protein – 4
BRE	Transcription factor IIB recognition element
BSA	Bovine serum albumin
CAD	Computer aided design
CDF	Cumulative distribution function
CER	Core enhancer region
ChIP	Chromatin immunoprecipitation
CNC	Computer numerical control
CODA	Computationally optimised DNA assembly
CRM	Cis-regulatory module
CSV	Comma separated variable
ct-DNA	calf thymus-DNA
DCE	Downstream core element
DCM	Dichloromethane

DMD	Digital micromirror device
DMSO	Dimethyl sulfoxide
DMT	Deoxyribonucleic acid methyltransferase
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DPE	Downstream promoter element
DRIE	Deep reactive ion etch
DRR	Distal regulatory region
EOF	Electroosmotic force
EOP	Electroosmotic pump
EWOD	Electrowetting on dielectric
FDM	Fused deposition modelling
Forkhead box	Foxo
FPS	Frames per second
FSC	Front scatter
GeMS	Gene morphing system
GFP	Green fluorescent protein
GRN	Gene regulatory network
GTF	General transcription factor
GUI	Graphical user interface
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
HPLC	High performance liquid chromatography
HPSF	High purity, salt free
HS	Horse serum
IC	Integrated circuit
IDT	Integrated DNA technologies

INR	Initiator
LB-Miller	Lysogeny broth – Miller variant
LCR	Ligase chain reaction
LENS	Laser engineered net shaping
LIGA	Lithographie, galvanoformung, abformung (lithography, electroplating and moulding)
LOM	Laminar object manufacturing
LSI	Large scale integration
MAP	Mitogen activated protein
MAPK	Mitogen activated protein kinase
MCR	MATLAB C runtime
Mrf4	Myogenic regulatory factor 4.
mRNA	Messenger ribonucleic acid
MSL	Microstereolithography
MTE	Motif ten element
Myf5	Myogenic factor 5
MyoD	Myogenic determination factor
Myog	Myogenin
NFY	Nuclear transcription factor Y
NF-κB	Nuclear factor-kappa B
NI-DAQ	National Instruments – Digital acquisition
NN	Nearest neighbour
OD	Outer diameter
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDF	Probability distribution function
PDMS	Poly(dimethylsiloxane)

PFPE	Perfluoropolyether
PIC	Pre-initiation complex
PMMA	Poly(methyl methacrylate)
PMT	Photomultiplier tube
PRR	Promoter
PS	Poly(styrene)
PSI	Pounds per square inch
PUMA	poly(urethane methacrylate)
PVC	Poly(vinyl chloride)
qPCR	Quantitative polymerase chain reaction
RIE	Reactive ion etch
RMSD	Root mean square difference
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
RNAP	Ribonucleic acid polymerase
RP	Rapid prototyping
RPM	Revolutions per minute
Runx2	Runt-related transcription factor 2
SAW	Surface acoustic wave
SDM	Site directed mutagenesis
SDS	Sodium dodecylsulfate
SEM	Scanning electron microscope
Shh	Sonic hedgehog
SLA	Stereolithography
SLS	Selective laser sintering
SRF	Serum response factor
SSC	Side scatter

TFIIB	Transcription factor IIB
TGFβ	Transforming growth factor β
TNFα	Tumour necrosis factor α
Tris	Tris(hydroxymethyl)aminomethane
TSS	Transcription start site
UV	Ultraviolet
VDR	Vitamin D receptor
WMISH	Whole mount in situ hybridisation

Chapter 1

1. Systems approaches of combinatorial dissection of *cis*-regulatory module function

1.1. Specification in development

Development is the process by which higher organisms transition from a single celled zygote to the complex, multicellular adult. During this transition the cells of the developing embryo will undergo periods of patterning, specification, migration, rapid division, apoptosis and differentiation. These processes are regulated in a precise and concerted manner by a remarkably small set of developmental genes[1,2]. The expression of these developmental genes must be tightly restricted to specific spatial and temporal locations within the developing embryo. The result of these processes is the specification of populations of cells, cell lineages, which will go on to form all of the >200 cell types found in the complex metazoans, known colloquially as 'higher animals'..

The specification of a cell lineage is the result of several processes acting in concert[3]. Firstly, the cell(s) must interpret the developmental cues from their surroundings to derive their 'identity'. Secondly, the expression of specific regulatory genes necessary for this identity must be activated and then stabilised. Thirdly, alternative regulatory genes for alternative identities must be excluded. Finally, various lineage specific genes necessary for proper development of the lineage within the context of the overall embryo must be activated. When a decision is made to express a given regulatory gene, a pleiotropic gene regulatory network (GRN) is initiated where a cascade of interregulating genes are expressed that result in the appropriate course of development for the given lineage[3].

1. Systems approaches of combinatorial dissection of *cis*-regulatory module function

The processes of development are initiated in response to a range of inter- and intracellular signalling cues. Many cues are derived from the overlapping gradients of signalling molecules throughout the embryo generated by specific groups of cells (such as dorsal-ventral orientation). Other cues, however, are transmitted via direct cell-cell contacts (such as Notch/delta signalling in pigmentation). A cell must integrate information from various competing and cooperating signals and determine the appropriate response. One mechanism of this integration process occurs during signal transduction: different signalling pathways might share common elements in their cascades and by affecting the activity of these elements, the information from different signalling sources is merged. The p38 mitogen activated protein kinase (MAPK) pathway is an example of a signalling pathway with multiple inputs[4]. The activation of specific regulatory factors within the nucleus, however, is the terminus of many signalling cascades. In the nucleus, integration of complementary and competing signals is achieved at the promoters, enhancers, silencers and other regulatory modules associated with specific target genes. The result of these regulatory interactions is the expression of genes that result in the assumption of a cellular identity. Although selected, in some cases the identity of a cell is still plastic, as demonstrated by tissue grafting experiments in the chick embryo[5]. Cells previously expressing genes specific for one location can be induced to express genes specific to another location once grafted to the new location and the alternative signalling cues are internalised and interpreted.

A cell lineage is maintained by a permanent alteration of the cell's response to signalling. Elements of a signalling cascade may be sequestered, degraded or expression deactivated in order to make a cell deaf to a specific signal. Activation of a specific regulatory factor may initiate a positive feedback loop, reinforcing its own expression that, by virtue of its effect on the expression of downstream genes, locks in the identity of the cell to a specific lineage. Furthermore, epigenetic modification of the DNA and histones is

known to lead to the silencing of whole regions of DNA, preventing the expression of regulatory genes therein[6–8]. Silencing is usually achieved by a combination of epigenetic and histone modification driven by the recruitment of DNA methylases and histone deacetylases that serve to favour the packaging of DNA into silent, non-expressing heterochromatin[9].

The spatial and temporal expression of genes necessary to specify a cell lineage is usually tightly controlled. Figure 1.1 exemplifies this using the specification of muscle progenitor cells in the developing embryo. The specification of a region of tissue in the embryo that will become the adult skeletal musculature is achieved by the overlapping presence of several signals; bone morphogenetic protein-4 (BMP4), noggin, Shh and the Wnt proteins. This process is discussed further in chapter 3.

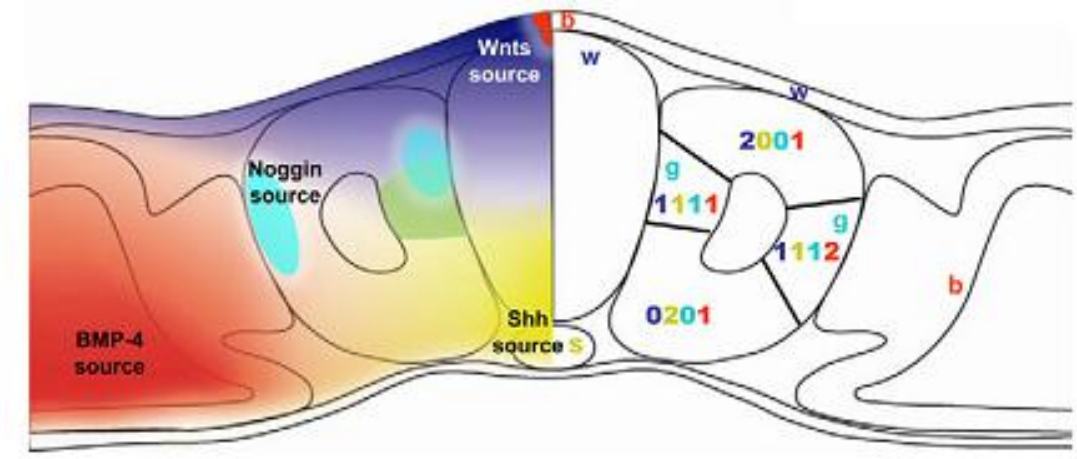


Figure 1.1: Overlapping signals that lead to the specification of muscle cells in the developing embryo. Image shows a transverse section through a developing embryo; A central neural tube (top centre) and notochord (bottom centre) are flanked by a pair of somites on each side followed by a pair of lateral limb buds. Left half of the image shows the morphogenic fields of various transcriptions factors: Red; BMP4, cyan; noggin, yellow; Shh and blue; Wnt proteins. Right side of image shows the relative concentration of these factors in the different portions of the somite. Letters denote the tissues that are the sources of the different signalling molecules. The section highlighted in green on the left and by 1111 on the right is the area where muscle cell progenitors are specified. Image taken from Piran *et al.*[10].

Bacteria are able to achieve a sufficiently sophisticated suite of regulatory control mechanisms by direct interactions between transcription factors and the core transcriptional machinery[11]. A population of bacteria may respond to changes in environment by evolving their responses appropriately, trimming excess genetic code and altering regulatory interactions. In contrast, the responses to environmental change of each of the >200 cell types of the higher animal must be encoded in each of the individual cell types. Remarkably, this function is achieved with relatively fewer transcriptional genes[2]. As a result, the regulatory interactions that ensure the appropriate responses occur are significantly more complex in higher animals than prokaryotes. In higher animals correct spatiotemporal gene regulation is achieved through the complex interactions of multiple DNA-binding proteins and their cognate binding sites in regulatory modules in the non-coding DNA. This is achieved with a relatively small amount of genes operating in pleiotropic networks and the mechanisms that result correct regulation of each gene are likely to be complex. To understand how genes are regulated, the mechanisms of their expression must first be understood.

1.2. Regulation of transcription

Several processes must occur in concert for a gene to be actively transcribed: The highly packaged chromatin immediately around the transcription start site (TSS) must be decondensed. The pre-initiation complex (PIC), containing ribonuceleic acid (RNA) polymerase II (RNAPII), must form at the sequence elements of the core promoter [12]. Finally, the assembled RNAPII is released and processive transcription occurs producing a messenger RNA (mRNA) transcript from which a functional protein is subsequently produced by translation. Although transcription constitutes the major point of regulation in the expression of most genes, other mechanisms of the regulation of protein expression and activity should also be noted; degradation of mRNA, inhibition of nuclear export,

sequestration/localisation of mRNA, degradation of protein, sequestration of protein and post-translational modification. Regulation of transcription is achieved by the interaction of enhancing and silencing regulatory modules with the core transcriptional machinery at the core promoter.

In contrast to the highly conserved prokaryotic promoter, the promoters of eukaryotic cells exhibit significantly more variation in the elements present therein. The diversity in promoter structure observed in eukaryotic organisms[13,14] is presumably a reflection of the range of regulatory conditions that must be represented in order to correctly control in a complex multicellular organism the spatiotemporal expression of the ~20000 genes present in the human genome with only an estimated <2000 DNA-binding transcription factors[1,2].

The promoter alone is not sufficient to facilitate the high levels of expression associated with some genes[6]. The core promoter is supplemented by additional elements that facilitate and regulate the expression of a given gene. These elements can be close (≤1 kb) to the promoter they regulate in the case of proximal regulatory elements or many 10s or even 100s of kb away in the case of distal regulatory elements[15]. Regulation of a target gene by a regulatory element is achieved in conjunction with the binding of a specific transcription factor. For example, activation of a target gene can be achieved by the binding of an activatory transcription factor to an element, which is called an enhancer[16]. Conversely, the binding of a repressive transcription factor to an element produces a silencer. Clusters of regulatory transcription factor binding sites that act on genes on the same chromosome, to bring about enhancement or silencing of a target gene, are termed *cis*-regulatory modules (CRMs)[17]. Further elements include insulators and locus control regions (LCRs). Insulators bound the influence of nearby enhancers and silencers and prevent the functionality of these elements from affecting the promoters beyond. LCRs, typified by the β-globin LCR, are collections of regulatory elements that cooperate to

achieve the regulation of a gene in a location-independent manner[18]. Similarly, groups of enhancers are able to associate into enhanceosome structures, the best characterised being the enhanceosome of the IFN-β gene[19] which has even been crystallised[20]. Groups of repressive regulatory elements can also cooperate to act as a repressosome[21] and, in the right circumstances, enhanceosomes can be converted to repressosomes[22]. The enhanceosome model of CRM action, therefore, involves the interaction of various factors bound to a CRM in with an overall structure and that each element of the structure is necessary in order for the whole to function as intended. In addition to this highly coordinated and cooperative model of CRM action, another model, the billboard model, also exists[23]. In this model, factors bound to binding sites do not interact with each other and instead regulate transcription through independent interactions with the core transcriptional machinery.

Regulation of the target promoter is brought about by the relative frequency with which these different elements are able to interact with the transcriptional machinery bound to the core promoter.

A diverse range of signals are integrated by interactions between and within CRMs that result in specific decisions about whether a gene is expressed or silenced[24]. A single decision made at a specific CRM can have dramatically wide ranging downstream effects that can affect the fate of a cell. For example, the expression of *myod* results in the commitment of a group of cells within a structure called the somite in the developing embryo to the myogenic lineage (see figure 1.1) during initial stages of myogenesis (see chapter 2 for a summary of the current knowledge of how *myod* regulation is achieved). Clearly, the mechanisms by which the appropriate response is computed as a result of integration of the diverse signalling inputs are necessarily complex.

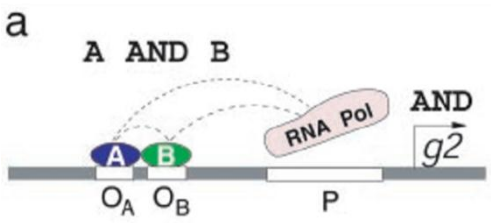
1.3. Signal interpretation at/by *cis*-regulatory modules

A vast number of regulatory inputs must be processed by the many cells of a developing embryo. The output, activation and/or silencing of gene expression at certain specific loci, will incorporate information about the history of the cell, or cell lineage with information from the environment. The effect of the integration of the external signalling influences and the internal lineage information is determined by CRMs that control the spatiotemporal expression of each gene. The regulatory responses to a given set of inputs by the cell is, therefore, 'hardwired' into the genome by the presence of these CRMs[25].

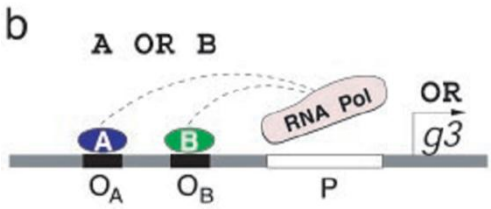
CRMs are typically between 100 and 1000 basepairs in length and contain a high concentration of transcription factor binding sites[26]. Some regulatory modules are capable of acting on the promoters of more than one gene and also across chromosomes[27]. Furthermore, whether the CRM is activatory or inhibitory depends what factors are bound to the sites within that CRM. Regulatory modules in both these contexts are herein referred to as CRMs. Because they are the site of integration between the signalling and internal regulatory state of the cell, CRMs, or combinations thereof, are the site where the appropriate response is determined.

A CRM receives input, binding of transcription factors, as a result of internal or external signalling events, operates upon them and derives an output: driving or halting gene expression as appropriate[28]. The CRM is, therefore, acting like a computer[29] with binding of transcription factors to DNA generating logic gates controlling the expression from the target promoter[30]. Cooperative and competitive binding form the basis of the interactions that underpin these logic gates. Cooperative binding is where two factors co-stabilise their interaction with the DNA by interacting with each other. Competitive binding is where two factors, which are independently able to bind the DNA, share overlapping binding sites and cannot bind simultaneously. If two activatory factors, A and B, must bind

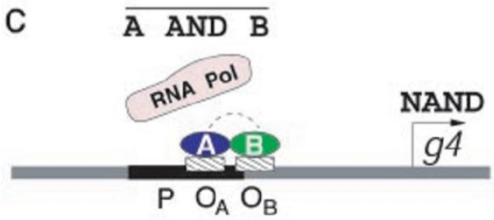
cooperatively to the DNA an AND gate is generated as the activatory effect is only observed when both factors are available (see figure 1.2, a). Similarly, if two independently binding activatory factors are able to perform the same activatory effect, an OR gate is generated (see figure 1.2, b). The NOT version of each of these gates, NAND and NOR, can be generated if the transcription factors compete with the RNAPII binding site. An example of a NAND gate is shown in figure 1.2, c.



Gate type	A	B	On/off
AND	✔	✔	ON
	✔	X	OFF
	X	✔	OFF
	X	X	OFF



Gate type	A	B	On/off
OR	✔	✔	ON
	✔	X	ON
	X	✔	ON
	X	X	OFF



Gate type	A	B	On/off
NAND	✔	✔	OFF
	✔	X	ON
	X	✔	ON
	X	X	ON

Figure 1.2: Gene regulatory logic gates. Three gates are demonstrated: a) represents an AND gate, where both factor A and B must be present for either bind and interact with the RNAP. b) represents an OR gate, where factors A and B binding to distinct, non-interacting sites and each interact with the RNAP independently. c) represents a NAND gate, where both factor A and B must be present for either to bind and thereby disrupt RNAP function. Image adapted from Buchler *et al.*[30].

Higher order interactions between transcription factors can result in more complex regulatory logic gates[31] including transistor-like latches[32]. Importantly, the same or similar modes or motifs of gene regulation appear repeatedly in disparate regulatory settings. Such motifs usually rely on the relative position of transcription factor binding sites, with

changes in position or orientation critically affecting function. Conversely, there are examples of considerable flexibility in the order and relative position of transcription factor binding sites[33].

Logic functions appear to play a central role in the regulation of many developmental genes across several multicellular species [34]. The mechanism by which a cell, finding itself in a given environment, determines what identity it should assume is hardwired into the genetic code by means of transcription binding sites within CRMs. Specific factors bound to these sites interact to bring about the expression of specific regulatory factors. The expression of these factors sets in motion a cascade of gene expression that brings about the appropriate response.

1.4. Evo-devo and CRM interactions

Unlike gene regulation in bacteria, where gene activation or repression frequently involves tightly binding σ factors[35], regulation in animals involves weakly interacting transcription factors binding to regulatory modules distinct from the promoters of genes. The separation of the regulatory and functional aspects of gene expression allows each to evolve separately. Gene duplication, mutual redundancy and subsequent divergence can lead to new functionality and/or spatiotemporal expression without necessarily restricted by either[36,37]. The study of how evolutionary change is brought about by modification of the regulation of development is called 'evo-devo'[38,39]. The evo-devo narrative, in response to the finding that significant differences in body plans between higher animals is not mirrored by significant differences in gene sequence, says that it is changes to the regulation of genes that drives morphogenic change rather than physicochemical change in protein structure/function[38,40]. For example, gene families such as the homeobox-containing *Hox* genes and Wnt genes are centrally involved with fundamental processes of specification and development, are highly functionally similar but possess dramatically

different expression patterns. It is the differences in expression patterns of these and other influential gene families that results in the differences in the body plans between higher animals.

Duplication of whole GRNs can lead to alteration of the body plan by patterning of previously adapted organs or tissues[41]. The flexibility and variation seen in the vertebrate body plan, as compared to the relatively insignificant variation in genome sequence, is probably due, at least in part, to the flexibility of a system where regulatory modules are divorced from the genes they regulate[40,42]. This conclusion is supported by the fact that there is significant variation in the relative number of members of DNA-binding protein families across different animal phyla, although each phyla does contain a common set of families overall[43]. The master muscle regulatory factors (MRFs) seem to exemplify this paradigm: A family of related transcription factors that exert high level control over the processes of muscle specification during development across all higher vertebrate animals. The MRFs are a family composed of four genes; *myod*, *myf-5, myogenein* and *mrf4* [44].

1.5. Mechanisms of CRM-promoter communication: Billboard vs. Enhanceosome

CRMs can be classified into either enhanceosomes, which feature highly cooperative transcription factor binding sites, or billboards (also known as information display), which are more flexible in terms of transcription factor binding site arrangement[23]. Enhanceosome CRMs involve the interaction of many proteins to generate a complex that can have either an enhancing or repressive effect. Billboard CRMs consist of distinct sites, each of which is competing for a limited number of target sites within another regulatory complex. The position and orientation of the multiple transcription factor binding sites that constitute an enhanceosome is vital to enhanceosome function. As a result, enhanceosomes are usually conserved evolutionarily. Billboard-type CRMs are, by contrast, not sensitive to the position or orientation of transcription factor binding sites as each site

operates on the core transcriptional machinery independently. The output from a billboard-type CRM is more likely to be stochastic, with output from the gene depending on the relative likelihood of each subset of transcription factors interacting with the transcription machinery. The best known example of enhanceosome CRM is virus inducible enhancer in the human *IFN-β* gene[19]. Conversely, the stripe 2 enhancer in *Drosophila* exemplifies the billboard-type CRM[45,46]. The two types can be distinguished by the independence of the billboard-type from position and/or orientation sensitivity. Sites within the billboard-type operate essentially independently or in pairs and are therefore resistant to individual sites being changed. In contrast, sites in the enhanceosome model are highly cooperative, often requiring architectural looping interactions[47,48]and removal of one site likely result in the silencing of the CRM. Long distance interactions between CRMs that constitute enhanceosome structures can be identified through chromatin conformation capture (3C) assays[49,50] that provide information about the 3D structure of interacting sites. More recently, high throughput modifications of the 3C protocols, such as circularised 3C (4C) and carbon-copy 3C (5C), have provided genome wide interaction maps[51]. Due to the involvement of multiple modules undergoing complex interactions, the enhanceosome model is likely to demonstrate context sensitivity, whereas an individual module can either be activatory, repressive or silent depending on the presence and factors bound to the other modules.

CRMs can either operate in either rheostatic or binary manners. In the former, the CRM affects the rate of transcription at the target promoter quantitatively. Whereas in the latter, the CRM increases the likelihood that transcription from the target promoter will take place without affecting the overall rate of transcription. Rheostatic enhancers will increase the overall quantity of the gene product in a given population of cells. Binary enhancers, by contrast, increase the likelihood that a given cell within a population will express the gene, but not the final concentration of the gene product[52]. To distinguish

between these two types measurements of individual cells within large populations must be made. If the whole cell population state is measured by, for example a Western blot or qPCR, a moderate level of protein expression might be indicative of either all the cells expressing a moderate amount of protein, rheostatic type, or half the cells expressing a high level of protein and half the cells expressing none, binary type. To measure large populations of cells individually, either live cell imaging or flow cytometry can be used. Statistics about individual cells can then be determined and the two types of CRM, rheostatic or binary, distinguished.

1.6. Investigation of CRMs

CRMs are typically identified by either computational or perturbation experiments. In the former type, which is much faster but less conclusive and requires aligned genomic sequence information, sequences from the genomes of difference species are compared for conservation or simply scanning for high densities of predicted binding sites. Regions that are strongly conserved between species are considered to be important as there is a strong evolutionary pressure to resist changes in these regions. The CRMs putatively identified thusly must then be confirmed by *in vivo/in vitro* experimentation either in culture or in whole organisms. The latter type, which is slower but more conclusive and requires many constructs, the DNA sequence around a promoter of interest is manipulated. Specific regions or binding sites can be removed and their effect on expression of a gene can be monitored. Because transcription factors do not appear to bind sites with affinity directly proportional to the sequence of the site, as would be expected[53], it is not possible to use solely predictive, computational methods for CRM identification.

Several bioinformatics tools have been developed to detect CRMs based upon sequence conservation across species[54–56] or the density of predicted transcription factor binding sites[26,57,58]. The vast amount of information arising from whole genome sequencing

and microarray expression studies have led to attempts to predict the number and interactions of CRMs necessary to achieve given gene expression patterns [59–61]. In situations where comparable sequence information is not available, CRMs can still be identified by integrating information from chromatin immunoprecipitation on chip (ChIP-chip) and ChIP-sequencing (ChIP-seq) that can identify regions where specific factors bind or where the chromatin is in the open conformation across the whole genome. Information from such experiments can be used to make sequence search models more accurate[62]. In relatively simple systems, such as segmentation in drosophila, a knowledge of the spatiotemporal expression of a range of factors and a clearly discernible output can be combined to make accurate models of CRM interactions based on probabilistic models of site occupancy[63]. Confirmation of networks predicted by such systems still requires some form of specific site ablation or knockdown.

The spatiotemporal expression pattern of a given gene can be determined by a combination of whole mount *in situ* hybridisation (WMISH) and quantitative PCR (qPCR), respectively. Ablation of whole CRMs, or sites within CRMs, that target the gene of interest can be followed by a mixture of WMISH, qPCR and microarray analysis[64]. A gold standard of confidence a CRM mode of action can be obtained by performing either *in vivo* gene ablation or knock-down and rescue of effect. A combination of approaches can be used to identify gene regulatory networks including statistical mechanics, *in vivo* knock outs and knowledge of protein-protein interactions[65]. An example of how prior knowledge of relevant GRNs, inter-species sequence comparison, gene knock-downs and reporter constructs can be used to elucidate mechanisms of gene regulation is presented by Ransick *et al.*[66]. In this case, the mechanism of how Notch signalling is integrated in the decision during sea urchin mesoderm specification to express *glial cells missing* (*gcm*) is described. The development of additional combinatorial approaches could increase the rate at which GRNs such as these can be discovered in higher animals[24].

1. Systems approaches of combinatorial dissection of *cis*-regulatory module function

Regulatory networks that control aspects of *Drosophila* and sea urchin[67] development have been elucidated[68]. The relatively simple body plans and easily handled embryos of these organisms mean they are well suited for developmental studies. Simple logic systems derived from eukaryotic regulatory networks have been successfully expressed and operated in bacteria[69,70]. Figure 1.3 shows the most up to date understanding of this GRN as elucidated by work from the Davidson lab and others.

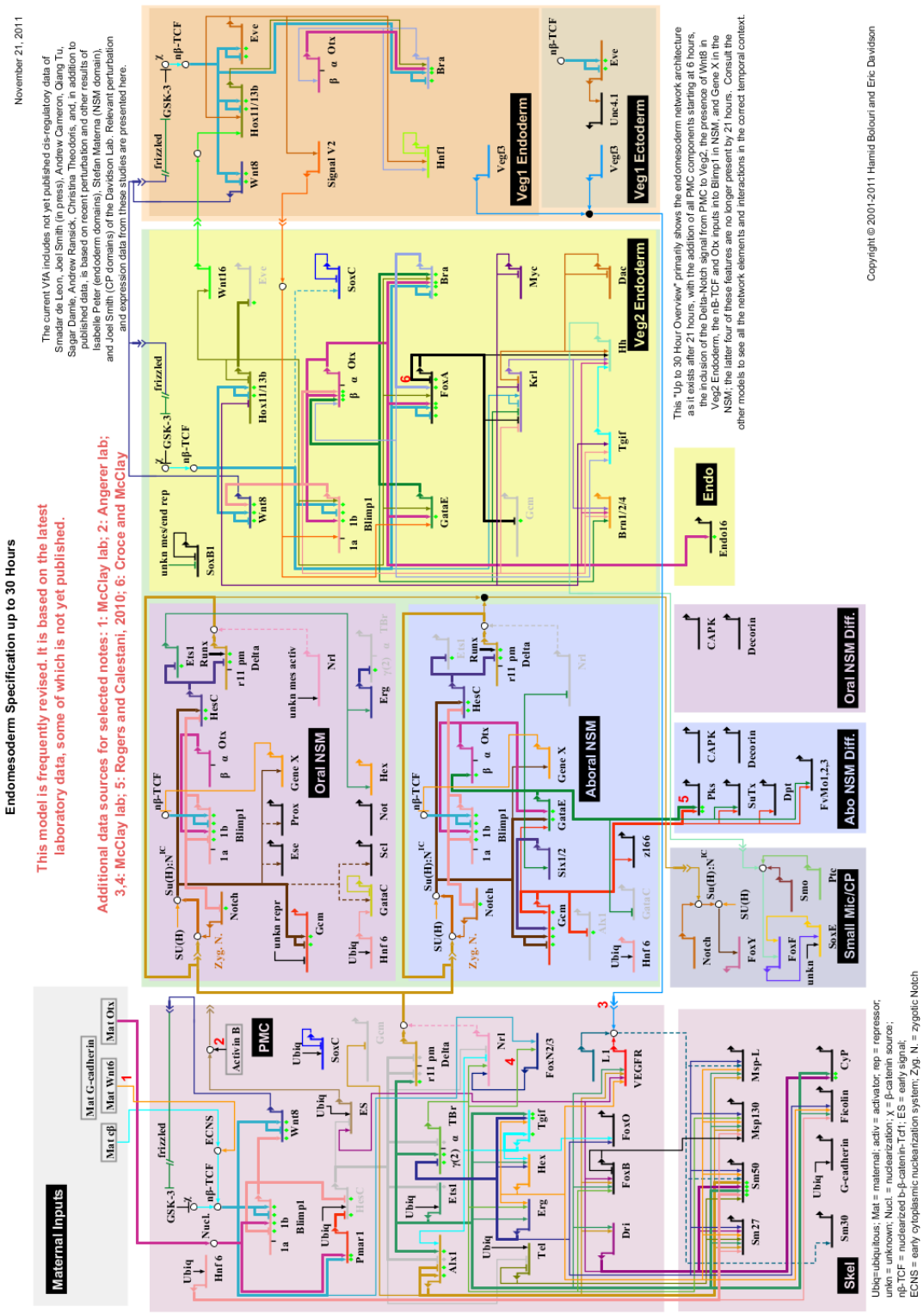


Figure 1.3: GRN for endomesoderm in sea urchin development. Labelled horizontal lines indicate genes. Arrowed lines indicate where a gene product of one gene affects the regulation of another. Network obtained for the whole sea urchin genome at up to 30 hours. from http://sugp.caltech.edu/endomes/. Refer to figure text for explanation of abbreviations.

1. Systems approaches of combinatorial dissection of *cis*-regulatory module function

A crucial concept in CRM function is activator synergy[71] and the regulatory logic gates are a form of transcription factor synergy. Cooperative binding is an obvious mechanism of activator synergy. This direct type of interaction may be transmitted short distances by one or more third party factors, the binding of which stabilises each of the DNA binding proteins. Indirect synergies can also occur: A factor bound to one site might recruit a chromatin remodelling complex or cause nucleosome slippage that adjusts the chromatin structure so as to expose another site[7,72,73]. Similarly, some effects that would otherwise be observable by single site sequence modification could be masked due to redundancy in the regulatory systems[37]. To observe and understand these effects, a combinatorial approach must be taken where pairs and whole sets of sites are simultaneously mutated and the effects observed. In this manner, an 'alphabet' of common networks or motifs of *cis*-regulatory functionalities can arise. Such a compendium would prove invaluable for future developments in the field of synthetic biology that might lift systems whole sale from those that already exist.

Investigation of CRM function by combinatorial investigation of transcription factor binding can be performed by examining the activity of CRMs that have sites mutated in pairs, triplets or more. By carefully comparing expression of reporter genes in the context of every combination of the 'on' (wild type) and 'off' (mutated) sites, a clear picture can be obtained of the effect and function of each site within the context of the presence or absence of each other site. This type of manipulation is usually achieved by gene ablation where sections or sites within the regulatory regions are removed.

Gene ablation is performed by deleting whole sections of the sequence surrounding a promoter and observing the effect via expression of a visualisable product such as GFP or β-galactosidase[74]. This is a typically low resolution approach (100's to 1000's of bp) that depends on the presence of unique sites for restriction enzymes. At the other end of the scale, site directed mutagenesis (SDM) can be used to target specific sites by

making short (<5 bp) changes to the sequence[75,76]. SDM cannot make multiple changes in a single reaction and is limited in terms of the length of the change. Therefore SDM is suitable for targeting individual sites, but cannot make large scale changes. Thus SDM is only suitable for the generation of limited, small scale libraries. More recently, zinc-finger nucleases have been employed to make alter sequences at specific sites in the sea urchin genome[77]. Zinc finger nucleases rely on accurate targeting via DNA binding domains and therefore suffer the same limitations of restriction enzymes. These techniques offer a method to determine whether a given sequence is capable of affecting the expression of a target gene.

Error-prone polymerase chain reaction (PCR) could be used to produce a library of variants of a single sequence[78,79], with this library then being assessed for activity. This method would, however, not make use of any of the available *a priori* information available in this situation and would make changes to all parts of the sequence at random rather than at specific sites as desired. The various methods of library production are discussed in more detail in section 8.1.

To produce a set of sequences, therefore, that are capable of interrogating the apparent complex, higher order interactions between previously identified CRMs (see section 3.5) a combinatorial approach must be taken which existing techniques are not suitable to provide. Such an approach would require the development of new technologies and techniques necessary for the efficient generation of the library.

1.7. Thesis aims and objectives

The time consuming methods of knockout studies and genomic deletion mapping do not always return successful results and do not allow for the combinatorial analysis of the many potential inputs of a *cis*-regulatory system. By combining information from microarrays, ChIP, bioinformatic binding site prediction and highly parallel, microfluidic

synthesis techniques useful information about how regulation of gene expression can be obtained[24]. Combinatorial libraries can be used to supplement traditional approaches of construct generation for transient expression studies in suitable cell culture models to rapidly elucidate the complex interactions between factors bound to discrete sites in the CRMs and the promoter of the *myod* gene. This type of investigation can also determine whether the mechanisms by which the previously identified CRMs of *myod* act in a manner consistent with either the billboard or enhanceosome model.

De novo gene assembly is such a technology that could be used to produce a library of combinatorial CRM variants in a parallelisable manner. The use of gene assembly to produce such a library requires the development of optimisation algorithms, of assembly protocols and microfluidic systems. The development of each of these necessary enabling technologies is described herein and applied to the investigation of the mechanisms regulating *myod*, a master regulator of myogenesis. This thesis aims to develop technologies necessary to investigate the mechanisms of CRM interaction within the context of previously identified CRMs of the *myod* gene. The technologies and systems developed here could then be applied to other systems for the rapid acquisition of information about CRM regulation in different gene systems. Figure 1.4 shows a high level overview of the different aspects of the project.

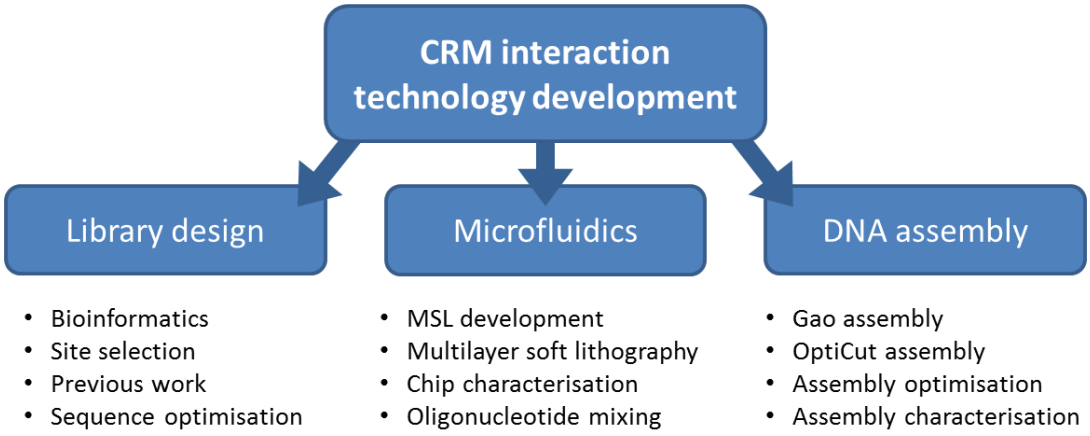


Figure 1.4: Chart showing a high level breakdown of the different aspects required for the development of novel technology for the investigation of CRM interactions presented in this thesis. The thesis breaks down into three main subject areas; library design, microfluidics and DNA assembly. Each of these areas involves several distinct aspects that are split amongst the chapters as appropriate.

A combinatorial mutant library was designed by taking into account information from a variety of sources. Chapter 2 outlines the diverse signalling pathways involved in muscle specification and the regulation of *myod*. How these signalling events are integrated at the *myod* promoter is less clear, despite the considerable effort in the last two decades by various labs, including the one in which this project is based. The information currently known about the contribution of several suspected CRMs is also summarised in chapter 2. In addition to the information available in the literature, data obtained by previous researchers within the group in which this project was based was incorporated. An argument for the prioritisation of certain binding sites is then made in chapter 2 and a set of these sites were selected for combinatorial mutation in a specific CRM context. Figure 1.5 summarises the process for combining the knowledge into a list of sites of interest from which the sequences composing combinatorial mutant library was generated.

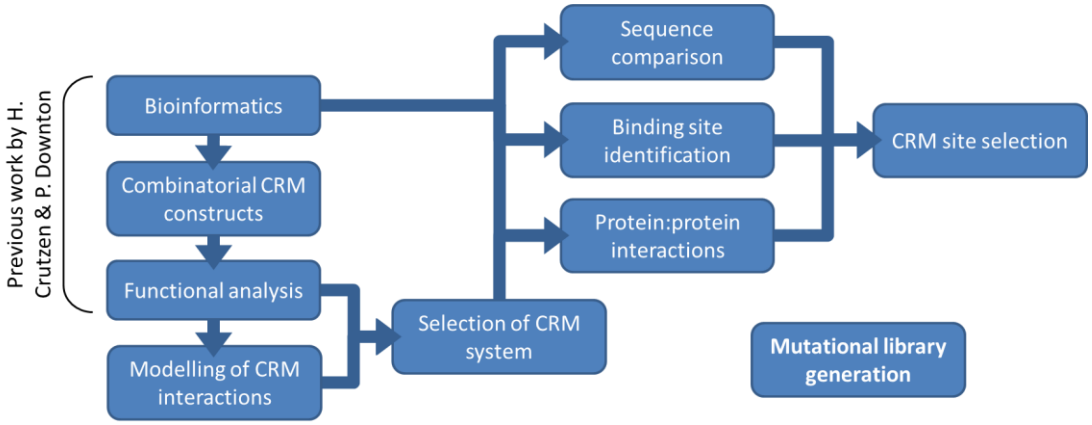


Figure 1.5: Flow chart describing the process of determining the sequences that make up the mutational library (see chapter 3). The process starts with previous work by firstly by H. Crutzen and secondly by P. Downton. Using information from this previous work and modelling of the CRM interactions by J. Reid and others, a CRM system of a set of sites to mutate within a specific CRM were selected. These sites were then used to produce a mutant library that consists of a complete set of combinatorial mutations of these sites.

Chapter 3 reviews the application of microfabrication technologies to the production of microfluidics that could be used in this project. Whilst suitable for methods development and small scale library generation, molecular biology on the bench top scale is unsuitable for the generation of large scale, productive development. As a result, a method for the fabrication of microfluidic devices for the contamination free mixing of DNA assembly substrates is described in chapter 6. The microfluidic devices were fabricated by multilayer soft lithography, with each layer of the device composed of polydimethylsiloxane (PDMS) casts of moulds fabricated by microstereolithography (MSL). The process of optimising MSL for the production of microfluidic devices and its use as a mould to cast PDMS is described in chapter 5. Figure 1.6 shows a flow chart that describes the process employed here for the development, assembly and testing of the microfluidics in this project.

Chapter 4 contains the materials and methods for the experimental work presented throughout this thesis.

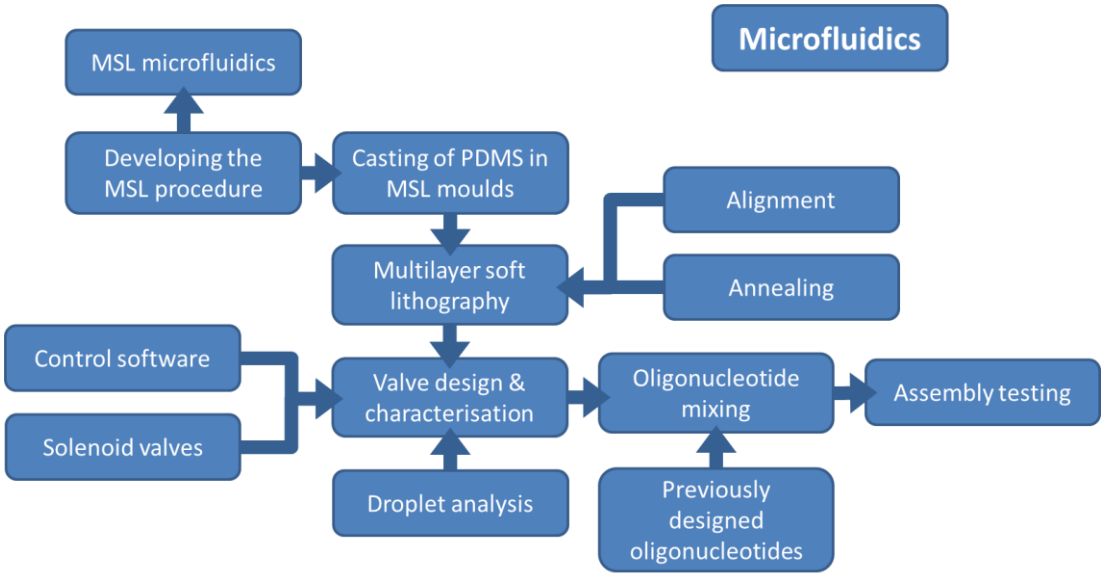


Figure 1.6: Flow chart describing the process for the development, assembly and testing of the microfluidic oligonucleotide mixing chip used in this project (see chapters 5 and 6). The MSL procedure was characterised and used to make MSL moulds from which PDMS casts were obtained. Several processes were necessary for the successful operation of the microfluidic chip: Assembly of the chip by multilayer soft lithograph and fabrication and writing of electronics and control software. The successfully assembled and characterised microfluidic chip was then employed to make oligonucleotide mixtures.

The combinatorial mutant library was generated by a DNA assembly process. Two methods were used: Gao assembly and OptiCut assembly. Both methods involve the ligation of specifically designed oligonucleotides followed by the amplification of the full length ligated DNA assembly. The methods are distinguished by the sequence optimisation method, whether all or a subset of the oligonucleotides were present in the assembly reaction and the need for intermediate amplification steps. The development of the OptiCut optimisation algorithm is described in chapter 7. The optimisation of the CRM assembly process is described in chapter 8. The overall process for the CRM assembly described herein is summarised in figure 1.7.

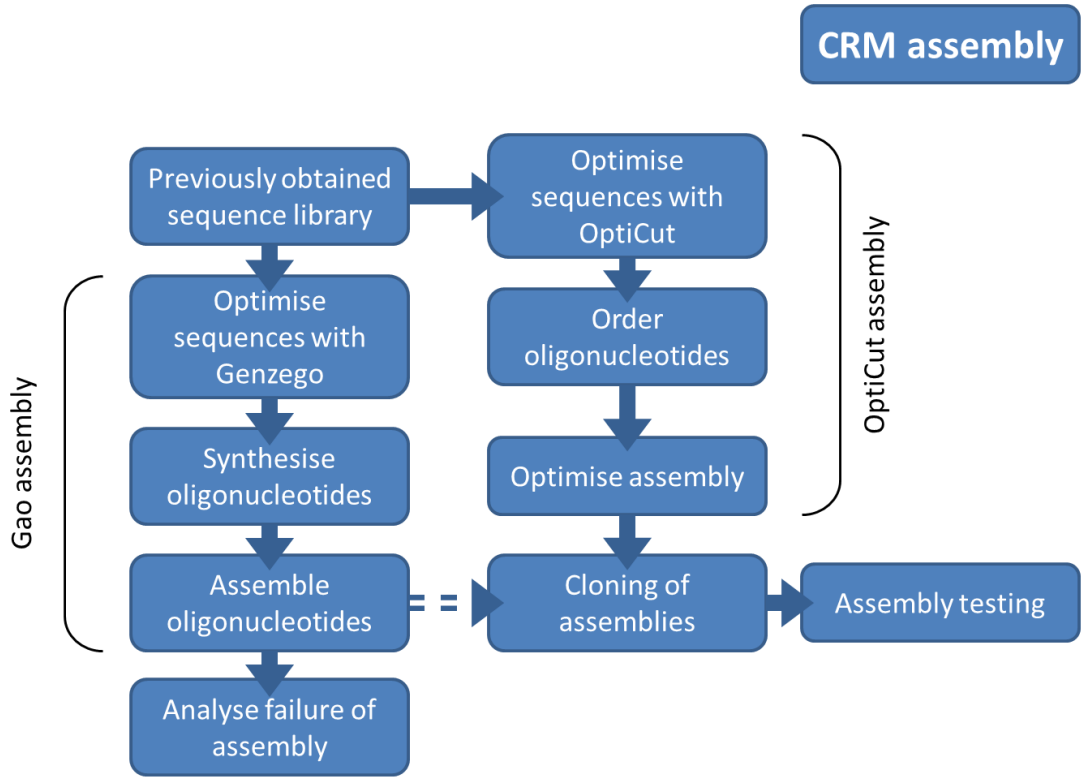


Figure 1.7: Flow chart describing the CRM assembly process used in this project (see chapter 8). The Gao assembly (section 8.2) process was the first attempted and failed at the assembly of oligonucleotides step. An alternative approach, OptiCut assembly (section 8.3), was then used as the source CRM mutant sequences that were then partially tested.

Whilst knock-down of a specific factor can yield high quality evidence of that factor's importance to the regulation of a factor, knockdown alone does not indicate whether the observed effect is direct or indirect. Initially, observation of effects on the basis of ablation of specific sites or combinations of sites is a useful first step to identifying the factors to knock-down in further analysis.

Finally, chapter 9 describes the testing of both the position-orientation dependence of the CRMs (see section 9.3) and of a small portion of the mutant library (see section 9.4) generated in chapter 8. The position-orientation investigation shows whether the observed effects are an artificial product of the plasmid environment into which they are place or representative of the CRMs acting as independent regulatory entities.

Chapter 10 then draws the principle conclusions of the previous results chapters together and discusses potential future work.

1.8. References

1. Davidson, E.H. *The regulatory genome: Gene regulatory networks in development and evolution*. (Academic Press: San Diego, CA., 2006).
2. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S. a & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nature reviews. Genetics* **10**, 252-63 (2009).
3. Levine, M. & Davidson, E.H. Gene regulatory networks for development. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4936-42 (2005).
4. Ono, K. & Han, J. The p38 signal transduction pathway: activation and function. *Cellular signalling* **12**, 1-13 (2000).
5. Pourquié, O., Coltey, M., Teillet, M. a, Ordahl, C. & Le Douarin, N.M. Control of dorsoventral patterning of somitic derivatives by notochord and floor plate. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 5242-6 (1993).
6. Svejstrup, J.Q. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica et biophysica acta* **1677**, 64-73 (2004).
7. Bai, L. & Morozov, A.V. Gene regulation by nucleosome positioning. *Trends in genetics : TIG* **26**, 476-83 (2010).
8. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature reviews. Genetics* **10**, 295-304 (2009).
9. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-32 (2007).
10. Piran, R., Halperin, E., Guttmann-Raviv, N., Keinan, E. & Reshef, R. Algorithm of myogenic differentiation in higher-order organisms. *Development (Cambridge, England)* **136**, 3831-40 (2009).
11. Payankaulam, S., Li, L.M. & Arnosti, D.N. Transcriptional repression: conserved and evolved features. *Current biology : CB* **20**, R764-71 (2010).
12. Smale, S.T. & Kadonaga, J.T. The RNA polymerase II core promoter. *Annual review of biochemistry* **72**, 449-79 (2003).
13. Juven-Gershon, T. & Kadonaga, J.T. Regulation of Gene Expression via the Core Promoter and the Basal Transcriptional Machinery. *Developmental biology* **339**, 225-229 (2010).
14. Gershenzon, N.I. & Ioshikhes, I.P. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics (Oxford, England)* **21**, 1295-300 (2005).
15. Maston, G. a, Evans, S.K. & Green, M.R. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* **7**, 29-59 (2006).
16. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569-577 (1997).
17. Blackwood, E.M. Going the Distance: A Current View of Enhancer Action. *Science* **281**, 60-63 (1998).
18. Li, Q., Peterson, K.R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077-86 (2002).
19. Panne, D. The enhanceosome. *Current opinion in structural biology* **18**, 236-42 (2008).
20. Panne, D., Maniatis, T. & Harrison, S.C. An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111-23 (2007).
21. Gowri, P.M., Yu, J.H., Shaufl, A., Sperling, M.A. & Menon, R.K. Recruitment of a Repressosome Complex at the Growth Hormone Receptor Promoter and Its Potential Role in Diabetic Nephropathy. *Molecular and Cellular Biology* **23**, 815-825 (2003).
22. Lee, B. *et al.* From an enhanceosome to a repressosome: molecular antagonism between glucocorticoids and EGF leads to inhibition of wound healing. *Journal of molecular biology* **345**, 1083-97 (2005).
23. Arnosti, D.N. & Kulkarni, M.M. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of cellular biochemistry* **94**, 890-8 (2005).

24. Jeziorska, D.M., Jordan, K.W. & Vance, K.W. A systems biology approach to understanding cis-regulatory module function. *Seminars in cell & developmental biology* **20**, 856-62 (2009).
25. Ben-Tabou de-Leon, S. & Davidson, E.H. Gene regulation: gene control network in development. *Annual review of biophysics and biomolecular structure* **36**, 191 (2007).
26. Berman, B.P. *et al.* Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 757-762 (2002).
27. Arnone, M.I. & Davidson, E.H. The hardwiring of development: Organization and function of genomic regulatory systems. *Development* **124**, 1851-64 (1997).
28. Kulkarni, M.M. & Arnosti, D.N. Information display by transcriptional enhancers. *Development (Cambridge, England)* **130**, 6569-75 (2003).
29. Istrail, S., De-Leon, S.B.-T. & Davidson, E.H. The regulatory genome and the computer. *Developmental biology* **310**, 187-95 (2007).
30. Buchler, N.E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5136-41 (2003).
31. Bintu, L. *et al.* Transcriptional regulation by the numbers: models. *Current opinion in genetics & development* **15**, 116-24 (2005).
32. Zartman, J.J. & Shvartsman, S.Y. Enhancer organization: transistor with a twist or something in a different vein? *Current biology* **17**, R1048-50 (2007).
33. Cameron, R.A. & Davidson, E.H. Flexibility of transcription factor target site position in conserved cis-regulatory modules. *Developmental biology* **336**, 122-35 (2009).
34. Istrail, S. & Davidson, E.H. Logic functions of the genomic cis-regulatory code. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4954-9 (2005).
35. McAdams, H.H., Srinivasan, B. & Arkin, A.P. The evolution of genetic regulatory systems in bacteria. *Nature reviews. Genetics* **5**, 169-78 (2004).
36. Teichmann, S.A. & Babu, M.M. Gene regulatory network growth by duplication. *Nature genetics* **36**, 492-6 (2004).
37. Barolo, S. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *BioEssays* **34**, 135-141 (2011).
38. Carroll, S.B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36 (2008).
39. Hoekstra, H.E. & Coyne, J. a The locus of evolution: evo devo and the genetics of adaptation. *Evolution; international journal of organic evolution* **61**, 995-1016 (2007).
40. Wagner, G.P., Pavlicev, M. & Cheverud, J.M. The road to modularity. *Nature reviews. Genetics* **8**, 921-31 (2007).
41. Davidson, E.H. & Erwin, D.H. Gene regulatory networks and the evolution of animal body plans. *Science* **311**, 796-800 (2006).
42. Litvin, O., Causton, H.C., Chen, B.-J. & Pe'er, D. Modularity and interactions in the genetics of gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 6441-6 (2009).
43. Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. & Teichmann, S. a Structure and evolution of transcriptional regulatory networks. *Current opinion in structural biology* **14**, 283-91 (2004).
44. Pownall, M.E., Gustafsson, M.K. & Emerson, C.P. Myogenic Regulatory Factors and the Specification of Muscle Progenitors in Vertebrate Embryos. *Annual Reviews in Cell and Developmental Biology* **18**, 747-783 (2002).
45. Kulkarni, M.M. & Arnosti, D.N. cis -Regulatory Logic of Short-Range Transcriptional Repression in Drosophila melanogaster. *Molecular and Cellular Biology* **25**, 3411-3420 (2005).
46. Ludwig, M.Z. *et al.* Functional evolution of a cis-regulatory module. *PLoS biology* **3**, e93 (2005).

47. Tolhuis, B., Palstra, R.-jan, Splinter, E., Grosveld, F. & Laat, W.D. Looping and Interaction between Hypersensitive Sites in the Active b -globin Locus. *Molecular Cell* **10**, 1453-1465 (2002).
48. Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes & development* **13**, 2465-77 (1999).
49. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science* **295**, 1306-1311 (2002).
50. Dekker, J. Gene Regulation in the Third Dimension. *Science* **319**, 1793-1794 (2008).
51. Sexton, T., Bantignies, F. & Cavalli, G. Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. *Seminars in cell & developmental biology* **20**, 849-55 (2009).
52. Fiering, S., Whitelaw, E. & Martin, D.I. To be or not to be active: the stochastic nature of enhancer action. *BioEssays : news and reviews in molecular, cellular and developmental biology* **22**, 381-7 (2000).
53. Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E.E.M. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65-70 (2009).
54. Aerts, S., Van Loo, P., Thijs, G., Moreau, Y. & De Moor, B. Computational detection of cis -regulatory modules. *Bioinformatics* **19**, ii5-ii14 (2003).
55. Sharan, R., Ben-Hur, A., Loots, G.G. & Ovcharenko, I. CREME: Cis-Regulatory Module Explorer for the human genome. *Nucleic acids research* **32**, W253-6 (2004).
56. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338-345 (2005).
57. Rebeiz, M., Reeves, N.L. & Posakony, J.W. SCORE : A computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 9888-93 (2002).
58. Yu, X., Lin, J., Zack, D.J. & Qian, J. Identification of tissue-specific cis-regulatory modules based on interactions between transcription factors. *BMC bioinformatics* **8**, 437 (2007).
59. Noto, K. & Craven, M. Learning probabilistic models of cis-regulatory modules that represent logical and spatial aspects. *Bioinformatics* **23**, e156-62 (2007).
60. Zhang, J., Yuan, Z. & Zhou, T. Synchronization and clustering of synthetic genetic networks: A role for cis-regulatory modules. *Physical Review E* **79**, 1-12 (2009).
61. Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nature genetics* **37 Suppl**, S38-45 (2005).
62. Won, K.-J. *et al.* An integrated approach to identifying cis-regulatory modules in the human genome. *PloS one* **4**, e5501 (2009).
63. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature* **451**, 535-40 (2008).
64. Dean, A.K., Harris, S.E., Kalajzic, I. & Ruan, J. A systems biology approach to the identification and analysis of transcriptional regulatory networks in osteocytes. *BMC bioinformatics* **10 Suppl 9**, S5 (2009).
65. Geier, F., Timmer, J. & Fleck, C. Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge. *BMC systems biology* **1**, 11 (2007).
66. Ransick, A. & Davidson, E.H. cis-regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification. *Developmental biology* **297**, 587-602 (2006).
67. Yuh, C.-H., Bolouri, H. & Davidson, E.H. Genomic Cis-Regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene. *Science* **279**, 1896-1902 (1998).
68. Howard, M.L. & Davidson, E.H. cis-Regulatory control circuits in development. *Developmental biology* **271**, 109-18 (2004).
69. Fritz, G., Buchler, N.E., Hwa, T. & Gerland, U. Designing sequential transcription logic: a simple genetic circuit for conditional memory. *Systems and synthetic biology* **1**, 89-98 (2007).

70. Guet, C.C., Elowitz, M.B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science (New York, N.Y.)* **296**, 1466-70 (2002).
71. Levine, M. Transcriptional enhancers in animal development and evolution. *Current biology : CB* **20**, R754-63 (2010).
72. Jiang, C. & Pugh, B.F. Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews. Genetics* **10**, 161-72 (2009).
73. Ay, A. & Arnosti, D.N. Nucleosome positioning: an essential component of the enhancer regulatory code? *Current biology : CB* **20**, R404-6 (2010).
74. Goldhamer, D., Faerman, A., Shani, M. & Emerson Jr, C. Regulatory elements that control the lineage-specific expression of myoD. *Science* **256**, 538-542 (1992).
75. Sucov, H.M., Hough-Evans, B.R., Franks, R.R., Britten, R.J. & Davidson, E.H. A regulatory domain that directs lineage-specific expression of a skeletal matrix protein gene in the sea urchin embryo. *Genes & Development* **2**, 1238-1250 (1988).
76. Kappel, A. *et al.* Role of SCL / Tal-1 , GATA , and Ets transcription factor binding sites for the regulation of Flk-1 expression during murine vascular development. *Blood* **96**, 3078-3085 (2000).
77. Ochiai, H. *et al.* Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes to cells : devoted to molecular & cellular mechanisms* **15**, 875-85 (2010).
78. Fujii, R., Kitaoka, M. & Hayashi, K. Error-prone rolling circle amplification: the simplest random mutagenesis protocol. *Nature protocols* **1**, 2493-7 (2006).
79. Stephens, D.E., Singh, S. & Permaul, K. Error-prone PCR of a fungal xylanase for improvement of its alkaline and thermal stability. *FEMS microbiology letters* **293**, 42-7 (2009).

Chapter 2

2. Traditional and combinatorial investigations into the regulation of *myod*

The regulation of the muscle regulatory factor (MRF) *myod* was investigated to examine the mechanisms underpinning the expression of this tightly regulated factor. This chapter aims to summarise the key findings in the literature about the mechanisms of *myod* regulation in mouse (*Mus muculus*) in a variety of developmental contexts. Information from the literature, summarised in sections 2.1 to 2.4, is combined with data obtained by previous researchers within the group that this research project is based to present a current understanding of *myod* regulation. By developing on this previous work, the position/orientation of the previously identified CRMs can be determined. Furthermore, combinatorial mutant libraries can be developed and tested in order to deduce the activity of specific sites within a CRM.

2.1. Muscle specification *in vivo*

During vertebrate development, structures called somites are formed from the paraxial mesodermal tissue on either side of the neural tube and notochord. Within the somites, four compartments are defined that will become the dermis (dermatome), the vertebrae (schlerotome), the tendons (syndetome) and the skeletal musculature (myotome). The dermatome and myotome are frequently referred to together as the dermomyotome as they are both specified within the dorsal somite. Surgical grafting revealed that specification of the regions within the somite was determined by exogenous signals from 'organisers' in neighbouring tissues[1,2].

2. Traditional and combinatorial investigations into the regulation of *myod*

Signals from the adjacent structures in the developing embryo result in the specification of the regions of the somite[3]. Shh signalling from the notochord and floor plate and Wnt signalling from the neural tube stimulate Myf5 expression in the epaxial somite [4]. Bone morphogenetic protein (BMP) signalling from the limb bud inhibits MyoD expression in the lateral portion of the somite and is counteracted by noggin expression originating from the dorsal medial lip of the hypaxial dermamyotome[5]. Together these signals specify a region of the developing somite that expresses either MyoD or Myf5 and will go on to form the myotome[142]. Figure 1.1 shows the interactions of these signalling molecules diagrammatically.

Specification of cells into the myogenic lineage is defined by expression of either MyoD or Myf5. The fact that homozygous *myod* and *myf5* double knockout mice completely lack skeletal muscle illustrates the crucial role of these two genes to the process of myogenesis[7]. Interestingly, homozygous knockouts for either *myod* or *myf5* exhibit largely normal muscle development indicating that these factors are able to rescue muscle development in each other's absence. It should be noted, however, that *myf5* null mice do exhibit a reduced capacity for regeneration in the adult[8]. Myogenin appears to act during the terminal stages of the differentiation cascade as disruption of the myogenin gene prevents proper muscle differentiation *in vivo* whilst MyoD expression is unaffected [9]. Contrastingly, Mrf4 has been shown to be involved in both muscle specification and terminal differentiation [10]. These factors, MyoD, Myf5, myogenin and Mrf4, are collectively known as the muscle regulatory factors (MRFs). The MRFs are members of the basic helix-loop-helix (bHLH) transcription factors[11] and the MRF family of genes is thought to have arisen out of the duplication of a single ancient gene[12].

Other factors are also associated with muscle cell specification, such as Pax3 and Pax7[13,14]. The Pax genes act upstream of the MRFs and appear to act to maintain a specified

but uncommitted population of muscle stem cells, satellite cells, that are discussed in section 2.4.

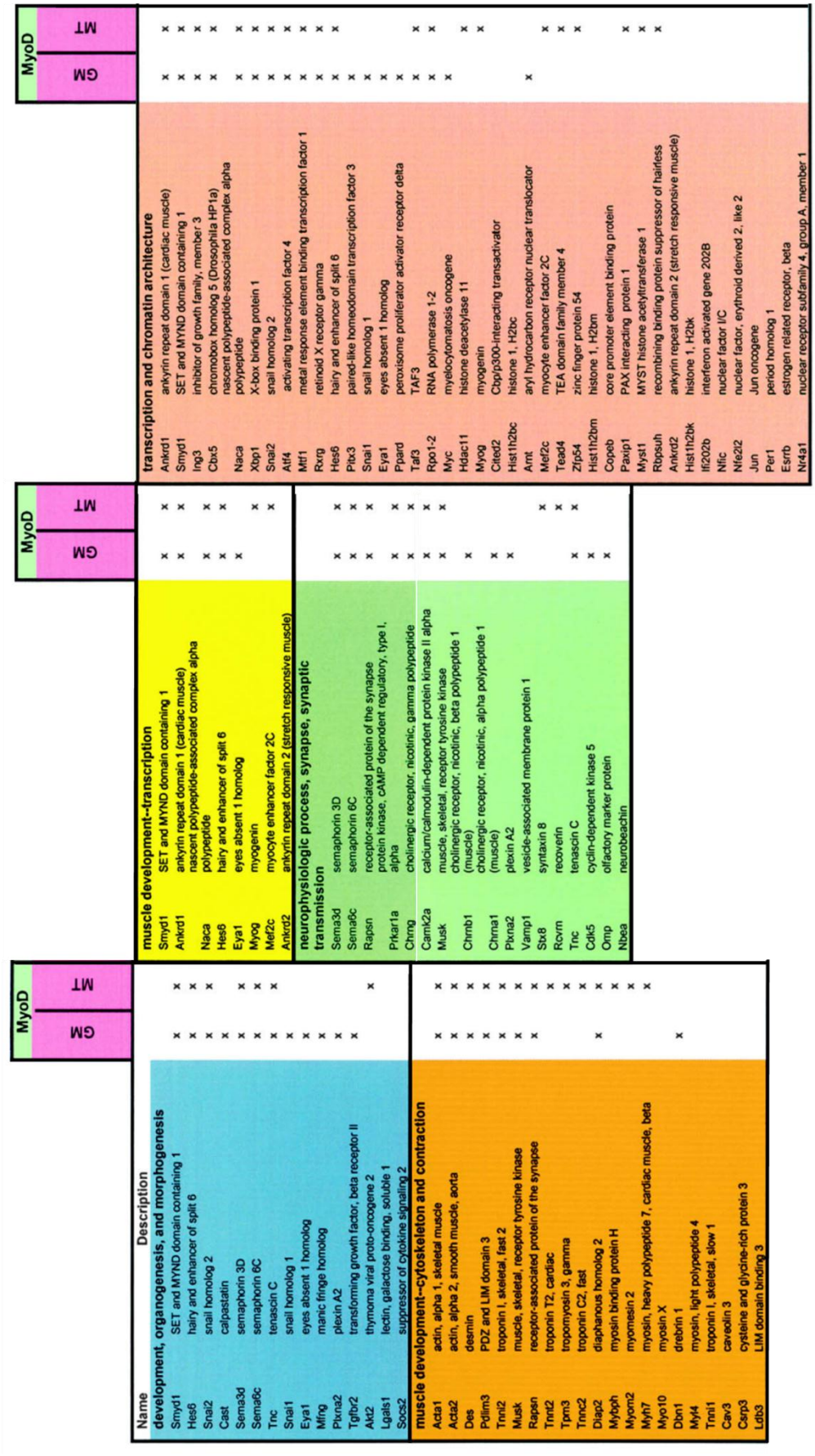
2.2. The MyoD protein

The first evidence for MyoD was found in the fibroblast cell line 10T1/2. Treatment with the demethylating agent 5-aza-cytidine results in the expression of muscle specific genes in these cells[15]. This finding strongly suggests that methylation-dependant gene silencing is responsible for preventing MyoD expression in these cells. Chromatin remodelling is, therefore, most likely necessary for correct expression of the MyoD *in vivo*. Subsequently, MyoD was shown to be able to force a variety of cell types to express muscle specific genes[16]. This evidence indicates that MyoD is a master regulator of cell fate; expression is tightly controlled by gene silencing and is necessary and sufficient for activation of a slew of genes that are associated with the muscle phenotype.

MyoD is a member of the bHLH family of transcription factors[17]. The bHLH structural motif consists of two α-helices linked by a short loop[18]. One end of each of the antiparallel helices associates with the other to form a cross-shape. The basic region at the end of one helix is responsible for binding to the DNA. The rest of the HLH domain is involved in allowing MyoD to form heterodimers with other HLH domain-containing proteins, such as the ubiquitously expressed Ebox proteins. When dimerised with a suitable partner, such as itself or a member of the Ebox family of proteins, the consensus binding site for MyoD is the motif CANNTG, called an E-box. An activation domain on the MyoD protein is then responsible for activating gene expression, in concert with activation or repression domains on the appropriate Ebox protein[19]. The binding of MyoD appears to be cooperative with a binding observed at paired E-boxes or where another site can substitute for the second E-box[20,21].

2.3. Regulation of MyoD

MyoD is known to regulate many downstream genes, a feature that is typical of a master regulatory factor[22]. Microarray analysis of mRNA expression of C2C12 cells (a widely used mouse myoblast model cell line) shows that the expression level of a wide range of genes vary during differentiation[23]. Table 2.1 shows a list representing the range of genes that MyoD is known to regulate [24]. A more comprehensive list, including targets of other MRFs, can be found in Blais *et al.*[24]. Additionally, post-translational regulation of MyoD by ubiquitin-dependant degradation is also involved [25], which is not included within the scope of this study. As well as targeting a range of genes involved in muscle specification, MyoD also targets genes involved in chromatin remodelling. As discussed previously, chromatin remodelling is an important step in the activation and expression of a gene (see section 2.1). By activating factors responsible for chromatin remodelling, such as the histone deacetylases, MyoD is able to indirectly affect gene expression for genes which it does not itself bind to. MyoD itself is expressed in specified and proliferating myoblasts, but is downregulated as differentiation occurs. The closely related member of the MRF family, *myf5*, is able to overcome the absence of MyoD expression in homozygous knockouts for *myod*[26]. Figure 2.1 shows how the expression profile of MyoD and Myf5 are linked.



Name	Description	MyoD GM	MyoD MT
development, organogenesis, and morphogenesis			
Smyd1	SET and MYND domain containing 1	x	x
Hes6	hairy and enhancer of split 6	x	x
Snai2	snail homolog 2	x	x
Cast	calpastatin	x	
Sema3d	semaphorin 3D	x	x
Sema6c	semaphorin 6C	x	x
Tnc	tenascin C	x	x
Snai1	snail homolog 1	x	
Eya1	eyes absent 1 homolog	x	
Mfng	manic fringe homolog	x	
Plxna2	plexin A2	x	
Tgfbr2	transforming growth factor, beta receptor II	x	
Akt2	thymoma viral proto-oncogene 2		x
Lgals1	lectin, galactose binding, soluble 1		
Socs2	suppressor of cytokine signaling 2		
muscle development--cytoskeleton and contraction			
Acta1	actin, alpha 1, skeletal muscle	x	x
Acta2	actin, alpha 2, smooth muscle, aorta	x	x
Des	desmin	x	x
Pdlim3	PDZ and LIM domain 3	x	x
Tnni2	troponin I, skeletal, fast 2	x	x
Musk	muscle, skeletal, receptor tyrosine kinase	x	x
Rapsn	receptor-associated protein of the synapse	x	x
Tnnt2	troponin T2, cardiac		x
Tpm3	tropomyosin 3, gamma		x
Tnnc2	troponin C2, fast		x
Diap2	diaphanous homolog 2	x	x
Mybph	myosin binding protein H		x
Myom2	myomesin 2		x
Myh7	myosin, heavy polypeptide 7, cardiac muscle, beta		x
Myo10	myosin X	x	
Dbn1	drebrin 1		
Myl4	myosin, light polypeptide 4		
Tnni1	troponin I, skeletal, slow 1		
Cav3	caveolin 3		
Csrp3	cysteine and glycine-rich protein 3		
Ldb3	LIM domain binding 3		
muscle development--transcription			
Smyd1	SET and MYND domain containing 1	x	x
Ankrd1	ankyrin repeat domain 1 (cardiac muscle)	x	x
Naca	nascent polypeptide-associated complex alpha polypeptide	x	x
Hes6	hairy and enhancer of split 6	x	x
Eya1	eyes absent 1 homolog	x	
Myog	myogenin		x
Mef2c	myocyte enhancer factor 2C		x
Ankrd2	ankyrin repeat domain 2 (stretch responsive muscle)		
neurophysiologic process, synapse, synaptic transmission			
Sema3d	semaphorin 3D	x	x
Sema6c	semaphorin 6C	x	x
Rapsn	receptor-associated protein of the synapse	x	x
Prkar1a	protein kinase, cAMP dependent regulatory, type I, alpha	x	x
Chrng	cholinergic receptor, nicotinic, gamma polypeptide	x	x
Camk2a	calcium/calmodulin-dependent protein kinase II alpha	x	x
Musk	muscle, skeletal, receptor tyrosine kinase	x	x
Chrnb1	cholinergic receptor, nicotinic, beta polypeptide 1 (muscle)	x	
Chrna1	cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)	x	
Plxna2	plexin A2	x	
Vamp1	vesicle-associated membrane protein 1		
Stx8	syntaxin 8		x
Rcvrn	recoverin		x
Tnc	tenascin C	x	x
Cdk5	cyclin-dependent kinase 5	x	
Omp	olfactory marker protein	x	
Nbea	neurobeachin		
transcription and chromatin architecture			
Ankrd1	ankyrin repeat domain 1 (cardiac muscle)	x	x
Smyd1	SET and MYND domain containing 1	x	x
Ing3	inhibitor of growth family, member 3	x	x
Cbx5	chromobox homolog 5 (Drosophila HP1a)	x	x
Naca	nascent polypeptide-associated complex alpha polypeptide	x	x
Xbp1	X-box binding protein 1	x	x
Snai2	snail homolog 2	x	x
Atf4	activating transcription factor 4	x	x
Mtf1	metal response element binding transcription factor 1	x	x
Rxrg	retinoid X receptor gamma	x	x
Hes6	hairy and enhancer of split 6	x	x
Pitx3	paired-like homeodomain transcription factor 3	x	
Snai1	snail homolog 1	x	
Eya1	eyes absent 1 homolog	x	
Ppard	peroxisome proliferator activator receptor delta	x	
Taf3	TAF3	x	x
Rpo1-2	RNA polymerase 1-2	x	x
Myc	myelocytomatosis oncogene	x	
Hdac11	histone deacetylase 11		x
Myog	myogenin		x
Cited2	Cbp/p300-interacting transactivator		
Hist1h2bc	histone 1, H2bc	x	
Amt	aryl hydrocarbon receptor nuclear translocator		
Mef2c	myocyte enhancer factor 2C		x
Tead4	TEA domain family member 4		x
Zfp54	zinc finger protein 54		x
Hist1h2bm	histone 1, H2bm		
Copeb	core promoter element binding protein		
Paxip1	PAX interacting protein 1		x
Myst1	MYST histone acetyltransferase 1		x
Rbpsuh	recombining binding protein suppressor of hairless		x
Ankrd2	ankyrin repeat domain 2 (stretch responsive muscle)		
Hist1h2bk	histone 1, H2bk		
Ifi202b	interferon activated gene 202B		
Nfic	nuclear factor I/C		
Nfe2l2	nuclear factor, erythroid derived 2, like 2		
Jun	Jun oncogene		
Per1	period homolog 1		
Esrrb	estrogen related receptor, beta		
Nr4a1	nuclear receptor subfamily 4, group A, member 1		

Table 2.1: List of downstream targets for MyoD in both growing cells (GM) and myotubes (MT). Table adapted from [24].

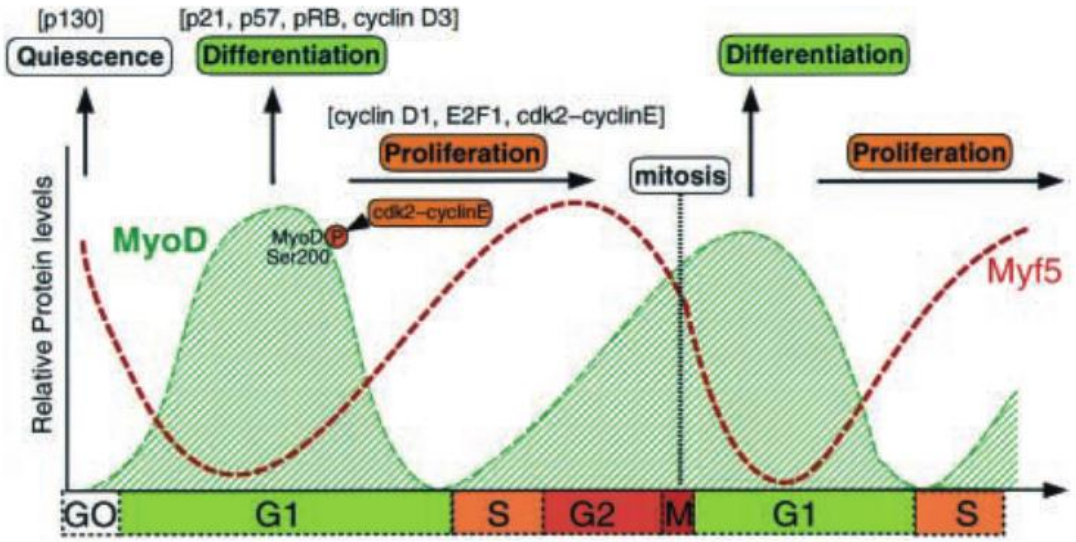


Figure 2.1: Postulated expression dynamics of MyoD and Myf5 in myoblasts through the cell cycle. Signalling from the factors shown in brackets has been implicated in the decision between differentiation, proliferation or quiescence. Image taken from Kitzmann *et al.*[25].

Signalling from diverse pathways, such as p38[27], Notch[28,29], TNFα[30], Shh[31], Insulin[32], TGFβ[33] and Wnt[34], have been implicated in the regulation of *myod*. A more complete list of pathways and implicated factors, with appropriate references can be seen in table 2.1. Two regions are known, the core enhancer and distal regulatory regions (CER and DRR, respectively)[35–37], that are able to regulate MyoD expression. The mechanism of how these signalling pathways integrate at the *myod* promoter is currently unknown. Deletion of the CER or DRR in mice results in altered expression of a *lacZ* reporter gene[38] and replacement of the *myod* promoter with a heterologous promoter results in an expression profile similar to native expression[39]. The CER is responsible for the correct timing of MyoD expression in the limb buds and branchial arches[40]. Targeted mutagenesis of the DRR shows that the DRR is not necessary for myogenic differentiation[41]. The DRR is important for integrating the effect of innervation of adult muscle on MyoD expression but not sufficient to describe all[42]. Either directly or indirectly, the information from these diverse signalling pathways must be integrated at the *myod* locus, thus determining MyoD expression.

Signalling pathway	Activator/repressor	Candidate effector proteins
EDAR	Activator Repressor	AP1 (c-Jun/Fra2, JunD/Fra2) AP1 (c-Jun/cFos, JunD/cFos)
Wnt	Activator	Wnt6, Wnt7a[43], Pax3[44]
PPAR	Repressor	PPARγ[45,46]
Insulin	Activator	CREB[47]
p38	Activator	E12,E47[27,30,48]
Fas	Activator	TRAIL receptor DR5/FADD[49,50]
TNFα	Repressor	NF-κB[30]
Epo	Repressor	Stat3[51], GATA1[52]
MyoD	Activator	MyoD[53]
Notch	Activator Repressor	Hes6[54] Hes1[55]
AhR	Repressor	ARNT[56]
TGF-β	Repressor	Mef2[33], Smad3[57]
BMP	Repressor	Runx2[58]

Table 2.2: Signalling pathways, their qualitative effect on MyoD expression and candidate downstream effectors. See text in section 2.6 for explanation of acronyms.

The regulation of *myod* is not completely understood[6]. The identification of additional regions that contribute to the regulation of the *myod* promoter activity is likely to elucidate key regulatory mechanisms that occur during the developmental specification of skeletal muscle as well as how the repair of adult muscle tissue is regulated. MyoD is itself a regulatory transcription factor and is able to effect the transcription of many downstream genes which could, by virtue of various gene regulatory networks (GRNs), feedback onto the *myod* promoter.

2.4. Satellite cell specification

During development a population of cells are specified that are responsible for repair and growth of muscle in the adult. This population of normally quiescent adult stem cells,

satellite cells, reside between the sarcolemma and the endomysium of muscle fibres and are necessitated by the fact that differentiated myotubes are unable to proliferate. Satellite cells are defined by being positive for Pax7 expression[59–61] and proliferate slowly, maintaining their own population by asymmetric division[62]. The C2C12 cell line, which is a widely used model for adult stem cells, is used in this project for the *myod* regulation studies.

Upon injury or exercise-associated damage to the muscle, the satellite cells are activated by a variety of factors[63]. Activated satellite cells begin to express Myf5, coupled with a down regulation of Notch, and rapidly divide as cycling myoblasts before differentiating into muscle cells, which then fuse to form replication incompetent myotubes[4]. Interestingly, there is some evidence that Pax7 positive satellite cells express osteoblast-specific markers before terminal myogenic differentiation, suggesting a role in osteogenesis[64]. Finally, differentiation occurs when the proliferating myoblasts fuse to form nascent myotubes, which is associated with a down regulation of Notch signalling[29,65]. During differentiation, a cascade of MRF activation, reminiscent of the cascade during embryonic development is seen, where My5/MyoD expression gives way to myogenin expression.

The differentiation of satellite cells in adult muscle is similar to the differentiation of specified cells in somite in the developing embryo. The regulatory interactions that control the expression of MyoD in the context of differentiating satellite cell are likely to be similar to those that occur in the expression of MyoD during development. Thus, investigation of the differentiation of satellite cells is likely to shed light on the process of myogenesis in the developing embryo.

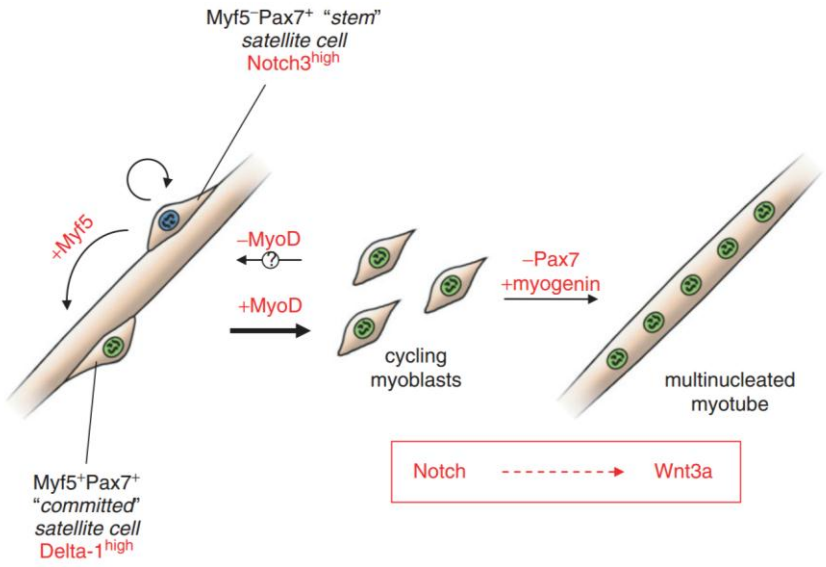


Figure 2.2: Regulation of satellite cell activation, proliferation and differentiation. Quiescent satellite cells, shown by blue nuclei, are activated, shown by green nuclei, and proliferate before finally differentiating and fusing. Important factors are shown in red. Image taken from [4].

2.5. Previous work

The master muscle regulatory gene *myod* is known to be regulated by a number of previously identified CRMs, the DRR and CER, mentioned above. In addition to the published literature, work has been undertaken by previous researchers working on similar projects within the group that this project was based. This previous work identified further putative CRMs that regulate the activity of the *myod* promoter in plasmid constructs. This section of the thesis seeks to summarise this work and bring in additional insight by looking at the sites found within their CRMs and their potential relevance to the regulation of *myod* in a development setting. By the end of the next section, the selection of sites within a specific CRM is described and justified using information from previous expression (section 2.5.1) and bioinformatics studies (section 2.5.2), ChIP analysis (section 2.5.3) and literature review. The generation of the combinatorial mutant library containing sequences that possess each possible combination of the mutated sites is then described in the rest of this project.

2.5.1. Previous identification of cis-regulatory modules of myod

A previous Ph.D student, H. Crutzen, identified several putative CRMs in the region upstream of the *myod* promoter (PRR) transcription start site (TSS). A bioinformatics approach was taken to find regions of conservation upstream of the *myod* promoter by comparing 100 bp sliding window sections across several vertebrate species (work by Dr. Sascha Ott). Similar searching methods followed by functional analysis has been successfully used to find enhancers in several genes[66–68].

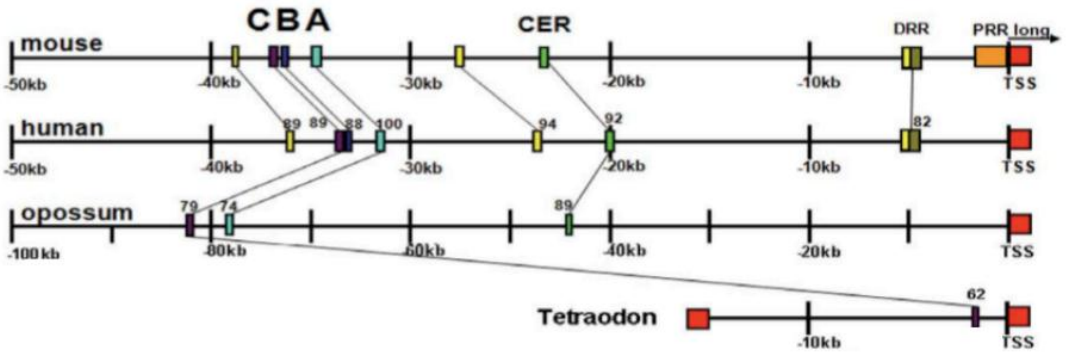


Figure 2.3: Diagram of regions of homology found upstream of the myod promoter (PRR) and transcription start site (TSS) in several species. DRR, CER, A, B and C are noted. Numbers next to homologous regions denote the percentage similarity.

Conservation was tested across diverse vertebrate species for which genome sequences were then available; Humans, mouse, opossum and fish (see figure 2.3). Several regions were identified, including the CER and DRR that had been previously identified, thus validating the approach. Interestingly, the promoter itself was not conserved between species, supporting the previous finding that the promoter was not essential for expression of MyoD in mouse[36].

CRMs have been investigated previously by a process called promoter bashing and targeting mutagenesis of specific sites within regions that have been then determined to have an important regulatory effect[69]. Enhanceosome-type CRMs are highly sensitive to the position and orientation of the transcription factor binding sites within them.

Rearrangement of the binding sites, therefore, will obliterate the activity of the enhanceosome. In contrast, the rearrangement of a billboard-type CRM will not significantly alter their affect transcription from the linked gene. These differences are described in further detail in section 1.5.

Dr. Crutzen's work indicated that the mouse versions of these regions were able to regulate the promoter of mouse *myod* gene in transient transfection experiments. A total of 16 constructs were made for these experiments: All possible combinations of CRM A, B, C and the CER. The CRMs in each construct were either present or absent. The distance between each CRM and the promoter and each of the other CRMs is changed in each construct depending on which combination of CRMs precedes it. The order and orientation of each CRM, however, remains unchanged in the constructs. Expression testing of these combinations indicates whether the CRMs are capable of modulating the expression from the *myod* promoter.

Figure 2.4 shows how combinatorial expression data indicates how different combinations of CRMs interact to bring about regulation of the *myod* promoter. This data includes the DRR in the combinations, yielding a total of 32 constructs, and represents the most up to date combinatorial expression data for the CRMs available (data obtained by P. Downton, unpublished).

The normalised green fluorescent protein (GFP) expression data for the different CRM combinatorial constructs shown in figure 2.4 show a wide range of expression levels. Some constructs, including the PRR alone, express at very low levels, which reflects previous findings[36]. In contrast, other constructs, such as A-CER-DRR-PRR, C-B-A-CER-PRR, C-A-CER-DRR-PRR and B-A-CER-DRR-PRR, express at very high levels, up to 3 times higher than the SV40 promoter, which is constitutively active. Furthermore, output of the CRM combinations does not appear to be a sum of the effects of each individual CRM. Thus, the

CRM combinations appear interacting in a more complex manner. In terms of individual CRMs, the DRR appears to be a non-specific enhancer that increased expression in nearly every case it was present, as compared to a construct containing the same CRMs absent the DRR. The combination of A-CER also appeared to have a relatively high expression level expressing highly in nearly every case construct that this pair is present in.

A simple additive model, where the expression of a more than one CRM is equal to the sum of the expression of each CRM alone, is clearly not sufficient to explain the synergies observed in the data presented in figure 2.4. The CRMs appear, therefore, to be exhibiting context sensitivity, where expression level is determined via interaction of the CRMs. Alternative modelling approaches were then used to attempt to determine the overall function of each individual CRM.

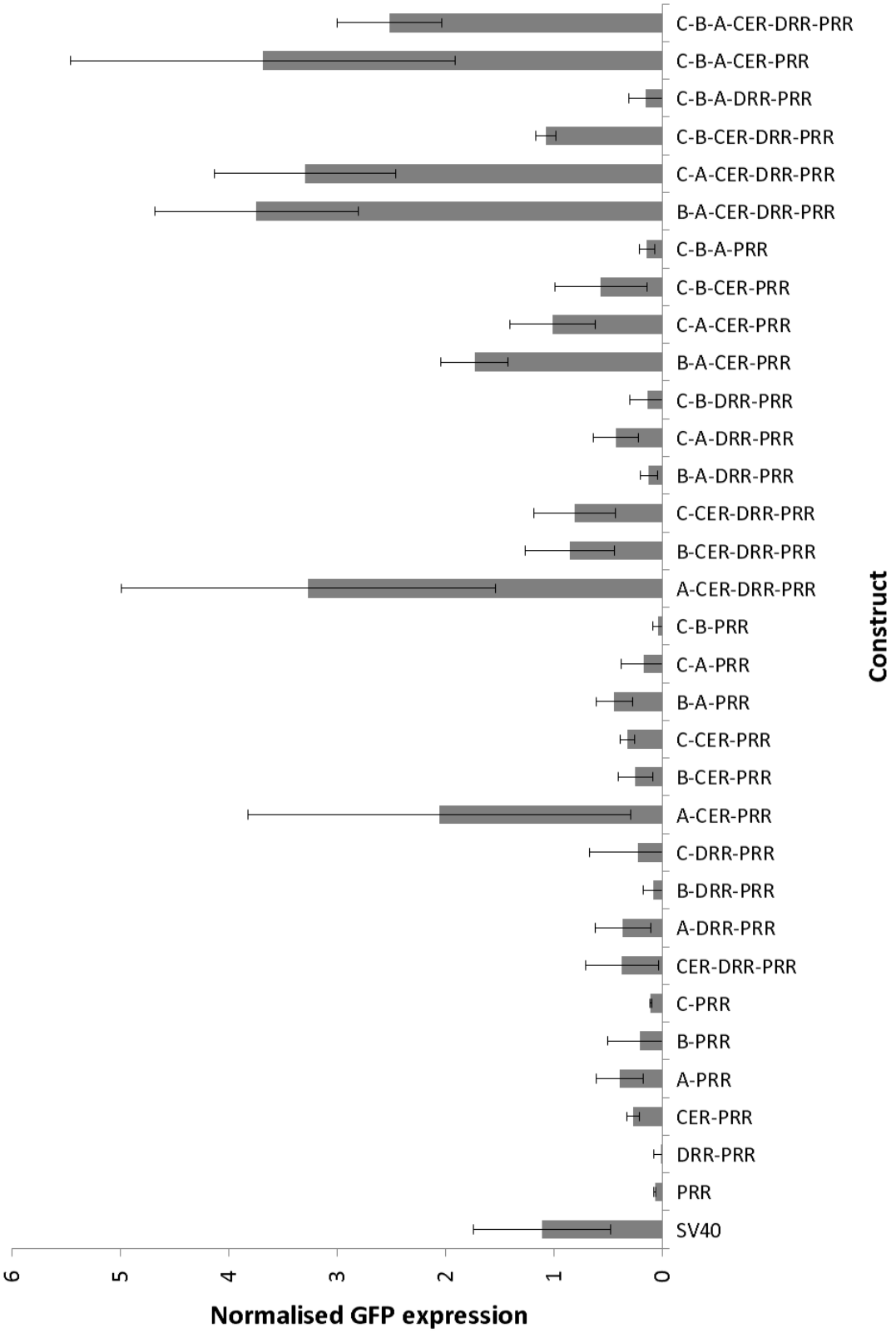


Figure 2.4: Combinatorial plasmid expression for previously identified CRMs in differentiating C2C12 cell cultures. Each bar represents the average of four identical repeats with each plasmid, error bars represent one standard deviation of the data. Data obtained by P. Downton.

2.5.2. Modelling of higher order interactions

Thermodynamic models can be used to model how CRMs interact in order to bring about regulation of a target promoter[70–72]. In the future, a complete physical model of the interactions of CRMs by virtue of transcription factors bound to specific sites in each module is envisaged. A related approach to modelling the complex CRM interactions revealed by H. Crutzen and P. Downton was undertaken by J. Reid, S. Mukajee and M. Nicodemi[73]. In this model, each CRM combination is assigned a 'V-term' that represents the extent to which each combination differs from the expectation based solely upon the additive model. This difference is defined as the effect, in addition to each CRM's additive contribution to overall expression, that is derived from each CRM's interaction with each of the other CRMs in the combination. The model represents the CRMs as a variable, $\boldsymbol{n}_k$ where $k \in$ {DRR,CER,A,B,C}, that can be either 1 or 0, representing their presence in a given CRM combination. A furter set of variables, σ_k, is 1 when a given CRM is in a conformation where it is able to interact with the transcriptional machinery and 0 when the CRM is not able to interact. In this notation, exp[ν(**n**;σ) is the statistical weight of the conformation σ of construct **n**. The partition function of the system can be written as;

$$Z(\boldsymbol{n}) = Z_{int}(\boldsymbol{n}) + Z_0(\boldsymbol{n}) \qquad \text{Eqn. 1}$$

Where Z_{int} is the sum of the weights of the states where the CRMs in construct n interacts with the transcriptional machinery and Z_0(**n**) is the remaining states where there is no interaction. Z_{int}(**n**) can be expressed as;

$$Z_{int}(\boldsymbol{n}) = \sum_{\sigma} \exp\{\nu(\boldsymbol{n}; \sigma)\} \qquad \text{Eqn. 2}$$

ν(**n**;σ) represents the relative free energy of the CRMs of construct **n** folded in the state σ and the interaction with the ranscriptional machinery. ν(**n**;σ) is the summation of the free energies of the CRMs interacting with the trancriptional machinery individually or as

ensembles with any number of the CRMs in construct **n**. Parameter fitting was performed by simulated annealing[74].

A high V-term, therefore, indicates that, whilst the individual CRMs themselves might cause a lower level of expression, the combination of the CRMs result in a higher level of expression, which is presumed therefore to be a result of their interactions. Conversely, a negative V-term indicates that the interactions between the CRMs is repressive of the individual CRMs' enhancing effect.

The CRM interaction terms in the model are fitted to the experimental data and the resulting interaction perameters are then used to develop the predicted expression data. The closeness of the fit between the experimental and predicted model data indicates the acccuracy of the reproduction. To reflect the real *in vivo* situation, this type of modelling requires that the system is complete and all the interacting CRMs are accounted for in the model. The CRM combinations with the most positive or negative V-terms are the most likely to have the strongest interactions, or the expression levels which are most likely to be changed by preventing or breaking the interactions. These combinations were targetted for combinatorial study as they exhibited the strongest interaction dependant effects.

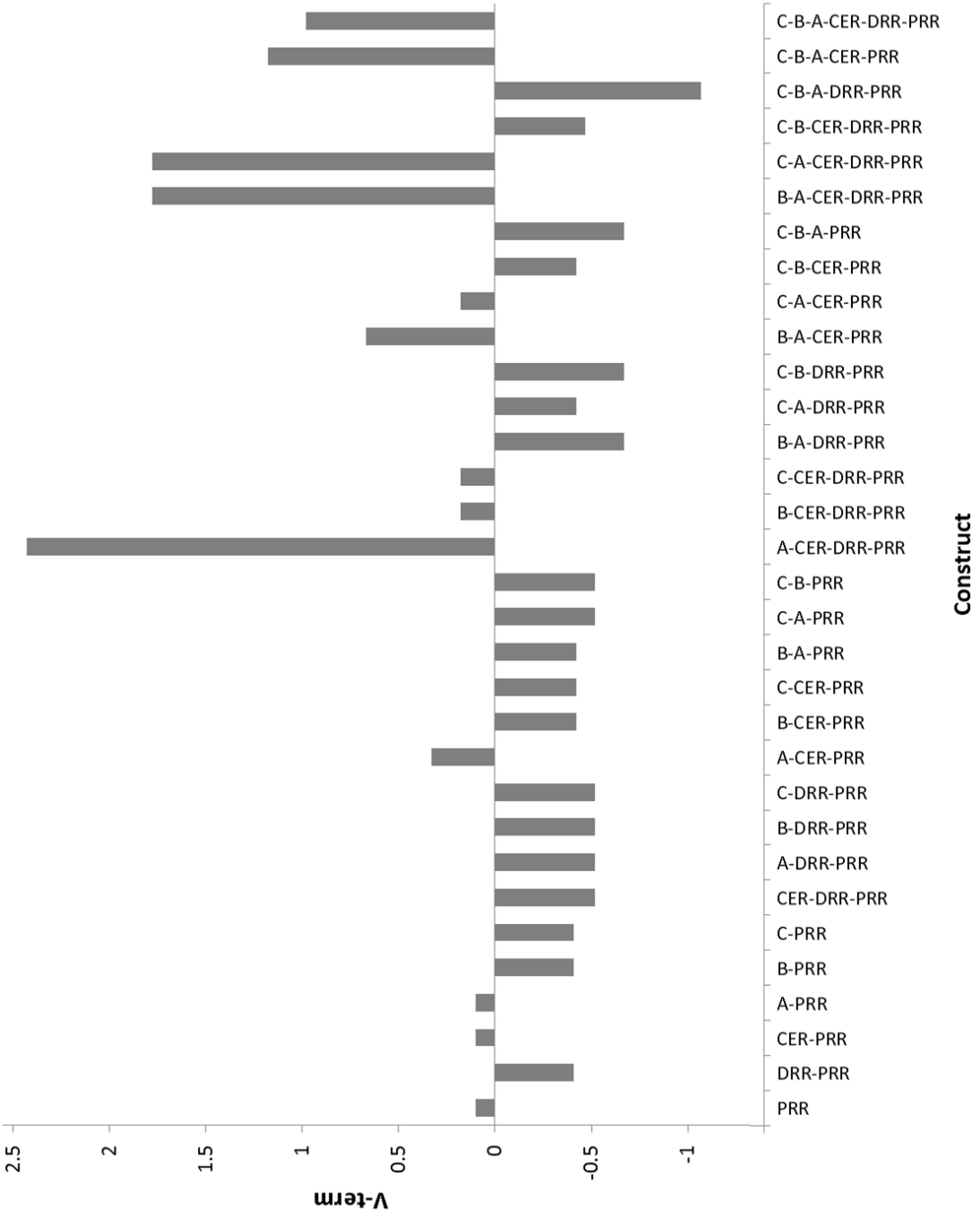


Figure 2.5: V-terms obtained from models of the obtained expression data. Each V-term represents the extent to which a given construct differs from the expected expression. Expected expression is based upon an additive model of CRM action. Data was obtained by P. Downton. V-terms were obtained by J. Reid.

The V-terms derived for each of the CRM combinations is shown in figure 2.5. The majority of interactions are weakly negative (ie these constructs expressed less than was expected), but some are stongly positive (ie these constructs express much more strongly than was

expected). The systems chosen for position/orientation effect studies, B-A-PRR, C-B-PRR and C-B-A-PRR all have negative V-terms and low level expression level except for C-B-A-PRR which has a high expression level. The system chosen for the CRM-B modification study was C-B-A-CER-PRR as this system has an overall high expression level and a positive V-term. A high expression level is desired as it is resonable to suspect that, in the background of a positive V-term, that modifications to the system that might prevent the synergistic interactions between CRMs would result in a measureable drop in expression level.

2.5.3. Chromatin immunoprecipitation of factors on CRMs

In addition to relative expression data for the different CRM combinations, chromatin immunoprecipitation (ChIP) experiments were performed to inform whether specific sites would be prioritised for investigation (data obtained by Dr. K. Vance). ChIP involves the following steps: Genomic DNA from cells is sheared by sonication. Antibodies to specific factors are used to pull those factors and any DNA that they are bound to out of solution. Lastly, specific primers are used to determine whether a specific DNA sequence is present in the pulled down DNA. Hence, ChIP can be used to determine whether a specific factor binds directly to a given DNA sequence *in vivo*.

Investigation of the acetylation and methylation state of histone proteins in the region of the CRMs before and after differentiation indicated that the chromatin state of each changes during this process (K. Vance, unpublished work). This observation is consistent with the CRMs participating in the regulation of the *myod* gene, which exhibits dramatic changes in expression during this period.

Gene name	PRR1	PRR2	CER	A	B/C
MyoD		✔	✔	✔	
Myf5			✔		
Mrf4					
Myogenin		✔	✔	✔	
Jun	✔		✔	✔	✔
cFos		✔	✔		
Ets1		✔			
Pbx1&Meis1					✔
VDR					

Gene name	PRR1	PRR2	CER	A	B/C
NF-κB		✔		✔	
Runx2		✔			✔
Smad3					✔
GR		✔			✔
Foxo1a	✔		✔	✔	
Foxo3a		✔			
Lef1					
CREB					
Bach1					

Table 2.3: Summary of results from ChIP experiments. Tick marks represent a positive result indicating binding of the factor to the indicated regulatory region. Data obtained by K. Vance and P. Downton.

2.6. Prioritisation of binding sites within CRMs

To investigate the mechanisms of regulation of the previously identified CRMs, a combinatorial gene ablation approach was employed (link to chapter 1 and the benefits of such approaches). Figure 2.6 shows a prioritised list of sites within the CER and CRM-A, -B and -C developed using several sources of information. The various sources of information and the literature evidence for factors that bind to specific sites and their contribution to *myod* regulation is now discussed.

The integration of signalling inputs that regulate *myod* begins with the cross talk of different signalling pathways as their signalling cascades converge on common elements. Knowledge of these signalling pathways, particularly of their downstream effectors, is useful in the prioritisation of binding sites for further study. Figure 2.6 represents a set of prioritised sites developed using multiple sources of information: Hits using the BiFa tool (which searches for sequences that match transcription factor binding sites from the TRANSFAC database[75]), microarray data (H. Crutzen), phylogenetic conservation of site

relative order (P. Downton), ChIP experiments (K. Vance and P. Downton), prior knowledge of downstream effectors of signalling pathways and, in the case of the CER only, previous DNase footprinting[39].

Sites were also assessed according to whether the factors bound were capable of forming bridges between CRMs, the relative score of the binding site hit (how well the identified sequence matches the consensus sequence), whether they were a downstream target (and therefore possible autoregulatory target for MyoD) and whether they were known to induce DNA bending. The potential for a transcription factor to form bridges and cause bending in the DNA is important as these are essential features of CRM interaction. Removing or otherwise interfering with these factors could affect CRM function even though the factors affected do not directly interact with the transcriptional machinery.

2.6.1. Summary of contributing factors

The factors that were taken into account during the prioritisation of the binding sites on the CRMs are summarised in tables 2.4 to 2.7. These factors include: binding site location, as defined by the start and stop position of the site. The category, such as activator-repressor (AR), bridge or competitor. Whether the site binds a protein that is itself a downstream target of MyoD. The pathway by which the factors controlling the factors that bind to the site are controlled by. Whether previous ChIP experiments have shown that a factor is or is not bound to the CRM. Whether the site appears to be or not to be phylogenetically conserved. The extent to which factors bound to the site could induce DNA bending important in the formation of regulatory structures composed of several CRMs and their binding proteins. The presence of hypersensitivity or footprinting sites in DNase assays is also indicated for the CER only.

Site name	Binding proteins	Location		Category	Downstream target	Pathway	Chip		Phylogenetic	Architectural	DNase footprint
TBP	TBP	70	78	AR		Wnt				90-120°	
Ets	SPNB2, ETS1	76	88			Wnt					
STAT	STAT6, STAT1, STAT3, STAT5A, SOCS6, STAT5B.	82	101	Bridge		Epo					Hypersensitive
TBP	TBP	85	92	AR/ competitor		Wnt			✔	90-120°	
Ebox	TCFE2A, E12, TCFE2A	140	156						✔		
MyoD	MYOD	142	154	AR	✔		✔	MyoD			
Ikaros	IKAROS	168	177	AR					✔		Footprint
BACH	BACH1	192	207				X	BACH	✔		Footprint
AP1	JUNB, JUN, C-FOS	193	206	Bridge	✔	EDAR	✔	Jun/ cFos	✔	0-40°	Footprint
MAF	p45, BACH2, MAF, MAFK, NRF2, BACH1, MAFG, MAFB	165	206				X	BACH	✔		Footprint
Ebox	TCFE2A, E12, TCFE2A, MYOD	205	219		✔	p38	✔	MyoD			Footprint
HIF1a	ARNT, HIF1A	239	251		✔	AhR			✔		
MAF	P45, BACH2, MAF, MAFK, NRF2, BACH1, MAFG, MAFB	250	261				X	BACH	✔		
AP1	JUNB, JUN, C-FOS	250	263	Bridge	✔	EDAR	✔	Jun/ cFos	✔	0-40°	Hypersensitive
MyoD	TCFE2A, E12, TCFE2A, MYOD	252	270	AR/ competitor	✔	p38			✔		
Ets	SPNB2, ETS1	258	270	AR		Wnt	X	Ets			
Gata	GATA4	275	287	AR							Hypersensitive
STAT	STAT5A, STAT1, STAT3, STAT6, SOCS6, STAT5B	316	229	Bridge		Epo					
Ets	SPNB2, EPHRIN-A2, GABPB1, GABPA, EPHB1, ETS1	319	327	AR/ competitor		Wnt					
Ets	SPNB2, EPHRIN-A2, GABPB1, GABPA, EPHB1, ETS1	382	390	AR		Wnt					
Pax6	PAX6	391	405	Bridge/AR		p38			✔		

Table 2.4: Factors affecting the prioritisation of binding sites in the CER. DNase footprinting data obtained by Goldhamer et al.[76].

Site name	Binding proteins	Location		Category	Downstream target	Pathway	Chip		Phylogenetic	Architectural
E2F	E2F1	10	16	AR		TGFβ				25°
Ets	ETS1	18	34				X	Ets1		
GABP	GABPB1, GABPA	19	31							
VDR	VDR	50	65				X	VDR		
E2F	E2F1	64	70	AR		TGFβ				25°
AR	AR	70	79	AR						
MyoD	TCFE2A, E12, TCFE2A, MYOD	93	111	AR	✔	p38	✔	MyoD	✔	
GR	NR3C1	112	127				X	GR		
PITX2	PITX2	128	139	AR						
KROX	EGR1	149	163	Bridge						
Ets	ETS1	177	192				X	Ets1	✔	
GR	NR3C1	181	208				X	GR	✔	
AR	AR	187	202	AR					✔	
CREB	ATF4, ATF4, ATF2, CREB1	204	210	Bridge/AR	✔	p38			✔	
Ets	ETS1	220	235				X	Ets1	✔	
Foxo1	FKHR1	221	231	AR/ Competitor		Insulin	✔	Foxo1	✔	
KROX	EGR1	229	243	Bridge					✔	
SP1	SP3	232	245	AR/ Competitor					✔	
E2F	E2F1	240	246	AR/ Competitor		TGFβ			✔	25°
AP1	JUNB, JUN, C-FOS	241	253	Bridge	✔	EDAR	✔ X	Jun cFos	✔ ✔	0-40°
MAF	P45, BACH2, MAF, MAFK, NRF2, BACH1, MAFG, MAFB	243	254				X	BACH		
ER	ER	245	264	AR/Bridge						55°
DR4	RAR-α, MRAR-γ, LXR-α	256	273	AR		PPAR				
TFIII	TFII-I	286	295	AR						

Table 2.5: Factors affecting the prioritisation of binding sites in CRM-A.

Site name	Binding proteins	Location		Category	Downstream target	Pathway	Chip		Phylogenetic	Architectural
HAPF1	CEBPB	1	7	AR						
NF-κB	P50, SHC, CD40, GAB2, P65, DCTN2, P50B, VCP	4	20		✔	TNF-α				
NF-κB	P50, SHC, CD40, GAB2, P65, DCTN2, P50B, VCP	25	41		✔	TNF-α				
AR	AR	50	59	AR/ Competitor						
PXR	LXR-α	54	66	AR/ Competitor					✔	
AP1	JUNB, JUN, C-FOS	58	66	Bridge	✔	EDAR	✔ X	Jun cFos	✔	0-40°
SRF	SRF	62	80	AR/ Competitor		Fas				80°
Ets	ETS1	67	81				X	Ets1	✔	
NFY	NFYA, CEBPA, CBFA1, RUNX2	103	116	Bridge/AR		Notch	✔	Runx2	✔	
Ets	SPNB2, EPHRIN-A2, GABPB1, GABPA, EPHB1, ETS1	115	123	AR/ Competitor		Wnt	X	Ets1	✔	
Ebox	TCFE2A, MYOD, MAX, MAD4, N-MYC, MAD3, TCFE2A, TCF12, ARC, TCF4, MITF, C-MYC, MAD1, E12	146	156	Bridge/AR	✔	p38	X	MyoD	✔	80°
Foxo	FKHR1	149	159	AR/ Competitor		Insulin	X	Foxo1		
GATA	GATA1, GATA6, GATA3, GATA4	153	160	AR/ Competitor					✔	50°
Ebox	TCFE2A, E1,2 TCFE2A	160	176				X	MyoD	✔	
MyoD	MYOD	163	173	AR	✔				✔	
GATA	GATA1, GATA6, GATA3, GATA4	179	186	AR					✔	50°
TCF	TCF4, TCF7L2	190	198	AR		Wnt				
LEF1	TCFE2A, E12, LEF1, TCF1	191	197	AR		Wnt	X	Lef1		117-130°

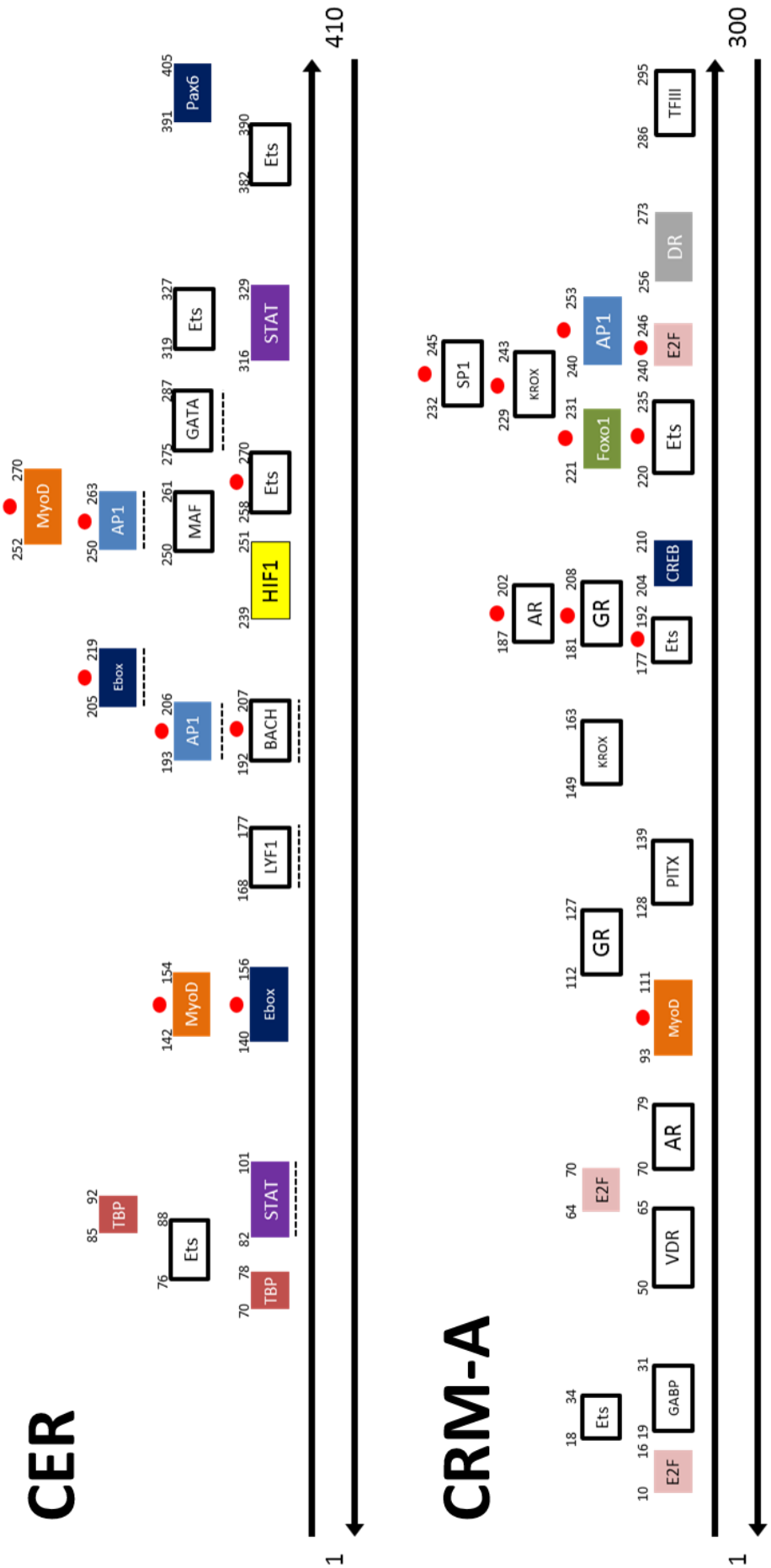
Table 2.6: Factors affecting the prioritisation of binding sites in CRM-B.

Site name	Binding proteins	Location		Category	Downstream target	Pathway	Chip		Phylogenetic	Architectural
ER	ER, ESR2	73	84	AR						55°
CEBP	CEBPB	85	99	AR						
SREBP	SREBP1C, SREBF2	129	141						✔	
Meis1	MEIS1	131	143	AR/ competitor			✔	Meis1	✔	
AP1	JUN, C-FOS	132	143	Bridge	✔	EDAR	✔	Jun	✔	0-40°
							X	cFos		
DR	VDR	150	171	AR		PPAR	X	VDR	✔	
KROX	EGR1	184	198	Bridge					✔	
Hand1	TCFE2A, E12, TH1, TCFE2A	186	202	AR					✔	
NFAT	NFATC2, NFATC3, NFATC4	192	204	AR		PPAR				
HAPF1	CEBPB	194	201	AR						
HMGIY	HMGIY	196	203	AR					✔	
AHR	AHR	211	228	AR/ competitor		AhR			✔	
Ebox	TCFE2A, MYOD, MAX, MAD4, N-MYC, MAD3, TCFE2A, TCF12, ARC, TCF4, MITF, C-MYC, MAD1, E12	216	226	Bridge	✔	p38	X	MyoD		80°
T3R	RAR-α, MRAR-γ, TRANSDUCIN	230	239	AR/ competitor		PPAR			✔	
PBX	PDX-1, PBX1, PKNOX1, HOXB7	238	250	Bridge/ AR			✔	PBX1		

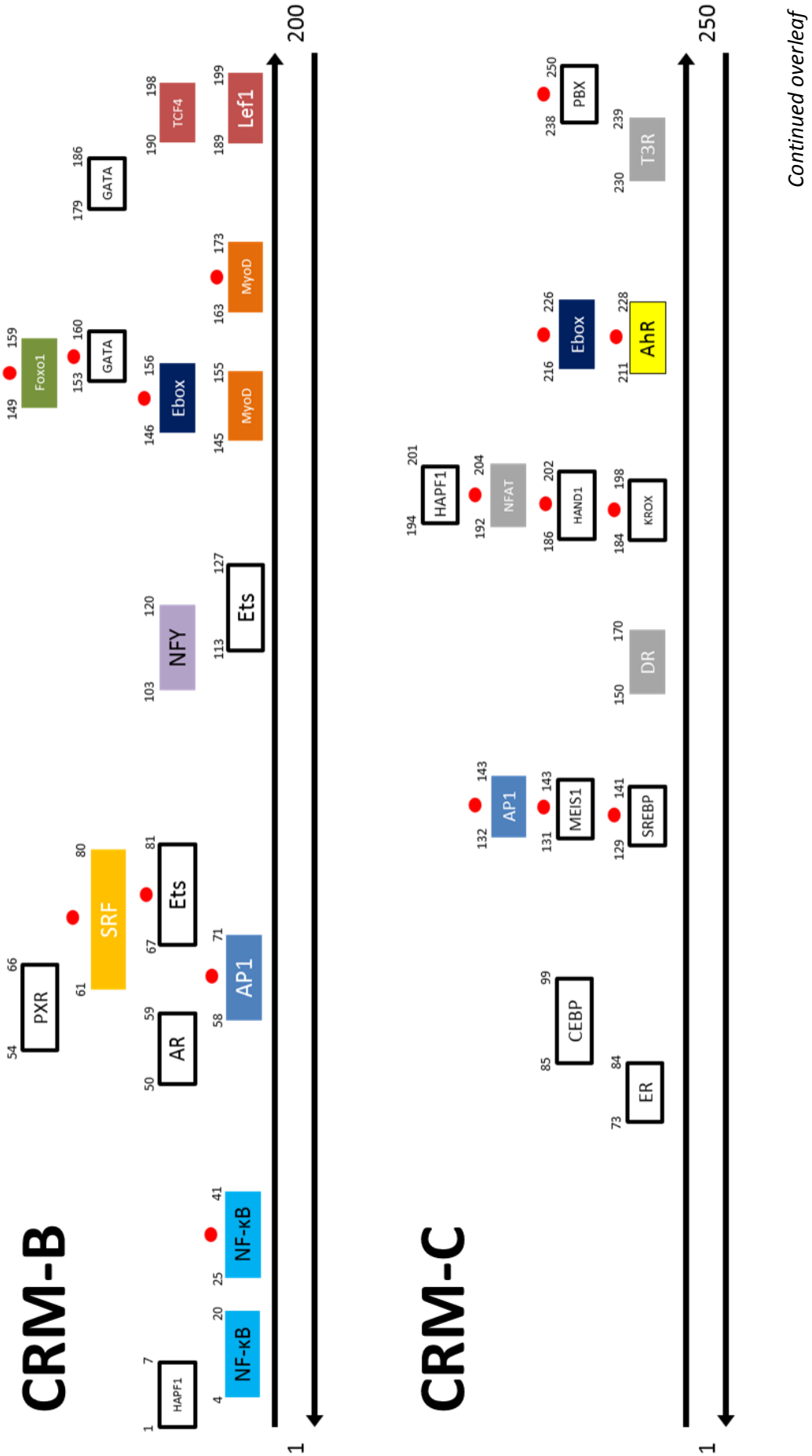
Table 2.7: Factors affecting the prioritisation of binding sites in CRM-C.

2.6.2. CRM binding site maps

The cooperative and competitive interactions between transcription factors with neighbouring and overlapping binding sites can be more easily seen in diagrammatic form. Figure 2.6 shows scale diagrams of the binding sites that appear in tables 2.4 to 2.7. In this figure, the binding sites are coloured according to the signalling pathways that are known to regulate the transcription factors that are known to bind to each site.



Continued overleaf



Continued overleaf

Figure 2.6: Schematic of prioritised transcription factor binding sites found within the CER (top), A (2nd top), B (2nd bottom) and C (bottom). Red dots correspond to phylogenetic conservation (P. Downton). Dotted lines indicate either footprinting or DNase hypersensitivity was found at these positions by Goldhamer *et al.*[39]. Numbers indicate the start and end of each site, which are colour coded according to the pathways index. Note the different sequence lengths for each CRM: The lengths of each binding site are approximately to scale relative to the length of each CRM. NB Vertical position of the site relative to the CRM is for illustrative purposes only.

Clusters of binding sites become apparent when viewing the binding site map presented in figure 2.6 that are not so readily in table form (see tables 2.4 to 2.7). Multiple close or overlapping binding sites provide the basis on which competitive and cooperative binding can work. If possible, sites within the clusters should be targeted for individual mutation.

Interesting patterns emerge through examining the information in tables 2.4 to 2.7 and in figure 2.6. The presence of common binding sites on many or all CRMs, such as AP1 and Eboxes and Ets sites, and the limitation of other sites to specific CRMs, such as NF-κB on CRM-B and PBX on CRM C, suggests that these sites might be involved in facilitating the interactions that result in the regulatory output of these modules. The information from the preceding sections is combined with information from the literature in the following section to discuss the relative important of most of the sites discussed above.

2.6.3. Discussion of factors contributing to the prioritisation of binding sites

Nuclear factor-κB (NF-κB) is a well-known transcription factor that regulates a wide variety of genes[77] and is furthermore known to regulate *myod* as a downstream effector of the tumour necrosis factor-α (TNFα) signalling pathway[30]. A pair of sites specific for NF-κB was found on CRM-B and the fact that no other sites were observed on these CRMs was particularly interesting as this made it likely that manipulation of NF-κB by specific

inhibition or knockdown would influence the system through these sites. To further validate the interest in NF-κB, presence of the factor on any of the CRMs was investigated by ChIP. Interestingly, the ChIP experiment indicated that NF-κB does not bind to the CRM-B/-C pair, but does bind in CRM-A and to part of the PRR. This result does not necessarily mean, however, that the two identified sites do not participate in NF-κB binding at some other point of differentiation other than the 20 hour time point analysed. This evidence was sufficient to conclude that, since the mutagenesis experiments were to be performed at the same time point as previous experiments for consistency, the NF-κB sites would not be mutated in this project.

Serum response factor (SRF) has been implicated to be involved with DRR activity[78], in conjunction with Mef2[79]. A SRF site is found in CRM-B where is occludes an Ets site and partially overlaps with an activator protein (AP1) and a PXR site. As mentioned previously (see section 1.2), competitive interactions are likely to be involved in logical computation on CRMs. Interestingly, the SRF site is adjacent to an androgen receptor (AR) site. Coexpression of AR and SRF in C2C12 cells is capable of coactivating the skeletal α-actin promoter by interacting with SRF[80]. It is possible that a complex containing SRF and AR could compete with the binding site for AP1. In addition to their sites on CRM-B, further sites for both AR and AP1 are present on other CRMs: One additional AR site is present on CRM-A and AP1 sites are present on all other identified CRMs.

The AP1 heterodimer, for instance, is composed of members of the c-Fos and c-Jun protein families and the composition of the heterodimer is altered by multiple signalling inputs[81]. Furthermore, the composition of the AP1 heterodimer has been shown to be relevant to the regulation of *myod*[82,83]. Two potential subunits of the AP1 dimer were analysed by ChIP, Jun and c-Fos. Jun was found to be bound to the CER, CRM-A and CRM-B/-C as well as the PRR. c-Fos, by contrast, was found to be bound to the CER and the PRR only. The AP1 binding sites could be occupied by AP1 heterodimers composed of other

factors, however. AP1 also appear to bend DNA dependant on the composition of the heterodimer[84] and the sequence of the site[85]. This DNA bending effect could be important in the formation of larger complexes from of multiple CRMs, such as those in enhanceosomes.

Runt-related transcription factor 2 (Runx2) is capable of binding to DNA via nuclear factor-Y (NFY) sites. Runx2 is involved in osteogenic development and is likely to be involved in the suppression of MyoD expression in bone-forming tissues[86]. Furthermore, BMP-2 signalling is able to convert C2C12 myoblasts[87] and primary skeletal myoblasts to the osteoblastic lineage[88]. Runx2 is capable interacting with a variety of factors including AP1[89] and Smad3[90]. The presence of a single NFY site in CRM-B was potentially interesting, as was the fact that this site was partially overlapping with an Ets site. Furthermore, Runx2 was found to be bound to CRM-B/-C as well as the PRR by ChIP. Another site related that binds a factor related to bone development is the vitamin D receptor (VDR)[91] site in CRM-A. $VDR^{-/-}$ mice exhibit smaller muscle fibres and deregulated MRF expression[92] indicating the VDR has an important role in myogenic development as well. VDR was, however, found not to be bound to any of the CRMs or the PRR.

Forkead box subtype O (Foxo) 1 and 3 factors are implicated in the suppression of proliferation and erythrocyte differentiation[93,94], are involved in insulin signalling[95] and are implicated in the regulation of myogenesis[96]. Foxo sites are found on CRM-A and CRM-B and appear to be both phylogenetically conserved and overlap with other factor binding sites in both cases. The foxo site overlaps with the Ets site in CRM-A and the GATA, Ebox and MyoD sites in CRM-B. Two members of the Foxo family were analysed by ChIP: Foxo1a was found to be bound to the CER and CRM-A as well as the PRR. Foxo3a was found to be bound to the PRR only. The GATA-binding factor (GATA) family of proteins is involved with a range of functions including erythroid differentiation[97] and cell growth. Changes in GATA site occupancy have been linked to changes in chromatin looping[98] and is capable of

inducing a bend in DNA[99], both of which are an important requisites for enhanceosome function (see section 1.5). Two GATA sites are present in CRM-B and one on CER. Furthermore, the GATA site in the CER is hypersensitive to DNase[39], indicating that the DNA in this region is decondensed and easily accessible.

Aryl hydrocarbon receptor (AhR) is a bHLH factor (like MRFs) involved with a range of developmental and adaptive response contexts[100,101]. AhR and is thought to be involved in regulation of myogenesis through the Ah receptor nuclear translocator (ARNT) homodimer[56]. The peroxisome proliferator-activated receptor (PPAR) signalling pathway is also involved in adipogenic development[102] and is potentially involved in regulation of *myod* through the PPARγ factor[45].

Myeloid ecotropic viral integration site 1 (Meis1) and Pre-B-cell leukemia transcription factor 1 (PBX1) are two DNA binding proteins that are known to interact with homeobox (HOX) proteins that play an important role in morphogenesis in all animals[103,104]. The PBX and Meis1 sites in CRM-C are potential points of integration of morphogenic signals from these factors early in development. The interaction of PBX1 and Meis1 proteins with Hox genes is associated with the recruitment of coactivating factors that are necessary for chromatin remodelling associated with the activation of gene expression (see chapter 1, gene regulation)[105]. Pbx and Meis1 are also thought to be involved in the regulation of *myod* activity[106–108]. Pbx and Meis1 were found to be bound to CRM-B/-C by ChIP. The presence of both a Pbx and a Meis1 site in CRM-C indicate that this CRM could be responsible for the opening of further CRMs or possibly the *myod* promoter itself. The Meis1 site in CRM-C is phylogenetically conserved and occluded by an AP1 and sterol regulatory element-binding protein (SREBP) binding sites. SREBP is itself implicated in the activation of gene expression via recruitment of chromatin remodelling complexes[109].

The hypoxia inducible factor-1 (HIF1) binding site in the CER is interesting despite the lack of phylogenetic conservation and appearance in the DNase sensitivity experiments, as factors binding this site are involved in hypoxia sensing[110]. HIF1 appears to be involved in the regulation of *myod*[111].

Lymphoid enhancer-binding factor 1 (LEF1) is transcription factor usually associated with T- and B-cell development[112] but has also been associated with regulation of somite myogenesis via an interaction with Pitx2[113]. LEF1 is regulated via a functional interaction with β-catenin[114], which is an integral component of the developmentally important Wnt pathway[115]. Furthermore, LEF1 is capable of inducing a significant bend in the DNA of 117-130° when bound to its specific site[116], which appears to be functionally relevant in the regulation of the T-cell receptor-α (TCRα) promoter[117]. ChIP analysis indicated that LEF1 was not bound to any of the CRMs or the PRR, however. Architectural transcription factors are likely to play important roles in the formation and function of enhanceosome-type structures. As a result, the presence of a LEF1 site close to two GATA sites (also capable of inducing DNA bending) in CRM-B is interesting. Although these sites occur at the end of the identified sequence, it should be noted that there is only a short stretch of sequence (<100 bp) between CRM-B and CRM-C. This site could, therefore, have an important role in any the interactions between CRM-B and CRM-C.

E-twenty six (Ets) sites are present on the CER (4), CRM-A (3) and CRM-B (2). This large family of transcription factors is involved in regulating a variety of functions[118–120]. None of the Ets sites in the CER were identified in DNase footprinting or hypersensitivity assays. Several of the Ets sites were identified as being phylogenetically conserved, as shown in figure 2.6. SRF is known to undergo cooperative binding with Ets family members[120] and both sites are found in CRM-B. Ets1 was analysed by ChIP and found to be bound the PRR, but to none of the CRMs. The Ets family is large, however, and it is possible that another member of the family is responsible for binding to the Ets sites in this context.

Enhancer-box (Ebox) sites are present in the CER, CRM B and CRM C. A large number of factors have the potential to bind to Ebox sites such as E12, E47, c-Myc and MyoD[121]. Eboxes can bind a variety of heterologous transcription factor complexes, the identity of which has implications for enhancer function[122]. The fact that many of the factors that bind to Eboxes are able to form bridges to factors bound on other sites means that this sites might be necessary for correct CRM interaction and therefore function. Interestingly a non-canonical Ebox site has been recently identified in the CER that appears to be involved in regulation of *myod* activity[123]. ChIP analysis indicated that MyoD, which is capable of binding some Ebox sequences, was bound to the CER and CRM-A as well as the PRR. Other potential Ebox binding proteins were not examined by ChIP.

ChIP data (see section 2.5.3) is able to provide strong positive indication of the presence of a transcription factor on a DNA sequence. It should be noted, however, that ChIP will yield a positive result if and only if the antibody used is able to bind to the specific protein. Binding of the antibody to the target protein could be affected or prevented by post translational modification or alternative splicing of the protein. Furthermore, many consensus binding sites can be bound by several members of the same protein family, each with a similar effect. An antibody specific to a single member of that family will not detect the binding of the other members of the same family to the same site.

2.6.4. Selection of binding sites for further analysis

As a result of the initial transient expression studies performed by H. Crutzen (see section 2.5.1) CRM B was selected for further analysis through the generation of a combinatorial mutant library. This was mainly because the presence/absence of CRM-B appears to be correlated with substantial changes in expression level in several constructs, but also because it was the smallest of the three CRMs, and therefore the easiest to assemble. The following sites within CRM-B were selected for further study on the basis of their presence

in C2C12 cells at the differentiation time point selected (20 hours) as determined by microarray, ChIP presence/absence, phylogenetic conservation and evidence in the literature for involvement in muscle regulation (numbers in brackets are the start-stop positions of each site): AP1 (58-71), Ets (67-81), NFY (103-120), Ets (113-127), Ebox (146-156), FOXO1 (149-159), MyoD (163-173), Lef1 (189-199). See figure 2.6 for a map of the sites with CRM-B. Each site was selected for the reasons discussed above in section 2.6.4 and on the basis of the information in table 2.6. In addition to these sites, a bacterial binding site (BBS) was added to the flanking sequence upstream of CRM-B. This sequence was added as a way of driving artificial CRM-CRM interactions in future experiments. The total number of sites was therefore 9, so the library size was 2^9 = 512 sequences long.

Some sites could not be modified without affecting neighbouring sites, such as PXR (54-66), the SRF (61-80) site and GATA (153-160). These sites, whilst potentially interesting for the reasons discussed above, were not included. Some sites, such as the MyoD (145-155) / Ebox (146-156) and TCF4 (190-198) and Lef1 (189-199) rely on most heavily on the same bases, as a result modification of one site inevitably affects the binding to the other.

2.7. Generation of mutant sites

Once a specific site was selected for mutation, changes to the consensus sequence were investigated *in silico* using the BiFa tool. Mutant sequences were required to be at least four consecutive base pairs, due to the anticipated method of separation of the sequences that would have been required had the Gao assembly proven successful. A successful mutation sequence was one that significantly or completely reduced the predicted binding affinity of the specific transcription factor without significantly altering the binding affinity of the adjacent factors.

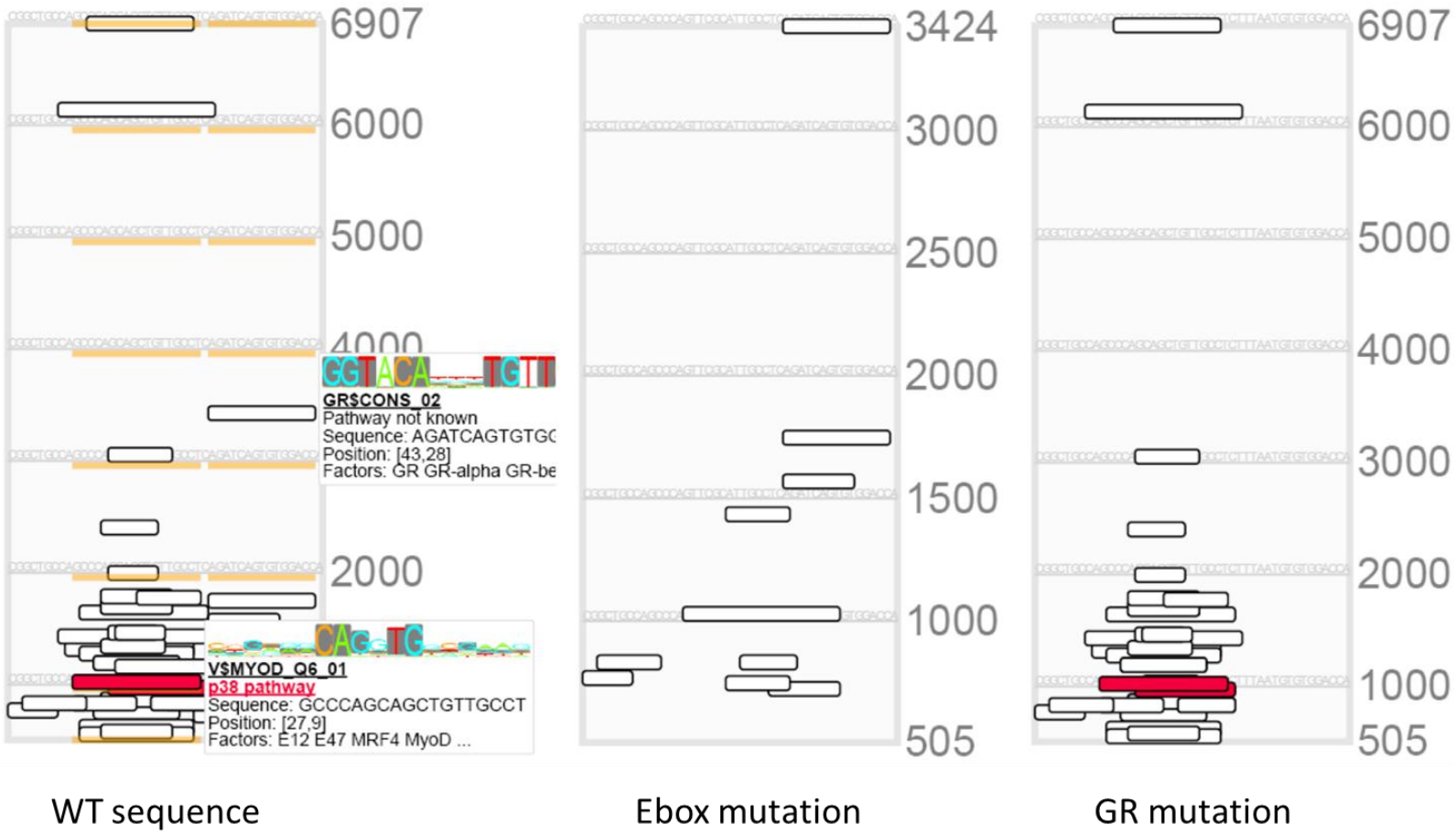


Figure 2.7: Diagram illustrating the use of the BiFa tool to design sequences that significantly reduce the predicted binding affinity of the targeted transcription factor. The left pane shows the wild type sequence with two consensus sequences highlighted. The middle and right hand pane indicate the predicted binding affinity with mutant sequence replacing part of Ebox or GR sites, respectively. Vertical axis indicates strength of predicted binding.

The consensus sequences for several of the targeted transcription factors were adjacent to or overlapped with other consensus sequences. As a result the mutant sequences had to be carefully selected to ensure that only the targeted binding site was affected. Figure 2.7 shows an example using two adjacent predicted binding sites, GR and Ebox. As can be seen in figure 2.7, the replacement of part of the consensus sequence for either site with mutant sequence did not affect the predicted binding of the other as determined by the predicted binding strength score of the relevant sites. This score is determined by the extent, measured in arbitrary units, to which the binding site matches the consensus binding site motif for a given factor.

2.8. Unaddressed issues with CRM investigation

As discussed in section 2.5, experiments on the previously identified CRMs have been limited to combinatorial presence or absence experiments by P. Downton and H. Crutzen, microarray experiments with C2C12 cells by H. Crutzen, ChIP experiments by K. Vance and P. Downton as well as investigations of the CER and DRR in the literature[40]. No experiments have been performed to ascertain whether the position and/or orientation of the CRMs have a functional effect on their ability to regulate the *myod* promoter. These experiments are likely to yield information useful for determining whether these CRMs function through the various mechanisms described in chapter 1 (enhanceosome vs. billboard, see section 1.5).

The identified CRMs contain multiple potentially interesting factors that are capable of undergoing interactions, integrating information from signalling pathways and facilitating gene expression from the *myod* promoter. The potential interaction space in such a multivariate system requires combinatorial analysis to elucidate the mechanisms behind the CRM synergies that have been previously observed. As discussed in section 1.6,

current methods do not allow for the rapid, parallel generation of libraries of combinatorial CRM mutants.

Current techniques such as ChIP-chip have allowed the binding sites for a given protein throughout the genome to be identified and together with ChIP-seq have revealed useful information for the modelling of CRM function. One step gene assembly, rather than step wise mutagenesis represents the most plausible solution to this issue. An assembly of this type requires sequence optimisation algorithms to be developed as well as methods for the assembly of optimised sequences once optimised. An established method such as flow cytometry could then be employed to investigate whether the resulting constructs had an effect on the function of the different CRMs and could elucidate the mechanisms of CRM function.

The work presented here was used to select the binding sites of CRM B that were to be targeted in the combinatorial mutant library. The mutant library was meant to address the fact that investigation of CRM interactions via single pertrubation studies might not assess the full contribution of each site if they are acting in a cooperative or competitive manner. The library that was designed as a result of the work presented in this chapter was assembled by ligative assembly in chapter 8 following optimisation of the assembly sequences as described in chapter 7. The use of several members of the library is then described in chapter 9.

2.9. References

1. Pourquié, O., Coltey, M., Teillet, M. a, Ordahl, C. & Le Douarin, N. M. Control of dorsoventral patterning of somitic derivatives by notochord and floor plate. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 5242–6 (1993).
2. Pourquié, O., Fan, C. M., Coltey, M., Hirsinger, E., Watanabe, Y., Bréant, C., Francis-West, P., Brickell, P., Tessier-Lavigne, M. & Le Douarin, N. M. Lateral and axial signals involved in avian somite patterning: a role for BMP4. *Cell* **84**, 461–71 (1996).
3. Tajbakhsh, S. & Sporle, R. Somite Development : Constructing the Vertebrate Body. *Cell* **92**, 9–16 (1998).
4. Punch, V. G., Jones, A. E. & Rudnicki, M. A. Transcriptional networks that regulate muscle stem cell function. *Systems Biology and Medicine* **1**, 128–140 (2009).
5. Marcelle, C., Stark, M. R. & Bronner-fraser, M. Coordinate actions of BMPs, Wnts, Shh and Noggin mediate patterning of the dorsal somite. *Development* **124**, 3955–3963 (1997).
6. Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132**, 2685–2695 (2005).
7. Rudnicki, M. A., Schnegelsberg, P. N., Stead, R. H., Braun, T., Arnold, H. H. & Jaenisch, R. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351–9 (1993).
8. Gayraud-Morel, B., Chrétien, F., Flamant, P., Gomès, D., Zammit, P. S. & Tajbakhsh, S. A role for the myogenic determination gene Myf5 in adult regenerative myogenesis. *Developmental Biology* **312**, 13 – 28 (2007).
9. Hasty, P., Bradley, A., Morris, J. H., Edmonson, D. G., Venuti, J. M., Olson, E. N. & Klein, W. H. Muscle definciency and neonatal death in mice with targeted mutation in the myogenin gene. *Nature* **364**, 501–506 (1993).
10. Kassar-Duchossoy, L., Gayraud-Morel, B., Gomes, D., Rocancourt, D., Buckingham, M., Shinin, V. & Tajbakhsh, S. Mrf4 determines skeletal muscle identity in Myf5 : Myod double-mutant mice. *Nature* **431**, 466–471 (2004).
11. Bismuth, K. & Relaix, F. Genetic regulation of skeletal muscle development. *Experimental cell research* **316**, 3081–6 (2010).
12. Sharman, A. C. & Holland, P. W. H. Conservation, duplication and divergence of developmental genes during chordate evolution. *Netherlands Journal of Zoology* **46**, 47–67 (1996).
13. Buckingham, M. & Montarras, D. Skeletal muscle stem cells. *Current Opinion in Genetics & Development* **18**, 330–336 (2008).
14. Buckingham, M. Skeletal muscle progenitor cells and the role of Pax genes. *Comptes rendus biologies* **330**, 530–3 (2007).
15. Lassar, A. B., Paterson, B. M. & Weintraub, H. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* **47**, 649–56 (1986).
16. Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B. & Miller, A. D. Activation of muscle-specific genes in pigment, nerve, fat, liver and fibroblast cell lines by forced expression of MyoD. *Developmental Biology* **86**, 5434–5438 (1989).
17. Massari, M. E. & Murre, C. Helix-Loop-Helix Proteins : Regulators of Transcription in Eucaryotic Organisms. *Molecular and Cellular Biology* **20**, 429–440 (2000).
18. Ma, P. C. M., Rould, M. A., Weintraub, H. & Pabo, C. O. Crystal Structure of MyoD bHLH Domain-DNA Complex : Perspectives on DNA Recognition and Implications for Transcriptional Activation. *Cell* **77**, 451–459 (1994).
19. Weintraub, H., Dwarki, V. J., Verma, I., Davis, R., Hollenberg, S., Snider, L., Lassar, a & Tapscott, S. J. Muscle-specific transcriptional activation by MyoD. *Genes & Development* **5**, 1377–1386 (1991).
20. Biesiada, E., Hamamori, Y. & Kedes, L. Myogenic Basic Helix-Loop-Helix Proteins and Sp1 Interact as Components of a Multiprotein Transcriptional Complex Required for Activity of the Human Cardiac α -Actin Promoter. *Molecular and Cellular Biology* **19**, 2557–2584 (1999).

21. Knoepfler, P. S., Bergstrom, D. A., Uetsuki, T., Dac-korytko, I., Sun, Y. H., Wright, W. E., Tapscott, S. J. & Kamps, M. P. A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with Pbx – Meis1 / Prep1. *Nucleic Acids Reseach* **27**, 3752–3761 (1999).
22. Bergstrom, D. A., Penn, B. H., Strand, A., Perry, R. L. S., Rudnicki, M. A. & Tapscott, S. J. Promoter-Specific Regulation of MyoD Binding and Signal Transduction Cooperate to Pattern Gene Expression. **9**, 587–600 (2002).
23. Moran, J. L., Li, Y., Hill, A. a, Mounts, W. M. & Miller, C. P. Gene expression changes during mouse skeletal myoblast differentiation revealed by transcriptional profiling. *Physiological genomics* **10**, 103–11 (2002).
24. Blais, A., Tsikitis, M., Acosta-alvear, D., Sharan, R., Kluger, Y. & Dynlacht, B. D. An initial blueprint for myogenic differentiation. *Genes & Development* **19**, 553–569 (2005).
25. Kitzmann, M. & Fernandez, A. Cellular and Molecular Life Sciences Crosstalk between cell cycle regulators and the myogenic factor MyoD in skeletal myoblasts. *Cellular and molecular life sciences : CMLS* **58**, 571–579 (2001).
26. Rudnicki, M. a, Braun, T., Hinuma, S. & Jaenisch, R. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell* **71**, 383–90 (1992).
27. Cuenda, A. & Rousseau, S. p38 MAP-kinases pathway regulation, function and role in human diseases. *Biochimica et biophysica acta* **1773**, 1358–75 (2007).
28. Luo, D., Renault, V. M. & Rando, T. a The regulation of Notch signaling in muscle stem cell activation and postnatal myogenesis. *Seminars in cell & developmental biology* **16**, 612–22 (2005).
29. Buas, M. F. & Kadesch, T. Regulation of skeletal myogenesis by Notch. *Experimental cell research* **316**, 3028–33 (2010).
30. Glass, D. J. Skeletal muscle hypertrophy and atrophy signaling pathways. *The International Journal of Biochemistry & Cell Biology* **37**, 1974–84 (2005).
31. Buckingham, M. Skeletal muscle formation in vertebrates. *Current opinion in genetics & development* **11**, 440–8 (2001).
32. Conejo, R., Valverde, A. M., Benito, M. & Lorenzo, M. Insulin produces myogenesis in C2C12 myoblasts by induction of NF-kappaB and downregulation of AP-1 activities. *Journal of cellular physiology* **186**, 82–94 (2001).
33. Bryan, B. a, Li, D., Wu, X. & Liu, M. The Rho family of small GTPases: crucial regulators of skeletal myogenesis. *Cellular and molecular life sciences : CMLS* **62**, 1547–55 (2005).
34. Cossu, G. & Borello, U. Wnt signaling and the activation of myogenesis in mammals. *EMBO Journal* **18**, 6867–6872 (1999).
35. Goldhamer, D., Faerman, A., Shani, M. & Emerson Jr, C. Regulatory elements that control the lineage-specific expression of myoD. *Science* **256**, 538–542 (1992).
36. Tapscott, S. J., Lassar, A. B. & Weintraub, H. A Novel Myoblast Enhancer Element Mediates MyoD Transcription. *Molecular and Cellular Biology* **12**, 4994–5003 (1992).
37. Pownall, M. E., Gustafsson, M. K. & Emerson, C. P. Myogenic Regulatory Factors and the Specification of Muscle Progenitors in Vertebrate Embryos. *Annual Reviews in Cell and Developmental Biology* **18**, 747–783 (2002).
38. Chen, J. C. J., Love, C. M. & Goldhamer, D. J. Two upstream enhancers collaborate to regulate the spatial patterning and timing of MyoD transcription during mouse development. *Developmental Dynamics* **221**, 274–88 (2001).
39. Goldhamer, D. J., Brunk, B. P., Faerman, A., King, A., Shani, M. & Emerson Jr, C. P. Embryonic activation of the myoD gene is regulated by a highly conserved distal control element. *Development* **649**, 637–649 (1995).
40. Chen, J. C. J. & Goldhamer, D. J. The core enhancer is essential for proper timing of MyoD activation in limb buds and branchial arches. *Developmental Biology* **265**, 502 – 512 (2004).

41. Chen, J. C. J., Ramachandran, R. & Goldhamer, D. J. Essential and Redundant Functions of the MyoD Distal Regulatory Region Revealed by Targeted Mutagenesis. *Developmental Biology* **245**, 213–223 (2002).
42. Chargé, S. B., Brack, A. S., Bayol, S. A. & Hughes, S. M. MyoD- and nerve-dependent maintenance of MyoD expression in mature muscle fibres acts through the DRR / PRR element. *BMC Developmental Biology* **8**, (2008).
43. Tajbakhsh, S., Borello, U., Vivarelli, E., Kelly, R., Papkoff, J., Duprez, D., Buckingham, M. & Cossu, G. Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5. *Development (Cambridge, England)* **125**, 4155–62 (1998).
44. Brunelli, S., Relaix, F., Baesso, S., Buckingham, M. & Cossu, G. Beta catenin-independent activation of MyoD in presomitic mesoderm requires PKC and depends on Pax3 transcriptional activity. *Developmental biology* **304**, 604–14 (2007).
45. Hu, E., Tontonoz, P. & Spiegelman, B. M. Transdifferentiation of myoblasts by the adipogenic transcription factors PPARγ and C / EBPα. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9856–9860 (1995).
46. Yeow, K., Phillips, B., Dani, C., Cabane, C. & Zoubir, E. Inhibition of myogenesis enables adipogenic trans-differentiation in the C2C12 myogenic cell line. *FEBS Letters* **506**, 157–162 (2001).
47. Wilson, E. M. & Rotwein, P. Control of MyoD function during initiation of muscle differentiation by an autocrine signaling pathway activated by insulin-like growth factor-II. *The Journal of biological chemistry* **281**, 29962–71 (2006).
48. Xu, Q., Yu, L., Liu, L., Cheung, C. F., Li, X., Yee, S., Yang, X. & Wu, Z. p38 Mitogen-activated Protein Kinase Calcium-Calmodulin – dependent Protein Kinase- , and Calcineurin-mediated Signaling Pathways Transcriptionally Regulate Myogenin Expression. *Molecular Biology of the Cell* **13**, 1940 –1952 (2002).
49. Freer-Prokopa, M., O'Flaherty, J., Ross, J. A. & Weymana, C. M. Non-canonical role for the TRAIL receptor DR5/FADD/caspase pathway in the regulation of MyoD expression and skeletal myoblast differentiation. *Differentiation* **78**, 205–212 (2009).
50. O'Flaherty, J., Mei, Y., Freer, M. & Weyman, C. M. Signaling through the TRAIL receptor DR5/FADD pathway plays a role in the apoptosis associated with skeletal myoblast differentiation. *Apoptosis* **11**, 2103–13 (2006).
51. Yang, Y., Xu, Y., Li, W., Wang, G., Song, Y., Yang, G., Han, X., Du, Z., Sun, L. & Ma, K. STAT3 induces muscle stem cell differentiation by interaction with myoD. *Cytokine* **46**, 137–41 (2009).
52. Ogilvie, M., Yu, X., Nicolas-Metral, V., Pulido, S. M., Liu, C., Ruegg, U. T. & Noguchi, C. T. Erythropoietin stimulates proliferation and interferes with differentiation of myoblasts. *The Journal of biological chemistry* **275**, 39754–61 (2000).
53. Thayer, M. J., Tapscott, S. J., Davis, R. L., Wright, W. E., Lassar, a B. & Weintraub, H. Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**, 241–8 (1989).
54. Delgado, I., Huang, X., Jones, S., Zhang, L., Hatcher, R., Gao, B. & Zhang, P. Dynamic gene expression during the onset of myoblast differentiation in vitro. *Genomics* **82**, 109–121 (2003).
55. Gao, X., Chandra, T., Gratton, M. O., Quélo, I., Prud'homme, J., Stifani, S. & St-Arnaud, R. HES6 acts as a transcriptional repressor in myoblasts and can induce the myogenic differentiation program. *The Journal of cell biology* **154**, 1161–71 (2001).
56. Sogawa, K., Nakano, R., Kobayashi, a, Kikuchi, Y., Ohe, N., Matsushita, N. & Fujii-Kuriyama, Y. Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 1936–40 (1995).
57. Liu, D., Black, B. L. & Derynck, R. TGF-B inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3. *Genes & Development* **15**, 2950–2966 (2001).

58. KIM, H., LEE, K., KOMORI, T., LI, Q., Wozney, J. M., Chi, X.-Z., Ryoo, H.-M., BAE1, S.-C., KIM, E., Ueta, C., Choi, J.-Y. & Bae, S. Runx2 Is a Common Target of Transforming Growth Factor β 1 and Bone Morphogenetic Protein 2 , and Cooperation between Runx2 and Smad5 Induces Osteoblast-Specific Gene Expression in the Pluripotent Mesenchymal Precursor Cell Line C2C12. *Molecular and Cellular Biology* **20**, 8783–8792 (2000).
59. Buckingham, M. & Relaix, F. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annual review of cell and developmental biology* **23**, 645–73 (2007).
60. Seale, P., Sabourin, L. a, Girgis-Gabardo, a, Mansouri, a, Gruss, P. & Rudnicki, M. a Pax7 is required for the specification of myogenic satellite cells. *Cell* **102**, 777–86 (2000).
61. Oustanina, S., Hause, G. & Braun, T. Pax7 directs postnatal renewal and propagation of myogenic satellite cells but not their specification. *EMBO Journal* **23**, 3430–3439 (2004).
62. Kuang, S., Kuroda, K., Grand, F. Le & Rudnicki, M. A. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **129**, 999–1010 (2007).
63. Ten Broek, R. W., Grefte, S. & Von den Hoff, J. W. Regulatory factors and cell populations involved in skeletal muscle regeneration. *Journal of cellular physiology* **224**, 7–16 (2010).
64. Hashimoto, N., Kiyono, T., Wada, M. R., Umeda, R. & Goto, Y. Osteogenic properties of human myogenic progenitor cells. *Mechanisms of Development* **125**, 257–269 (2008).
65. Hirsinger, E., Malapert, P., Dubrulle, J., Delfini, M., Duprez, D., Henrique, D., Ish-horowicz, D. & Pourquié, O. Notch signalling acts in postmitotic avian myogenic cells to control MyoD activation. *Development* **128**, 107–116 (2001).
66. Griffin, C., Kleinjan, D. a, Doe, B. & van Heyningen, V. New 3' elements control Pax6 expression in the developing pretectum, neural retina and olfactory region. *Mechanisms of development* **112**, 89–100 (2002).
67. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning Human gene deserts for Long-Range Enhancers. *Science* **302**, 413 (2003).
68. Kleinjan, D. Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. *Developmental Biology* **265**, 462–477 (2004).
69. Kirchhamer, C. V, Yuh, C. H. & Davidson, E. H. Modular cis-regulatory organization of developmentally expressed genes: two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 9322–8 (1996).
70. Gertz, J., Siggia, E. D. & Cohen, B. a Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**, 215–8 (2009).
71. He, X., Samee, A. H., Blatti, C. & Sinha, S. Thermodynamics-Based Models of Transcriptional Regulation by Enhancers : The Roles of Synergistic Activation , Cooperative Binding and Short-Range Repression. *PLoS Computational Biology* **6**, (2010).
72. Reid, J. E., Ott, S. & Wernisch, L. Transcriptional programs: modelling higher order structure in transcriptional control. *BMC bioinformatics* **10**, 218 (2009).
73. Nicodemi, M., Reid, J. E. & Mukherjee, S. Statistical mechanics model. (2011).
74. Scott Kirkpatrick Optimization by Simulated Annealing : Quantitative Studies. *Journal of Statistical Physics* **34**, 975–986 (1984).
75. Matys, V. TRANSFAC(R): transcriptional regulation, from patterns to profiles. *Nucleic Acids Research* **31**, 374–378 (2003).
76. Goldhamer, D. J., Brunk, B. P., Faerman, A., King, A., Shani, M. & Emerson Jr, C. P. Embryonic activation of the myoD gene is regulated by a highly conserved distal control element. *Development* **649**, 637–649 (1995).
77. Gilmore, T. D. Introduction to NF-kappaB: players, pathways, perspectives. *Oncogene* **25**, 6680–4 (2006).
78. L'honore, A., Lamb, N. J., Vandromme, M., Turowski, P., Carnac, G. & Fernandez, A. MyoD Distal Regulatory Region Contains an SRF Binding CArG Element Required for MyoD Expression in Skeletal Myoblasts and during Muscle Regeneration. *Molecular Biology of the Cell* **14**, 2151–2162 (2003).

79. L'honore, A., Rana, V., Arsic, N., Franckhauser, C., Lamb, N. J. & Fernandez, A. Identification of a New Hybrid Serum Response Factor and Myocyte Enhancer Factor 2-binding Element in MyoD Enhancer Required for MyoD Expression during Myogenesis. *Molecular Biology of the Cell* **18**, 1992–2001 (2007).
80. Vlahopoulos, S., Zimmer, W. E., Jenster, G., Belaguli, N. S., Balk, S. P., Brinkmann, A. O., Lanz, R. B., Zoumpourlis, V. C. & Schwartz, R. J. Recruitment of the androgen receptor via serum response factor facilitates expression of a myogenic gene. *The Journal of biological chemistry* **280**, 7786–92 (2005).
81. Hess, J., Angel, P. & Schorpp-Kistner, M. AP-1 subunits: quarrel and harmony among siblings. *Journal of cell science* **117**, 5965–73 (2004).
82. Andreucci, J. J., Grant, D., Cox, D. M., Tomc, L. K., Prywes, R., Goldhamer, D. J., Rodrigues, N., Bédard, P.-A. & McDermott, J. C. Composition and function of AP-1 transcription complexes during muscle cell differentiation. *The Journal of biological chemistry* **277**, 16426–32 (2002).
83. Bengal, E., Ransone, L., Scharfmann, R., Dwarki, V. J., Tapscott, S. J., Weintraub, H. & Verma, I. M. Functional antagonism between c-Jun and MyoD proteins: a direct physical association. *Cell* **68**, 507–19 (1992).
84. Kerppola, T. K. & Curran, T. DNA bending by Fos and Jun: the flexible hinge model. *Science* **254**, 1210–4 (1991).
85. Rajaram, N. & Kerppola, T. K. DNA bending by Fos-Jun and the orientation of heterodimer binding depend on the sequence of the AP-1 site. *The EMBO journal* **16**, 2917–25 (1997).
86. Lian, J. B., Stein, G. S., Javed, A., van Wijnen, A. J., Stein, J. L., Montecino, M., Hassan, M. Q., Gaur, T., Lengner, C. J. & Young, D. W. Networks and hubs for the transcriptional control of osteoblastogenesis. *Reviews in endocrine & metabolic disorders* **7**, 1–16 (2006).
87. Katagiri, T., Yamaguchi, a, Komaki, M., Abe, E., Takahashi, N., Ikeda, T., Rosen, V., Wozney, J. M., Fujisawa-Sehara, a & Suda, T. Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage. *The Journal of cell biology* **127**, 1755–66 (1994).
88. Gersbach, C. A., Byers, B. A., Pavlath, G. K. & Garcia, Andres, J. Runx2 / Cbfa1 stimulates transdifferentiation of primary skeletal myoblasts into a mineralizing osteoblastic phenotype. *Experimental Cell Research* **300**, 406 – 417 (2004).
89. D'Alonzo, R. C., Selvamurugan, N., Karsenty, G. & Partridge, N. C. Physical interaction of the activator protein-1 factors c-Fos and c-Jun with Cbfa1 for collagenase-3 promoter activation. *The Journal of biological chemistry* **277**, 816–22 (2002).
90. Zhang, Y. W., Yasui, N., Ito, K., Huang, G., Fujii, M., Hanai, J., Nogami, H., Ochi, T., Miyazono, K. & Ito, Y. A RUNX2/PEBP2alpha A/CBFA1 mutation displaying impaired transactivation and Smad interaction in cleidocranial dysplasia. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 10549–54 (2000).
91. Suda, T., Ueno, Y., Fujii, K. & Shinki, T. Vitamin D and bone. *Journal of cellular biochemistry* **88**, 259–66 (2003).
92. Endo, I., Inoue, D., Mitsui, T., Umaki, Y., Akaike, M., Yoshizawa, T., Kato, S. & Matsumoto, T. Deletion of vitamin D receptor gene in mice results in abnormal skeletal muscle development with deregulated expression of myoregulatory transcription factors. *Endocrinology* **144**, 5138–44 (2003).
93. Tuteja, G. & Kaestner, K. H. SnapShot: forkhead transcription factors I. *Cell* **130**, 1160 (2007).
94. Tuteja, G. & Kaestner, K. H. Forkhead transcription factors II. *Cell* **131**, 192 (2007).
95. Barthel, A., Schmoll, D. & Unterman, T. G. FoxO proteins in insulin action and metabolism. *Trends in endocrinology and metabolism* **16**, 183–9 (2005).
96. Kitamura, T., Kitamura, Y. I., Funahashi, Y., Shawber, C. J., Castrillon, D. H., Kollipara, R., Depinho, R. A., Kitajewski, J. & Accili, D. A Foxo/Notch pathway controls myogenic differentiation and fiber type specification. *The Journal of Clinical Investigation* **117**, 2477–2485 (2007).
97. Kaneko, H., Shimizu, R. & Yamamoto, M. GATA factor switching during erythroid differentiation. *Current Opinion in Hematology* **17**, 163–168 (2010).

98. Jing, H., Vakoc, C. R., Ying, L., Mandat, S., Wang, H., Zheng, X. & Blobel, G. a Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Molecular cell* **29**, 232–42 (2008).

99. Ghirlando, R. & Trainor, C. D. GATA-1 bends DNA in a site-independent fashion. *The Journal of biological chemistry* **275**, 28152–6 (2000).

100. Alexander, D. L., Ganem, L. G., Fernandez-Salguero, P., Gonzalez, F. & Jefcoate, C. R. Aryl-hydrocarbon receptor is an inhibitory regulator of lipid synthesis and of commitment to adipogenesis. *Journal of cell science* **111 (Pt 2**, 3311–22 (1998).

101. Shimba, S., Hayashi, M., Ohno, T. & Tezuka, M. Transcriptional regulation of the AhR gene during adipose differentiation. *Biological & pharmaceutical bulletin* **26**, 1266–71 (2003).

102. Feige, J. N., Gelman, L., Michalik, L., Desvergne, B. & Wahli, W. From molecular action to physiological outputs: peroxisome proliferator-activated receptors are nuclear receptors at the crossroads of key cellular functions. *Progress in lipid research* **45**, 120–59 (2006).

103. Ferrier, D. E. & Holland, P. W. Ancient origin of the Hox gene cluster. *Nature reviews. Genetics* **2**, 33–8 (2001).

104. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425–32 (2007).

105. Mann, R. S. & Affolter, M. Hox proteins meet more partners. *1Current Opinion in Genetics & Development* **8**, 423–429 (1998).

106. Berkes, C. A., Bergstrom, D. A., Penn, B. H., Seaver, K. J., Knoepfler, P. S. & Tapscott, S. J. Pbx Marks Genes for Activation by MyoD Indicating a Role for a Homeodomain Protein in Establishing Myogenic Potential. **14**, 465–477 (2004).

107. Maves, L., Waskiewicz, A. J., Paul, B., Cao, Y., Tyler, A., Moens, C. B. & Tapscott, S. J. Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. *Development* **134**, 3371–3382 (2007).

108. Liu, Y., Chu, A., Chakroun, I., Islam, U. & Blais, A. Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation. *Nucleic acids research* **5800**, 1–15 (2010).

109. Oliner, J. D., Andresen, J. M., Hansen, S. K., Zhou, S. & Tjian, R. SREBP transcriptional activity is mediated through an interaction with the CREB-binding protein. *Genes & Development* **10**, 2903–2911 (1996).

110. Semenza, G. Signal transduction to hypoxia-inducible factor 1. *Biochemical pharmacology* **64**, 993–8 (2002).

111. Ono, Y., Sensui, H., Sakamoto, Y. & Nagatomi, R. Knockdown of hypoxia-inducible factor-1alpha by siRNA inhibits C2C12 myoblast differentiation. *Journal of cellular biochemistry* **98**, 642–9 (2006).

112. Bain, G. & Murre, C. The role of E-proteins in B- and T-lymphocyte development. *Seminars in immunology* **10**, 143–53 (1998).

113. Abu-Elmagd, M., Robson, L., Sweetman, D., Hadley, J., Francis-West, P. & Münsterberg, A. Wnt/Lef1 signaling acts via Pitx2 to regulate somite myogenesis. *Developmental biology* **337**, 211–9 (2010).

114. Behrens, J., Kries, J. P. von, Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R. & Birchmeier, W. Functional interaction of B-catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).

115. Kikuchi, A. Regulation of beta-catenin signaling in the Wnt pathway. *Biochemical and biophysical research communications* **268**, 243–8 (2000).

116. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795 (1995).

117. Giese, K., Pagel, J. & Grosschedl, R. Functional analysis of DNA bending and unwinding by the high mobility group domain of LEF-1. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12845–50 (1997).

118. Oikawa, T. & Yamada, T. Molecular biology of the Ets family of transcription factors. *Gene* **303**, 11–34 (2003).

119. Verger, A. & Duterque-Coquillaud, M. When Ets transcription factors meet their partners. *BioEssays* **24**, 362–70 (2002).

120. Hollenhorst, P. C., McIntosh, L. P. & Graves, B. J. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annual review of biochemistry* **80**, 437–71 (2011).

121. Murre, C., Bain, G., Dijk, M. A. Van, Engel, I., Furnari, B. A., Massari, M. E., Matthews, J. R., Ouong, M. W., Rivera, R. R. & Stuiver, M. H. Structure and function of helix-loop-helix proteins. *Biochimica et biophysica acta* **1218**, 129–135 (1994).

122. Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V, Buskin, J. N., Hauschka, S. D. & Lassar, a B. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537–44 (1989).

123. Zhang, X., Patel, S. P., McCarthy, J. J., Rabchevsky, A. G., Goldhamer, D. J. & Esser, K. a A non-canonical E-box within the MyoD core enhancer is necessary for circadian expression in skeletal muscle. *Nucleic acids research* 1–12 (2011).doi:10.1093/nar/gkr1297

Chapter 3

3. Review of microfluidics

Large scale DNA mutant library synthesis should be fast, efficient and use minimal reagents. A microfluidic platform offers a solution that can meet all these criteria. Various methods are available for the fabrication of different classes of microfluidic systems this chapter seeks to summarise the key methods for the fabrication of microfluidic systems with special emphasis of methods and techniques which yield properties that are beneficial to this application. This chapter is a review of fabrication techniques, materials and designs which are used in modern microfluidic devices. The design of monolithic microfluidic devices made by MSL (chapter 5) and of devices assembled by multilayer soft lithography from polydimethylsiloxane (PDMS) layers cast in MSL moulds refer to the findings of this review (chapter 6).

3.1. Microfluidic overview

The growth of microfluidics as a discipline is reminiscent of the development of the integrated circuit (IC). Although today ICs are present in many of the devices that are indispensable to modern life, the adoption of microfluidics has been somewhat slower[1].

Miniaturisation of biological and chemical assays is associated with several benefits: Reduction in raw materials requirements, rapid mixing facilitating rapid reaction times, fast heat transfer, decreased analysis time and facilitates parallelisation of assays. Thus, the cost and volume of work done can be greatly decreased and increased, respectively.

Microfluidic devices in the context of this thesis are defined as devices which have features and channels where one dimension is typically 1 – 100 µm in size, through which fluids flow. The first microfluidic device was a gas chromatography system fabricated in silicon

and glass by Terry *et al.* in 1979[2]. This early work was not significantly expanded until the 1990's when techniques, such as photolithography, used to make micro electromechanical systems (MEMS) were applied to biological and chemical fields[3].. During the 1990's microfluidic devices were primarily fabricated using the techniques of lithography and micromachining[4].Since 2000, the fabrication of devices from polymeric materials such as PDMS[5,6] reduced the cost and production time required and allowed the field to experience strong growth.

3.2. Types of microfluidic devices

Microfluidic devices fall into three distinct categories that are distinguished by the combination of phases of fluid within the microfluidic channels. Continuous flow systems are single liquid phase, whereas droplet systems consist of two or more mutually immiscible phases. By contrast, digital devices do not possess channels in the traditional sense instead moving droplets of fluid between electrodes by means of dieletric forces. The following sections describe each of these types in more detail. Table 3.1 compares the relative utility of each of these types.

3.2.1. Continuous

Continuous flow microfluidics involves devices with channels made in suitable materials (see section 3.7) that have a single liquid phase flowing through them. A variety of methods can be used to control the flow of fluid through channels including on chip valves and pumps, off chip pumps, semi-permeable membranes, magnetic fields and electro-osmotic pumps (EOPs) (see section 3.12). Due to surface wetting effects, single phase devices are prone to contamination between reagents and must undergo washing or surface treatment to avoid these issues. Continuous flow devices are generally easier to fabricate and most widely applicable. Each of these mechanisms will be discussed during this introduction.

3.2.2. Droplet

Droplet microfluidics involves the formation of droplets of one phase, such as an oil phase in another phase, such as an aqueous phase. The hydrophobic oil does not mix with the water-based polar fluid and so droplets of one can be carried by a flow of the other. Droplets are kept separate from each other and do not contact the internal surfaces of the device, provided the device material and droplet fluid properties are appropriate. Droplet microfluidics makes use of the same flow control methods as continuous flow devices (see sections 3.11 and 3.14), but also makes use of the geometry of channels to control the mixing and merging of droplets (see section 3.16). Droplet microfluidic devices require accurate and consistent control of flow rates and often require stabilisation prior to running productively. Furthermore, the liquids may contain toxic oils or surfactants.

3.2.3. Digital

Digital microfluidics involves the movement of droplets of a polar phase across electrically active surfaces in air or non-polar/conducting medium by electrowetting[7,8]. Digital microfluidics is often referred to as electrowetting on dielectric (EWOD) microfluidics. Confusingly, because digital microfluidics employs the movement of droplets, digital microfluidics is sometimes referred to as droplet microfluidics. Digital microfluidic platforms can be prepared by patterning electrodes across parallel glass plates using a pattern mask. As a result EWOD does not require mould production or photolithography and so can be produced relatively easily. The required devices are complicated, however, and as droplets transit around the surface of the device, they may leave residue that can cause contamination of subsequent droplets. Furthermore, the range of suitable fluids is limited by the requirements of the method. Surface acoustic waves (SAWs) and ultrasound can be employed to produce and manoeuvre droplets.

Type	Risk of contamination	Ease of use	Sample volume	Flexibility
Continous	Moderate	High	Moderate	Low
Droplet	Low	Moderate	Low	Moderate
Digital	High	Moderate	Low	High

Table 3.1: Comparison of the relative utility of the three primary types of microfluidic device.

3.3. Microfabrication methods

At the heart of microfluidic devices is the production or microfabrication process. Techniques such as lithography, used to mass produce ICs, can be used to etch channels directly into materials such as glass or silicon that are relatively impervious to attack by organic solvents or aqueous solutions of neutral pH. Lithography is also frequently used to pattern spun layers of photoresist (usually SU-8). Micromachining is employed to create channels directly in softer materials, such as plastic. Patterns made in a master can then be reproduced in a secondary material, such as PDMS or poly(methyl methacrylate) (PMMA), through a process called lithography, electroplating and moulding (LIGA). In both the lithographic and micromachining cases, the channels are formed initially as grooves on a surface, the fourth wall of which is formed when the patterned surface is sealed against another, flat surface.

3.4. Photolithographic techniques

Photolithography allows the patterned etching or deposition of materials onto an underlying material. Typically, a layer of photoresist is applied to a surface, which is then patterned using a photomask in conjunction with exposure to light (usually UV, depending

on the photoresist). Photomasks can be produced be a range of techniques depending on the requirements of the application: Low resolution (>150 nm) masks can be prepared by printing onto a transparent medium and higher resolution masks generally require the employment of phase-shifting, he use of short wavelength light such as x-rays or immersion lithography. Once patterned photoresist is then developed, where the excess resist is removed leaving behind a patterned layer of photoresist. Although this photoresist can be used directly as part of a microfluidic device, it is more common to etch the underlying material using an etchant to which the photoresist is resistant. After this process, remaining photoresist can be removed leaving the negative of the photoresist pattern engraved into the underlying material.

Photolithographic techniques are capable of producing features and channels 100's of nm in size. The scale of features typically used in microfluidic devices (1-10 µm) are easily produced by photolithography. Furthermore, photomasks can be reused and patterns repeated for mass-manufacturing. Silicon is the most commonly used material in photolithography, but other materials can also be etched such as HF etching of glass. Accurate production of small devices with small feature sizes (<10 µm) by photolithography requires precisely manufactured photomasks, expensive equipment such as mask aligners and the use of toxic or caustic chemicals. This requires lithographic techniques be performed by experienced personal operating in a controlled laboratory environment. In addition to constituting the devices themselves, photolithographically patterned devices can be used as moulds to produce microfluidic devices by casting or hot embossing using suitable polymers such as PDMS and polymethylmethacrylate (PMMA).

3.5. Micromachining

Micromachining uses small drill bits or a laser to pattern a surface by directly removing material from the surface. Parts can be designed in 3D CAD software and produced using

CNC machines. Micromachining is best employed with hard materials such as PMMA, that will not deform during the milling process compared to softer materials, such as PDMS. Although variations in parts can be introduced as drill bits undergo wear, the cost of replacing drill bits is insignificant next to the cost of the CNC machine itself or the cost of an equivalent lithographic setup. Lasers, which do not undergo wear, but whose power may fluctuate, can be used to machine both hard and soft materials but may leave rough or damaged surfaces[9]. The primary limitation of micromachining is resolution; typical drill bit diameters are >100 μm which limits the minimum feature size. The primary benefit of micromachining is that parts can be produced rapidly without the need for toxic chemicals.

3.6. Additive layer manufacture

Additive layer manufacture (ALM) is a broad term which encompasses the range of technologies that form whole parts or devices by the gradual addition of successive layers. There are many ALM technologies available that can be categorised into the following basic groups: Photopolymerisation, powder fusion, extrusion, printing, sheet lamination and beam deposition[10]. A further group exists that consists of hybrid technologies that take elements from two or more of the above categories. Table 3.2 compares these manufacturing techniques and each is discussed in more detail below.

Photopolymerisation or stereolithography (SLA) involves the light activated reaction of a liquid monomer to form a complex polymer. Activation can be achieved by exposure with a patterned light source or with a laser which is rastered across a surface. Very high resolutions (< 10 nm) can be obtained by using two photon laser systems[11,12]. The range of materials which can be used in a photopolymerisation process is limited. The material must be able to polymerise specifically (i.e. not be polymerised when exposed to ambient light or be handled limited wavelength environment), quickly and at high resolution (with minimal spreading of the reaction). The leeching of uncured liquid

components, such as activators or inhibitors of polymerisation, into media can result in toxicity to biological systems. MSL flow cells have been successfully used in chemical applications[13–16].

Method	Build area (XYZ) (mm)	Post curing	Support material?	Material	Layer thickness (µm)	Resolution (µm²)	Commercial examples
SLA	500 x 500 x 600	Yes	No	Acrylate resin	25-100	20-100	EnvisionTec[17], 3D Systems[18]
SLS	500 x 500 x 400	No (Possibly firing)	Yes	Plastic, metal	100-200	200-300	3D Systems[19]
FDM	250 x 250 x 300	No	No	Thermoplastic (ABS)	150-300	100-200	Stratasys[20], BFB[21]
3DP	300 x 200 x 200	No	Yes	Plastic	25-100	25-50	3D Systems[22], Objet[23], Z corp[24]
LOM	160 x 210 x 135	Yes	No	PVC, PS	100-200	100	Solido 3D[25]
Beam deposition	N/A	No	Yes	Plastic, metal	400	200-300	N/A

Table 3.2: Comparison of ALM-based rapid manufacturing methods. Dimension units are as specified in the column headers. Resolution and layer thicknesses are representative of published and commercially available examples. NB Commercial suppliers of ALM machines are subject to rapid change at this time as the industry is undergoing consolidation

Powder fusion systems (also known as selective laser sintering, SLS) typically involve the selective sintering of a powder material by a heat source, such as a laser. The laser is rastered across a surface to produce a solid material from a powder, of either plastic or metal, which is briefly melted as a result of the action of the laser[19]. This method has the advantage that the unsintered material remains to support future layers, meaning that overhanging and suspended features, such as an arch or ball-on-chain, can be produced. Conversely the unsintered material must be removed after the building process before the part is complete. Unsintered material cannot be removed from fully enclosed hollows thus limiting the range of parts that can be made by this process.

Extrusion, also called “fused deposition modelling” (FDM), involves the heating of a solid thermoplastic, usually acrylonitrile butadiene styrene (ABS), the plastic becomes sufficiently pliable to push through a nozzle. Extruded material cools as it leaves the nozzle head and becomes solid once more. Parts can be built by extruding trails of thermoplastic onto a solid surface, building layers by rastering of the extruding nozzle. Successive layers can be built on top of previously formed layers provided the newly extruding material does not melt the previously extruded material. Companies such as Stratasys Inc.[20] and Bits from Bytes[21] utilise the extrusion method in their Dimension3D and BFB3000 respectively.

Printing systems use printing heads similar to those used in inkjet printers to place discrete drops of material onto a surface and previously formed layers to form a part[26], called 3D printing (3DP). Liquid material is ejected from the nozzle and is solidified on the part before the next layer is placed. The method used to cure the liquid polymer to a solid depends on the polymer involved. Due to the nature of the method, overhangs are impossible unless a separate support material is employed. The support material must be removed in a post-processing step before the part is ready for use. A prominent example of a printing system is the Objet PolyJet technology used in their Connex and Eden products[23].

Sheet lamination (or laminate object manufacturing, LOM) involves the cutting of thin sheets of material into specific shapes and then binding them together to make a part. A laser or other cutting tool can be used to cut a specific shape from the sheet material and a variety of methods can be used to bond successive layers together. Cubic technologies produce LOM machines used the Solidimension SD300 brand, which uses polyvinyl chloride (PVC)[25]. Sheet lamination of materials that contract under heating has also been performed[27]. Pre-stressed polystyrene (PS) sheets are scribed and/or punched as necessary then aligned, clamped together and heated. The sheets then relax back to their original conformation as a result of the heating, shrinking back to their pre-stressed shape.

Beam deposition involves the direction of a continuous stream of particles at a surface resulting in the accumulation of those particles at the surface. The beam can be a stream of ions or a laser focused to a point within a powder stream. In the latter case, the powder is heated and fused at the specific position. An example of laser mediated powder fusion in a beam deposition arrangement is laser engineered net shaping (LENS) [28].

Hybrid technologies such as a cross between printing and powder fusion, where a binder agent is printed from an inkjet head onto a powder substrate also exist[26]. Alternatively, a material which sets solid on a surface can be projected from an inkjet head, multiple heads are then capable of making builds comprising more than one material. Furthermore, biological materials including cells can be projected[29] which opens up the possibility of direct digital manufacture of living structures.

All the technologies described here share the advantage of requiring no 'tooling' and can be used to make objects that traditional multi-axis machining cannot, due to drill access. All the methods can take a 3D object *in silico* and produce a full 3D object that would not otherwise be possible to produce in a relatively short amount of time. Whilst some methods involve the use of potentially harmful glues or other chemicals, the level of

personal risk is below that of the chemicals associated with lithographic techniques. Furthermore, lithographic techniques can be applied to a relatively small range of materials, such as silicon and glass, which can be etched appropriately.

3.7. Microfluidic device materials

The material chosen for any microfluidic device depends on the application of the device[30]. Lithographic techniques are used to make devices comprising of channels in PDMS, silicon[31] or glass[32]. A form of RIE, deep reactive ion etching (DRIE) is usually employed to produce channel patterns in hard materials, such as silicon. Glass offers unparalleled optical transparency and quartz quality glass is clear into mid-UV wavelengths. The use of hard materials such of these precludes any type of actuatable valve features unless a flexible membrane layer is included[33–36].

Soft materials offer the advantage that parts can be made by imprinting or replica moulding[37]. PDMS is the predominant material for microfluidics due to its advantageous properties: PDMS can be easily cast as a liquid and cures to a rubber-like, transparent solid[38]. PDMS cures rapidly, in <1 hour at 80°C. PDMS layers can be bonded together to make single materials, a process called multilayer soft lithography (discussed in section 3.9)[6]. PDMS is sufficiently flexible that pneumatic valves are possible at relatively low pressures (<1 bar).. PDMS is, however, associated with disadvantages: PDMS is, however, susceptible to swelling in response to organic solvents such as decane and xylene. PDMS also absorbs small molecules from solution[39]. Overall, PDMS is good for producing multiple copies of a device by replica moulding, the final articles can incorporate actuatable features as a result of bonding multiple layers. Conversely, swelling in organic solvents and sequestration of small molecules from solution limit the range of applications in which PDMS can be used. To overcome solvent-based swelling, other materials are available:

'Liquid Teflon™' based on photocurable perfluoropolyethers (PFPEs)[40,41]. Chips can also be fabricated entirely from Teflon™ by imprinting[42].

A frequently used material for microfluidics is PMMA, also known as Plexiglass™. Like PDMS, PMMA can be cast as a liquid and set to a solid and is optically transparent[43]. PMMA can also be printed if pre-softened with acetonitrile[44]. Laser machining can be employed to create microfluidic channels in PMMA[45,46], SU-8[47] or even glass[48]. The geometry of channels produced by laser machining is limited as the laser will generate a channel with a V-shaped cross-section reflecting the power output across the diameter of the laser beam.

A major advantage of polymeric materials such as PMMA and polycarbonate (PC) is that they are susceptible to hot embossing, also known as thermoforming[49,50]. Hot embossing is a way of replica moulding with solid substrates by imprinting with a mould at above a critical temperature at which the substrate becomes pliable. Injection moulding is another, related method, but requires a completely enclosing mould.

Other materials include polyurethane-methacrylate (PUMA)[51], which is especially suited to high aspect ratio structures, thiolene[52], which can withstand a temperature range from -150 to 125°C and solvents such as toluene. Microfluidic devices can be fabricated out of pre-stressed polymer sheets (i.e. 'shrinky-dinks'), made from polystyrene[27] or cyclic olefins[53], layers of which can be shrunk and bonded with heat and pressure. Even paper can be used as material for microfluidic devices[54], which could be used as disposable point-of-care diagnostic aids.

3.8. Does smaller equal better in microfluidics?

The advantages of microfluidic systems over traditional bench top systems are plain: smaller reaction volumes require less substrate, the small scale of components means that the time between reaction and detection can be minimised and small devices mean less

power is required to drive them. Integration of small scale devices onto a chip facilitates parallelisation of the assays. Laminar flow within channels implies that only diffusive mixing is possible, enabling tight regulation over reactant localisation. As such much of the work in microfluidics has focussed on making channels and features smaller so as to fit more functionality onto one chip[55]. The methods of detection used in devices do not always scale efficiently. Very low amounts of analytes leads to a low signal to noise ratio[56]. Ideally, all microfluidic experiments would be wholly performed on one device. The terms micro-total analysis system (μTAS) and lab-on-a-chip were first created to describe such devices.

Circumstances, however, sometimes dictate that products of reactions on one chip must be transferred to other chips or to the bench top where further processing is performed[57]. In these cases, the minimum size of the devices must be carefully selected to ensure that a sufficient amount of reagents are available for the off-chip processing steps.

3.9. Multilayer soft lithography

Micropatterns in a hard surface can be transferred to a soft material by embossing or the hard surface can be used as a mould from which casts can be made. The alignment and assembly of multiple patterned layers into a single device is called multilayer soft lithography[6]. The most frequently used material for multilayer soft lithography is PDMS. When uncured, mixtures of monomeric dimethlysiloxane and the curing agent are liquid and can be poured easily with a consistency approximating honey. The temperature dependant curing of the mixture to PDMS yields an optically clear (absorbance increases rapid at wavelengths < 300 nm[58,59]) rubber-like solid. Devices made from PDMS casts of moulds made by a variety of microfabrication methods discussed above are described in the following examples.

The Quake lab at Stanford University has, for much of the last decade, been at the forefront of PDMS microfluidics[6]. During this time the paradigm of large scale integration

(LSI) was developed[60] that predicts that microfluidics as a platform will only take off once the design rules of critical components is decided upon. Thereafter, new devices can be created by bolting together components without each device having to be designed from the ground up. The core components can then be mass produced, bringing their cost down to an affordable level. Systems like SmartBuild[61] appear to be following this paradigm. Although the SmartBuild system is modular and uses clever design to accomplish an easy to assemble and use system, its utility is fundamentally limited by the applicability of static parts that possess user-operated mechanical valves.

3.10. Bonding of PDMS devices

There are several methods to bond the multiple layers of PDMS devices that are necessary to make functional microfluidic chips, such as oxygen plasma treatment, partial curing or coronal discharge[62]. Using an uncured PDMS mortar has been shown to yield the strongest bond, able to withstand a pressure of >600 kPa, followed by partially cured PDMS, oxygen plasma and coronal discharge considerably weaker by a factor of a half.

3.11. Flow in microfluidic channels

Microfluidic systems have large surface area:volume ratios. As a result, the flow of fluid through microfluidic channels is dominated by viscous forces meaning that flow through channels occurs laminally, with no chaotic mixing. Laminar flow is observed in low Reynold's number regimes, typically <2000. In regimes where a larger Reynold's number is calculated, turbulent flow is expected to dominate. The following equations can be used to determine whether flow through a pipe, relevant to microfluidic applications, will be laminar. Equation 3.1 and 3.2 are used to determine the Reynold's number (*Re*) of a given system

$$Re = \frac{\rho v D_H}{\mu} \qquad D_H = \frac{4A}{P} \qquad \text{Eqns. 3.1 and 3.2}$$

where ρ is the fluid density, v is the velocity, D_H is the hydraulic diameter and μ is the dynamic viscosity. The hydraulic diameter used for non-circular pipes and is equal to four times the cross sectional area (A) of the pipe divided by the perimeter (P). In short, a fast flowing, high density, low viscosity fluid flowing through a pipe which has a low surface area:volume is less likely to undergo laminar flow than a slow flowing, low density, high viscosity fluid moving through a pipe which has a high surface:volume.

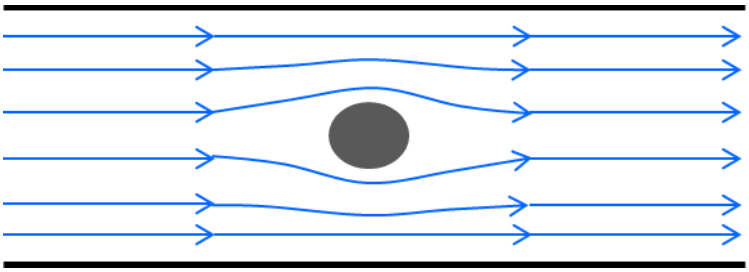


Figure 3.1: Flow lines (blue) of a fluid undergoing laminar flow as it moves through a tube and around an obstruction (grey oval). Laminar flow is characterised by non-intersecting flow lines depicted here, where fluid flows in laminar sheets with very little mixing.

Modelling of fluid dynamics can be employed to determine flow in microfluidic environments. Although various fluid dynamics packages are available, see section 4.6, the COMSOL modelling package will be focussed on here as it was used for the fluid dynamics models described in chapter 5. By performing fluid dynamic modelling it is possible to determine flow profiles through systems that are too complicated for a straight forward analytical solution to be applicable. Optimisation of flow profiles by modification of the model geometry means that the optimal design can be produced in fewer design iterations. It is also possible to simulate the generation, movement and merging of droplets in droplet microfluidic systems.

3.12. Pumping in microfluidic devices

Pumping in microfluidic devices is usually achieved with equipment external to the device itself. Syringe pumps are the most widely used as they are capable of the consistent, very

low flow rates (<10 µL/min) typically used for microfluidic devices. Maximum flow rates in microfluidic devices are limited by the viscosity of the liquid and the strength of the interaction between the liquid and the channel walls. Syringe pumps are typically expensive (costing several hundred to a few thousand pounds)[63–65]. Peristaltic pumps can also be used, but less commonly due to the fact that low flow rates are difficult to achieve with sufficient consistency (to prevent backflow)[66–68]. Selective application of air pressure to reservoirs connected to microfluidic chips is also a commonly used method to move fluid around microfluidic devices[69]. In contrast to these three types of pump, which all require equipment external to the microfluidic device itself, micropumps are micro-scale devices that are capable of moving fluid around microfluidic channels[70,71].

Peristaltic micropumps can be built by multilayer soft lithography[72,73]. Cycling the pressure in each of three or more valves by external pressure regulation causes flow in microfluidic channels. Instead of using a series a pneumatic valves, a single, biased valve can be used[74] to achieve peristaltic pumping. Although the three valves are typically controlled by three separate air lines, peristaltic pumping can be achieved using a single line and a serpentine pneumatic channel[75]. Only devices that consist of soft materials or incorporate flexible materials are capable of this type of pumping. Furthermore, flow from peristaltic micropumps is inconsistent and susceptible to backflow. Similarly, valveless micropumps also result in only biased flows, with back flow common without the use of check valves[76].

Monolithic mechanical micropumps can also be fabricated in MSL devices[77]. This air pressure driven membrane displacement pump could be easily driven by a magnetic or piezoelectric method. The principle disadvantage with all mechanical pumps is that they require an off chip power source to drive them. The equipment required to drive these pumps is likely to be orders of magnitude larger than the devices and pumps they drive, thus limiting their utility in true lab-on-a-chip applications. The equipment can be reused in

multiple microfluidic designs, however, unlike integrated pumps that cannot be separated from the device.

A promising method of non-mechanical microfluidic pumping is the use of electro-osmotic pumps (EOPs)[78]. EOPs utilise the electro-osmotic force (EOF) that is the movement of locally separated, solvated charges in response to an electrical field[79]. For example, when a surface that holds positively charged, covalently bound charges (stationary phase) is submerged in a solution containing charged solutes, the solutes will rearrange to correct the local charge imbalance around the surface (see figure 3.2). A layer of the solution close to the positively charged surface will be enriched in negatively charged ions. This layer is called the Debye layer, also known as an electrical double layer. Application of an electrical field along the charged surface results in the movement of the ions in the Debye layer, the movement of the ions is then imparted to the rest of the solution via the solvation layer around each ion.

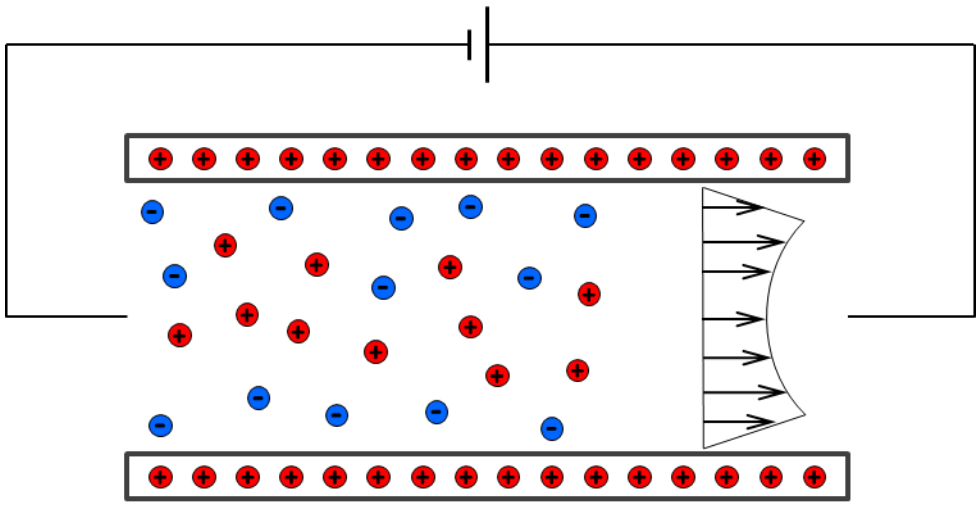


Figure 3.2: Diagram of how the EOF is created. A positively charged tube causes a charge imbalance in charged solutes across the tube as negatively charged solutes prefer to be closer to the positively charged walls and vice versa for positively charged solutes (forming a Debye, or double, layer). Application of an electric field along the tube pushes the net negatively charged surface layer of fluid to move away from the negative electrode. The combination of this forward force and of viscous drag results in the flow profile shown in the figure.

For example, an EOP based upon a silica monolith stationary phase is capable pumping deionised water at a rate of 2.9 µL/min against a back pressure of 3.1 atmospheres, the power consumption of the pump is low; although a potential of 6 kV is required, the current consumption is 1.5 nA meaning the pump consumes power in the mW range[80]. In addition to supplying such a high voltage, the application of such a voltage across aqueous solutions results in electrolysis of the water, so electrodes must be carefully placed to avoid bubble formation that would block channels[81]. Interestingly, stationary phases can be made by photopolymerisation of polyimides[82], which opens up the possibility of building microfluidic devices with integrated pumps using a multi-material MSL method. Furthermore, EOPs have been used in a functional DNA fingerprinting lab-on-a-chip previously reported by the Haswell group[83].

Fluid movement can also be achieved with only the entirely inactive mechanism of capillary action[84]. The semi-permeable nature of PDMS can be used to push fluids around microfluidic devices[85,86]. Another mechanism for microfluidic pumping is electrochemical pumping[87,88], where a gas is generated from a solution (usually by electrolysis) that causes fluid movement by virtue of pressure imbalance. Electrochemical pumping is relatively slow and can require complicated circuitry to perform. For applications where gas permeability of PDMS is not desired, PDMS can be treated to lower permeability such as with parylene coating[19,63,64] and wax treatment[91].

3.13. Measurement in microfluidic devices

Although microfluidic systems are capable of creating many reaction conditions per second, up to 100 Hz, the rate of the system can be limited by the maximum rate at which information can be extracted from the system[92]. The interrogation of microfluidic reactions can be performed either on- or off-chip and fall into two categories: continuous or discrete. This review will focus primarily on the on-chip methods as the off-chip methods could

include any measurement technology currently available and is too broad to cover here. The following information is summarised in table 3.3.

Several technologies are available for the on-chip interrogation of microfluidic reactions. The most widely used is optical measurements using visible or fluorescent dyes that track the progress of a reaction[93]. Raman and confocal microscopy can also be used to extract more information from each reaction, but the long scan times (several seconds) associated with these usually prevents their application in high speed systems (>1 Hz)[94]. Recently, however, ultrafast systems with scan times of 10 µs for single wavelength surface enhanced resonance Raman scattering (SERRS) has been demonstrated[95].

The output of microfluidic reactions can be passed to continuously operating detection systems such as HPLC[96] and mass spectroscopy[97]. The advantage of these techniques is that they are able to separate the products of the reaction, allowing more information to be obtained. To insure the fidelity of the measurements taken in these systems, the rate at which samples are introduced must not exceed the time taken for the slowest component to leave the detector, otherwise sample overlap will occur. As a result the detection system must be tuned to the reaction generation rate.

Electrical methods, defined as methods that involve an electrical field component, encompass electrical field distortion as a result of the passage of particles or cells[98] as well as electrochemical turnover of a species at an electrode[14,99].

Method	Mode	Speed	Applicability
Off-Chip	Discrete	Slow	Very broad.
Optical	Continuous	Rapid	Limited by dyes, sensitivity, selectivity.
Electrical fields	Continuous	Rapid	Limited: Particles only
Electrochemical	Continuous	Rapid	Limited: Suitable species only.
HPLC	Potentially continuous	Dependant on sample transit rate.	Broad
Mass spectroscopy	Potentially continuous	Rapid	Broad

Table 3.3: Summary on detection methods in microfluidic devices.

3.14. Methods of microfluidic valve actuation

To be able to perform more than one action reliably, microfluidic devices cannot solely rely on geometry and selective pumping alone. The process of controlling the flow of liquids within a microfluidic chip requires actuation of valves within the chip itself. Actuators can be extensively miniaturised for use in microfluidics[100]. A valve can be closed using a variety of mechanisms. Only valves in soft materials, such as PDMS, will be discussed here as PDMS is used in the fabrication of the microfluidic device described in chapter 6. Actuation of hard materials such as plastics or metals, such as solvent absorbtion[101] and shape memory alloys[102–104], is possible.

The most conceptually simple actuated valve is a membrane valve[100]. Here, a membrane separates two chambers; raising the pressure in one chamber relative to the other pushes the membrane into the lower pressure chamber. The lower chamber will collapse, blocking it, if the pressure difference becomes sufficient. If the lower chamber forms part of a channel, then the channel will be sealed. This system, commonly referred to

as the 'Quake valve', was first reported in Science in 2000[6]. Accurate flow control by this type of valve, by varying pressure on the valve membrane, is possible[105]. The method of assembly of a typical Quake valve can be seen in figure 3.3, part A. Two layers, a fluidic and pneumatic layer are fabricated from suitable moulds. Alignment and sealing of the two layers yields cross-over points where sealing of one chamber in response to a pressure being applied to the other can be achieved.

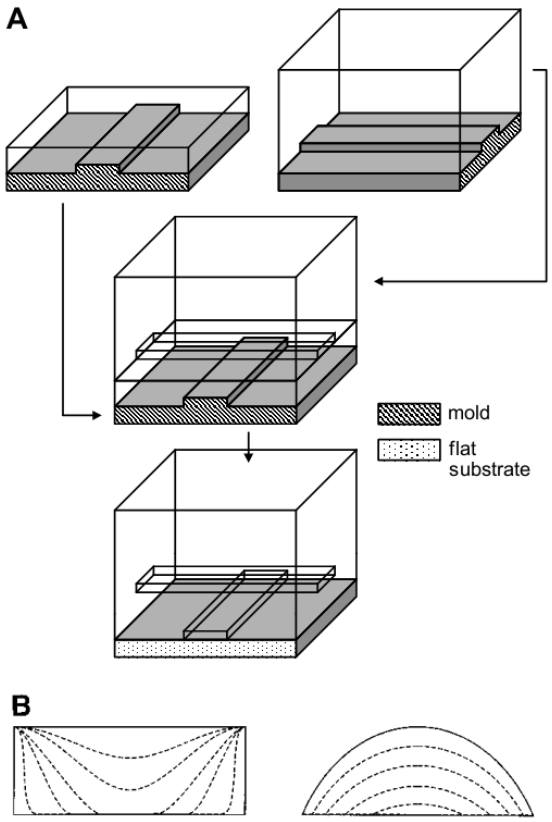


Figure 3.3: Diagram of Quake valve assembly process: Layers are separately cast and bonded together (A). How chamber deformation results in channel sealing in either rectangular or semi-circular cross section channels (B). Image taken from Unger *et al.*[6].

Generally, Quake valves are actuated by applying a positive air pressure to one chamber using off chip solenoid valves. Positive air pressure can also be created inside a sealed chamber by heating[31]. Actuation by heating in this manner is slower than by actuating using pressure applied through off chip valves; 150 ms closing time and 300 ms opening time for a 130 x 30 x 2.9 µm membrane using a 240 mW heater, due to the time required for heat

transfer to occur which presumably occurs at a similar rate to the valve actuation times stated above. The advantage of this method is that all the machinery necessary for the valve can be integrated onto the chip rather than necessitating off chip valves. The operation of the heating elements requires additional circuitry and equipment and will most likely render one side of the device opaque, preventing through device microscopy.

Local heating can be used to direct droplets within a flow due to the Marangoni effect[106]. Local viscosity changes as a result of applying a temperature gradient across a flow causes droplets within the flow to tend towards the path of least viscosity. Setting up the temperature gradient takes seconds and the maximal flow rate is limited to ~10 µL/min, which together means that the rate of sorting in such a device would be slow.

Local heating and cooling can be used to cause phase change of plugs of various kinds and can be used to seal valves: Generally, a material such as wax is heated, and forced into a channel. Once cooled, the wax then seals the channel. To remove the seal, the wax is then heated and forced out of the channel. Movement of the wax can be achieved by applying air pressure[107,108] or magnetic fields[109]. Temperature dependant hydrogels can also be employed[110]. The actuation time of such systems is considerable, requiring several seconds for melting. Furthermore, the application of external pressures or fields is usually necessary for sealing. The main advantage of this type of system is that it is latchable: Once a valve is sealed the external pressure can be removed and the valve will remain sealed.

Direct mechanical movement is also used to seal valves: Braille reader[111] and piezoelectric drivers[112] are most commonly used. Actuators of this type directly impinge on soft PDMS membranes, deflecting them into and sealing underlying channels. The main limitation associated with this valve actuator type is the maximum density of actuators that can be achieved. Braille readers, typically solenoid driven, are relatively large, with heads of diameters measuring in the millimetres. Valve densities using direct mechanical actuation

will, therefore, be an order of magnitude less than can be achieved using Quake valves[113]. Furthermore, the presence of a large, opaque object on one side of the microfluidic chip prevents any through chip spectrophotometric measurements that might be taken. Quake valves, on the other hand, consist entirely of uniform optically clear PDMS.

3.15. Methods of making droplets within microfluidic devices

Several methods exist to generate droplets within a microfluidic device[114]. Most depend on simple collision of two streams of fluid of different phases in junctions of specific geometry. The making and control of droplets requires functionality from the previously discussed microfluidic actuation and pumping sections.

Droplet generation can be achieved in a microfluidic device with a simple T-junction[115], a tube-within-a-tube or by flow-focussing. The size and frequency of the droplets can be controlled by the relative flow rates of the two phases, as well as their properties; addition of surfactants stabilises individual droplets. Generally such droplet generation is performed continuously, using off chip pumps to continuously generate droplets at a passive interface. Incorporation of a feedback mechanism can allow automatic changing of droplet size over a period of time[116]. Significant, instantaneous changing of the droplet size or rate of generation, however, may require a pause and/or reconfiguration of the pump. Droplet generation on demand, with most methods of fluid pumping and droplet generation, is not easily possible.

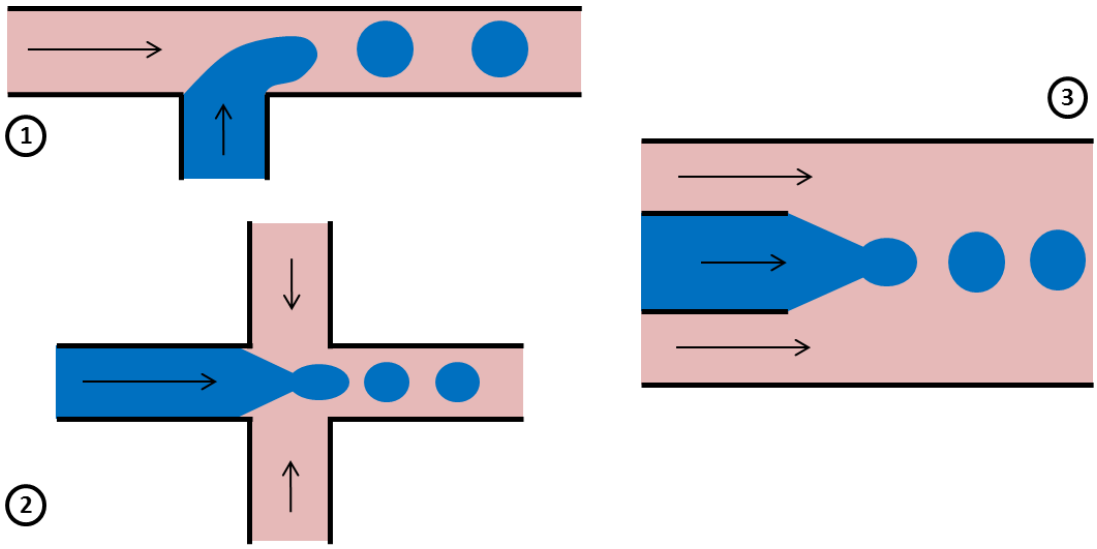


Figure 3.4: Three channels geometries capable of producing droplets from two continuously flowing phases (pink and blue). 1) Simple T-junction. 2) Flow focussing. 3) Tube within a tube. Flow of each phase is shown by arrows.

Similar to the principle that drives the print heads in inkjet printers, piezoelectric actuators can be used to create droplets on demand[117,118]. The piezoelectric actuators are relatively bulky and must be positioned directly above the reservoirs they actuate. The system is capable of producing single droplets with a range of sizes that depend on the size the aperture, the frequency of actuation and the actuation and relaxation time.

Droplets can be generated with a simple pneumatic 'chopper'[119]. In this system, two continuously co-flowing fluid phases flow through a series of valves. By actuating the valves, the stream is cut into pieces, generating droplets that depend on the space between the actuators and the flow rate of the two phases. This setup can generate droplets on demand, but the system requires running for a period before consistently sized droplets are produced. Furthermore, the size of the droplets is dependent on the spacing between the valves, which means that generation of significantly larger or smaller droplets would require a completely new device.

Off chip valves can be used to accurately control the flow of fluids into microfluidic chips resulting in different concentrations of two components in mixed droplets[120]. The valves used in these cases must be specifically modified to permit low fluid flow rate.

Another interesting method of making droplets is to use an oscillating needle that moves between the interface between two phases in a reservoir[121,122]. By varying the rate and position of the centre of oscillation, the droplet size and frequency can be easily varied. With sufficiently large reservoirs, continuous operation can be achieved without the centre of oscillation requiring adjustment throughout the experiment.

3.16. Methods of controlling droplets within microfluidic devices

Control of droplet direction is critical for microfluidic devices aimed at more than just one purpose. Typically microfluidic devices are designed for single experiments with geometries of channels being designed to perform single functions. If a droplet microfluidic device is required, for instance, to mix a subset of available substrates together then the device must have the ability to select or otherwise filter which droplets are allowed to merge. The methods for controlling droplets within microfluidic devices fall into two separate categories; passive and active[123].

Passive methods of controlling droplets involve clever design of channel geometry in order to derive a desired outcome. Splitters, mergers and sorters can be built by using selective junctions that respond to the pressure changes caused by droplet size and/or presence in a given channel[124–126]. Aqueous droplets will absorb onto hydrophilic wall sections in normally hydrophobic PDMS devices (akin to normal phase chromatography) and this effect can be used to mediate passive droplet merging[127]. These interesting designs show that a wide range of functionalities are possible using only passive effects. The design of the channels, however, depends on the size and frequency of droplets. A new chip is required if a significant change in the frequency or size of the droplets is necessary.

Active methods, by contrast, are flexible and are capable of delivering the control without need for reconfiguration of a whole device. Active membranes[128], magnetic[129,130], optical[131,132], selective flow[133] or electrical fields[134–136] can be employed to deflect droplets down different outlets. In addition to controlling droplet deflection, droplet merging can be achieved using electrical potentials[137].

The primary issue with droplet merging is ensuring that the wrong droplets do not merge or that a droplet is not contaminated when passing by an inlet. Multi-junction devices can be used to minimise droplet contamination[138].

3.17. Droplet monitoring

Measurements inside microfluidic devices are typically performed using colour changes monitored by external devices[139], such as confocal microscopes[140]. Monitoring systems of this type can be used to make high resolution measurements of continuous flow[141] and droplet mixing[142]. Integration of monitoring systems, such as optical fibres[143–145] or microlenses[146], shrinks the system considerably and can improve signal:noise ratios[147]. Integrated electrodes can be used to make droplet size estimations[148] or for cell culture monitoring[149].

On-chip integration of monitoring devices as described above is necessary in order to achieve a true lab-on-a-chip (LoC) or micro-total-analysis-system (µTAS). Investigation of biological systems often requires expensive enzymes and reagents that are often only available in 10's of µL. Therefore the benefits of miniaturisation in terms of reagent use and parallel reactions can be realised without all the components being shrunk to fit onto such a device.

3.18. Applications of microfluidics to synthetic and systems biology

3.18.1. Synthetic biology

Synthetic biology involves the use of artificial environments or reagents to generate biological products. One highly publicised example of synthetic biology is the production of an entire 0.5 Mbp genome for the bacteria *Myocoplasma genitalium*[150]. By miniaturising this synthetic process it is hoped that larger constructs can be synthesised quickly and accurately[151]. Parallelising assays will allow brute force, or boot-strap methods, to become viable in the context of synthetic biology[152,153]. Examples of application of microfluidics to synthetic biology include high throughput synthesis artificial DNA sequences[154] and directed evolution of DNA sequences[134].

The principle reaction of synthetic biology, PCR, requires temperature cycling as the steps of melting, primer annealing and extension occur optimally at different temperatures. The large surface area:volume relationship implicit in microfluidic devices allows for very efficient heat transfer suiting this reaction to miniaturisation[155]. Droplet microfluidics is particularly suited to PCR reactions, which are extremely sensitive to contamination[156,157].

PCR microfluidics involves two steps: 1) Appropriate primers and template must be brought together with common components such as enzyme, substrates and buffer and be subjected to temperature cycling. 2) Products must be either measured on chip, usually by capillary electrophoresis, or be taken off chip for further analysis.

Examples of PCR microfluidics from range simple plugs in capillary tubes being moved between heated locations[158] through convection driven flow[159] to the more complex droplets moving around segments of a circle arranged around a Peltier heating element[160]. These papers exemplify what is a commonly published theme; PCR, even single molecule PCR, is feasible on the microfluidic scale. Interestingly, microfluidic reactors have not, as

yet, permeated the actual biological lab bench. Systems like Raindance Technology's RainStorm and ThunderStorm devices[116,161] as well as Dolomite's custom etched glass and Mitos chips[162] can create large numbers of droplets in defined geometries but do not usually offer the ability to vary the composition of each droplet during an experiment.

Gene synthesis has been performed in a continuous flow microfluidic device[110]. Interestingly, this study used a mould created using Eden 350 (Objet Geometries) from which a PDMS cast was taken and involved several of the previously discussed methods to control fluid flow: A syringe pump, hydrogel valves, primers immobilised on beads and magnetic separation were employed.

3.18.2. Systems biology

Systems biology is the study of the complex, high order interactions that occur in biological systems such as cells. Careful comparison of combinatorial experiments, that involve pulling multiple levers within the system simultaneously, are required to find how each input interrelates to develop the output. Traditionally, many carefully controlled experiments must be performed in order to understand the many interactions that occur within a cell system[163]. Microfluidics based approaches are readily applicable to repetitive and parallelisable experiments of this nature[151,164] and for generating the large amount of input variance for combinatorial experiments.

Microfluidics is also suited to single cell analysis that is necessary in order to elucidate stochastic regulatory mechanisms[165]: Single cells can be trapped, subjected to flows containing active factors and the effect on sub-cellular protein localisation followed[166]. By trapping single cells, complexities arising from population effects of intercellular communication are avoided, resulting in a clear system to study. The signal processing performed by cells as they interpret an input can be investigated on a single cell

basis, yielding important information about how signal transduction can occur particularly with reference to oscillatory signals[167,168].

Continuous flow microfluidic devices from the Quake lab have been used to identify and characterise putative transcription factor binding sites by affinity analysis[169,170]. These devices were able to analyse an 8 bp oligonucleotide library containing ~1500 sequences for binding to target protein in one experiment.

3.19. Review conclusions

The production of a microfluidic device requires the selection of appropriate methods in order to shape and pattern the appropriate materials to produce the geometry and features necessary to achieve the desired effects. The material should possess the correct mechanical properties without absorbing from or dissolving into the media within the device. Active features such as valves and pumps require flexible membranes or other considerations (EOPs, EWOD) and must be incorporated in the appropriate material. Accurate manipulation of fluids within microfluidic devices is crucial and the mechanism, such as electric or magnetic fields, active membrane/valve etc., should be appropriate for the task. Microfluidic devices are typically produced by aligning and bonding multiple layers each with appropriate material properties. Bonding of the layers in this case must be sufficient to produce a device that can operate at the operating pressure expected. By careful selection of the appropriate choices in the above categories, an appropriate device can be fabricated.

New manufacturing methods, such as ALM, offer the opportunity to produce devices composed of multiple layers in a single manufacturing step. ALM, therefore, promises to make microfluidic fabrication more simple and rapid. The range of materials that are suitable for ALM, however, is more restricted than those that can be used in the more traditional lithographic or micromachining techniques. There are surprisingly few

examples in the literature of microfluidics made by ALM, either directly or indirectly. The key challenge is to develop the MSL, a type of ALM technology, for the fabrication of microfluidic systems. Although MSL microfluidic systems can be used directly, only a limited range of materials and therefore properties are suitable. Therefore, fabrication of microfluidic devices indirectly through casting of a secondary material is also necessary. Chapter 5 describes the use of MSL for the direct fabrication of microfluidic flow cells and also the development of the MSL process for the production of moulds for PDMS casting. Chapter 6 then describes the use of the techniques developed in chapter 5 for the fabrication of a PDMS microfluidic chip by multilayer soft lithography.

In the future, microfluidics will be able to be applied to diverse areas of molecular and systems biology. At one end of the scale, paper microfluidic devices are inexpensive and can be used in remote areas to obtain medical diagnoses whilst at the other end commercially available high throughput machines are capable of producing 1000's of droplets per second. The development of novel, high throughput gene synthesis and assembly technology as well as point-of-care disposable diagnostic devices is likely to have a significant effect on modern biological research as well as medicine[1].

3.20. References

1. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442**, 368–373 (2006).
2. Stephen C. Terry, Jerman, J. H. & Angell, J. B. A Gas Chromatographic Air Analyzer Fabricated in a Silicon Wafer. *IEEE* **26**, 1880–1886 (1979).
3. Gravesen, P., Branebjerg, J. & Jensen, O. S. Microfluidics. A Review. *Journal of Micromechanics and Microengineering* **3**, 168–182 (1993).
4. Beebe, D. J., Mensing, G. a & Walker, G. M. Physics and applications of microfluidics in biology. *Annual review of biomedical engineering* **4**, 261–86 (2002).
5. McDonald, J. C. & Whitesides, G. M. Poly(dimethylsiloxane) as a material for fabricating microfluidic devices. *Accounts of chemical research* **35**, 491–9 (2002).
6. Unger, M. A., Chou, H.-P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic Microfabricated Valves and Pumps by Multilayer Soft Lithography. *Science* **288**, 113–116 (2000).
7. Pollack, M. G., Shenderov, a D. & Fair, R. B. Electrowetting-based actuation of droplets for integrated microfluidics. *Lab on a chip* **2**, 96–101 (2002).
8. Cho, S. K., Moon, H. & Kim, C. Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits. *Journal of Microelectromechanical Systems* **12**, 70–80 (2003).
9. Huft, J., Da Costa, D. J., Walker, D. & Hansen, C. L. Three-dimensional large-scale microfluidic integration by laser ablation of interlayer connections. *Lab on a chip* 2358–2365 (2010).doi:10.1039/c004051g
10. Gibson, I., Rosen, D. W. & Stucker, B. *Additive Manufacturing Technologies: Rapid Prototyping to Direct Digital Manufacturing*. (Springer: 2009).doi:10.1007/978-1-4419-1120-9
11. Farsari, M. & Chichkov, B. N. Materials processing: Two-photon fabrication. *Nature Photonics* **3**, 450–452 (2009).
12. Lee, K.-S., Yang, D.-Y., Park, S. H. & Kim, R. H. Recent developments in the use of two-photon polymerization in precise 2D and 3D microfabrications. *Polymers for Advanced Technologies* **17**, 72–82 (2006).
13. Snowden, M. E., King, P. H., Covington, J. A., Macpherson, J. V & Unwin, P. R. Fabrication of versatile channel flow cells for quantitative electroanalysis using prototyping. *Analytical chemistry* **82**, 3124–31 (2010).
14. Fisher, R. D., Mbogoro, M. M., Snowden, M. E., Joseph, M. B., Covington, J. A., Unwin, P. R. & Walton, R. I. Dissolution Kinetics of Polycrystalline Calcium Sulfate-Based Materials: Influence of Chemical Modification. *ACS applied materials & interfaces* **3**, 3528–3537 (2011).
15. Snowden, M. E., Unwin, P. R. & Macpherson, J. V. Single walled carbon nanotube channel flow electrode: Hydrodynamic voltammetry at the nanomolar level. *Electrochemistry Communications* **13**, 186–189 (2011).
16. Leigh, S. J., Purssell, C. P., Bowen, J., Hutchins, D. A., Covington, J. A. & Billson, D. R. A miniature flow sensor fabricated by micro-stereolithography employing a magnetite/acrylic nanocomposite resin. *Sensors and Actuators A: Physical* **168**, 66–71 (2011).
17. EnvisionTec EnvisionTec Technology Overview. (2012).at <http://www.envisiontec.de/index.php?page=technology_overview>
18. 3D Systems 3D systems - SLA. (2012).at <http://production3dprinters.com/sla/stereolithography>
19. 3D Systems 3D Systems - SLS. (2012).at <http://production3dprinters.com/sls/selective-laser-sintering>
20. Stratasys Stratasys - FDM. (2012).at <http://www.stratasys.com/Products/Overview.aspx>
21. BFB BFB - FDM. (2012).at <http://www.bitsfrombytes.com/content/3dtouch-3d-printer>
22. 3D Systems 3D Sytems - 3D Printers. (2012).at <http://printin3d.com/3d-printers>
23. Objet Objet - 3D printers. (2012).at <http://objet.com/3d-printers>

24. ZCorp ZCorp - 3D printers. (2012).at <http://www.zcorp.com/en/Products/3D-Printers/spage.aspx>
25. Solido3D Solido 3D. (2012).at <http://www.solido3d.com/solidoContent.aspx?PageID=31>
26. Calvert, P. Inkjet Printing for Materials and Devices. *Chemistry of Materials* **13**, 3299–3305 (2001).
27. Chen, C., Breslauer, D. N., Luna, J. I., Grimes, A., Chin, W., Lee, P. & Khine, M. Shrinky-Dink microfluidics : 3D polystyrene chips. *Society* **8**, 622–624 (2008).
28. Griffith, M. Understanding thermal behavior in the LENS process. *Materials & Design* **20**, 107–113 (1999).
29. Lee, J. J.-H., Park, S.-A., Park, K., Kim, J.-H., Kim, K.-S. & Kim, W. Fabrication and characterization of 3D scaffold using 3D plotting system. *Chinese Science Bulletin* **55**, 94–98 (2010).
30. Becker, H. & Gärtner, C. Polymer microfabrication technologies for microfluidic systems. *Analytical and bioanalytical chemistry* **390**, 89–111 (2008).
31. Baechi, D., Buser, R. & Dual, J. A high density microchannel network with integrated valves and photodiodes. *Sensors and Actuators A: Physical* **95**, 77–83 (2002).
32. Castanoalvarez, M., Pozoayuso, D., Garciagranda, M., Fernandezabedul, M., Rodriguezgarcia, J. & Costagarcia, a Critical points in the fabrication of microfluidic devices on glass substrates. *Sensors and Actuators B: Chemical* **130**, 436–448 (2008).
33. Hua, Z., Xia, Y., Srivannavit, O., Rouillard, J.-M., Zhou, X., Gao, X. & Gulari, E. A versatile microreactor platform featuring a chemical-resistant microvalve array for addressable multiplex syntheses and assays. *Journal of Micromechanics and Microengineering* **16**, 1433–1443 (2006).
34. Mehta, G., Lee, J., Cha, W., Tung, Y.-C., Linderman, J. J. & Takayama, S. Hard top soft bottom microfluidic devices for cell culture and chemical analysis. *Analytical chemistry* **81**, 3714–22 (2009).
35. Paul, B. K., Abhinkar, B. S. & Lee, S. High pressure hermetic compression seals for embedding elastomeric membrane microvalves in polymer microfluidic devices. *Precision Engineering* **35**, 348–354 (2011).
36. Zeng, Y., Novak, R., Shuga, J., Smith, M. T. & Mathies, R. a High-performance single cell genetic analysis using microfluidic emulsion generator arrays. *Analytical chemistry* **82**, 3183–90 (2010).
37. Quake, S. R. From Micro- to Nanofabrication with Soft Materials. *Science* **290**, 1536–1540 (2000).
38. McDonald, J. C., Duffy, D. C., Anderson, J. R. & Chiu, D. T. General Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* **21**, 27–40 (2000).
39. Toepke, M. W. & Beebe, D. J. PDMS absorption of small molecules and consequences in microfluidic applications. *Lab on a chip* **6**, 1484–6 (2006).
40. Rolland, J. P., Dam, R. M. Van, Schorzman, D. A., Quake, S. R. & DeSimone, J. M. Solvent-Resistant Photocurable " Liquid Teflon " for Microfluidic Device Fabrication. *Journal of the American Chemical Society* **126**, 2322–2323 (2004).
41. Huang, Y., Castrataro, P., Lee, C. & Quake, S. R. Solvent resistant microfluidic DNA synthesizer. *Lab on a Chip* **7**, 24–26 (2007).
42. Ren, K., Dai, W., Zhou, J., Su, J. & Wu, H. Whole-Teflon microfluidic chips. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 8162–6 (2011).
43. El-Kalla, E. H., Sayyah, S. M., Abifi, H. H. & Saeed, A. F. Ultraviolet-visible spectroscopic studies of poly(methyl methacrylate) doped with some luminescent materials. *Acta Polymerica* **40**, (1989).
44. Sun, X., Peeni, B. a, Yang, W., Becerril, H. a & Woolley, A. T. Rapid prototyping of poly(methyl methacrylate) microfluidic systems using solvent imprinting and bonding. *Journal of chromatography. A* **1162**, 162–6 (2007).

45. Romoli, L., Tantussi, G. & Dini, G. Experimental approach to the laser machining of PMMA substrates for the fabrication of microfluidic devices. *Optics and Lasers in Engineering* **49**, 419–427 (2011).
46. Hong, T.-F., Ju, W.-J., Wu, M.-C., Tai, C.-H., Tsai, C.-H. & Fu, L.-M. Rapid prototyping of PMMA microfluidic chips utilizing a CO2 laser. *Microfluidics and Nanofluidics* **9**, 1125–1133 (2010).
47. Li, B., Gueit, A. & Sharon, A. Thickness management in three-dimensional laser manufacturing of suspended structures in a single SU-8 layer. *Review of Scientific Instruments* **77**, (2006).
48. Schafer, D., Gibson, E. A., Salim, E. A., Palmer, A. E. & Squier, J. Microfluidic cell counter with embedded optical fibers fabricated by femtosecond laser ablation and anodic bonding. *Optics express* **17**, 6068–6073 (2009).
49. Becker, H. & Heim, U. Hot embossing as a method for the fabrication of polymer high aspect ratio structures. *Sensors and Actuators A: Physical* **83**, 130–135 (2000).
50. Worgull, M. *Hot Embossing: Theory and Technology of Microreplication*. (Elsevier Academic Press: 2009).
51. Kuo, J. S., Zhao, Y., Ng, L., Yen, G. S., Lorenz, R. M., Lim, D. S. W. & Chiu, D. T. Microfabricating high-aspect-ratio structures in polyurethane-methacrylate (PUMA) disposable microfluidic devices. *Lab on a chip* **9**, 1951–6 (2009).
52. Hung, L.-H., Lin, R. & Lee, A. P. Rapid microfabrication of solvent-resistant biocompatible microfluidic devices. *Lab on a chip* **8**, 983–7 (2008).
53. Do, J., Zhang, J. Y. & Klapperich, C. M. Maskless writing of microfluidics: Rapid prototyping of 3D microfluidics using scratch on a polymer substrate. *Robotics and Computer-Integrated Manufacturing* 6–9 (2010).doi:10.1016/j.rcim.2010.06.004
54. Martinez, A. W., Phillips, S. T. & Whitesides, G. M. Three-dimensional microfluidic devices fabricated in layered paper and tape. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19606–11 (2008).
55. Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic Large-Scale Integration. *Science* **298**, 580–584 (2002).
56. Gorris, H. H. & Walt, D. R. Analytical Chemistry on the Femtoliter Scale. *Angewandte Chemie (International ed. in English)* 3880 – 3895 (2010).doi:10.1002/anie.200906417
57. Mazutis, L., Baret, J.-C., Treacy, P., Skhiri, Y., Araghi, F., Ryckelynck, M. & Griffiths, A. D. Multi-step microfluidic droplet processing: kinetic analysis of an in vitro translated enzyme. *Lab on a Chip* **9**, 2902–2908 (2009).
58. Cai, D., Neyer, a, Kuckuk, R. & Heise, H. Optical absorption in transparent PDMS materials applied for multimode waveguides fabrication. *Optical Materials* **30**, 1157–1161 (2008).
59. Fujii, T. PDMS-based microfluidic devices for biomedical applications. *Microelectronic Engineering* **61-62**, 907–914 (2002).
60. Melin, J. & Quake, S. R. Microfluidic Large-Scale Integration : The Evolution of Design Rules for Biological Automation. *Annual Reviews of Biophysical and Biomolecular Structures* **36**, 213–231 (2007).
61. Yuen, P. K. SmartBuild-a truly plug-n-play modular microfluidic system. *Lab on a chip* **8**, 1374–8 (2008).
62. Eddings, M. A., Johnson, M. A. & Gale, B. K. Determining the optimal PDMS–PDMS bonding technique for microfluidic devices. *Journal of Micromechanics and Microengineering* **18**, (2008).
63. SyringePump.com SyringePump.com - Microfluidics. (2012).at <http://www.syringepump.com/Micro.php>
64. Dolomite Dolomite - Syringe Pumps. (2012).at <http://www.dolomite-microfluidics.com/EUR/webshop/pumps-syringe-pumps-c-21_43>
65. Cole-Palmer Cole-Palmer - Mircofluidic Pumps. (2012).at <http://www.coleparmer.co.uk/Category/Microfluidic_Pumps/49217>
66. Dolomite Dolomite - Peristaltic Pumps. (2012).at <http://www.dolomite-microfluidics.com/webshop/pumps-peristaltic-pumps-c-38_40>

67. Apparatus, H. Harvard Apparatus - Syringe Pumps. (2012).at <http://www.harvardapparatus.com/webapp/wcs/stores/servlet/haicat1_10001_11051_37295_-1_HAI_Categories_N>
68. Cole-Palmer Cole-Palmer - Persistaltic Pumps. (2012).at <http://www.coleparmer.com/buy/category/microfluidic-peristaltic-pump>
69. Dolomite Dolomite - Pressure Pumps. (2012).at <http://www.dolomite-microfluidics.com/webshop/pumps-pressure-pumps-c-38_41>
70. Laser, D. J. & Santiago, J. G. A review of micropumps. *Journal of Micromechanics and Microengineering* **14**, R35–R64 (2004).
71. King, P. H. Towards Rapid 3D Direct Manufacture of Biomechanical Microstructures. (2009).
72. Chou, H.-P., Unger, M. A. & Quake, S. R. A Microfabricated Rotary Pump. *Biomedical Microdevices* **3**, 323–330 (2001).
73. Hong, J. W., Chen, Y., Anderson, W. F. & Quake, S. R. Molecular biology on a microfluidic chip. *Journal of Physics: Condensed Matter* **18**, S691–S701 (2006).
74. Wu, M.-H., Chang, Y.-H., Liu, Y.-T., Chen, Y.-M., Wang, S.-S., Wang, H.-Y., Lai, C.-S. & Pan, T.-M. Development of high throughput microfluidic cell culture chip for perfusion 3-dimensional cell culture-based chemosensitivity assay. *Sensors and Actuators B: Chemical* (2010).doi:10.1016/j.snb.2010.11.027
75. Wu, M.-H., Huang, S.-B., Cui, Z., Cui, Z. & Lee, G.-B. A high throughput perfusion-based microbioreactor platform integrated with pneumatic micropumps for three-dimensional cell culture. *Biomedical microdevices* **10**, 309–19 (2008).
76. Zhou, Y. & Amirouche, F. An Electromagnetically-Actuated All-PDMS Valveless Micropump for Drug Delivery. *Micromachines* **2**, 345–355 (2011).
77. King, P. & Covington, J. A novel monolithic microactuator fabricated by 3D rapid direct manufacture. *Procedia Chemistry* **1**, 1163–1166 (2009).
78. Wang, X., Cheng, C., Wang, S. & Liu, S. Electroosmotic pumps and their applications in microfluidic systems. *Microfluidics and nanofluidics* **6**, 145–162 (2009).
79. Probstein, R. F. *Physicochemical Hydrodynamics*. (1964).
80. Wang, P., Chen, Z. & Chang, H. A new electro-osmotic pump based on silica monoliths. *Sensors And Actuators* **113**, 500–509 (2006).
81. Mutlu, S., Yu, C., Selvaganapathy, P., Svec, F., Mastrangelo, C. H. & Frechet, J. M. J. Micromachined porous polymer for bubble free electro-osmotic pump. *IEEE* 19–23 (2002).
82. Walsh, Z., Abele, S., Lawless, B., Heger, D., Klán, P., Breadmore, M. C., Paull, B. & Macka, M. Photoinitiated polymerisation of monolithic stationary phases in polyimide coated capillaries using visible region LEDs. *Chemical communications (Cambridge, England)* **7345**, 6504–6 (2008).
83. Oakley, J. a, Shaw, K. J., Docker, P. T., Dyer, C. E., Greenman, J., Greenway, G. M. & Haswell, S. J. Development of a bi-functional silica monolith for electro-osmotic pumping and DNA clean-up/extraction using gel-supported reagents in a microfluidic device. *Lab on a chip* **9**, 1596–600 (2009).
84. Meyvantsson, I., Warrick, J. W., Hayes, S., Skoien, A. & Beebe, D. J. Automated cell culture in high density tubeless microfluidic device arrays. *Lab on a chip* **8**, 717–24 (2008).
85. Eddings, M. A. & Gale, B. K. A PDMS-based gas permeation pump for on-chip fluid handling in microfluidic devices. *Journal of Micromechanics and Microengineering* **16**, 2396–2402 (2006).
86. Johnson, M., Liddiard, G., Eddings, M. & Gale, B. Bubble inclusion and removal using PDMS membrane-based gas permeation for applications in pumping, valving and mixing in microfluidic devices. *Journal of Micromechanics and Microengineering* **19**, 095011 (2009).
87. Suzuki, H. & Yoneyama, R. Integrated microfluidic system with electrochemically actuated on-chip pumps and valves. *Sensors and Actuators B: Chemical* **96**, 38–45 (2003).
88. Xie, J., Miao, Y., Shih, J., He, Q., Liu, J., Tai, Y.-C. & Lee, T. D. An electrochemical pumping system for on-chip gradient generation. *Analytical chemistry* **76**, 3756–63 (2004).

89. Heo, Y. S., Cabrera, L. M., Song, J. W., Futai, N., Tung, Y.-C., Smith, G. D. & Takayama, S. Characterization and resolution of evaporation-mediated osmolality shifts that constrain microfluidic cell culture in poly(dimethylsiloxane) devices. *Analytical chemistry* **79**, 1126–34 (2007).

90. Lei, Y., Liu, Y., Wang, W., Wu, W. & Li, Z. Studies on Parylene C-caulked PDMS (pcPDMS) for low permeability required microfluidics applications. *Lab on a chip* **11**, 1385–8 (2011).

91. Ren, K., Zhao, Y., Su, J., Ryan, D. & Wu, H. Convenient method for modifying poly(dimethylsiloxane) to be airtight and resistive against absorption of small molecules. *Analytical chemistry* **82**, 5965–71 (2010).

92. DeMello, A. J. Control and detection of chemical reactions in microfluidic systems. *Nature* **442**, 394–402 (2006).

93. Dittrich, P. S. & Manz, A. Single-molecule fluorescence detection in microfluidic channels--the Holy Grail in muTAS? *Analytical and bioanalytical chemistry* **382**, 1771–82 (2005).

94. Leung, S.-A., Winkle, R. F., Wootton, R. C. R. & deMello, A. J. A method for rapid reaction optimisation in continuous-flow microfluidic reactors using online Raman spectroscopic detection. *The Analyst* **130**, 46–51 (2005).

95. Cecchini, M. P., Hong, J., Lim, C., Choo, J., Albrecht, T., Andrew, J. & Edel, J. B. Ultrafast Surface Enhanced Resonance Raman Scattering Detection in Droplet-Based Microfluidic Systems. *Analytical Chemsitry* **83**, 3076–3081 (2011).

96. Lavrik, N. V, Taylor, L. T. & Sepaniak, M. J. Nanotechnology and chip level systems for pressure driven liquid chromatography and emerging analytical separation techniques: a review. *Analytica chimica acta* **694**, 6–20 (2011).

97. Ohla, S. & Belder, D. Chip-based separation devices coupled to mass spectrometry. *Current opinion in chemical biology* **16**, 453–9 (2012).

98. Sun, T. & Morgan, H. Single-cell microfluidic impedance cytometry: a review. *Microfluidics and Nanofluidics* **8**, 423–443 (2010).

99. Ogburn, E. T., Dziewatkoski, M., Moles, D., Johnson, J. M. & Heineman, W. R. Microfabricated Electrochemical Detector for High-Performance Liquid Chromatography. *Analytical chemistry* (2011).doi:10.1021/ac200476j

100. De Volder, M. & Reynaerts, D. Pneumatic and hydraulic microactuators: a review. *Journal of Micromechanics and Microengineering* **20**, 043001 (2010).

101. Xia, C., Lee, H. & Fang, N. Solvent-driven polymeric micro beam device. *Journal of Micromechanics and Microengineering* **20**, 085030 (2010).

102. Krulevitch, P., Lee, A. P., Ramsey, P. B., Trevino, J. C., Hamilton, J. & N, M. A. Thin film shape memory alloy microactuators. *Journal of Microelectromechanical Systems* **5**, 270–282 (1996).

103. Makino, E., Mitsuya, T. & Shibata, T. Fabrication of TiNi shape memory micropump. *Sensors and Actuators A: Physical* **88**, 256–262 (2001).

104. Vyawahare, S., Sitaula, S., Martin, S., Adalian, D. & Scherer, A. Electronic control of elastomeric microfluidic circuits with shape memory actuators. *Lab on a chip* **8**, 1530–5 (2008).

105. Prentice-Mott, H., Toner, M. & Irimia, D. Microfluidic proportional flow controller. *Journal of Micromechanics and Microengineering* **20**, 115020 (2010).

106. Selva, B., Miralles, V. & Jullien, M. Thermocapillary actuation by optimized resistor pattern: Bubbles and droplets displacing, switching and trapping. *Lab on a Chip* **10**, 1835–1840 (2010).

107. Pal, R., Yang, M., Johnson, B. N., Burke, D. T. & Burns, M. a Phase change microvalve for integrated devices. *Analytical chemistry* **76**, 3740–8 (2004).

108. Yang, B. & Lin, Q. A Latchable Phase-Change Microvalve With Integrated Heaters. *Journal of Microelectromechanical Systems* **18**, 860–867 (2009).

109. Oh, K. W., Namkoong, K. & Park, C. AS phase-change microvalve using a meltablemagnetic material: Ferrowax. *9th International Conference on Miniturized Systems for Chemistry and Life Sciences* 554–556 (2005).

110. Huang, M. C., Ye, H., Kuan, Y. K., Li, M.-H. & Ying, J. Y. Integrated two-step gene synthesis in a microfluidic device. *Lab on a chip* **9**, 276–85 (2009).

111. Gu, W., Zhu, X., Futai, N., Cho, B. S. & Takayama, S. Computerized microfluidic cell culture using elastomeric channels and Braille displays. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15861–6 (2004).

112. Tamanaha, C. R., Whitman, L. J. & Colton, R. J. Hybrid macro-micro fluidics system for a chip-based biosensor. *Journal of Micromechanics and Microengineering* **12**, 347–347 (2002).

113. Gómez-Sjöberg, R., Leyrat, A. A., Pirone, D. M., Chen, C. S. & Quake, S. R. Versatile, fully automated, microfluidic cell culture system. *Analytical chemistry* **79**, 8557–63 (2007).

114. Christopher, G. F. & Anna, S. L. Microfluidic methods for generating continuous droplet streams. *Journal of Physics D: Applied Physics* **40**, R319–R336 (2007).

115. Thorsen, T., Roberts, R., Arnold, F. & Quake, S. Dynamic Pattern Formation in a Vesicle-Generating Microfluidic Device. *Physical Review Letters* **86**, 4163–4166 (2001).

116. Miller, E., Rotea, M. & Rothstein, J. P. Microfluidic device incorporating closed loop feedback control for uniform and tunable production of micro-droplets. *Lab on a Chip* **10**, 1293–1301 (2010).

117. Xu, J. & Attinger, D. Drop on demand in a microfluidic chip. *Journal of Micromechanics and Microengineering* **18**, 065020 (2008).

118. Shemesh, J., Bransky, A., Khoury, M. & Levenberg, S. Advanced microfluidic droplet manipulation based on piezoelectric actuation. *Biomedical microdevices* **12**, 907–14 (2010).

119. Chen, C.-T. & Lee, G.-B. Formation of Microdroplets in Liquids Utilizing Active Pneumatic Choppers on a Microfluidic Chip. *Journal of Microelectromechanical Systems* **15**, 1492–1498 (2006).

120. Churski, K., Korczyk, P. & Garstecki, P. High-throughput automated droplet microfluidic system for screening of reaction conditions. *Lab on a chip* **10**, 816–8 (2010).

121. Chabert, M., Dorfman, K. D., de Cremoux, P., Roeraade, J. & Viovy, J.-L. Automated microdroplet platform for sample manipulation and polymerase chain reaction. *Analytical chemistry* **78**, 7722–8 (2006).

122. Du, W.-B., Sun, M., Gu, S.-Q., Zhu, Y. & Fang, Q. Automated microfluidic screening assay platform based on DropLab. *Analytical chemistry* **82**, 9941–7 (2010).

123. Teh, S.-Y., Lin, R., Hung, L.-H. & Lee, A. P. Droplet microfluidics. *Lab on a chip* **8**, 198–220 (2008).

124. Tan, Y.-C., Fisher, J. S., Lee, A. I., Cristini, V. & Lee, A. P. Design of microfluidic channel geometries for the control of droplet volume, chemical concentration, and sorting. *Lab on a chip* **4**, 292–8 (2004).

125. Cristobal, G., Benoit, J.-P., Joanicot, M. & Ajdari, A. Microfluidic bypass for efficient passive regulation of droplet traffic at a junction. *Applied Physics Letters* **89**, 3–5 (2006).

126. Niu, X., Gulati, S., Edel, J. B. & deMello, A. J. Pillar-induced droplet merging in microfluidic circuits. *Lab on a Chip* **8**, 1837–1841 (2008).

127. Fidalgo, L. M., Abell, C. & Huck, W. T. S. Surface-induced droplet fusion in microfluidic devices. *Lab on a chip* **7**, 984–6 (2007).

128. Abate, A. R., Agresti, J. J. & Weitz, D. a. Microfluidic sorting with high-speed single-layer membrane valves. *Applied Physics Letters* **96**, 203509 (2010).

129. Lou, X., Qian, J., Xiao, Y., Viel, L., Gerdon, A. E., Lagally, E. T., Atzberger, P., Tarasow, T. M., Heeger, A. J. & Soh, H. T. Micromagnetic selection of aptamers in microfluidic channels. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2989–94 (2009).

130. Vojtísek, M., Iles, A. & Pamme, N. Rapid, multistep on-chip DNA hybridisation in continuous flow on magnetic particles. *Biosensors & bioelectronics* **25**, 2172–6 (2010).

131. Baroud, C. N., de Saint Vincent, M. R. & Delville, J.-P. An optical toolbox for total control of droplet microfluidics. *Lab on a chip* **7**, 1029–33 (2007).

132. Wakamoto, Y., Inoue, I., Moriguchi, H. & Yasuda, K. Analysis of single-cell differences by use of an on-chip microculture system and optical trapping. *Fresenius' Journal of Analytical Chemistry* **371**, 276–281 (2001).

133. Kruger, J., Singh, K., O'Neill, A., Jackson, C., Morrison, A. & O'Brien, P. Development of a microfluidic device for fluorescence activated cell sorting. *Journal of Micromechanics and Microengineering* **12**, 486–494 (2002).

134. Agresti, J. J., Antipov, E., Abate, A. R., Ahn, K., Rowat, A. C., Baret, J.-C., Marquez, M., Klibanov, A. M., Griffiths, A. D. & Weitz, D. a Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4004–9 (2010).

135. Link, D. R., Grasland-Mongrain, E., Duri, A., Sarrazin, F. & Cheng, Z. Electric Control of Droplets in Microfluidic Devices. *Angewandte Chemie (International ed. in English)* **45**, 2556–2560 (2006).

136. Ahn, K., Kerbage, C., Hunt, T. P., Westervelt, R. M., Link, D. R. & Weitz, D. a. Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices. *Applied Physics Letters* **88**, 024104 (2006).

137. Tan, W.-H. & Takeuchi, S. Timing controllable electrofusion device for aqueous droplet-based microreactors. *Lab on a Chip* **6**, 757–763 (2006).

138. Li, L., Boedicker, J. Q. & Ismagilov, R. F. Using a Multijunction Device to Inject Substrate into an Array of Preformed Plugs without Cross-Contamination. *Analytical Chemistry* **79**, 2756–2761 (2007).

139. Craighead, H. Future lab-on-a-chip technologies for interrogating individual molecules. *Nature* **442**, 387–93 (2006).

140. Fang, W.-F., Hsu, M.-H., Chen, Y.-T. & Yang, J.-T. Characterization of microfluidic mixing and reaction in microchannels via analysis of cross-sectional patterns. *Biomicrofluidics* **5**, 014111 (2011).

141. Atencia, J. & Beebe, D. J. Controlled microfluidic interfaces. *Nature* **437**, 648–655 (2005).

142. Casadevall i Solvas, X., Srisa-Art, M., DeMello, A. J. & Edel, J. B. Mapping of fluidic mixing in microdroplets with 1 microsecond time resolution using fluorescence lifetime imaging. *Analytical chemistry* **82**, 3950–6 (2010).

143. Chabinyc, M. L., Chiu, D. T., McDonald, J. C., Stroock, A. D., Christian, J. F., Karger, A. M. & Whitesides, G. M. An integrated fluorescence detection system in poly(dimethylsiloxane) for microfluidic applications. *2Analytical chemistry* **73**, 4491–4498 (2001).

144. Ramasubramanian, M. K. & Alexander, S. P. An integrated fiberoptic-microfluidic device for agglutination detection and blood typing. *Biomedical microdevices* **11**, 217–29 (2009).

145. Qi, S., Liu, X., Ford, S., Barrows, J., Thomas, G., Kelly, K., McCandless, A., Lian, K., Goettert, J. & Soper, S. a Microfluidic devices fabricated in poly(methyl methacrylate) using hot-embossing with integrated sampling capillary and fiber optics for fluorescence detection. *Lab on a chip* **2**, 88–95 (2002).

146. Neuzil, P., Cheng, F., Soon, J. B. W., Qian, L. L. & Reboud, J. Non-contact fluorescent bleaching-independent method for temperature measurement in microfluidic systems based on DNA melting curves. *Lab on a chip* **10**, 2818–21 (2010).

147. Mogensen, K. B., Klank, H. & Kutter, J. P. Recent developments in detection for microfluidic systems. *Electrophoresis* **25**, 3498–512 (2004).

148. Wang, F. & Burns, M. A. Multiphase bioreaction microsystem with automated on-chip droplet operation. *Lab on a Chip* **10**, 1308–1315 (2010).

149. Pavesi, A., Piraino, F., Fiore, G. B., Farino, K. M., Moretti, M. & Rasponi, M. How to embed three-dimensional flexible electrodes in microfluidic devices for cell culture applications. *Lab on a Chip* 1–3 (2011).doi:10.1039/c1lc20084d

150. Gibson, D. G., Benders, G. A., Andrews-pfannkoch, C., Denisova, E. A., Baden-tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Iii, C. A. H. & Smith, H. O. Complete Chemical

Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science* **319**, 1215–1220 (2008).

151. Szita, N., Polizzi, K., Jaccard, N. & Baganz, F. Microfluidic approaches for systems and synthetic biology. *Current opinion in biotechnology* **21**, 1–7 (2010).
152. Stahler, P., Beier, M., Gao, X. & Hoheisel, J. D. Another side of genomics : Synthetic biology as a means for the exploitation of whole-genome sequence information. *Journal of Biotechnology* **124**, 206–212 (2006).
153. Gulati, S., Rouilly, V., Niu, X., Chappell, J., Kitney, R. I., Edel, J. B., S, P. & DeMello, A. J. Opportunities for microfluidic technologies in synthetic biology. *Journal of the Royal Society Interface* **6**, S493–S506 (2009).
154. Tian, J., Ma, K. & Saaem, I. Advancing high-throughput gene synthesis technology. *Molecular Biosystems* **5**, 714–22 (2009).
155. Zhang, C. & Xing, D. Single-molecule DNA amplification and analysis using microfluidics. *Chemical reviews* **110**, 4910–47 (2010).
156. Tewhey, R., Warner, J. B., Nakano, M., Libby, B., Medkova, M., David, P. H., Kotsopoulos, S. K., Samuels, M. L., Hutchison, J. B., Larson, J. W., Topol, E. J., Weiner, M. P., Harismendy, O., Olson, J., Link, D. R. & Frazer, K. a Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nature biotechnology* **27**, 1025–31 (2009).
157. Theberge, A. B., Courtois, F., Schaerli, Y., Fischlechner, M., Abell, C., Hollfelder, F. & Huck, W. T. S. Microdroplets in Microfluidics: An Evolving Platform for Discoveries in Chemistry and Biology. *Angewandte Chemie International Edition* n/a–n/a (2010).doi:10.1002/anie.200906653
158. Chiou, J., Matsudaira, P., Sonin, a & Ehrlich, D. A closed-cycle capillary polymerase chain reaction machine. *Analytical chemistry* **73**, 2018–21 (2001).
159. Zhang, C. & Xing, D. Parallel DNA amplification by convective polymerase chain reaction with various annealing temperatures on a thermal gradient device. *Analytical Biochemistry* **387**, 102–112 (2009).
160. Schaerli, Y., Wootton, R. C., Robinson, T., Stein, V., Dunsby, C., Neil, M. a a, French, P. M. W., Demello, A. J., Abell, C. & Hollfelder, F. Continuous-flow polymerase chain reaction of single-copy DNA in microfluidic microdroplets. *Analytical chemistry* **81**, 302–6 (2009).
161. Bai, Y., He, X., Liu, D., Patil, S. N., Bratton, D., Huebner, A., Hollfelder, F., Abell, C. & Huck, W. T. S. A double droplet trap system for studying mass transport across a droplet-droplet interface. *Lab on a chip* **10**, 1281–5 (2010).
162. Cui, Y. & Campbell, P. A. Towards monodisperse microbubble populations via microfluidic flow-focusing. *2008 IEEE International Ultrasonics Symposium Proceedings* 1663–1666 (2008).
163. Feng, X., Du, W., Luo, Q. & Liu, B.-F. Microfluidic chip: next-generation platform for systems biology. *Analytica chimica acta* **650**, 83–97 (2009).
164. Breslauer, D. N., Lee, P. J. & Lee, L. P. Microfluidics-based systems biology. *Molecular bioSystems* **2**, 97–112 (2006).
165. Meyvantsson, I. & Beebe, D. J. Cell culture models in microfluidic systems. *Annual review of analytical chemistry (Palo Alto, Calif.)* **1**, 423–49 (2008).
166. Sims, C. E. & Allbritton, N. L. Analysis of single mammalian cells on-chip. *Lab on a Chip* **7**, 423–440 (2007).
167. Mettetal, J. T., Muzzey, D., Gómez-Uribe, C. & van Oudenaarden, A. The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science* **319**, 482–4 (2008).
168. Bennett, M. R., Pang, W. L., Ostroff, N. a, Baumgartner, B. L., Nayak, S., Tsimring, L. S. & Hasty, J. Metabolic gene regulation in a dynamically changing environment. *Nature* **454**, 1119–22 (2008).
169. Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science (New York, N.Y.)* **315**, 233–7 (2007).

170. Fordyce, P. M., Gerber, D., Tran, D., Zheng, J., Li, H., DeRisi, J. L. & Quake, S. R. De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nature Biotechnology* **28**, 962–967 (2010).

Chapter 4

4. Methods

The methods and materials used during the project are described within this chapter. Firstly, the microstereolithography (MSL) process is described, with reference to the build procedure and settings. The production of PDMS microfluidic devices by replica moulding of MSL moulds is then described. The development of the build settings, the post curing and replica moulding procedures will be referenced in chapters 5 and 6. Secondly, methods of molecular and cellular biology are described. Generally used methods and methods developed specifically during the project are described. These sections will be referenced in chapters 8 and 9. Finally, the modelling and programming aspects are then described and will be referenced in chapters 5, 6 and 7.

4.1. EnvisionTec Perfactory Mini microstereolithography machine

A modified EnvisionTec Perfactory Mini[1] was used to fabricate flow cells directly from 3D CAD files designed in SolidWorks 2009. This machine uses MSL, a type of additive layer manufacture (ALM) discussed in chapter 3, to build 3D objects by sequentially depositing successive layers. Figure 4.1 shows a schematic of the Pefactory machine whilst figure 4.2 shows how the EnvisionTec machine fits into the work scheme for creating flow cells.

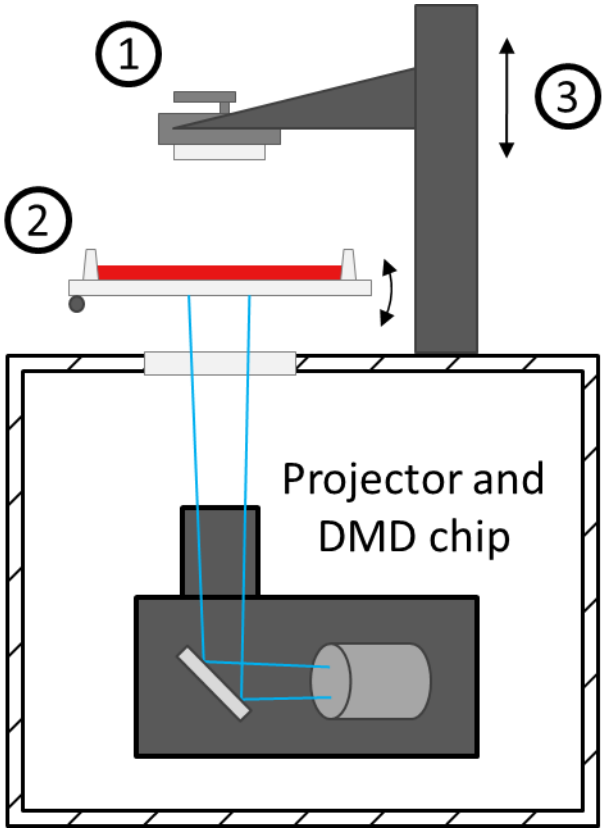


Figure 4.1: Schematic of EnvisionTec Perfactory MSL machine. 1: Removable build platform onto which completed parts are attached. 2: Tilting resin tray. 3: Z-stage driven by a lead screw. Computer is connected to the network for transfer of build jobs.

Parts fabricated on the EnvisionTec Perfactory are built onto the build platform, figure 4.1 (1). The build platform consists of a glass block and metal rails that facillitate attachement to the Z-stage, figure 4.1 (3) of the Perfactory machine. The liquid resin from which parts are cured is held in the resin tray, figure 4.1 (2). The resin tray of the EnvisionTec Perfactory Mini consists of a glass base onto which a thin (~2 mm) layer of transparant silicone rubber is attached. The silicone top surface is also treated with an agent to aid detachment of each layer. Walls around the resin tray hold in the liquid resin. The volume of resin can be varied depending on the task at hand.

The projector in the EnvisionTec Perfactory has been raised closer to the build platform to reduce the pixel size so that the minimum feature size can also be reduced. A modified version of the EnvisionTec Perfactory Mini firmware and software was kindly

provided by EnvisionTec. The resin tray, shown in figure 4.1, has a tilting mechanism. Thus, when a layer of resin is cured, one end of the resin tray is pulled down, away from the build platform. This tilting causes the part, including the newly formed layer, to peel from the resin tray. The peeling process is employed to reduce delamination of the new layer. Layer delamination is discussed in section 4.1.8.

4.1.1. EnvisionTec Perfactory workflow

The workflow for the EnvisionTec Perfactory is described in figure 4.2. STL files created in a CAD package, such as SolidWorks, are 'sliced' into layers by EnvisionTec RP software, compiled into 'job' files and transferred to the machine. On running a job file from the machine, the job file is unpacked and the build is executed. Once complete, parts are taken off the build platform using a scalpel or sharp knife. The part is then transferred to a 250 mL beaker and washed with acetone and/or isopropanol. Swilling of the beaker can then be employed if necessary, to desorb the uncured resin from the solid part. Acetone is a more aggressive solvent than isopropanol and will dissolve cured resin slowly. For this reason either acetone washing is employed sparingly or isopropanol is used. Difficult to clean parts can be soaked in isopropanol and continuously mixed using the horizontal rocker. High pressure air can also be used to clear uncured resin from tubes or holes where appropriate. Finally, the part is post-cured by flashing in a UV flasher box (see section 4.1.8

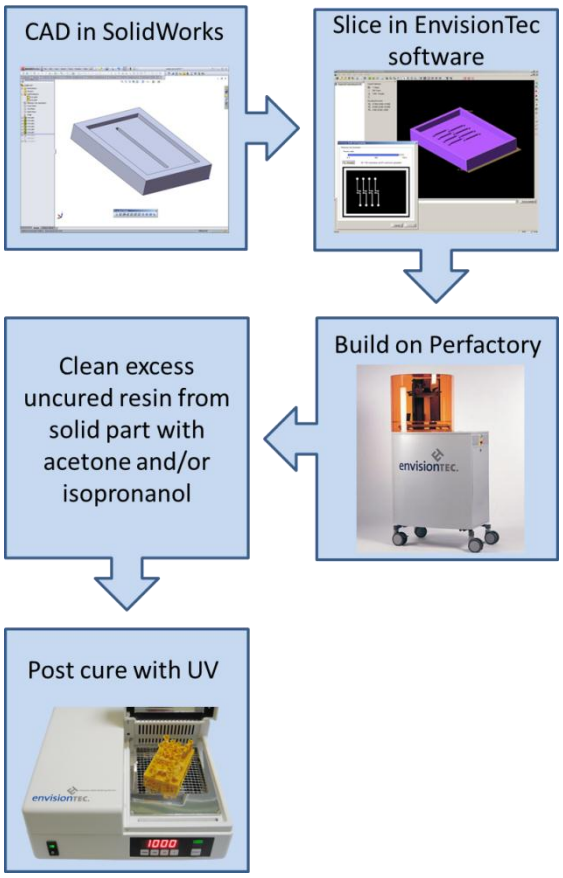


Figure 4.2: Workflow schematic of making parts with the EnvisionTec Perfactory Mini machine.

4.1.2. EnvisionTec Perfactory build material

The EnvisionTec Perfactory uses blue light (mainly between 250 and 550 nm) to cure photosensitive liquid resins into solid polymers. Although a variety of compatible resins are available, only one resin was used here, R11. R11 is a liquid resin composed of a di-acrylate monomer, tri-, penta- and hexa-acrylate crosslinkers, a free radical photoinitiator and a dye to control light penetration. The photoinitiator causes a radial reaction between acrylate groups in the mixture. Any of the acrylate group-containing species in the mixture can be covalently bound to any other. The result is a highly complex, disordered polymer. The dye limits the depth through the resin that the light can penetrate such that after around 25 µm the reaction is effectively prevented. The reaction is exothermic and polymerisation can

also be initiated by heating, so the dye also prevents a runaway reaction that could cure the entire of the available resin and damage the resin tray.

Several other materials are available for use with the Perfactory system[2]. These materials are formulated to include material properties for specific applications: Photosilver resin has high temperature resistance suitable for making moulds for vulcanising rubber (~130°C). PIC and WIC resins can be removed easily from an encasement by heating ('burning out') making these resins suitable for the production of moulds used in the jewellery industry. eShell resins are opaque and are formulated in a variety of colours for the manufacture of discrete hearing aids. NanoCure resin contains suspended nanoparticles that provide high stiffness and temperature resistance as well as being hard wearing.

4.1.3. Optical characterisation of R11 resin

R11 is formulated in several colours as a result of the dye molecule used, R11, red, blue, clear and rose. To determine the absorption profile of different colours of R11 resins available, samples of each resin were prepared. Sheets of each resin 0.5 ± 0.01 mm were cut from blocks of cured resin. Sheets were then polished using fine grain sandpaper followed by polishing with Wenol (Reckitt Benckiser, DE). Sheets were then trimmed to 9 mm wide pieces to fit into the sample holders (Hellmet spectrophotometer optical calibration filter holders). All pieces were tested in a Cary 100 Bio spectrophotometer (Agilent, UK) and the results presented in chapter 6.

4.1.4. EnvisionTec Perfactory capability

The maximum build envelope of the EnvisionTec Perfactory is 28 x 21 x 250 mm which is defined by the projected area and the maximum travel of the Z-stage. The resolution of the projector in the EnvisionTec Perfactory is 1400 x 1050 pixels which means that each pixel is

20 µm square. Theoretically, the minimum feature size is the same as the pixel size. In practice, the minimum feature size is around 100 micrometres for experienced users in good conditions. Furthermore, a wall is more likely to build successfully than a tower of the same diameter as the wall width. Single towers are not attached to the rest of the part by a sufficiently large surface area leading to delamination during the peeling step.

The key parameters affecting minimum feature size are exposure time, projector brightness and resin condition. Exposure times from 3.7 to 9.5 seconds will result in successful builds. Delamination increases with shorter exposure times. The same exposure and peel settings must be used throughout the part to get an even finish with minimal delamination. Projector brightness settings from 560 to 620 will result in successful builds. The effect of projector brightness on builds is very similar to exposure time. Resin condition deteriorates through use or time. The best builds (minimal delamination, smallest features) are made with new resin. Build failure results in significant deterioration of resin condition, but successful builds will also result in slow deterioration of the resin. Poor resin condition is defined by a high viscosity of relative to new resin and the presence of lumps which accumulate from failed and delaminated builds. Resin can be filtered through a 1 mm^2 steel mesh in a Buchner funnel attached to a vacuum pump (Caper 2D, Charles Austin Pumps, UK) to remove the larger lumps. Smaller lumps are soft enough, however, to pass through the filter. Filtered resin significantly improves the build quality. Filtering is only performed once per batch of resin as the accumulation of small, unfilterable lumps renders the resins unable to produce good quality builds.

4.1.5. Burn-in range settings

A cured layer of R11 resin will make a conformal bond with the glass surface of the build platform and the silicone rubber of the resin tray. To ensure that the nascent part is securely fastened to the build platform rather than the resin tray, the first several layers

(~400 µm) are built with 'burn-in' settings. Burn-in settings comprise a longer exposure time (9.5 seconds) and a slower peel speed (800 µm/s) relative the standard range.

4.1.6. Build range settings

The remaining layers of the part are built with standard build settings; exposure time of 3.7 seconds and a peel speed of 1200 µm/s. These are sufficient to cure layers and allow parts to be built rapidly.

Thin parts tend to warp after they are removed from the machine. The warping process is most likely a result of the change in build parameters between the burn-in and the build ranges. Warping can be reduced by clamping thin parts flat during the post curing process. Warping can also be reduced by building the build range with the same settings as the burn-in range. Using the same settings for the burn-in and the build ranges is also used in taller parts where a consistent finish is required.

4.1.7. WYKO build characterisation

To characterise parameters such as surface roughness and layer thickness a Wyko (Microprecision Instruments, UK) optical profiling system was used. The Wyko uses interferometry to vertically scan a surface returning a 3D representation of the surface which can then be analysed. Analysis was carried out using Gwyddion (version 2.25-1); an open source scanning probe data analysis software[3].

4.1.8. Post curing

Parts are post cured in the EnvisionTec flasher box (Otoflash, EnvisionTec, UK) after building. The flasher box has two metal halide tubular arc lamps which emit flashes of bright white light at ~10 Hz. EnvisionTec recommend at least 3000 flashes per part. The parts become notably hot during flashing so parts are flashed 1000 times, turned, left for a minute and flashed for another 1000 times. The heating of a part is probably at least in part

due to the exothermic reaction of the R11 monomer as it reacts. Flashing is necessary to ensure that all the uncured monomer that is trapped within the part is fully reacted. Uncured monomer can have a significant impact on the properties of finished parts. Post curing is necessary in order to maximise the strength of R11 parts and to prevent inhibition of the PDMS curing reaction. Post curing also reduces, but does not eliminate, warping in thin parts.

4.2. Microfluidic device fabrication

4.2.1. PDMS casting

Some flow cells described in chapter 5 required material properties not possessed by R11 (see section 4.1.2). Poly dimethoxysilane (PDMS) met these material requirements well. A R11 mould, that held the negative shape of the desired PDMS shape, was produced. PDMS (sylgard 184, DowCorning, US) was mixed in a 10:1 proportion of polymer base to curing agent by weight. Thorough mixing of the PDMS was necessary to ensure that the cured material had consistent properties throughout. The mixing process, however, results in the incorporation of significant amounts of air bubbles. The mixture is poured into the mould ensuring that a the top of the liquid PDMS is level with the top of the walls of the mould. Degassing of the mixed, uncured polymer was achieved using a vacuum desiccator attached to a vacuum pump (Caper 2D, Charles Austin Pumps, UK). A vacuum was then applied to the cast for around 30 minutes, which caused the majority of the bubbles to pop or merge into large bubbles. Stubborn bubbles could be lanced using a dry knife. It was important to degas all uncured PDMS samples, even samples which did not have evident bubbles as bubbles could develop during the heated curing process.

PDMS cures continually once the two components are mixed. The mixture will cure at room temperature overnight or within an hour at 60°C. In this work all PDMS curing was performed at 60°C after degassing. It was discovered that moulds which have never been

used before will slow the curing process significantly such that full curing will only be achieved in 24 hours at 60°C. Subsequent casts of the same mould will then cure as normal. Flashing the mould for more than 3000 flashes lessens the inhibition of the curing process.

Once cured, the PDMS casts were removed from the moulds using a scalpel blade. The back of the blade was used to prevent damage to the mould. The sides of the cast were first loosened from the side walls of the cast. The blade could then be used to gently lever the cast out of the mould taking care that the blade did not cut the cast or that the cast was not bent sufficiently to tear.

4.2.2. Membrane thickness determination

The relationship between layer thickness and spin speed was determined by spinning uncured PDMS onto glass squares cut from glass slides using a carbide scribe. Spinning was performed with a spin coater (G3P-8, Speciality Coating Systems, US). Glass squares were stuck to a silicon wafer using double sided sticky tape, the silicon wafer was held to the spinning chuck by a vacuum force applied through the chuck. ~100 µL of uncured PDMS was applied to the centre of each glass square. The spin coater was then run in the following manner: 1) Spin at 500 for 20 seconds. 2) Ramp to selected spin speed over 20 seconds. 3) Remain at specified spin speed for 20 seconds. 4) Stop spinning without ramping. Duplicates were performed at each spin speed. Coated glass squares were transferred to a pre-heated hot plate (KW-4AH, Chemat Technology, UK) set to 60°C for 1 hour to cure. A scalpel was then used to cut through the coated layer, one half of which was then peeled away to leave a step feature.

To measure the layer thickness, the TalySurf (Taylor-Hobson, UK) was used. Three measurements were made across the step feature made previously at different positions along the step. A script was written in MATLAB to convert the raw data from the TalySurf into .xlsx format for analysis in MATLAB.

4.2.3. Multilayer soft lithography

Microfluidic chips were assembled by multilayer soft lithography. Each layer was cast in MSL moulds as will be described in Chapter 6. Chips consisted of two layers, the pneumatic layer and fluidic layer, bonded together with a PDMS membrane between them.

Membranes were made and bonded to layers by a partial curing. ~100 µL of uncured liquid PDMS was spun on silicon wafers at 2500 RPM and partially cured by placing at 60°C for 12±1 minutes. Partial curing of the PDMS prevents the liquid PDMS from flowing into, and blocking, the channels of a layer whilst ensuring that the liquid PDMS is sufficiently tacky to bond to the applied layer. Alignment of the layers was performed by hand using a dissection microscope. A black background and a strong light source (halogen lamp) projecting along plane of the chip were used to increase the contrast between the transparent PDMS layers. Alignment was performed using the valve membranes and valve seats as references (see chapter 6). This method ensured correct alignment of two layers in approximately 80% of cases. Realignment of layers bonded using the partial curing method is not usually possible due to filling of features with uncured resin.

Following alignment, the layers were placed at 60°C for at least one hour in order to cure fully. Full curing is important to ensure that uncured PDMS monomer and/or oligomers do not leach into solutions carried within channels[4].

4.2.4. Microfluidic device control

Three way solenoid valves (12 V DC, ES-3W-12, Clippard, USA) were used to control the flow of air in the pneumatic channels of the microfluidic devices. The valves were attached to a custom built manifold consisting of machined plastic and MSL adaptors. The valve ports were connected in such a way as to ensure that the device was exposed to either regulated air-line pressure (at around 20 PSI) or atmospheric pressure during valve closing or opening. The regulated air pressure was connected to each of the solenoid valves by

1/16” ID TYGON® tubing (Cole-Palmer, UK). The solenoid valves were connected to the microfluidic device by 1/50” ID TYGON® tubing (Cole-Palmer, UK). The 1/50” tubing then connected to 90° bent needles (Fishman, UK) via an interference fit. The needles interfaced with the microfluidic device via a barbed end adaptor built from MSL specifically for the task. Needles were glued into channels built within the MSL adaptor with cyano-acrylate superglue.

LuerLoc® fittings (Cole-Palmer, UK) were placed in-line between the solenoid valves and the device to allow filling of the pneumatic channels in the microfluidic device with water containing food dye. Filling the PDMS channels of the pneumatic layer with water prevents the formation of air bubbles in the fluidic channels due to the relatively high permeability of PDMS to air. To fill the pneumatic channel, the LuerLoc® is disconnected from both the solenoid valve line and the microfluidic device adaptor. A 1 mL syringe was used to fill the 1/50” tubing with water containing food dye such that the tube was half full. The LuerLoc® fitting was thoroughly dried with a paper towel to minimise the risk that liquid would flow back towards the solenoid valves. The filled tubing was then reconnected to the microfluidic device adaptor followed by being reconnected to the LuerLoc® adaptor. Connection in this order was to ensure that the liquid in the tubing did not flow back towards the solenoid valve, which could result in valve failure. Once connected, a steady air pressure (20 PSI) was applied through the solenoid valves into the pneumatic channels. The air pressure behind the liquid in the channels forces the air ahead of the liquid through the PDMS walls of the device. Because of the relatively high permeability of PDMS to air, the air in the pneumatic channels can be fully removed within 15 – 30 minutes. Fluid channels will remain filled with liquid whilst the system is closed (I.E. the solenoid valves are shut and the channels full of fluid). If the channels are opened to the air then evaporation and movement of water vapour through the PDMS will render the channels dry overnight.

4.2.5. Microfluidic device operation

Mixing of oligonucleotides was performed in a PDMS microfluidic chip. A droplet system was created using mineral oil (M5904, Sigma-Aldrich, UK) as a carrier phase and oligonucleotide-containing deionised water as the immiscible droplet phase. Reservoirs in 1.5 mL Eppendorf tubes were attached to the chip via 1/50" ID TYGON® tubing (Cole-Palmer, UK). Two holes were drilled in the lids of the Eppendorf tubes and the tubes were then routed to the bent needle adaptors in the chip manifold. Cleaning of the system was performed with 100% ethanol. ~1 mL of ethanol was loaded into each reservoir and was pushed through each of the valves in turn.

Loading of each oligonucleotide was performed by using a P20 pipette (Pipetman®, Anachem, UK). 10 µL of each oligonucleotide was placed in the bottom of each reservoir and the pneumatic valve actuated whilst a constant pressure was applied to the reservoir to cause the 10 µL bolus to enter the tubing and travel to the chip valve. The valves were then actuated in the order and timing as previously determined (see chapter 6) to achieve mixing of the oligonucleotide-containing droplets.

Collection of the chip eluent was then performed with a 1.5 mL Eppendorf tube. Both carrier and droplet phases were captured. Once in the Eppendorf tubes, the droplets merged to form a single, large droplet in the bottom of the tube. A 0.1-10 µL pipette (Eppendorf, UK) was then used to sample from the large droplet, which then formed the template for an assembly reaction as described in chapter 6.

4.3. DNA manipulation

Pipetting was performed using micropipettes (Eppendorf, UK) with appropriately sized tips (1000 µL, 200 µL , 10 µL, graduated, filtered, StarLab, UK). Weighing of substrates was performed on an weighing scales (CP225D, Sartorius,UK).

4.3.1. Restriction digests

Analytical digests were performed rapidly with minimal substrate (~200 ng) and cloning digests were performed slowly (i.e. until completion) with mg quantities of substrate. All digests were performed with restriction enzymes obtained from New England Biolabs (NEB, UK). Digests were performed with 10 units of the appropriate enzyme in volumes of either 25 or 50 µL supplemented with the appropriate buffer and Bovine serum albumin (BSA) (NEB, UK) as necessary. Incubation was performed in a 37°C incubator (Sanyo, UK) or a PCR machine. Heat inactivation was always performed after digests for cloning and usually performed after analytical digests. Heat inactivation was performed in either a 65°C water bath (OLS200, Grant, UK) or an appropriately programmed PCR machine for enzymes requiring different temperatures. Once heat inactivated, digests can be stored at 4°C before downstream use.

4.3.2. PCR amplification

PCR amplification typically consisted of: 1x PCR buffer, 2.5 mM $MgCl_2$, ~20 ng plasmid DNA/other template, 0.5 µM forward and reverse primers, 0.5 mM of each deoxynucleotide triphosphate (dNTP), 5 units of AmpliTaq DNA polymerase and ~5% dimethyl sulfoxide (DMSO). Total reaction volumes for PCR was 20 µL. Reactions were carried out in 0.2 mL PCR tubes and conditions controlled by a Mastercycler Gradient PCR Machine (Eppendorf, UK). Unless otherwise specified the PCR program was: Initial denaturation at 95°C for 3 minutes. 30 cycles of 92°C for 30 seconds (denaturation), 55°C for 30 seconds (primer annealing) and 72°C for 1 minute per kilobase (elongation). Followed by a final elongation step at 72°C for 5 minutes.

4.3.3. Gel electrophoresis

Gel electrophoresis of DNA was performed on agarose gels in horizontal gel electrophoresis systems (VWR, UK). Agarose gels of appropriate concentration (0.7 – 2%) were prepared by mixing appropriate quantities of agarose with 100 mL of TEA buffer. TEA buffer consists of 40 mM Tris-acetate, 1 mM ethylenediaminetetraacetic acid (EDTA) at pH 8.0. To dissolve the agarose the mixtures were microwaved and mixed by swirling until homogeneous. 0.3 µg/mL ethidium bromide was added after the solution had cooled to around 60°C and mixed. The solution was poured into moulds and gel combs inserted to form wells. Electrophoresis was carried out in 1x TEA buffer with DNA ladder as appropriate and samples mixed with 1x loading buffer loaded into individual wells. Voltages of 40 – 80 V were applied during the electrophoresis; low voltages were used for samples requiring gel extraction and higher voltages used for analytical gels. Following electrophoresis, gels were visualised using a G:Box transilluminator (Syngene, UK) and analysed using GeneSnap software (Syngene, UK).

4.3.4. DNA ladder

The ladder markers used in gel electrophoresis were either low molecular weight marker (NEB, UK) or were generated by combining 0.2 mg/mL of *lambda* DNA and *phi*174 DNA ladder markers with 1x loading dye. 0.3 µL was added to each gel so that band sizes in test lanes could be accurately determined.

4.3.5. Gel extraction

For gel extraction, bands on high quality electrophoresed SeaKem® GTG® agarose (Lonza Biologics, UK) gels were visualised using a High Performance UV Transilluminator (Syngene, UK). Agarose percentages were tailored to the expected size of the fragment. Bands were excised with a scalpel. DNA extraction was achieved using the QIAquick Gel Extraction Kit

(Qiagen, UK) according to the manufacturer's instructions. Briefly, the gel piece was dissolved in a high salt binding buffer. Several drops of sodium acetate (3M) can be added if the pH of the binding buffer changes during the solvation, as indicated by a yellow to purple colour change by the indicator in the buffer. The dissolved gel piece, in binding buffer, was then bound to the silica QIAquick spin column. For high concentration gels (≥ 2%), a further wash with 500 µL binding buffer was performed to remove remaining agarose traces. The bound DNA was then washed to remove impurities with wash buffer containing 70% ethanol. Elution of the DNA was achieved using water, DNA concentration quantified using a NanoDrop ND-1000 Spectrophotometer (Nanodrop Technologies, US) and a quality checked by agarose gel electrophoresis.

4.3.6. Bacterial culture

Escherichia coli (*E. coli*) was used as a culture organism to grow up and select clones resulting from a cloning reaction or for when growing up a previously prepared plasmid. Preparation of the culture was performed in 15 mL round bottomed, snap capped tubes (BD falcon, UK) with 3 mL of Lysogeny Broth – Miller variant (LB-Miller). LB-Miller consists of 1% tryptone, 1% NaCl, 0.5% yeast extract (all w/v), was not pH adjusted prior to use and no buffers were added. LB-Miller is prepared by Warwick Life Sciences Preparation Room. The media was supplemented with 1 mg/ml of amplicillin to act a selective marker for the presence of the transfecting plasmid. All cultures were incubated at 37°C overnight in an orbital incubator (Sanyo E&E Europe BV, UK) at 180 RPM.

4.3.7. Ligation for cloning

DNA, which was used for cloning into target vectors, was derived exclusively from digests. Cultures of *E. coli* containing the insert DNA, carried in a temporary plasmid vector, such as TOPO, were grown up and miniprepped. The plasmid DNA was isolated by miniprep using a Miniprep kit (Qiagen, US). Plasmid DNA containing the insert of interest is then digested

with the appropriate enzymes (see section 4.3.1.) to liberate the insert. The digest was then run on a gel and the insert DNA is purified from the cut temporary vector DNA by gel purification (see section 4.3.3). The recipient vector was similarly digested but only gel purified if the digest cuts out a portion of the recipient vector, in all other circumstances, where the vector is simply being opened, a PCR purification was performed using a PCR purification kit (Qiagen, US). The recipient vector DNA was dephosphorylated using rAPid alkaline phosphatase (Roche, UK). Ligation was performed with a Quick Ligation kit (NEB, UK). A 3:1 molar ratio of insert to vector was mixed with an appropriate volume of 2x buffer and topped up to a volume of 20 µL with deionised water. To this mixture, 1 µL of ligase was added and incubated at room temperature for 5 minutes. The ligation mixture was then put on ice for transfection.

4.3.8. TOPO Cloning

Cloning of PCR products, such as those of the assembly/amplification reaction, was performed using the TOPO TA 2.1 kit (Invitrogen Ltd, UK). TOPO vector employs topoisomerases covalently attached to T/A overhangs in a standard vector. The topoisomerases ligate linear DNA into the host vector quickly and easily. TOP10 chemically competent cells (Invitrogen, UK) were transfected with an aliquot of the TOPO TA cloning reaction according to the manufacturer's instructions. Briefly, insert DNA was quantified by running a small aliquot of the purified PCR product on an appropriate gel with a marker of known concentration. Based on this quantification an aliquot of the PCR product preparation was added to a volume of diluted salt solution and vector DNA was added. The ligation reaction was allowed to proceed for 5 minutes at room temperature before the reaction tube was placed on ice. The total volume of the TOPO ligation reaction is 6 µL.

4.3.9. Bacterial transfections

TOP10 (Invitrogen, UK) were thawed on ice. A 1-3 µL aliquot of the ligation reaction was added to TOP10 cells and mixed gently without pipetting. The cells and ligated DNA were incubated for 30 minutes on ice before being heat shocked at 42°C for 30 seconds. The aliquot of cells was then ready for short term growth prior to plating. Retransfection of a previously prepared plasmid was performed in the same manner as above, without TOPO ligation steps and with an aliquot of the TOP10 cells. Up to 4 different plasmids were retransfected in independent transfection reactions from a single tube containing 50 µL of TOP10 cells.

Short term growth prior to plating of the transfected cells was performed by addition of 250 µL of Super Optimal broth with Catabolite repression (SOC) to the transfected cells after the heat shock step. SOC media consists of 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose pH adjusted to pH 7.0 and autoclaved or filter sterilised. SOC media was prepared by the Warwick Life Sciences Preparation Room. Transfected cells in SOC were incubated at 37°C in the orbital shaker (Sanyo, UK) at 180 RPM for up to 1 hour. Transfected cells were then transferred to agar plates for clonal selection.

Selection of successful transfectants was performed on the basis of ampicillin resistance and blue/white colony selection (the latter in the case of TOPO ligation only). Transfected cells were plated onto LB-agar plates. LB-agar consists of LB supplemented with 1.5% Bacto-agar (see section 4.3.6 for description of LB) and 1 mg/mL of ampicillin. Further, for blue-white selection 20 µL of 50 mg/ml X-gal in dichloromethane (DCM) was spread over the surface of the agar and allowed to dry. LB-agar was prepared by Warwick Life Sciences Preparation Room. Short term growth cultures were transferred to plates by pipette and spreading across the plate was performed using glass beads sterilised by

autoclaving at 121°C. The plates were then incubated upside down at 37°C overnight in an incubator (Sanyo, UK). Bacteria which pick up a plasmid containing the ampicillin resistance gene are able to grow on the surface of LB-agar supplemented with ampicillin. Insertion into the TOPO vector, without relegation of the vector alone, disrupts the *lacZ* gene which encodes a β-galactosidase enzyme which is able to digest the x-gal into an insoluble, blue-coloured product. Thus cells which were transfected with a relegated TOPO vector which did not receive an insert were able to survive on ampicillin supplemented agar but were unable to produce a blue coloured product when the agar was also supplemented with X-gal. Hence, picking white colonies allows TOPO vector containing the insert to be cultured.

4.3.10. Sequencing

Sequencing of successful transfectants was performed by the Genome Facility at Warwick University by the Sanger method using a Prism 7000 sequencer (Applied Biosystems, UK). DMSO was added to some sequencing reactions to reduce secondary structure formation due to palindromic sequences present in some plasmids and assembled sequences. All insert sequences were checked to be 100% correct sequencing prior to continuing to the next cloning step.

4.4. DNA assembly

4.4.1. Phosphoramidite synthesis

The DNA used to make the combinatorial mutant libraries was synthetically derived. The most widely used method of producing synthetic oligonucleotides is the phosphoramidite cycle[31,32]. A key limitation in DNA assembly is arises due to the method by which the DNA is first synthesised. The cycle involves the repetition of three steps; deblocking, coupling and capping. The three steps are depicted in figure 4.3. The building blocks of the process are nucleoside phosphoamidites which are analogues of the naturally occurring DNA base

pairs. Interestingly, nascent strand is synthesised from 3' to 5' in contract to the biological polymerase enzyme which synthesises 5' to 3'.

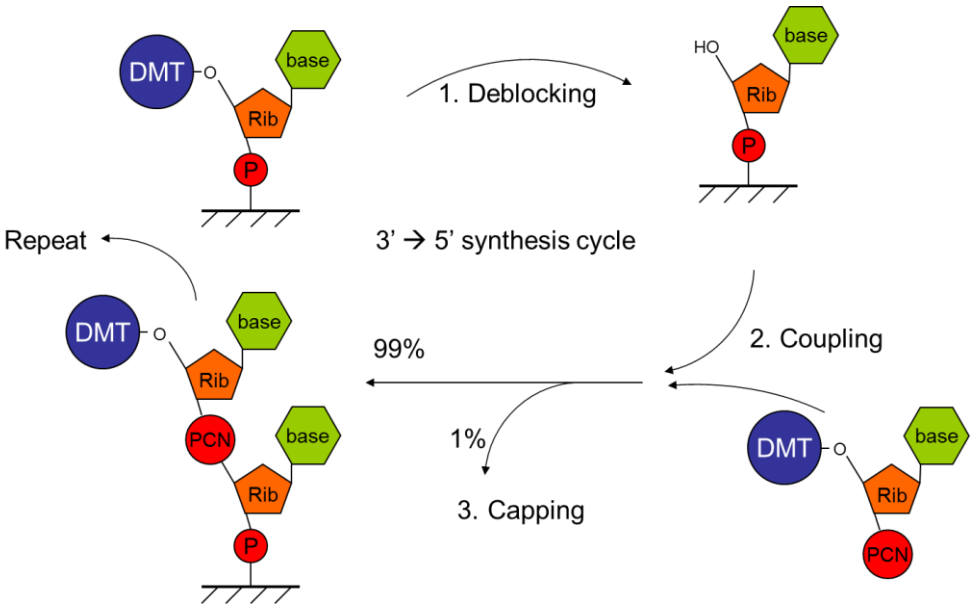


Figure 4.3: Steps of the phosphoramidite synthesis cycle for synthesis of DNA chains on a solid support. This method is widely used in the production of synthetic DNA.

Each step of the process is not 100% efficient, about 0.1 - 1% of the nascent strands fail to be coupled with each round of the reaction. These failed couplers are capped during the capping step to prevent the production of a deletion mutant. Instead, these capped species will form truncation mutants which are easier to separate downstream by, for instance, high performance liquid chromatography (HPLC). Deletion mutants can arise from incomplete deblocking, although this reaction is comparatively much more successful. The frequency of mutations in the nascent strands limits the overall length of contiguous, correct sequence that can be produced by DNA synthesis. The maximum synthesisable length whilst still obtaining sufficient full length, correct sequence is around 100 bp.

4.4.2. Assembly of genes by the Gao method

The method of gene synthesis that is performed by Prof. X. Gao in the University of Houston will be referred to from here on as 'Gao synthesis'[33–36]. DNA sequences are 'grown' in microfluidic chambers using a light activated decoupling step and normal

phosphoramidite chemistry. The projected light is patterned, using a digital micromirror device (DMD) to activate only the wells which are to add the next base. The microfluidic chip consists of 4096 individual wells each connect to the inlet and outlets such that all the chambers in the chip can be flushed simultaneously. This method has been successfully used to produce microarrays directly on glass surfaces (ie without spotting)[37].

By careful design and assembly of the synthesised oligonucleotides, synthetic genes can be made. The length of these assembled sequences is not bound by the limitations of the phosphoramidite synthesis process. Assuming that both strands are synthesised and assembled ligatively, the 4096 well chip could theoretically synthesis oligonucleotides sufficient to assemble a sequence just over 200 kbp long. Synthetic genes have been by assembling oligonucleotides which themselves were made by Gao synthesis previously[35].

4.4.3. Gao assembly

The oligonucleotides created during Gao synthesis are limited to <100 base pairs in length mainly due to failure of the coupling step (see figure 4.3, phosphoramidite synthesis). Hence, to synthesise the ~300 base pair CRM, several overlapping oligonucleotides were necessary.

The synthesis scale of oligonucleotides made by this method is very small. Each nucleotide is produced in femtomolar amounts. To obtain useful quantities of oligonucleotides the product of the synthesis must be amplified. A generic primer sequence is added to the start and end of each of the synthesised oligonucleotides to amplify all the sequences at once. This is step 2, ligator PCR, in figure 4.4. Gel electrophoresis is then used to check the success of the ligator PCR. Gel electrophoresis will confirm the size of the product of the PCR reaction, but will not confirm that every member of the oligonucleotide set has been amplified. In fact, variations that already exist in the relative amounts of individual oligonucleotides before the ligator PCR step will be amplified afterwards.

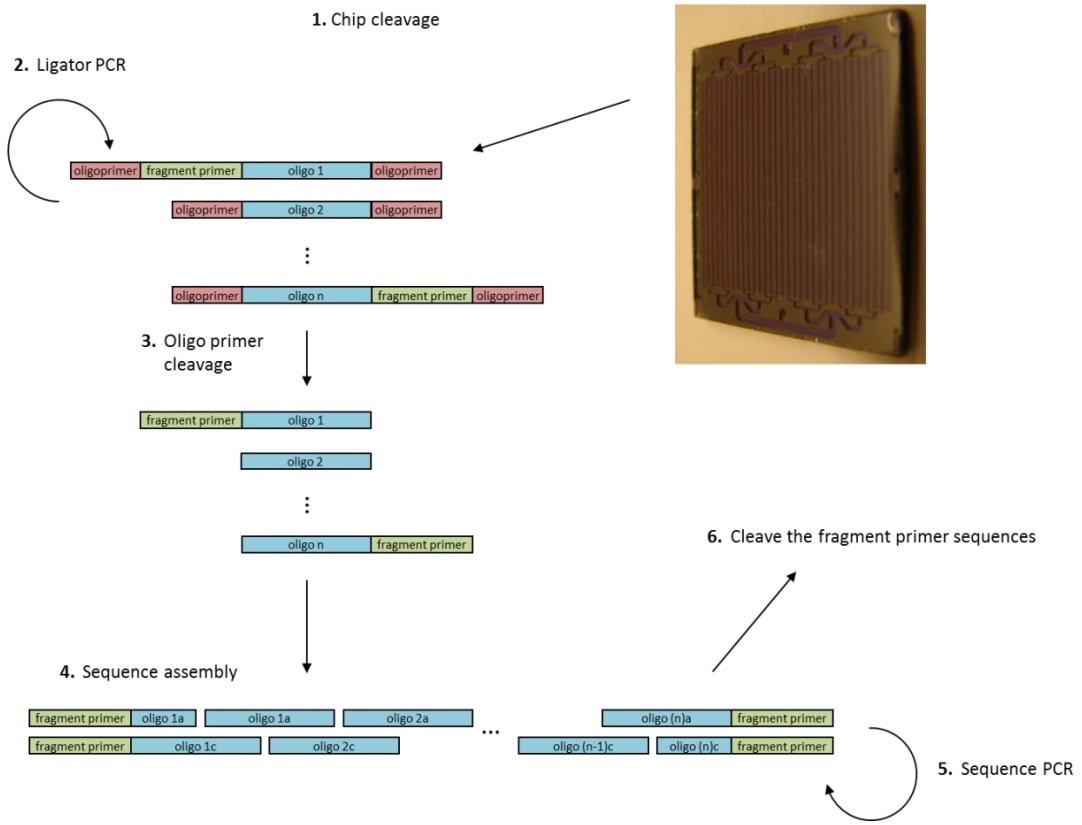


Figure 4.4: Schematic of the assembly process developed by Gao. Oligonucleotides are synthesised by Gao synthesis, amplified, and assembled into longer sequences by PCR.

Once the ligator PCR has been satisfactorily completed, the generic primer sequences must be cleaved off. Type IIS restriction enzymes are a type of restriction enzymes which recognise a palindromic sequence but then cut the DNA double strand several base pairs away from the recognition site[38]. By placing the recognition site of a blunt cutting type IIS appropriately in the generic primer binding sequence on each oligonucleotides can be cleaved off cleanly leaving a blunt end.

Once the generic oligonucleotide amplification sequences have been cleaved off, the oligonucleotide set is ready to be assembled. It should be noted that following the ligator PCR amplification of the set of all oligonucleotides that each oligonucleotide *and* its complementary pair is present in solution. For simplicity, figure 4.4 shows only the forward direction oligonucleotides. The presence of complimentary partners to each oligonucleotide in solution presents a problem during the next step, where each

oligonucleotide is to find its appropriate partners as described in figure 4.4, step 4. By adding ligase and putting the mixture through several rounds of heating and cooling it is hoped that at least a small proportion of the full length sequences will assemble. Because the assembly of the full length sequence is unlikely, several lesser interactions must occur on the same strand and be successfully ligated, the number of oligonucleotides in a single sequence is limited.

After assembling and ligating the oligonucleotides, the small proportion of full length sequences must be once again amplified by a PCR reaction. For the full sequence amplification a second short generic sequence is added to the first and last oligonucleotides in the sequence. Once again a type IIS restriction enzyme is used to cleave off the generic fragment primer.

For the assembly of longer sequences, an additional assembly step can be employed: The assembled sequences from the ligation reaction can be designed so they themselves contain overlaps suitable for PCR-based assembly. For this step, an overlap PCR is employed where each of the ligated sequences becomes a primer to another ligated sequence. The polymerase then fills in the gaps to produce a single complete sequence.

4.4.4. Chip cleavage

During synthesis of the microfluidic synthesis chip, the oligonucleotides were covalently attached to the solid silicon surface. To cleave the oligonucleotides from the chip a procedure developed by Qi Zhu (University of Houston) was followed.

The following list describes the ammonia cleavage of chip oligonucleotides. In this protocol, ‘dummy chip’ refers to an empty microfluidic chip which has not been used in a synthesis reaction. The ‘holder’ is a plastic device which channels the fluid from the tubing system to the inlets on the synthesis chip. Washing steps require routing the chip outlet into the reservoir, allowing the volume to be cycled indefinitely. Reservoirs consist of 1.5

mL Eppendorf tubes with two holes drilled to allow access by tubing. Flow speeds refer to the custom built peristaltic pump that was used In the Gao lab. The protocol now follows:

1. With dummy chip in the holder, clean the system in following sequence: 1% sodium dodecylsulfate (SDS) (1 mL) circulate for 10 minutes at low speed, 5 mM Tris (1 mL), pH 6.8, circulate for 10 minutes at low speed. Flow through with 3 mL 5 mM Tris, pH 6.8. In a new tube, add 1ml 5 mM Tris, pH 6.8, flow through about 300 µL.
2. Replace the dummy chip with the synthesis chip.
3. Flow through about 300 µL 5 mM Tris, pH 6.8, circulate with the rest of 5 mM Tris, pH 6.8 for 5 minutes. Then let the 5 mM Tris, pH 6.8, flow through completely.
4. Replace the solution to 1000 µL of 5 mM Tris, 1% BSA, pH 6.8. Flow through about 300 µL and then circulate the rest of solution for 20 minutes.
5. Flow through about 500 µL of 5 mM Tris, pH 6.8; flow through completely.
6. Circulate 250 µL ammonia hydroxide with the chip at room temperature for 10 minutes, 37°C for 10 minutes, 45°C for 10 minutes, and 50°C for 1.5 hours.
7. Collect all 250 µL of the ammonia solution (be careful not to run air into the chip). Wash the chip with 250 µL of 5 mM Tris, pH 6.8. Collect the first 100 µL in the tube containing the ammonia hydroxide. Circulate with the remaining 150 µL Tris buffer for 10 minutes at 50°C and collect.
8. Speed dry the sample in a heated vacuum centrifuge (heated to 45°C) to reduce the volume (ensure the caps are not closed!). Combine two solutions and continue until dry. The drying process will usually take 1.5-2 hours.
9. Resuspend the pellet in 20-50 µL deionised water.

The most difficult aspect of the cleavage process was avoiding trapping bubbles in the synthesis chip. If a particular well was blocked by a bubble, the wash solution would not fully reach that well which could lead to improper cleavage of the oligonucleotide synthesised therein. Avoiding and/or removing bubbles was a major concern.

After the synthesis process was complete, the chip was provided dry. Bubbles in the aqueous media would form spontaneously when the chip was heated. To remove bubbles, techniques such as changing the direction of the flow and putting the chip through heating and cooling cycles were employed.

4.4.5. Gao oligonucleotide digests

A Type IIS restriction enzyme which produces blunt ends, MlyI (NEB, UK), was used to cleave the oligonucleotide amplification primers from the oligonucleotide sequences prior to assembly. Reactions were carried out in a volume of 50 µL with 1x NEBuffer 4 and 1x BSA. Each digest was performed upon a whole amplification reaction volume (5-7 µg), minus an aliquot taken for comparative gel electrophoresis. Digests were performed overnight at 37°C.

Comparative gel electrophoresis was employed to determine whether the digest was successful. Gels of 2.5% agarose were used as the expected product size was 50-100 base pairs.

4.4.6. Gao assembly ligation reaction

The Gao assembly ligation reaction was carried out in a volume of 20 µL using *Taq* ligase (NEB, UK). The template, digested oligonucleotides obtained by the Gao synthesis chips, was added at either a 'high' or 'low' concentration. The low concentration contained ~0.5 µg of DNA whereas the high concentration contained ~4 µg. The cleaved oligonucleotide primers will contribute to this total but will not be actively ligated. As such, the actual concentration of active oligonucleotide substrate is around 30% less than these values. Because some oligonucleotides were used in more than one sequence it is difficult to determine each oligonucleotide's expected concentration.

4.4.7. Optimisation of the Gao assembly

Several variations of the assembly protocol were performed. The basic method was as follows: Mix oligonucleotides with appropriate buffers. Initial slow cool from 95°C to 60°C, add 1 µL of ligase. Heat to 95°C for 3 minutes, touch down from 76°C to 60°C at one degree per cycle with a hold at each step of the touchdown for 20 minutes. Hold at 60°C overnight. Alternatively, the touch down was also performed slowly whereby the mixture was heated to 95°C then cooled quickly to a 'topline' temperature 76°C then slowly to a 'bottomline' 60°C over 20 minutes, the cycle was repeated but the 'topline' temperature was reduced by 1°C per cycle. The rate of the slow cool did not change so the total time spent in the cooling cycle also diminished with each cycle.

In one case, an additional cycle was included on the end of the touchdown cycle: Heat to 95°C for 3 minutes and cool quickly to 50°C for 5 minutes and repeat for 30 cycles. This cycle was tagged onto the end of the appropriate touch down.

4.4.8. Gao assembly amplifications

Two amplifications were performed during the Gao assembly process: The first was the amplification of the oligonucleotide substrate. The second was the amplification of the assembled full length product.

The oligonucleotide amplification was carried out at on a large scale to provide a uniform starting material for the assembly reactions. Separate reactions were prepared from an identical master mix using *pfu* polymerase (Stratagene, US), 1x buffer (containing 2 mM $MgCl_2$, 200 µM dNTPs, 5 microM primer and 1% DMSO). A sample of each amplification was run on a 2.5% agarose gel to check that the reaction had been successful. The reaction volumes were then frozen and kept separate to avoid repeated freeze-thawing cycles.

The assembled products were amplified using a specific PCR for the full length sequences. Primers for this reaction were obtained from the Gao lab, which has DNA synthesis facilities.

4.4.9. OptiCut oligonucleotides

The oligonucleotide substrate (Eurofins MWG, Ger) used in the OptiCut assemblies were synthesised at the 0.01 micromole synthesis scale. All oligonucleotides were modified by the manufacturer to contain a phosphate on the 5' OH and were HPLC purified. Eurofins claim that the purity, as measured by capillary electrophoresis, of their standard high purity, salt free (HPSF) purification method is >70% whereas the HPLC purity is >80%.

4.4.10. Opticut assembly protocols

Oligonucleotide sequences were optimised using the OptiCut program described in chapter 7. Assembly of the OptiCut optimised oligonucleotides was performed by a similar assembly method as published previously[5]. The protocol is a two-step assembly-amplification. First, oligonucleotides were mixed in equimolar amounts to completely describe the entire sequence. Second, a PCR was performed on a sample of each of the ligase reactions. Each PCR was then gel purified. Assemblies were carried out in 96-well quantitative PCR (qPCR) plates (Beckman Coulter, UK) that could be used with the Eppendorf PCR thermocycler. Master mixes were used to ensure consistent reaction conditions.

Ligative assembly was performed on each equimolar mixture of HPLC-purified oligonucleotides (EuroFINS MWG, Germany) at high temperature (65 - 45°C) in the presence of *Taq* Ligase (NEB, USA): *Taq* ligase (20 units per reaction), 1x *Taq* ligase buffer, oligonucleotides to a concentration of 20 nM each. 5 minutes at 95°C, a further 1 minute at 95°C, cool rapidly to 65°C. Cool slowly to 45°C over 15 minutes. Hold at 45°C for 15

minutes. Once the assembly reaction is complete, the 96-well plates can be sealed with either adhesive film (ThermoSeal RT2, Alpha Labs, UK) or suitable sealing strips (Domed cap strips, ThermoScientific, UK) and be stored at 4°C.

To amplify the products of the ligation reaction, a small aliquot of the ligation is used as a template for a PCR with AmpliTaq Gold DNA polymerase (Applied Biosystems, US): *Taq* polymerase (2.5 units per reaction), primers (800 nM each), dNTPs (10 mM of each), DMSO (2% v/v), 1x PCR buffer, $MgCl_2$ (2 mM). The maximum template concentration (assuming 100% ligation) is 800 pM. Initial denaturation at 95°C for 3 minutes, 30 seconds at 95°C, 30 seconds at 55°C, 1 minute at 72°C and a final extension at 72°C for 3 minutes.

4.4.11. Purification protocol

Gel purification was employed to isolate the PCR products of the appropriate size. High quality 1.5% (w/v) agarose (SeaKem® GTG® agarose, Lonza Biologics, UK) gels were used in purification. Purifications were performed using a Qiagen Gel Purification Kit with MinElute columns (Qiagen, US), which are designed to maximise retention of products 70bp < x < 4 kbp.

4.4.12. Cloning and sequencing

TOPO® cloning was employed using TOPO® TA cloning® kits (Invitrogen Ltd, UK). Cloning was performed according to the manufacturer's protocol using gel purified PCR products. Selection of successful transfectants was performed on the basis of ampicillin resistance and blue/white colony selection. Sequencing of successful transfectants was performed by the Genomic Facility at University of Warwick.

4.5. Tissue culture

4.5.1. Cell culture and passage

All tissue culture was carried out in Class II Microbiology Safety Cabinets (Walker Safety Cabinets Ltd, UK). C2C12 cells were maintained in Dulbecco/Vogt modified Eagle's minimal essential medium (DMEM) without sodium pyruvate, with glutamax and supplemented with 10% fetal bovine serum (FBS) (Invitrogen, UK). All culture was performed in T75 flasks (BD Falcon) at 37°C humidified Galaxy R CO_2 incubators (New Brunswick, UK). Once cells reach ~80-90% confluence they were passaged into new flasks. C2C12 cells will differentiate when confluent and so to avoid differentiation and to maintain cells in a fast growing state, cells are passaged.

To passage, the medium was removed by aspiration and the cell surface washed gently with ~13 mL of PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4 at pH 7.4) warmed to 37°C. Cells were detached from the cell culture surface by adding 3 mL of room temperature trypsin solution (0.05% trypsin-EDTA, Invitrogen, UK) was added and allowed from 30 seconds to 2 minutes to work. The time spent in the trypsin solution should be minimised so as to minimise the damage done to cells. Flasks can be smacked several times to lift of cells once the trypsin has weakened their interaction with the culture surface. 12 mL of warmed (37°C) DMEM + 10% FBS was used to remove the cells and wash the flask to a suitable tube. The 10% FBS serves to saturate the trypsin and effectively halt the reaction. The tube was then spun at 1000 revolutions per minute (RPM) for 4 minutes to pellet the cells in suspension, the supernatant aspirated and the pellet re-suspended in a known volume of warmed (37°C) DMEM + 10% FBS. The cell suspension was then split between the desired number of flasks.

4.5.2. Cell freezing

For long term storage, stocks of cells are frozen in liquid nitrogen dewars at around -198°C. Cells were grown to 80-90% confluency, similar to the passage procedure, in an incubator set to 37°C with 5.0% CO_2 in DMEM + 10% FBS in T75 flasks. Flasks containing cells were removed from the incubator, media removed by aspiration and washed 2 times with PBS warmed to 37°C in an appropriate water bath. Care was taken so that the jet of PBS from the PipetteBoy (Integra Biosciences, UK) does not impinge on the growth surface so as to avoid the unnecessary dislodging of cells. 3 mL of trypsin is added to the flask and the flask incubated at 37°C, 5.0% CO_2 for 2-3 minutes. The culture area of the flask was observed, using a stereo light microscope at 100x magnification (TS100, Nikon, UK) to check that the cells are rounded up. The flasks were then vigorously slapped to dislodge the remaining attached cells. A further observation was performed with the light microscope to check that significant attached cells do not remain on the culture area of each flask.

The trypsin reaction was halted by addition of 12 mL of DMEM + 10% FBS to the flask. The FBS contains sufficient non-specific protein that it effectively blocks the trypsin enzyme active site. Cells were washed from each surface of the flask to ensure that a homogeneous suspension of all the cells in the flask is obtained. Care was taken not to introduce bubbles to the solution. The cell suspension was then transferred to a 50 mL falcon tube (BD Biosciences, UK) which was then centrifuged at 1000 RPM for 4 minutes to pellet the cells. The media was removed by aspiration and the cells re-suspended in a known volume of freezing solution. Freezing solution consists of a warmed (37°C) solution of 1:4:5 DMSO:FBS:DMEM. The volume of freezing solution used to re-suspend the pellet depends on the desired concentration of cells to be prepared. Cryopreservation requires 1-10 x 10^6 cells with a total volume of 1.8 mL per vial, typical cell concentrations used in the course of this project were 2.5 x 10^6 cells per vial, which is approximately ¼ of a T75 flask. 2 mL cryovials for freezing are obtained from VWR Jencons.

DMSO is cell toxic and once the cells are suspended in freezing solution they must be chilled and frozen as soon as possible. Freezing by steps is preferred over snap freezing as step freezing results in less loss of viability of the cells compared to snap freezing. Cryovials are placed at 4°C and chilled before being transferred to -20°C overnight. Frozen vials of cells are then transferred to -80°C for a further night and finally transferred to liquid nitrogen dewars (-198°C). Cells can be stored practically indefinitely (up to 20 years) although a loss of cell viability is usually experienced with storage greater than 1 year.

4.5.3. Cell Thawing

C2C12 cells in this study were prepared from stocks of cells in 1:4:5 DMSO:FBS:DMEM frozen in liquid nitrogen at ~-196°C. Thawing was achieved by bathing the tube containing the cells in sterile PBS at 37°C until the tube contents were completely thawed. The contents were removed by pipette and diluted to 10 mL with DMEM + 10% FBS warmed to 37°C. After dilution, the cells were spun at 1000 RPM for 4 minutes to pellet the cells. The supernatant was then aspirated leaving the cell pellet which was then diluted to ~15 mL with warmed (37°C) DMEM + 10% FBS and plated into a T75 flask. Cells typically settle to the bottom of the flask after 15 minutes and attach after 2 hours. Cells will then firmly attach and begin migratory behaviour after 4-5 hours. The medium of the cells is replaced 24 hours after the initial plating and growth monitored using a stereo light microscope to ensure the cells do not overgrow. Further feedings as necessary are performed every 48 hours until the cells are ready for passage. Thawed cells are typically slow growing and may take several days to reach passaging density.

4.5.4. Cell seeding

C2C12 cells for seeding were obtained from 80-90% confluent T75 flasks retrieved from incubation at 37°C, 5.0% CO_2. The process of removing the cells from the cell culture surface of the flask is described previously (see section 4.5.1). Several flasks may be

combined to increase the cell count sufficiently to allow the seeding of multiple 6-well plates as necessary. A single T75 flask at 80-90% confluency can seed approximately four 6 well flasks. The cell count derived from resuspension of cell pellets was assayed using a haemocytometer (Neubauer Improved). The cell suspension was diluted so as to obtain 2 x 10^5 cells per well of a 6 well plate (BD Biosciences, UK) and each well was seeded to a total volume of 2 mL. The seeded 6 well plates were placed in an incubator at 37°C, 5.0% CO_2 overnight.

4.5.5. Transient transfection

Transient transfection was used to determine the level of expression a specific CRM combination as capable of eliciting. Plasmids containing GFP under the control of the myod promoter and combinations of CRMs and their mutants were transfected into C2C12 cells. The workflow associated with this experiment is shown in figure 4.5.

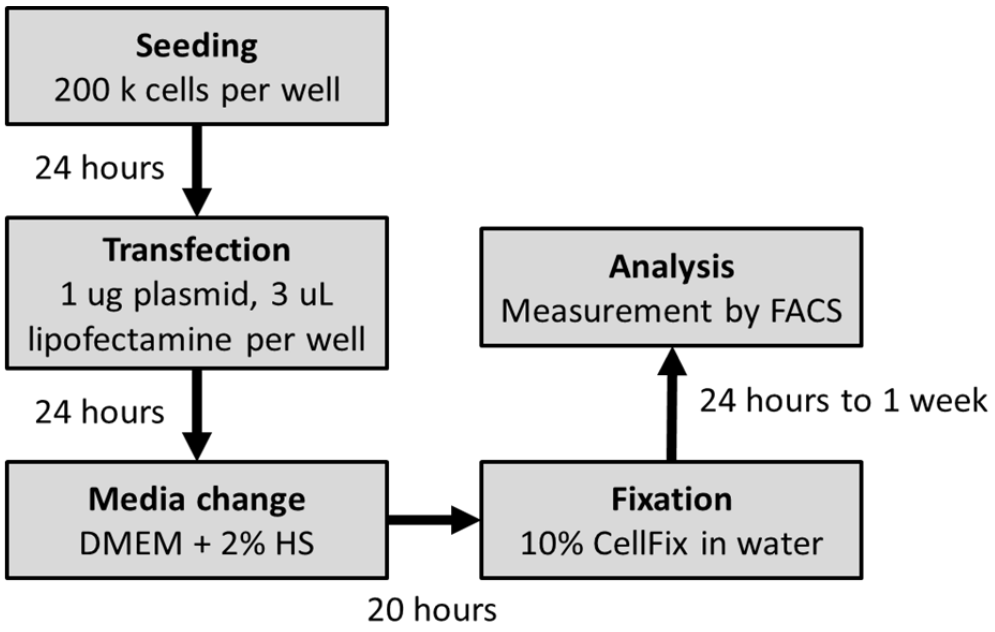


Figure 4.5: Workflow of a typical transient transfection experiment.

Transient transfections were performed 24 hours after seeding. Transfections were performed by adding OptiMEM (Invitrogen, UK) supplemented with lipofectamine™ 2000

(Invitrogen, UK) containing the plasmid DNA of interest. Test constructs, containing GFP under the control of a specific CRM combination, were transfected onto at least two wells of cells simultaneously. Transfection control was performed by adding a plasmid expressing the red mCherry fluorescent protein under the control of the same promoter into all the cells at a concentration of ¼ that of the test plasmid.

For each well (multiply by the number of wells to be treated simultaneously), 1 µg and 0.25 µg of the test and mCherry control plasmids, respectively, were combined with 100 µL of OptiMEM® (Invitrogen, UK) and incubated at room temperature for 5 minutes. In another tube, 3 µL of lipofectamine™ 2000 (Invitrogen, UK) was combined with a further 100 µL of OptiMEM and incubated for 5 minutes. The tube containing OptiMEM supplemented with lipofectamine was then transferred to the tube containing the OptiMEM and DNA and mixed gently by pipetting. The combined mixture was then incubated at room temperature for 30 minutes. The incubated mixture was then added to the appropriate well of a 6 well plate containing C2C12 cells that have been incubating overnight. The plate was then swilled gently by hand to ensure proper mixing of the OptiMEM transfection mixture. Plates were returned to the 37°C, 5.0% CO_2 incubator for 24 hours.

4.5.6. Differentiation

The media covering the cells since their seeding was removed and replaced with differentiation media 24 hours after the transfection mixture was added. Media and transfection mixture was removed from plate wells by aspiration and each was washed with 2 mL of PBS warmed to 37°C. The PBS was added slowly by PipetteBoy so as to minimise the disturbance to the cell monolayer. The PBS was then removed by aspiration and the wash procedure repeated twice to ensure the removal of the DMEM + 10% FBS. Differentiation medium consisting of DMEM + 2% Horse Serum (HS, Oxoid microbiology

products, Thermo Scientific, UK) was added to each of the wells so that each well receives a total of 2 mL media. The plates were then replaced at 37°C, 5.0% CO_2 for 20 hours so that differentiation may take place.

4.5.7. Cell fixing

Differentiated, transfected C2C12 cells were fixed after the cells have been under differentiating conditions (DMEM + 2% HS) for 20 hours. Plate wells were washed twice with PBS in a manner identical to that employed before the media change described above. The adherent cells were treated with 0.5 mL of trypsin to allow them to be removed from the culture surface. Room temperature trypsin solution was added to wells and the plates are incubated for a short period of 30 seconds to 2 minutes. Cell rounding up was observed on a stereo light microscope. Once the cells were partially rounded up, representing a significant loosening of cell-surface attachments, 1 mL of warm (37°C) DMEM + 10% FBS is added to block the trypsin reaction. A 1 mL pipette was used to wash the cell surface with the added media to generate a cell suspension. The cell suspension was then transferred to 1.5 mL Eppendorf tubes. Tubes containing cell suspensions were spun at 5000 RPM for 1 minute to pellet the cells. The supernatant was then removed by aspiration and the cell pellet resuspended in 1 mL of PBS using a 1 mL pipette to break up the cell pellet. The cell pellet washing procedure is repeated twice. The colour of the pellet usually changes from a brown-beige to a slightly grey white during the wash steps and becomes progressively easier to break up. After the final cell pellet washing step, the pellet was resuspended in 800 µL of 10% CellFix (BD Biosciences) in sterile water. The cells were then left in suspension, transferred to 4°C and covered to prevent photobleaching of the fluorophores prior to flow cytometry analysis.

4.5.8. Flow cytometry

To analyse the effect of changing CRM combinations and mutations thereof flow cytometric analysis was performed. Flow cytometry operates on the principle of encapsulating single cells within a droplet of liquid within a fast moving carrier, or sheath, fluid. The droplets can then be interrogated for fluorescence by several lasers in rapid succession thus exciting several fluorophores that might be present. The emitted light from the fluorophores can be split via successive filter sets so that multiple fluorophores can be detected simultaneously though independent channels. Measurements can be taken from many hundreds of individual cells per second making this technique a rapid and statistically accurate way of determining variance within a population.

Fixed cells are used in flow cytometry experiments where live cells are not required to be cultured subsequently. Cell fixing is necessary to allow analysis to be postponed as they can be stored for up to a month at 4°C. The quantum yield of fluorescent proteins, such as GFP will decrease over time, however, so the quantification of even fixed cells must be performed within this length of time. Both live and fixed cells can be labelled with antibodies to enable the identification of subpopulations of cells.

Cell fixing allows the later addition of antibody labelling, which can be used to identify sub-populations of cells within the overall population. The quantum yield of fluorescent proteins, such as GFP will decrease over time, however, so the quantification of even fixed cells must be performed within a month of storage at 4°C.

Standard fluorescent beads can be used to calibrate the efficiency of the machine in making fluorescent measurements. Untransfected cells are also used to normalise flow cytometry measurements. Normalising measurements against these two standards is necessary to obtain consistent measurements over the lengths of time necessary for a single experiment (30-40 samples).

Flow cytometry is not suitable to obtain expression profiles of single cells as cells cannot be tracked between consecutive measurements. To obtain an expression profile of a single cell in consistent culture conditions a through time live cell study is required. This type of analysis requires a cell to be followed continuously in culture and is usually performed by cell culture and microscope robots. Sophisticated software is required for cell tracking through successive frames, which is complicated by cells merging, dividing or passing over one another. Furthermore, different cells in confluent cell cultures are difficult to distinguish from their neighbours. As a result, the conditions that the cells are in, confluent and merging during differentiation, are not conducive to accurate through time measurement.

Flow cytometric analysis was performed on a BD Influx flow cytometry system (BD Biosciences, UK) running Spigot 6.1.4 (BD Biosciences, UK). The BD influx contains four lasers with which to interrogate the cells in suspension. In the experiments described herein a 488 nm laser is used to excite the GFP fluorophore whilst a 561 nm laser is used to excite the mCherry fluorophore. Filter sets 530/40 and 593/40 were used for the 488 nm and 561 nm laser respectively. Fixed cells suspended in 10% CellFix diluted in water are removed from storage at 4°C and kept on ice. Aliquots of the fixed cell suspension are diluted with sterile water to achieve a particle count rate of around 200 per second on each machine. The dilution factor was typically 1/5. Between 10000 and 50000 cells are counted during each run to ensure the statistical significance of the results. Results of a FACS run are analysed using FlowJo 7.5.5 software (FlowJo, UK) and Microsoft Excel (Microsoft, US).

4.5.9. Analysis of flow cytometry data

Flow cytometry measures several metrics about each of the several tens of thousands of particles that pass the detectors. Flow cytometry data must be appropriately analysed as misleading conclusions might be otherwise drawn from the large amount of data available.

In this section, a single example data set is examined in detail in order to determine the appropriate handling procedure. Once ascertained, this procedure is then applied, without modification, to the data set for each CRM construct.

The principle flow cytometry data handling procedure is 'gating'. The principle metrics upon which gating is performed, forward and side scatter (FSC and SSC, respectively), are discussed in section 4.5.8. Particles are gated according to these metrics to filter out particles of the inappropriate size or granularity. Gating allows the removal of cell debris and clumps that might inappropriately skew the results. Figure 4.6 shows 'raw' data from the flow cytometer. Particle detection during analysis was triggered by FSC measurement above a threshold of 5500 as seen in figure 4.6. The threshold was set at this level to exclude detection of small particle debris.

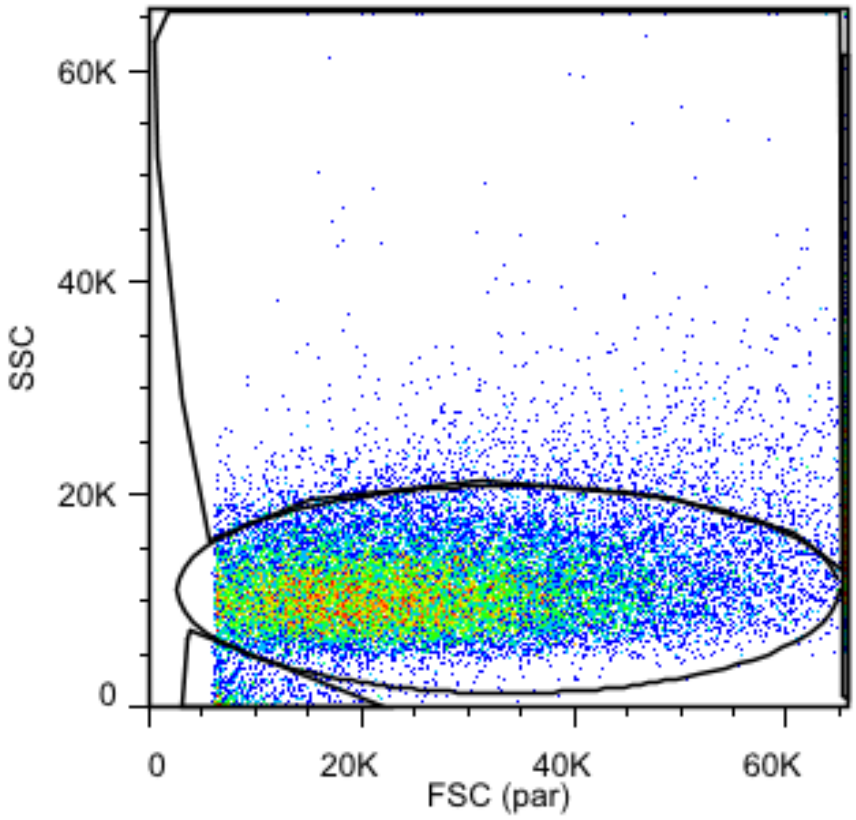


Figure 4.6: Raw flow cytometry data from a population of untransfected, differentiating C2C12 myoblasts with 'gates' indicated black lines from FlowJo software. The following gates are indicated: 1. The main gate. 2. Low FSC, low SSC. 3. High SSC. Colour indicates frequency with blue being lowest and red being highest.

By applying ‘gates’, like the one seen in figure 4.6., particles which are not cells can be removed on the basis of size or complexity. The C2C12 cell population shows a wide range of FSC values with the majority of cells falling into a peak around with a mean FSC value of 20000. The differentiating C2C12 cells will have stopped growing and started to fuse into myotubes, which might explain the large variation in FSC seen here as fused cells are larger than non-fused cells. The gate also removes the large number of low FSC, low SSC (small size, low complexity) particles that are most likely cell debris generated as a result of the fixing process. FSC detection becomes saturated at very high levels, so a small number (~0.1%) of very high FSC particles are present at the end of the x-axis. These particles are essentially off the scale and should not be included as their size cannot be accurately determined. Reducing the gain setting on the FSC channel would bring these particles into the range of the main axes, but would also risk dropping some of the valuable particles in the main peak below the detection threshold.

The gate is defined by an oval, as shown in figure 4.6, and reduces the number of particles to 88.5% of the original. Other gate polygons are available, but an oval is most likely to approximate the variance of the underlying population; two orthogonal normal distributions of size and complexity. It is critical that the same gating procedure be applied to all the samples to avoid any bias as a result of the gating.

The histogram of the relative GFP expression of each of the particle population gates shown in figure 4.6 is shown in figure 4.7. The mean values of the ungated, main, low FSC/low SSC and high SSC are 54.9, 48.4, 2.46 and 130 respectively. Inappropriate gating can, therefore, alter expression values by ~10%. These values clearly show how gating can affect the determined results. The gates used on the CRM construct expression experiments presented herein are a combination of the ‘high SSC’ and ‘main gates’ (gates 3 and 1 in figure 4.6, respectively).

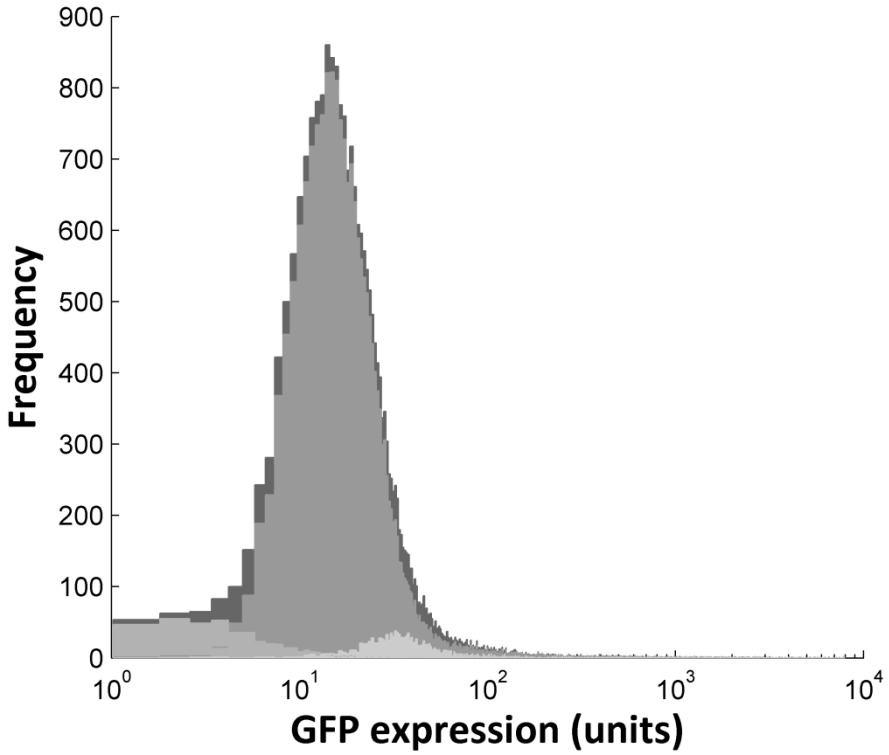


Figure 4.7: Histogram of GFP expression in particles in the gated populations shown in figure 4.6. The darkest shade is the ungated population, the second darkest is the main oval gate, second lightest is the low FSC, low SSC gate and the lightest is the high SSC population. The units of GFP expression are arbitrary.

The low FSC, low SSC gate consists of a population of with a predominantly low GFP expression (relative to the overall mean), supporting the hypothesis that these are cell fragments. Conversely, the high SSC gate consists of a population with predominantly high GFP expression (relative to the overall mean).

The high SSC population could be differentiating myoblasts. As discussed in the introduction (see section 1.5), myoblasts fuse to form myotubules during differentiation. A high SSC indicates granularity or complexity, which would be the case for multinucleate myotubules. This hypothesis could be tested by putting differentiating myoblasts through the flow cytometer at several time points after starting differentiation and monitoring any change in the proportion of high SSC particles. For consistency, all experiments on differentiating myoblasts were performed 20 hours after initiating differentiation.

4.6. COMSOL modelling

COMSOL modelling was undertaken to determine the likely behaviour of fluid in flow cell channels. COMSOL Multiphysics v4.2 (COMSOL AB, UK) is a popular finite element analysis software package that has a relatively intuitive interface and also interfaces well with MATLAB. COMSOL is a modelling package that incorporates tools for each stage of the modelling process, is designed to be easy to use and was available through one of the project collaborators. Other suitable modelling packages include FLOW-3D[6], OpenFOAM[7] and Ansys[8]. All COMSOL modelling was undertaken on a computer with a Phenom II x4 3.0 GHz processor, 4 GB of RAM and a 64-bit Windows 7 operating system.

In all cases, laminar flow physics were used as the flow through the channels was determined to be of low Reynold's number. Geometries were either designed in SolidWorks 2009, converted to .stl format and imported into COMSOL or were generated directly inside COMSOL using the in-built 3D geometry tool. The flow through the inlet was defined as a constant velocity normal to the plane of the inlet. Outlets were defined as zero pressure. Other boundaries were all defined as non-slip.

4.7. Programming

4.7.1. MATLAB

The optimisation of oligonucleotide sequences was performed in MATLAB 2009a (MathWorks, US) using a program, OptiCut, described in chapter 7. The Bioinformatics Toolbox was installed and functions from this toolbox were used or modified. MATLAB is an anagram for matrix laboratory and as such is optimised for handling and manipulating large matrices, as required in image analysis. The problem at hand here, see section 7.1.1 on the definition of the problem, potentially involves large matrices of values which must be manipulated rapidly.

MATLAB is easy to use by non-technical users due to its weakly dynamically typed nature. i.e. variable types can be dynamically defined without being assigned beforehand. Furthermore variable types can be redefined implicitly. This flexibility is very useful to a programmer who is not familiar with statically typed programming languages such as C++.

All optimisations were carried out on a Dell Latitude D630 laptop. This computer is equipped with a T7100 Intel™ Core 2 Duo™ 1.8 GHz CPU, 800 MHz front side bus (FSB) with 2 GB of random access memory (RAM).

4.7.2. DNA melting point determination

The Bioinformatics Toolbox within MATLAB was the source of several functions used in the program. For instance, the DNA melting temperature determination function, TmNNSanta98.m, was modified from the oligoprop.m function which determines many properties of a given DNA sequence. Putative overlap temperatures are estimated using the SantaLucia method[9]. The NN method is described in the chapter 7 (see section 7.1.3).

4.7.3. Algorithm scaling efficiency testing

To test how well the algorithm would scale with larger sequences and greater complexity as defined by a larger number of mutation sites, a testing environment was programmed. The environment generates randomised sequences using the MATLAB random number generator (uniform distribution of bases) to generate randomised sequences. A defined number of mutation sites are then created within the random sequence. The mutation sites are non-overlapping and between 4 and 6 base pairs in size. Each base of the mutation site is different from the corresponding base of the normal site.

Several hundred random sets were tested. Complexity was varied from a single mutation site (2 sequences in the set) to 10 mutation sites (1024 sequences in the set). Lengths varied from 200 to 600 base pairs in 50 base pair steps. The number of

oligonucleotides per strand was restricted to 8 for all examples tested. This value was chosen so that the results would be relevant for the optimisation performed in chapter 7.

4.7.4. LabVIEW

The graphical programming environment LabVIEW (Laboratory Virtual Instrumentation Engineering Workbench) (v2010, National Instruments, US) was used to automate the control of the off chip valves. A program was written which was capable of controlling the valves via a GUI or input text file. This program is presented in section 6.13-16. A key benefit of LabVIEW is the extensive support for communication with various types of instrumentation hardware. Another benefit is that LabVIEW is packaged with several large libraries of functions which make manipulation of data easy. Execution of text-based code via the MathScript node is possible and is generally compatible with MATLAB. Since MATLAB was used in other areas of this project, LabVIEW was a natural fit.

All LabVIEW programs were run on a Pentium 4 2.6 GHz with 1 GB RAM and communicate with a National Instruments – Digital acquisition (NI-DAQ) box (USB-6009, National Instruments, UK). Since the NI-DAQ box has 5 V output voltages, which is insufficient to actuate the 12 V valves, a high current circuit was prepared.

Each of the valves was connected to a 5V pin from the NI-DAQ box via a BDX33C transistor (Darlington Transistors, supplied by Farnell, UK). A 1 kΩ resistor was placed in between the NI-DAQ box and the base of the transistor to limit the current flow through the NI-DAQ box. A 1N4001 diode was placed in parallel to the solenoid valve to ensure that the current would only flow one way through the transistor.

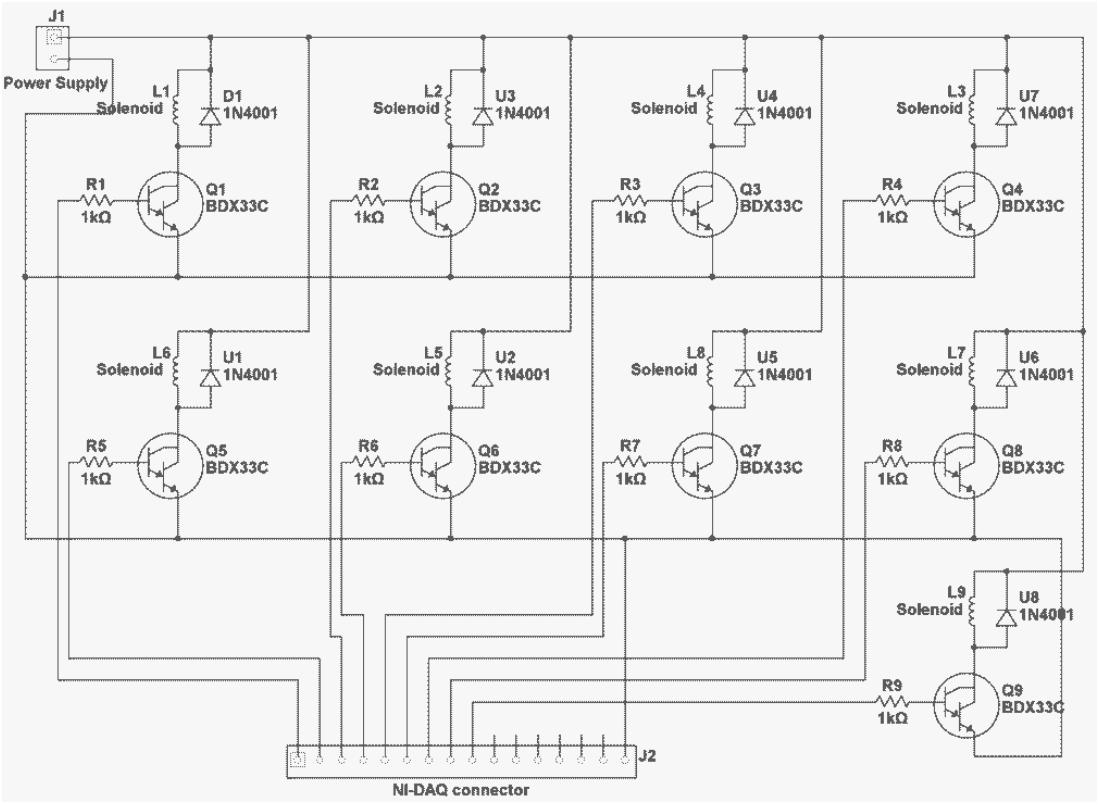


Figure 4.8: Schematic diagram of high voltage switch circuit used to drive 12V solenoid valves. A total of 9 units can be seen that apply the 12V potential of the power supply to the solenoid valves in response to the activation of the 5V pins from the NI-DAQ box.

4.7.5. Flow rate analysis

Flow rate through valves in microfluidic devices was achieved using two methods: The first method used a calibrated gas flow meter (ASF1430, Sensirion, CH) which was communicated with using a program written in LabVIEW. The second method was used to confirm the first. The fluid flowed through the valve into a piece of tubing approximately 1 m long. The tubing was placed against a 1 m ruler that was graduated in millimetres. The fluid position in the tube before and after each valve actuation was recorded and the volume of the flow was determined. The displaced air from the tube was then channelled through the flow meter and measured in the form of total counts recorded.

A similar experiment was performed to determine the flow rate through a valve in response to increasing pressure on the valve. In this experiment the flow rate was

measured by the flow meter alone and recorded as the average number of counts during the time the valve was actuated for.

4.7.6. Droplet size analysis

To determine the relationship between valve opening time and droplet size in microfluidic devices an automated program to analyse droplet size from live video of the microfluidic chip was written in MATLAB. A digital microscope (Veho VMS-001) was used to monitor droplet production. The MATLAB program was used to pull frames from the digital microscope, find droplets and record them in a database according to user instructions. Analysis was performed on live video running at approximately 20 frames per second (FPS)

4.8. References

1. EnvisionTec EnvisionTec Perfactory Mini MultiLens. (2012).at <http://www.envisiontec.de/index-page=machines&id=57.php.html>
2. EnvisionTec EnvisionTec Materials Summary. (2012).at <http://www.envisiontec.com/index-page=materials_list.php.html>
3. Klapetek, P. & Nečas, D. Gwyddion. *Sourceforge.net* (2011).at <http://gwyddion.net/>
4. Regehr, K. J., Domenech, M., Koepsel, J. T., Carver, K. C., Ellison-Zelski, S. J., Murphy, W. L., Schuler, L. A., Alarid, E. T. & Beebe, D. J. Biological implications of polydimethylsiloxane-based microfluidic cell culture. *Lab on a chip* **9**, 2132–2139 (2009).
5. Smith, H. O., Hutchison III, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly : phi X174 bacteriophage from synthetic oligonucleotides. *PNAS* **100**, 15440–15445 (2003).
6. FLOW-3D FLOW-3D. (2012).at <http://www.flow3d.com/>
7. OpenFOAM OpenFOAM. (2012).at <http://www.openfoam.com/>
8. Ansys Ansys. (2012).at <http://www.ansys.com/>
9. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1460–5 (1998).

Chapter 5

5. Flow cells by microstereolithography

As described in chapter 2, there are various techniques for the fabrication of different types of microfluidic devices. For this project, microfluidic devices were fabricated by a process called multilayer soft lithography that requires the production of patterned layers of material that are then annealed together to make a functioning device. This process is described in the following chapter, chapter 6. Before the layers can be assembled they must be patterned and, in this case, the layers were patterned by casting of the material into a prefabricated mould. This chapter describes the characterisation of the EnvisionTec Perfactory microstereolithography (MSL) machine build process for the purpose of making 3D microfluidic devices and moulds for casting of the commonly used polymer poly dimethylsiloxane (PDMS). The flow cells described herein were each produced in order to investigate how best to apply the MSL process to specific problems in biology and chemistry.

MSL was chosen as the fabrication technique because it is possible to rapidly produce 3D devices that would not be possible through other techniques such as micromachining or injection moulding. In the case of the thin layer flow cells it would not be possible to produce the required geometry by a non-ALM technique. Previous work, to produce relevant geometries for the thin layer flow cell, and their short comings is discussed in section 5.4. In the cases of the flow cell for microbiology (section 5.3.1), the MSL mould for PDMS flow cell (section 5.3.2) and the optical flow cell (section 5.5), viable devices could have been assembled from machined components. Here the rapid translation of the design to a finished monolithic product or mould facilitated by MSL allowed for rapid design revisions in minimal time. Table 5.1 summarises the advantages and disadvantages

of three relevant fabrication techniques; machining, ALM and injection moulding. It is important to note that, depending on the design of the device to be fabricated, further assembly may be necessary with all of the techniques. This further assembly should be determined on a part-by-part basis and should be factored into the build time.

Method	Tooling time	Build time	Repeatability	Unit cost
Machining	**Low**	**Moderate**	**High**	**Low**
ALM	**Very Low**	**Low**	**High**	**Moderate**
Injection moulding	**High**	**Very low**	**High**	**Depends on run size**

Table 5.1: Table showing the advantages and disadvantages of different fabrication methods that could be used to fabricate the microfluidic devices in chapter 5.

5.1. EnvisionTec build characterisation

The MSL process is a form of additive layer manufacture (ALM) and is described in chapter 3 (see section 3.6). Briefly, ALM involves the sequential deposition of patterned layers that together produce an object that might contain geometries, such as internal voids and channels, that traditional manufacturing techniques cannot. This section seeks to characterise the limitations of the build process by investigating the pixel size, layer thickness and incidence of build artifacts. This characterisation is performed by analysing example builds using a variety of techniques including interferometry and scanning electron microscopy (SEM).

5.1.1. Pixel size

The EnvisionTec Perfactory uses projected light from a digital micromirror device (DMD) to cure liquid resin to a solid in 25 μm layers. The digital nature of the DMD means that the resulting layers are ‘pixelated.’ As mentioned in the methods chapter (see chapter 4) the

pixel size is around 20 μm square. Measurements made between regular features on the xy plane supports this with pixel measurements of 20±2 μm (as inferred from interferometry results shown in figure 5.1). Periodic surface roughness on all flat surfaces can also be seen in figure 5.1 and 5.2. The roughness has a square pattern of positively embossed ridges. Each ridge is most likely the result of an overlap in the curing region of two neighbouring pixels. This periodic surface roughness is ±1 μm in the xz and yz planes and ±0.5 μm in the xy plane. The different surface roughness in the three orthogonal planes is most likely due to the fact that the surface in the xz and yz planes is unconstrained, unlike the xy plane which is constrained by the resin tray surface.

5.1.2. Layer thickness

One advantage of the ALM process is that individual layers can be very thin. The EnvisionTec Perfactory machine produces parts where each layer is around 20 ± 2 μm thick as determined by Wyko (Microprecision Instruments, UK) measurement. These measurements are consistent across several parts built in a range of conditions.

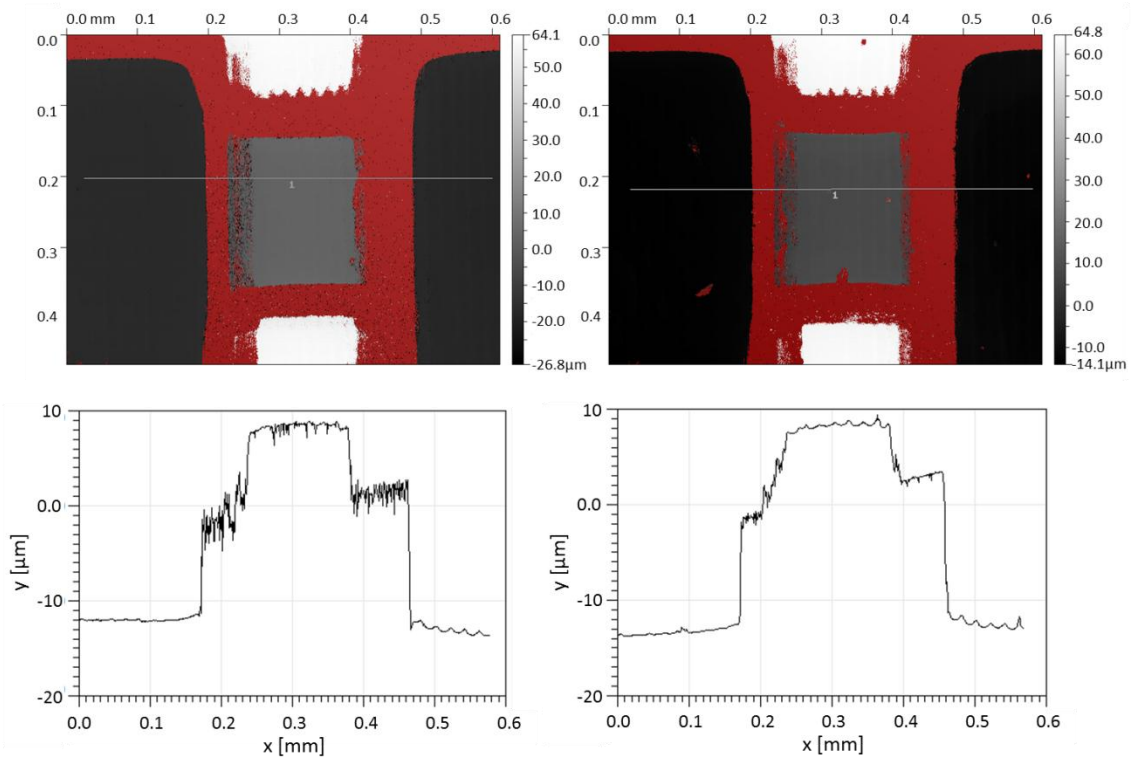


Figure 5.1: Example single layer thicknesses as measured by interferometry. Wyko data was analysed using Gwyddion. Grey bars indicate the position of the profiles shown in the lower axes. Red areas denotes for which height data could not be obtained.

Interestingly, the measured layer thickness of 20±2 µm differs from the EnvisionTec Perfactory's stated layer thickness of 25 µm. Performing interferometry across layers, in the xz plane, indicates a layer thickness of 25 µm. The interferometry data is supported by microscope image measurements. The difference between layer thickness on the surface and within a part could be due to layers only being completed once a subsequent layer is formed. The final layers, the layers measured here, are not followed by subsequent layers and therefore are shorter than the 25 µm layer thickness seen elsewhere. Another cause could be shrinkage of the layer during post-curing. The top layer is unconstrained by other layers and therefore is able to shrink more significantly. The implication of this observation is that features on the top surface of a MSL mould are significantly shorter than expected.

The vertical layer thickness is consistent due the position accuracy that the z-axis stepper motor possesses. A step feature was built in the x and y dimensions to test the accuracy of the EnvisionTec Perfactory when reproducing structures in these dimensions.

5.1.3. Model slicing

3D CAD models are sliced into layers by the EnvisionTec RP software. Pixels of a slice that are within the part are turned on, curing the layer according to the desired pattern. Parts, however, do not necessarily conform to the pixel grid as some features will not be properly represented. The example of a curve is a particularly good one. The full sweep of a curve cannot be resolved by this method, instead a series of stepped layers are produced that approximate the curve as best as possible. Edge features, such as that shown in figure 5.2 can be used to investigate the accuracy of the pixel assignments of the Perfactory RP software.

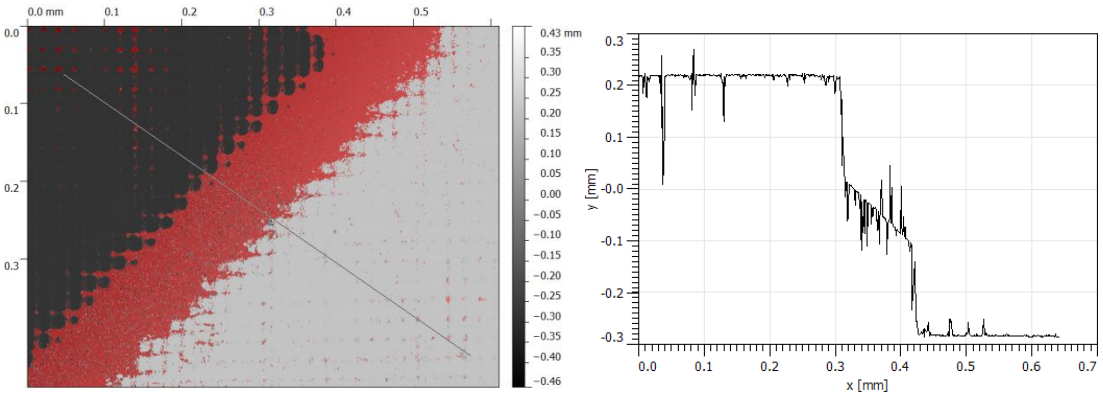


Figure 5.2: Edge feature interferometry measurements (left pane) of an MSL part. The line indicates the path of the profile shown in the right pane.

The pixilation of parts can be directly measured: The feature vertical difference between the top and bottom of the feature shown in figure 5.2 was specified to be 500 µm. The actual vertical height of the feature was measured as 527 µm. The difference between the specified and measured values is around the resolution of a single pixel and probably due to pixel assignments during the RP program layer slicing process.

5.1.4. EnvisionTec build capability

Several test parts were built to investigate the limits of the EnvisionTec Perfactory MSL system. Two parts with fine features are shown in figure 5.3. The machine is capable of minimum feature sizes of 100 µm in complex 3D shapes made in single builds over a matter of hours.

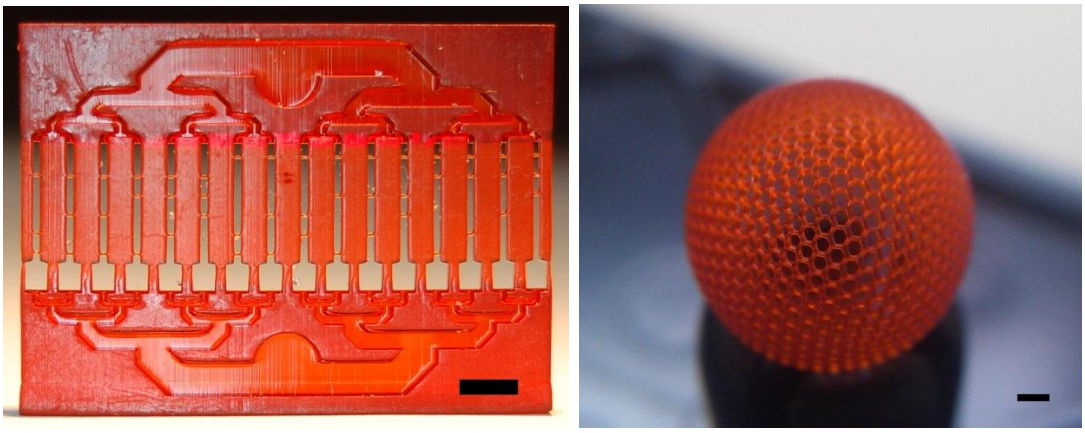


Figure 5.3: Example builds with the EnvisionTec Perfactory. Left hand pane (scale bar: 3mm) shows microfluidic channels with each chamber possessing a unique pattern of valve walls. Valve walls are 100 µm long and 250 µm wide (photo credit to C. Purssell). The right hand pane (scale bar: 1mm) shows a ball made from hexagons. The complete ball is 1 cm in diameter, the hexagons are 250 µm wide and the struts are 100 µm in diameter (photo credit to S. Leigh).

Chambers and channels can be made inside of MSL parts, as demonstrated by the mesh sphere shown in figure 5.3 and the channels in figure 5.4. Theoretically, the smallest size of these features is a single voxel, with dimensions of 20 x 20 x 25 mm. Single pixels can be resolved into features as can be seen in figure 5.4. A dye (discussed in section 4.1.2) present in the resin prevents the passing of light through the layer and into voids left behind by previous layers. Overflow of light in this manner can cause filling of the void, particularly if the void is only a few layers thick. The minimum totally enclosed vertical void height is ~500 µm or 10 layers. It was discovered through experimentation voids less than 8 layers tall will not be resolved by the EnvisionTec Perfactory, instead blocked channels will be produced.

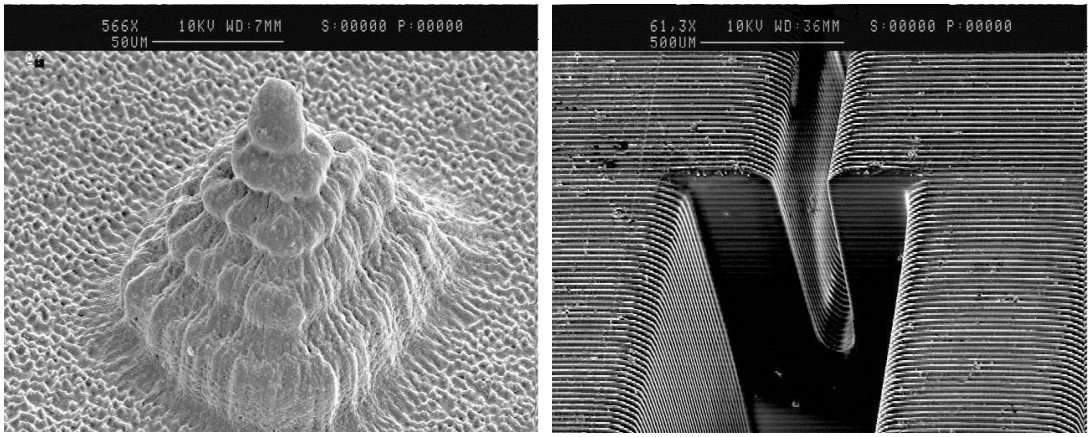


Figure 5.4: Scanning electron micrograph of a pyramidal microstructure (left) and channels (right) rendered in MSL resin. The pyramid is 100 µm wide at the base and 125 µm tall. The point of the pyramid consists of a single pixel. Part and photo credit to P. King.

The material blocking the channel in these cases, however, is usually not as solidified as the material of the rest of the part. The material in such channels can, therefore, be pushed out mechanically (using 1-2 bar air pressure) or by a soaking in isopropanol for 1-4 hours. Using such methods, enclosed channels with minimum vertical height of 400 µm have been resolved.

5.1.5. EnvisionTec build artifacts

During the manufacture of parts using the EnvisionTec Perfactory mini, artifacts were consistently present in the final builds. These artifacts were ridges or mounds in the final layer of the part. Since the flow cell chamber was composed of the last layers to be built, the presence of the ridges is likely to have negatively impacted the fluid dynamics within the flow cell chamber.

A square surface would yield a single pyramid peak in the middle, whereas a rectangular surface would yield a ridge along the middle of the long dimension of the rectangle. Figure 5.5 shows a Wyko image of once such ridge artifact.

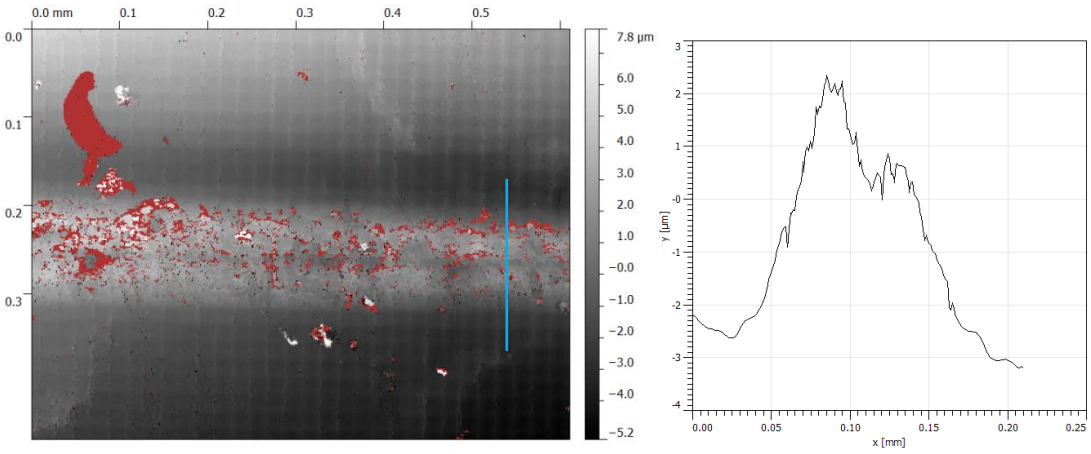


Figure 5.5: Wyko characterisation of a ridge-type artifact seen on flat rectangular surfaces. Blue bar indicates the position of the profile data shown in the right hand axes

Wyko measurement of the middle of the ridge indicates it is 6 ± 2 µm tall and 100 µm wide. At the ends of the ridge, the artifact reaches a maximum of 20 ± 5 µm in height and 280 µm wide. The presence of such an artifact in the flow cell chamber would significantly perturb the fluid dynamics.

Interestingly, using a faster peel speed reduced the appearance of the peak build artifacts, possibly because, at a faster peel speed, the part is pulled off the resin tray at once rather than being able to tear when peeled slowly. Furthermore, the presence of the artifact was correlated with the age of the R11 resin: Older resin produce more pronounced artifacts. Parts will require more force to pull them out of older resin, which is more viscous, therefore increasing the likelihood of detachment from the build platform and deformation of the final layer. It is possible that the artifacts are due to a tearing of the final build layer during the peeling process. For most build layers, the tear is then subsumed in the subsequent layer. In the final layer, however, the layer remains exposed.

The chamber depth of the radial flow cell was characterised using a Wyko interferometer. Three flow cells were tested, each with an expected chamber height of 100 µm. The determined chamber heights were 107, 99 and 96 µm. The height of the chamber of the radial flow cells can, therefore, be said to be 100 ± 4 µm.

5.2. EnvisionTec MSL fluidic parts

A number of flow cells were produced to evaluate the ability of the system to produce complex fluidic parts. The first set of flow cells discussed is for use in conjunction with living biological samples. The second set of flow cells utilise the minimum layer thickness to produce thin layer chambers. Finally, a flow cell that consists of multiple MSL parts designed to fit together into an easily reconfigurable optical flow cell is discussed. The aim of producing these parts was to determine the design and build parameters of parts used later in this thesis (see chapter 6) and also the suitability of parts built by the EnvisionTec Perfactory MSL process for a range of applications in biology and chemistry.

5.3. Flow cells for biological applications

Flow cells for biological applications are commonly made by casting a biocompatible or inert material, such as PDMS, into a mould made by micromachining or lithography. The following sections describe a flow cell for microbiological biofilm analysis made directly from MSL and also a PDMS flow cell cast from a mould that was made by MSL.

5.3.1. Flow cell for microbiology

The ability of the MSL process to create complex internal geometries can be applied to semi-permanent microbial cell culture. Microbial cell cultures are sensitive to insult by a variety of chemicals. The MSL resin is known to dissolve in a variety of solvents such as ethanol, isopropanol, acetone and toluene, as shown in previous work by Dr. Fauzan Harun, performed at the School of Engineering, University of Warwick. It was possible that the a component of the cured resin part is able to leech into these solvents and thus would leech into the culture medium and be toxic to the cells under culture. Biocompatibility testing of the R11 resin is necessary before flow cells using this resin can be used in culture experiments.

Biocompatibility of MSL resin was initially determined by soaking MSL parts in warm phosphate-buffered saline (PBS) overnight and then applying the PBS solution to microbial cultures. The biocompatibility testing and microbial cell culture work was undertaken in collaboration with Dr. Rich Boden, School of Biomedical & Biological Sciences, University of Plymouth. The design and fabrication of the microbial flow cell, including modelling of the proposed design, was performed by the author as a result of discussions with Dr. Boden. The assembly of the flow cell into a device for confocal measurements and all cell viability experiments were performed was performed by Dr. Boden

Cells were not adversely affected by the PBS that had been used to soak the MSL test part. MSL parts can, therefore, be used in semi-permanent cultures of bacterial biofilm.

By incorporating glass into the device, confocal measurements through a flow cell are possible. Furthermore, progress of reactions being performed by microbial cultures can be followed through time after the addition of the necessary substrates. A short chamber is required for microscopy due to the small depth of field available. An inlet and outlet are required for the exchange of media and the addition of a substrate. Ideally the addition of media to the chamber should be performed as uniformly as possible for two reasons: 1) So that the flow rate does not become sufficient to dislodge cells in any one area and 2) So that the whole culture area receives a treatment at the same time.

The confocal microscope used here (Leica SP2) incorporates a motorised stage that is used to move the sample in 3D. The interfacing tubing was inserted into the inlet/outlet ports vertically to avoid the tubing becoming dislocated or leaking during the movement of the stage. A shallow groove was introduced into the top surface of the device to facilitate sealing of the glass against the device.

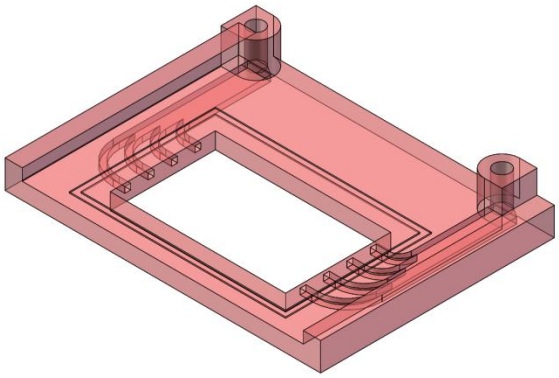


Figure 5.6: Isometric projection of the CAD model of the microbial flow cell. Note the vertical inlet ports, the groove on the top surface for sealing and the arrangement of the inlet ports. Image not to scale.

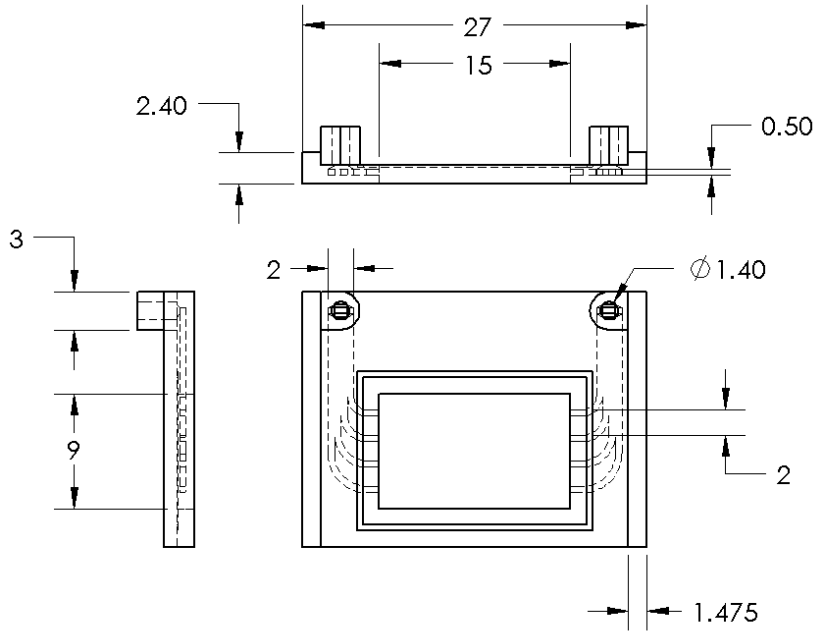


Figure 5.7: Schematic diagram of the CAD model of the later iteration of the microbial flow cell. All dimensions are in millimetres.

COMSOL modelling (see section 4.6 for details) was undertaken to investigate the fluid profile into the chamber. A laminar flow model was created that combined the microfluidic channels and a portion of the chamber (see figure 5.8). Several inlet flow rates (1, 0.1 and 0.01 m/s) were simulated with the boundary condition of the outlet defined as zero pressure at the outlet.

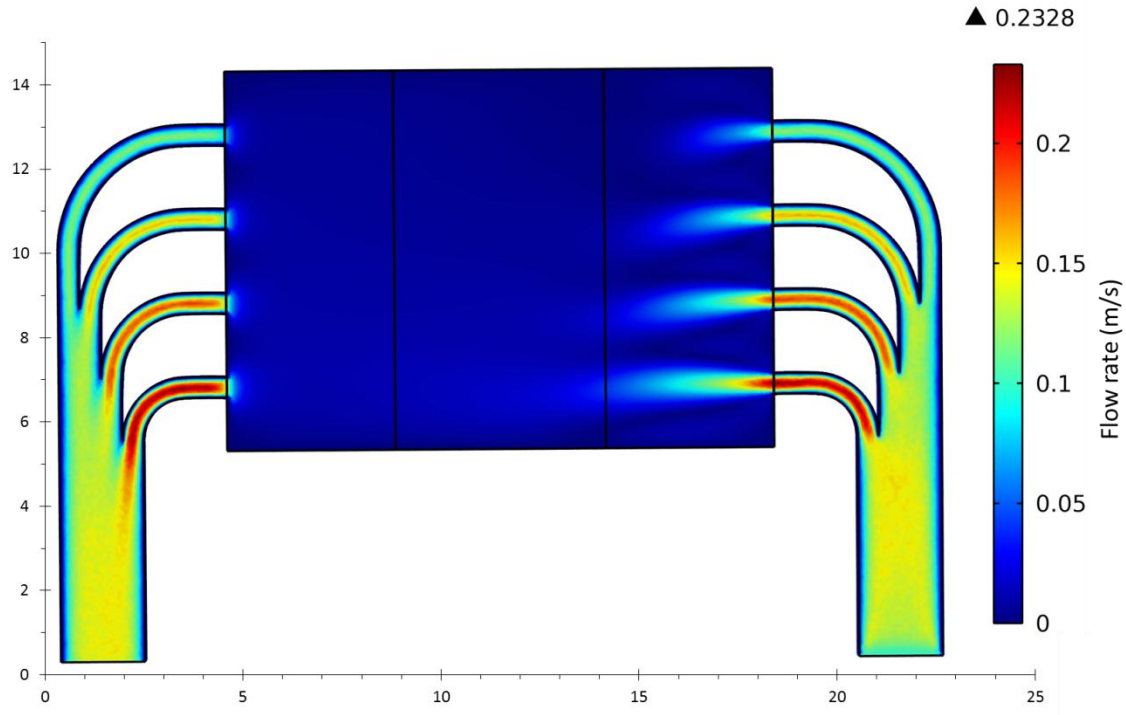


Figure 5.8: COMSOL modelling flow through MSL flow cell for microbiology. Fluid enters one of from the inlet at the bottom right of the figure, is split into four channels that then enter the main chamber. The same process then happens in reverse on the left hand side. Fluid flow is represented with the heatmap as indicated. Inlet flow rate is 0.1 m/s. X-Y scale is in millimetres. See text and methods for details of simulation.

The flow rate through each of the channels is proportional to each channels length; the shortest channel has the largest flow rate whilst the longest channel has the shortest. The relative velocity profiles at three inlet flow velocities can be seen in figure 5.9. The flow rates in the fastest flow channel, at 6.5 mm in each case, overlap because the measurements are relative to the maximum, which occurs in this flow channel. There is a greater range in the flow rates at a lower overall flow velocity (~25%) than with a higher overall flow rate (~45%). The profiles in figure 5.9 are interpolated from the mesh used in the simulation.

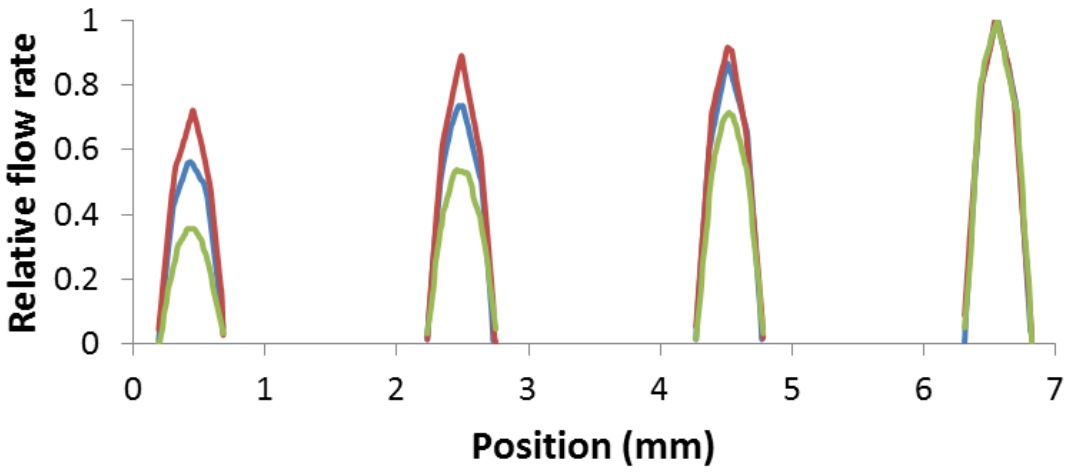


Figure 5.9: Relative velocity profiles of each of the four inlets as they enter the main chamber. Red line denotes inlet flow velocity of 1 m/s, the blue line denotes an inlet flow velocity of 0.1 m/s and the green line denotes an inlet flow velocity of 0.01 m/s.

The internal surfaces of the microfluidic channel can be considered to be rough (see section 5.1), but the simulation was performed assuming smooth walls. In laminar flow with non-slip walls the velocity at the walls is zero, so implementing a surface roughness will not reduce the velocity at the wall. Significant roughness, however, might cause the stationary layer to extend further into the channel than when the surface is smooth. Observed roughness is <2 µm, which is not significant with respect to a channel that is ~500 µm wide.

Interfacing with the microbiological flow cell was achieved using a needle inserted into the vertical inlet ports and glued in place with silicone cement. Given this fact and that the fluid exchange was performed by hand using a syringe, a low flow rate was expected. The lowest inlet flow velocity (0.01 m/s) corresponded to a flow rate of 5 mL/s and is representative of the flow rate expected during a hand syringe transfer.

The flow cell discussed here shows that small channels, with a 500 x 500 µm cross section, can be built in MSL using the EnvisionTec Perfactory machine. This flow cell was successfully employed to obtain real-time measurements of environmental microbial biofilm samples using a confocal microscope.

5.3.2. MSL moulds for PDMS flow cell

It is possible to make moulds out of MSL parts which can then be used to make flow cells by casting another material into the MSL mould. One of the key weaknesses of the MSL method is the limitation in material: only materials that can be combined with a suitable photocrosslinker can be used. Polydimethoxysilane (PDMS) is a plastic rubber material, which is transparent for visible light, cures from a liquid to a solid in a heat dependant reaction and is easy to handle and pour into moulds at room temperature. PDMS has been used in many cell-based assays previously with no deleterious effect on a variety of cell types (HepG2[1], CHO-K1[2], HeLa[3]) although PDMS has been found to selectively absorb solutes from solution[4]. PDMS was chosen as it is soft enough to be sealed against a surface with minimal force. Furthermore, many reports exist indicating that PDMS is non-toxic to cells although there is literature to suggest that PDMS can absorb small molecules that might affect cells in culture indirectly[5].

The following section describes the casting of a PDMS flow cell from an MSL mould. For evaluation purposes, a surface acoustic wave (SAW) device was added and used to make measurements of insect cells cultured in the flow cell chamber in order to determine whether the flow cell could be successfully used for this type of measurement. The design and fabrication of the mould and the casting of PDMS into the mould to produce the PDMS flow cell was performed by the author as a result of discussions with Dr. Zoltan Racz. The assembly of the PDMS flow cell onto the SAW device was carried out by Mr. Sanju Thomas, School of Engineering, Warwick University.

Creating inlet and outlet ports in monolithic flow cells cast from a soft material is difficult. The ports are essentially suspended voids within the mould and so two options are available; drilling after curing or partial disassembly of the mould prior to removal of the cast flow cell. Since drilling of soft materials is not reliable, partial disassembly of the mould

was performed. A mould with rods connecting the outer walls to the chamber walls would create a cast with a permanent access tube, but the cast would be destroyed when removal was attempted as the cast would have to travel through the rods. The mould could be made in multiple parts and/or broken to yield the final part but since smooth sealing surfaces and reusable moulds were desired this option was not preferred. The solution presented in figure 5.10 was used to avoid this problem. Removable rods were inserted through holes made in the mould for the casting. Once curing was complete, the rods could then be removed and the cast lifted from the mould.

The rods used to make the voids were 0.5 mm outer diameter (OD) Teflon tubing. Holes to allow for rod insertions were made in the mould with a 50 µm tolerance (see figure 5.10 for locations of holes). Rods were cut to length using a diamond scribe and inserted before the PDMS was cast. PDMS curing was as described in the methods section (see section 4.2.1). Once curing was complete the glass capillaries could be removed with a slight twisting motion. A scalpel was used to loosen the flow cell from the mould and then also used to lever the cast from the base of the mould. Care was required to ensure that the mould was not damaged by the scalpel during cast removal.

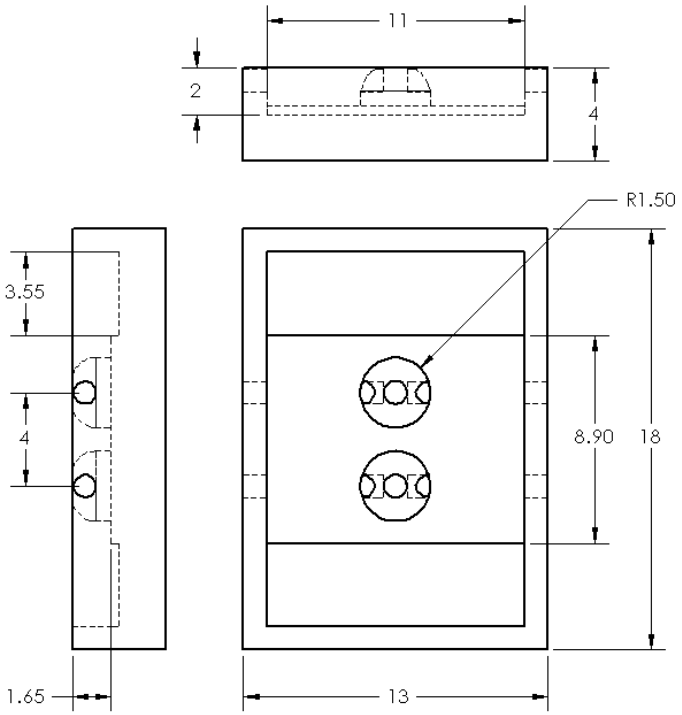


Figure 5.10: Schematic diagram of mould for SAW device flow cell. All dimensions are in millimetres.

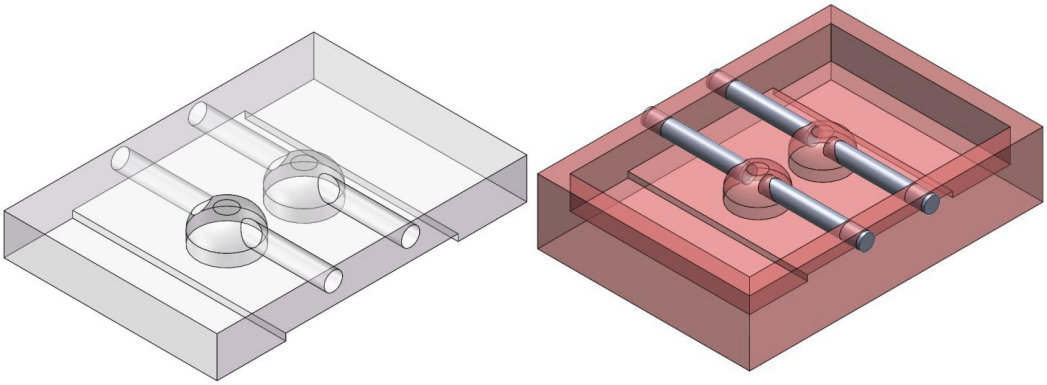


Figure 5.11: Isometric projections of CAD models of PDMS cast (left) made from MSL mould (right). Teflon rods are placed in the holes during curing of the PDMS. Once cured, the rods are removed prior to the cured PDMS flow cell so that tubes are left in the resulting PDMS part. Images are not to scale.

To interface with the finished cast short, flat ended needles with LuerLoc® fittings were used. The needles had the same diameter as the tubing used to make the holes and so no further sealing was required. Finally, the chambers themselves were given domed tops to prevent the trapping of bubbles in the corners of a cuboid chamber. These considerations represent a significant improvement from previous devices that have been made by

micromachining[6]. Furthermore, making a similar mould by traditional stereolithographic techniques would require multiple masks and thick layers of photoresist. The mould described here was made in one step in a matter of hours. The time taken to produce a first finished part, including time taken to make the mould itself was 8 hours.

5.3.3. Effect of uncured R11 on PDMS curing

Unexpectedly, PDMS curing time was found to be closely correlated with extent of R11 post curing. R11 was normally post cured after building in the EnvisionTec flasher box. It was found that flashing the R11 mould for less than 3000 flashes resulted in a significant reduction in curing rate: An insufficiently post-cured R11 mould extends the time that a PDMS cast will cure in from <1 hour to 24 hours at 60°C. Post curing of the R11 part for a minimum of 3000 flashes resulted in PDMS curing occurring within 6 hours at 60°C. Curing for 5000 flashes reduced curing time to 3-4 hours at 60°C. Interestingly, curing of PDMS in an R11 mould, which had already been used at least once, resulted in a return to normal curing time of 1-2 hours for PDMS at 60°C. Also, incompletely post cured R11 moulds could be conditioned to allow PDMS to cure in the standard 1-2 hours by heating of the R11 part. 24-48 hours of heating at 60°C was necessary for this conditioning to take place.

It was hypothesised that a chemical component of the uncured R11 resin, which was present in incompletely cured R11 parts, inhibited the curing process in PDMS. Post curing in the EnvisionTec flasher box for several thousand flashes or heating to 60°C for 24-48 hours completes the curing process, which removes or locks in the component of R11 resin that is inhibitory to PDMS curing. The most likely candidate is active, unterminated acrylate chains on the surface of the MSL mould which are removed by prolonged curing.

The SAW sensor used here consisted of a conductor patterned onto a piezoelectric surface. The conductor pattern is a pair of interdigitated combs. When a cycling voltage is applied across the combs the piezoelectric surface contracts and expands producing a wave

that travels across the surface. The frequency and amplitude of the wave depends on parameters of the applied cycling voltage, the material being used, the size of the combs and, importantly, any loading that might be present on the surface. The SAW device can be used, therefore, to detect changes in the properties, such as viscosity and density, of the medium at the surface. The PDMS flow cell used in combination with SAW devices was used successfully to obtain experimental data in conjunction with the Sf9 insect cell line (work by Z. Racz, S. Thomas and S. Pathak, unpublished). This work shows how high quality PDMS casts can be made from MSL moulds in the shortest time possible.

5.4. Thin layer flow cells

Thin flow cells are advantageous in some applications as all of the analyte or sample can be brought to the sensor. The MSL process, where parts are made by addition of successive layers, is ideally suited to this application. To determine how best to fabricate flow cells where the flow chamber is <100 μm tall, two flow cells were made. These flow cells were tested by performing electrochemical measurements within them. The work described in this section contributed to two papers that are summarised in appendix B.2 and B.3

Previous thin layer flow cells are usually assembled from multiple parts where a spacer is placed between two plates, each containing suitable inlet and outlet channels. Some example spacer sizes include the following: 16 μm gasket[7], 80 μm adhesive tape[8] and ~50 μm Kalrez gasket[9]. An advantage of the MSL process is that the device can be constructed monolithically; the gasket and the inlet/outlet channels are built in the same part. The gasket and the inlet/outlet will always be perfectly aligned in every run and dead volumes can be minimised.

Radial flow cells are so called as they create a radially symmetric flow through a chamber. The flow within the chamber is said to be radial as the two ports are arranged so that the inlet is in the centre of the circle and the outlet is towards the periphery. The non-

uniformity of the outlet(s), however, means that the flow is very unlikely to be radial, instead the flow will tend to find the shortest path through the chamber resulting in, at best, a partially radial flow. Radial flow cells are available from a variety of manufacturers: DropSens use a direct digital manufacturing method similar to MSL[10] whereas BASi use computer numerical control (CNC) to manufacture a flow cell, which requires assembly before use. The assembly of the flow cell makes it difficult to quickly swap out structures from the assembly, such as electrodes[11]. Furthermore, previously published work tends to use flow cells in which the inlet and the outlet are discrete points within a circular flow chamber[11,12].

Radial flow cells typically suffer from issues of recirculation[13] where flow becomes turbulent instead of laminar. The comparatively large chamber dimensions (1 mm height and 30 mm radius) in this case probably contributed to the fact that turbulent flow occurred. A higher flow rate and a smaller chamber would most likely limit this effect.

The two test chambers that were created here were designed as linear or radial flow cells and were tested by Dr. Eleni Bitzou and Dr. Mike Snowden, respectively, Department of Chemistry, University of Warwick. The design and fabrication of the flow cells, including modelling of the proposed design for the radial flow cell, was performed by the author as a result of independent discussions with Drs. Bitzou and Snowden. Once fabricated, the flow cells were then assembled into sensing devices by sealing against electrodes was performed by Drs. Bitzou and Snowden. Results from all flow characterisation experiments, presented in figures 5.16 and 5.17, were obtained by Dr Bitzou and Dr Snowden respectively.

Figure 5.12 shows two 3D models of 3D flow cells made to evaluate the EnvisionTec Perfactory MSL machine's ability to build parts using the minimum layer thickness. Both models include recessed inlets into which tubing can be glued. The outlet of both parts also

included openings to allow for the necessary reference and ground electrodes to be inserted.

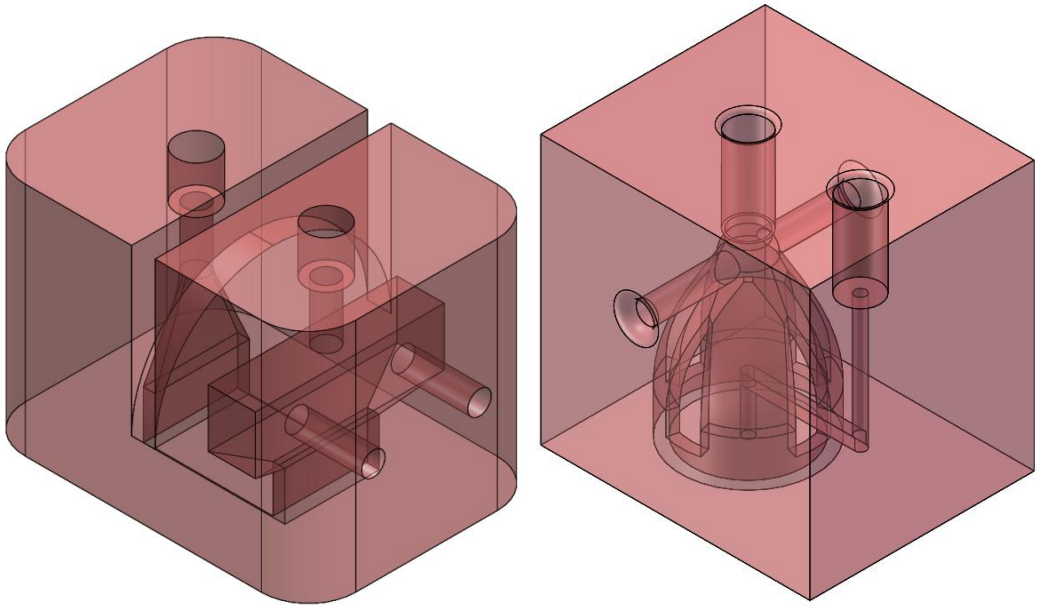


Figure 5.12: Isometric projection of the CAD model of the thin layer flow cell (left) and radial flow cell (right). Recesses on the inlet and outlet are for inserted tube interfaces. The fluid flows from the inlet to the outlet through complex geometries that are designed to perform different functions. Image not to scale

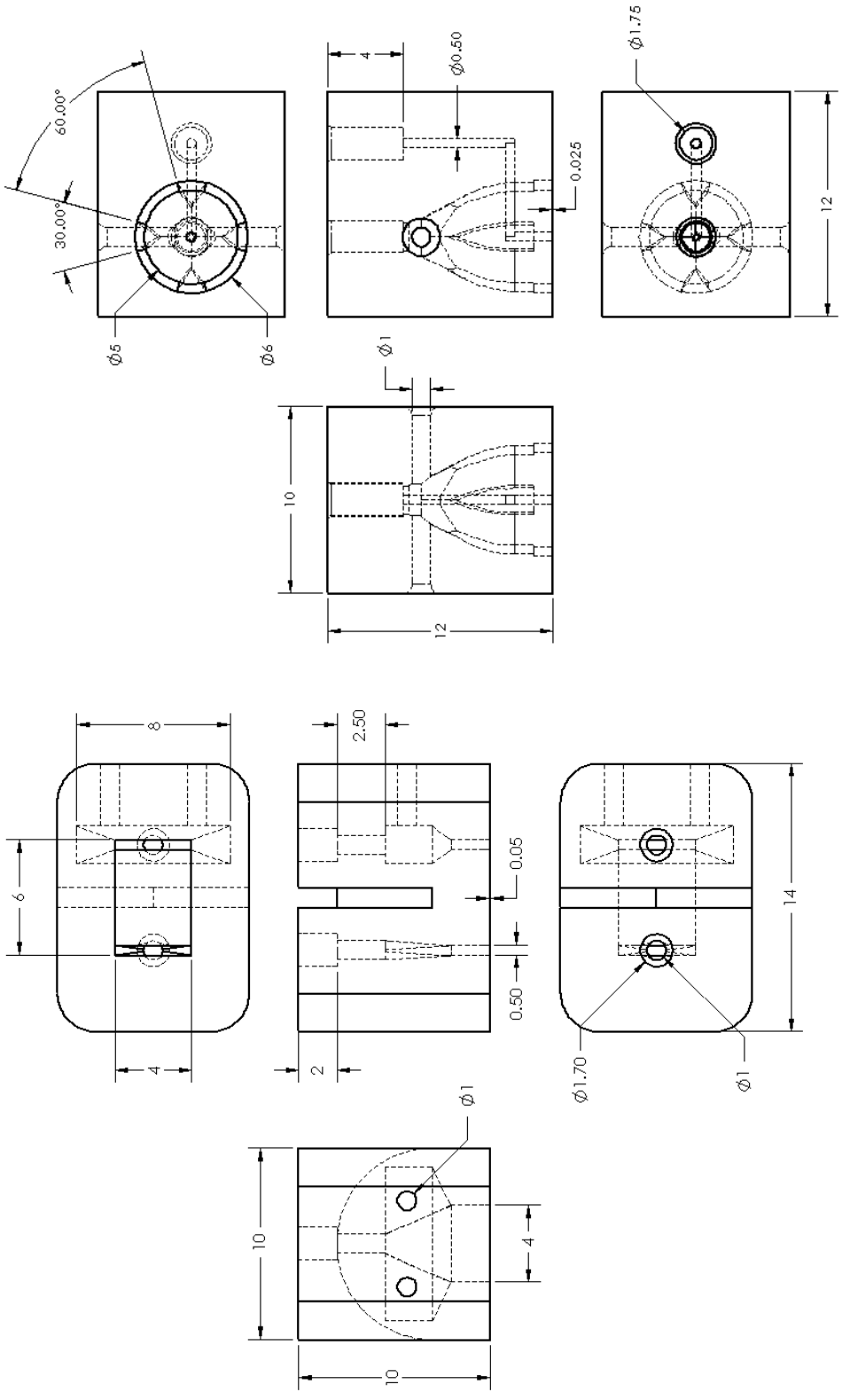


Figure 5.13: Schematic diagrams of the linear (left) and radial (right) flow cells built by MSL. Note the height of the flow chambers in each schematic (50 µm in the left and 25 µm in the right). All dimensions are in millimetres.

COMSOL (v4.2) modelling was undertaken to compare the radial flow of the radial flow cell design and the DropSens flow cell. A direct comparison between the two can be seen in figure 5.14. A Navier-Stokes convection simulation was carried out with a chamber height

of 200 µm and an inlet flow rate of 10 mL min^{-1}. Whereas the radial flow cell has axially symmetrical flow from the inlet to the circumference, the DropSens flow cell has a clearly biased flow. This result indicates that the DropSens flow cell will not exhibit true radial flow.

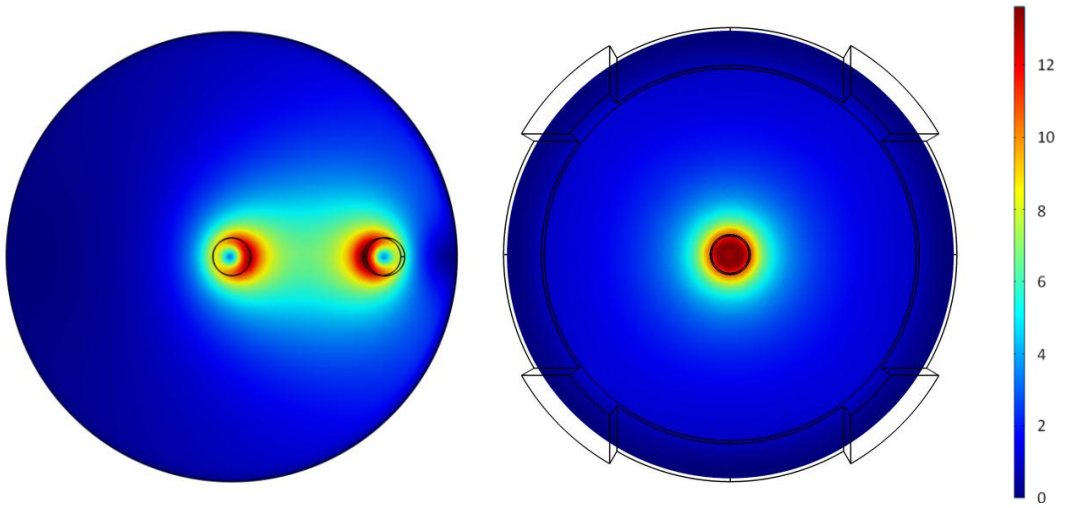


Figure 5.14: COMSOL Modelling of the currently available DropSens radial flow cell (left) verses relevant modelling of the MSL Radial flow cell (right). The fluid velocity profile shown is taken 50 or 100 µm from the base of the chamber for the DropSens and Radial flow cell respectively. The colour bar indicates the flow rate in mL/s.

An example of how the flow cells described in this section are interfaced is shown in figure 5.15. A small amount of cyanoacrylate super glue is applied evenly to the outside of tubing of the appropriate diameter. The tubing is then inserted into the inlet and outlet ports and twisted gently to ensure the glue coats the entire surface.

Both flow cells were used to detect specific analytes in solution, amphoterically. The following figures, 5.16 and 5.17, show representative amphometric traces of the linear and radial flow cells respectively. Data in these figures was obtained by E. Bitzou and M. Snowden.

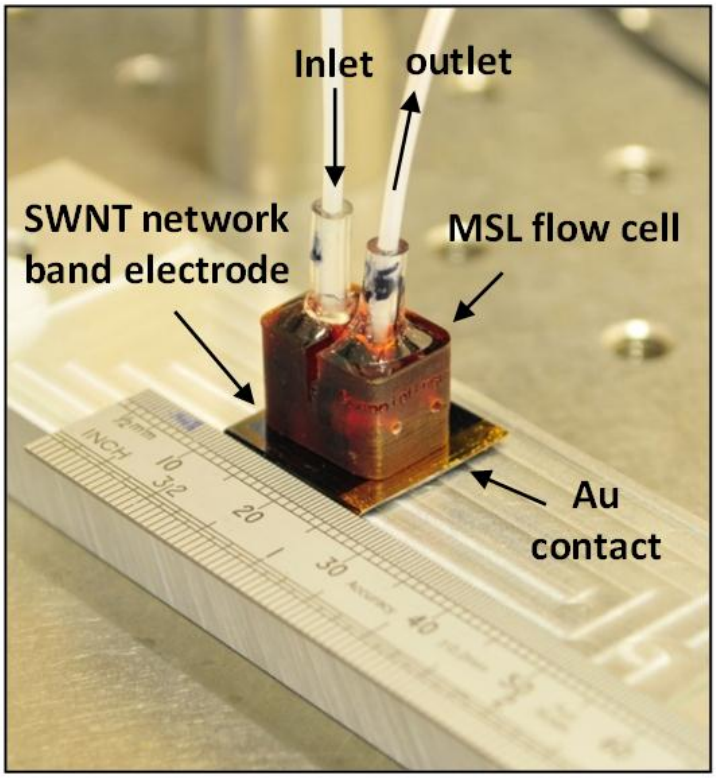


Figure 5.15: Image of thin layer flow cell on electrode surface with inlet and outlet connectors. The fixing string used to seal the flow cell against the electrode is not shown. A ruler with millimetre graduations is shown beside the flow cell for scale purposes. SWNT in this figure refers to single walled nano tubes that compose the electrode. Photo credit to E. Bitzou.

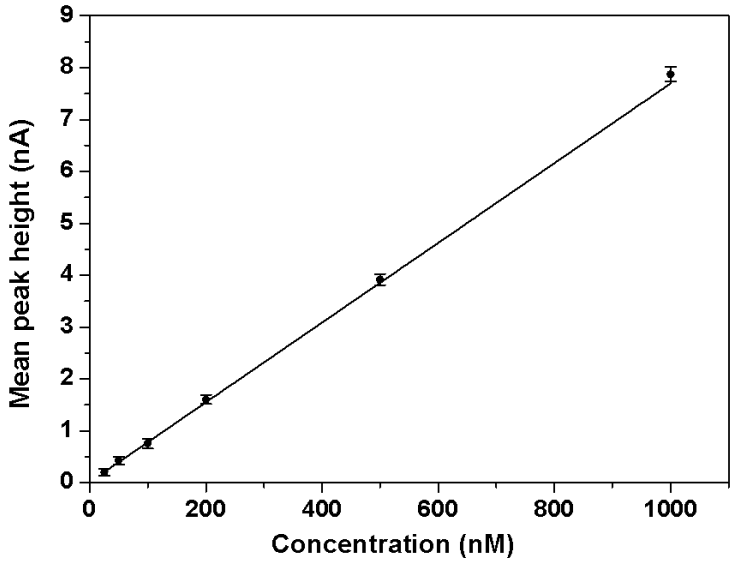


Figure 5.16: Data obtained from the thin layer flow cell. Different concentrations of dopamine were applied to the flow cell/electrode set up and the amphometric trace shown above was obtained. The inset shows detail of the lowest concentration tested. Data obtained by E. Bitzou.

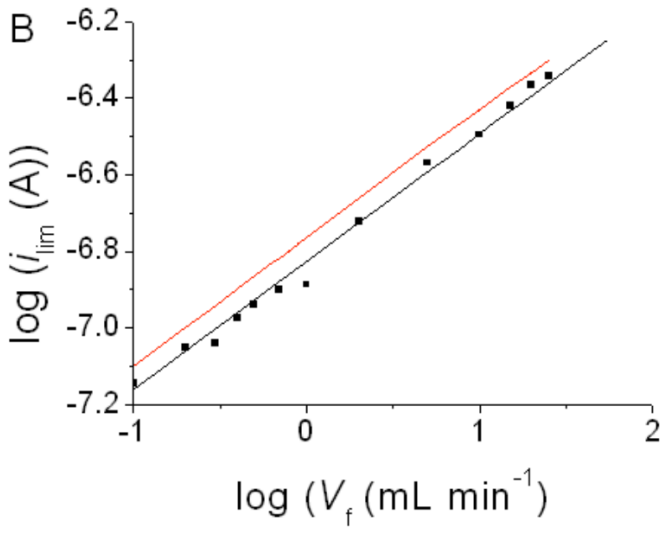


Figure 5.17: Data obtained from the radial flow cell. Graph of current flow (i_{lim}) against flow rate (V_f) for comparison of experimentally obtained (black dots) and expected (red line) current flow as determined by COMSOL modelling. Data obtained by M. Snowden.

The flow rate-current response reaction shown in figure 5.17 is the oxidation of 10 µM $FcTMA^+$ in 1.0 M KNO_3 at an Au disc electrode. The flow cell was placed over the electrode using a jig. The expected response line (figure 5.17, red) was determined from the Levich equation[14] and differs from the experimentally determined response line (figure 5.17, black). The cause of this difference could be due to variation in the positioning of the flow cell relative to the electrode or inconsistencies in the inlet or chamber geometries, both of which could serve to increase the stagnant region above the electrode that would reduce the effective concentration at the electrode in a manner similar to that seen in figure 5.17.

This work demonstrates that the minimum layer thickness of the EnvisionTec Perfactory MSL machine can be used to make a variety of flow cells. Furthermore, these flow cells can be used to make novel measurements that closely match theory. Work using MSL flow cells has formed the basis of three publications[15–17] with a further two on the thin layer flow cells in preparation.

5.5. Optical flow cell

Optical flow cells are used extensively in biology as light can be used to observe changes in proteins within tissues, cells or *in vitro*. Light is capable of interrogating tissues rapidly, relatively unobtrusively, involves no radioactive components and has become increasingly popular over the last two decades since the discovery of green fluorescent protein (GFP), a widely used protein label[18]. Light and fluorescent microscopy can be used to characterise, find and follow proteins that have been specifically labelled with a marker, such as GFP. The fact that certain protein side chains absorb UV light (principally tyrosine and phenylalanine) can be used to monitor changes in proteins, as the fluorescence intensity of these side chains changes depending on their environment (ie hydrophilic vs. hydrophobic), a process called dichroism.

Although monolithic parts are possible with the MSL process, it is necessary to build parts out of multiple sections when the desired geometry or material properties of the build material are not suitable. R11, for example, contains a dye which partially blocks ultraviolet (UV) light below 500 nm which permits close control of layer thickness. Several alternative resins, however, are available for use with the EnvisionTec Perfactory system. Spectrophotometry was employed to determine whether any of these materials are suitable for the building of optically transparent parts.

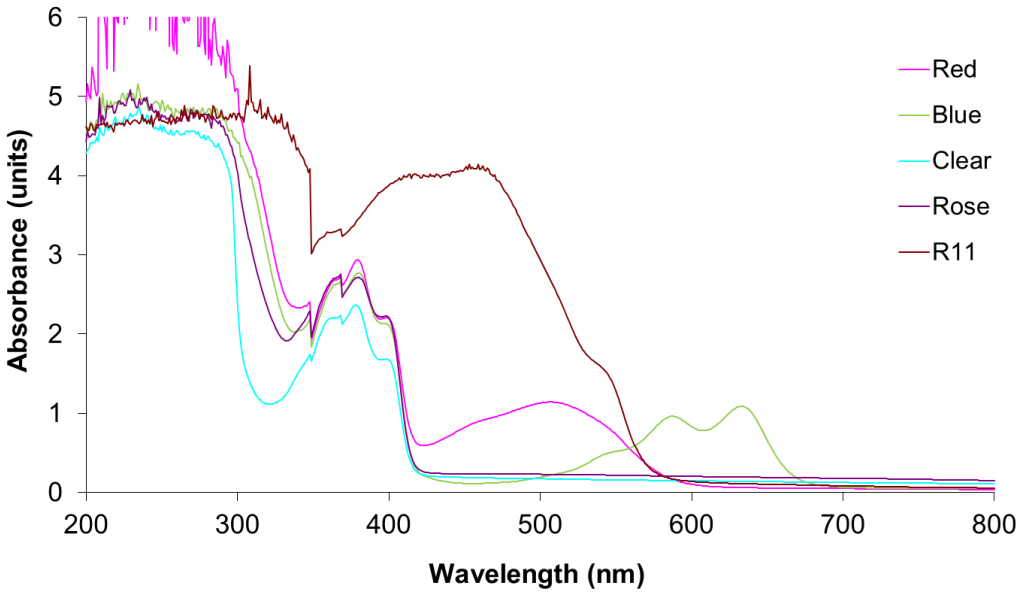


Figure 5.18: Spectrophotometric analysis of different resin materials from 200 to 800 nm. For clarity the absorbance axis has been limited to 6 units.

The absorbance spectra of five resin formulations were obtained and are shown in figure 5.18. All the tested resins had high absorbance values (>2) from 200 to 300 nm. An absorbance value of >2 is essentially opaque. The resins tested were nearly or totally opaque throughout the range of UV light (200 – 400 nm). R11 is also practically opaque up to 500 nm. Whilst two of the resins (clear and rose) could be used for wavelengths from 420 nm upwards, most would not be suitable for such applications. The cause of this opacity is most likely to be the specific dye molecule used in each material.

The opacity of all types of resin suitable for use with the EnvisionTec Perfactory machine meant that, for optical measurement through an MSL flow cell, another material must be used. Multiple component parts can be built by MSL and used in conjunction with materials that do possess suitable material properties to make a whole microsystem. To demonstrate this, a flow cell suitable to optical measurements was fabricated and tested using linear dichroism (LD).

LD is a method for determining the secondary structure of proteins in solution. Flow cells are used widely in linear dichroism, such as the Couette flow cell[19]. The Couette flow is complicated and requires multiple parts to work, including a quartz rod rotating at more than 3000 rpm. Fluid travelling through small channels also experiences a shear force as the walls drag on the moving fluid. Given sufficient shear force, molecules in solution can be induced to align allowing LD measurements without a complicated flow cell involving moving parts. Testing of the flow cell for LD was undertaken in collaboration with Miss Xi Cheng, Department of Chemistry, University of Warwick. Design and fabrication of the device was undertaken by the author as a result of discussions with Miss Cheng's PhD supervisor Prof. Alison Rodger. Miss Cheng then performed all the characterisation of the flow cell including obtaining the experimental data presented in figure 5.20.

Two pieces of high quality quartz (UV fused silica, UQG optics) were incorporated to achieve a sufficient level of optical clarity through the channel. The glass was pre-cut into 10 mm and 5mm square pieces 1 mm thick. The glass needed to form two opposite surfaces of the flow channel. To do this there would have to be a square of unsupported build of at least 5 x 5 mm with a 1 x 5 mm channel running through the middle. Although it is possible to build small overhangs with the Perfactory machine, the size of the overhang and the importance of the uniformity of the channel cross-section led to the decision to build the flow cell in three separate components. By building the device as three separate components, the overhangs could be avoided entirely (see figure 5.19). The channel was as thick as the middle part and size uniformity of this piece was critical to the final channel size. This part was built with the same settings for the burn-in and build layers. The small thickness of the part, however, meant that it was difficult to handle and would warp if left unrestricted during curing. To prevent warping, the middle part was sandwiched between two glass slides. The weight of a single glass slide was sufficient to prevent the middle part

from warping during the flashing process. A tolerance of 50 μm was sufficient to allow the top and middle parts to fit into the base part.

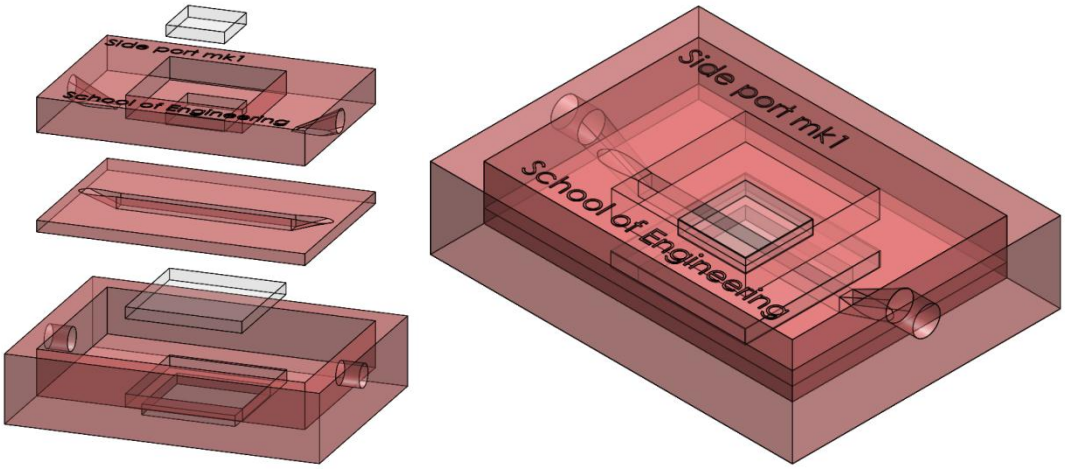


Figure 5.19: A dimetric (left) projection of a CAD model of the optical flow cell. The flow cell is exploded into the three component parts (left) and an isometric projection of the same three part optical flow cell shown assembled (right). Image not to scale.

The final device was sealed with silicone sealant to allow reconfiguration after assembly. Parts could then be used interchangeably, allowing top, middle or bottom parts to be swapped out without requiring an entire device to be built from scratch.

A 100 μL bolus of calf thymus-DNA (ct-DNA) (250 μM) was injected into the optical flow cell at time zero. As the DNA solution passed through the channel, shear forces cause the long DNA molecules to align resulting in an observable change in the LD absorbance at 260 nm. Figure 5.20 shows the relationship between flow rate and change in LD absorbance at 260 nm. Increasing the flow rate increases the observable signal as the shear force the solution experiences increases with the speed of the solution through the channel.

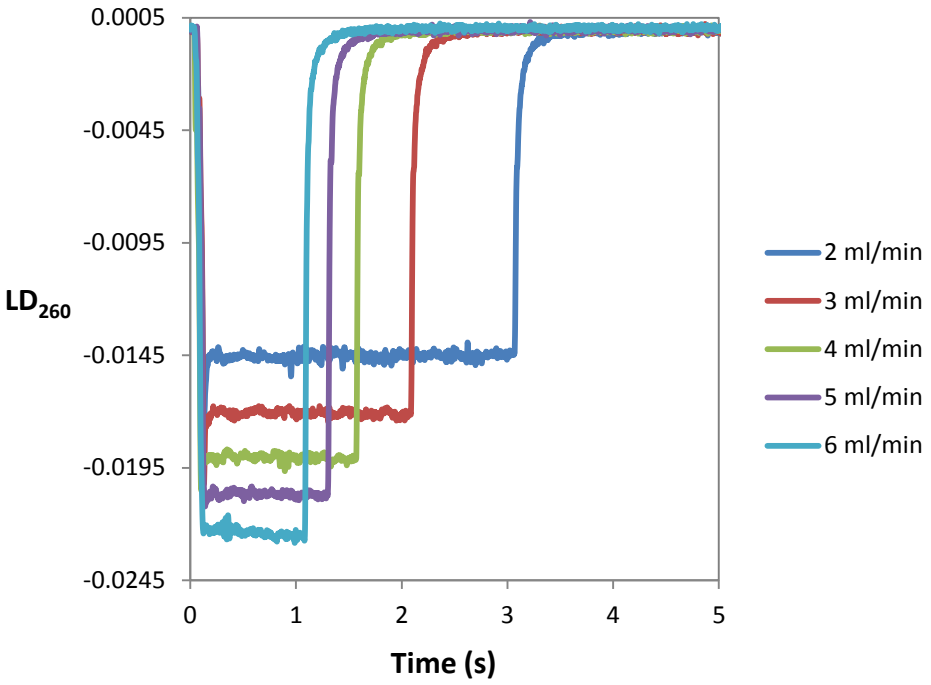


Figure 5.20: The relationship between continuous channel flow LD and flow rate. 100 µL of 250 µM ct-DNA was injected at time 0 and LD was measured continuously at 260 nm. Data obtained by X. Cheng.

It is unlikely that molecule aligning shear flow will occur throughout the channel of the device, due to its large size. Therefore Couette flow is not likely to occur thoughout the channel. Figure 5.20, however, shows that sufficient analyte does undergo shear flow-dependant alignment to be detectable by LD. The optical flow cell demonstrates the ability of the EnvisionTec Perfactory to fabricate components of a flow cell rapidly and accurately. These parts can be assembled and disassembled easily to change channel sizes and configurations. The utility of the resulting flow cell was demonstrated by aligning molecules of ctDNA.

5.6. Conclusions

Flow cells, principally made from PDMS, are used widely in biological microfluidics. Flow cells have been made from a variety of materials and by a variety of methods. Generally flow cells are composed of multiple parts, but are difficult to assembly repeatably and possess non-ideal fluid dynamics. This chapter shows that:

- Direct digital manufacturing can be used to make a range of monolithic or multi-part flow cells.
- PDMS flow cells cast from R11 moulds can be used in experiments with insect cells.
- R11 flow cells can be used directly for microbiological culture assays.
- R11 flow cells with complex internal geometries can be used to facilitate electrochemical and microbiological experiments.
- Optical components can be incorporated into R11 parts made by MSL for the purposes of assembling multiple parts into a single flow cell.

The work presented herein has pushed the ability of the EnvisionTec Perfactory to produce devices which were necessary for experiments to be possible, producing novel and interesting scientific publications across a diverse set of fields. The diverse range of applications described herein required close collaboration with several different departments. Design and fabrication was performed by the author as a result of discussions with various individuals during the project. Apart from modelling of flow cell designs during the design period, which was performed by the author, all characterisation and operation of the produced flow cells was performed by the collaborators in each case.

The potential for moulding of PDMS fluidic components in R11 moulds made by MSL is only briefly explored here. In chapter 6 the experience gained here was used to explore multilayer devices which incorporating moving parts within them, allowing a precise and repeatable level of fluidic control.

5.7. References

1. Leclerc, E., Sakai, Y. & Fujii, T. Cell Culture in 3-Dimensional Microfluidic Structure of PDMS. *Biomedical Microdevices* **5**, 109-114 (2003).
2. Hufnagel, H. *et al.* An integrated cell culture lab on a chip: modular microdevices for cultivation of mammalian cells and delivery into microfluidic microdroplets. *Lab on a chip* **9**, 1576-82 (2009).
3. Hung, P.J. *et al.* A novel high aspect ratio microfluidic design to provide a stable and uniform microenvironment for cell growth in a high throughput mammalian cell culture array. *Lab on a chip* **5**, 44-8 (2005).
4. Mukhopadhyay, R. When PDMS isn't the best. *Analytical chemistry* **79**, 3249-3253 (2007).
5. Toepke, M.W. & Beebe, D.J. PDMS absorption of small molecules and consequences in microfluidic applications. *Lab on a chip* **6**, 1484-6 (2006).
6. Rácz, Z. *et al.* Cell-based surface acoustic wave resonant microsensor for biomolecular agent detection. *Transducers '11* (2011).
7. Deng, H. & Berkel, G.J.V. A Thin-Layer Electrochemical Flow Cell Coupled On-Line with Electrospray-Mass Spectrometry for the Study of Biological Redox Reactions. *Electroanalysis* **11**, 857-865 (1999).
8. Horii, D., Atobe, M., Fuchigami, T. & Marken, F. Self-supported paired electrosynthesis of 2,5-dimethoxy-2,5-dihydrofuran using a thin layer flow cell without intentionally added supporting electrolyte. *Electrochemistry Communications* **7**, 35-39 (2005).
9. Jusys, Z., Kaiser, J. & Behm, R.J. A novel dual thin-layer flow cell double-disk electrode design for kinetic studies on supported catalysts under controlled mass-transport conditions. *Electrochimica Acta* **49**, 1297-1305 (2004).
10. Fanjul-Bolado, P., Queipo, P., Lamas-Ardisana, P.J. & Costa-García, A. Manufacture and evaluation of carbon nanotube modified screen-printed electrodes as electrochemical tools. *Talanta* **74**, 427-33 (2007).
11. Soucaze-Guillousi, B. & Kutner, W. Flow Characteristics of a Versatile Wall- Jet or Radial-Flow Thin-Layer Large-Volume Cell for Electrochemical in Flow-Through Analytical Systems. *Electroanalysis* **9**, 32-39 (1997).
12. Liu, Z., Niwa, O., Kurita, R. & Horiuchi, T. Miniaturized thin-layer radial flow cell with interdigitated ring-shaped microarray electrode used as amperometric detector for capillary electrophoresis. *Journal of chromatography. A* **891**, 149-56 (2000).
13. Detry, J., Deroanne, C., Sindic, M. & Jensen, B. Laminar flow in radial flow cell with small aspect ratios: Numerical and experimental study. *Chemical Engineering Science* **64**, 31-42 (2009).
14. Bard, A.J. & Faulkner, L.R. *Electrochemical Methods: Fundamentals and Applications, 2nd Edition*. 856 (Wiley: 2000).
15. Snowden, M.E., King, P.H., Covington, J.A., Macpherson, J.V. & Unwin, P.R. Fabrication of versatile channel flow cells for quantitative electroanalysis using prototyping. *Analytical chemistry* **82**, 3124-31 (2010).
16. Fisher, R.D. *et al.* Dissolution Kinetics of Polycrystalline Calcium Sulfate-Based Materials: Influence of Chemical Modification. *ACS applied materials & interfaces* **3**, 3528-3537 (2011).
17. Snowden, M.E., Unwin, P.R. & Macpherson, J.V. Single walled carbon nanotube channel flow electrode: Hydrodynamic voltammetry at the nanomolar level. *Electrochemistry Communications* **13**, 186-189 (2011).
18. Tsien, R.Y. The green fluorescent protein. *Annual review of biochemistry* **67**, 509-44 (1998).
19. Rodger, A. *et al.* Flow oriented linear dichroism to probe protein orientation in membrane environments. *Physical Chemistry Chemical Physics* **4**, 4051-4057 (2002).

Chapter 6

6. Design and operation of PDMS microfluidic device

The production of a combinatorial mutant library using bench top methods is time consuming and materially inefficient. Ideally, all or part of the production process should be transferred to a microfluidic device so that the assembly can occur using the minimal raw materials, highest rate and maximal efficacy. This chapter describes the design and fabrication of a microfluidic device from PDMS by the process of multilayer soft lithography. The fabrication process required characterisation of microfluidic valves and the development of electronic hardware and control software that is described herein. The flow chart in figure 6.1 represents how each necessary process relates to the overall goal of the fabrication of the microfluidic chip.

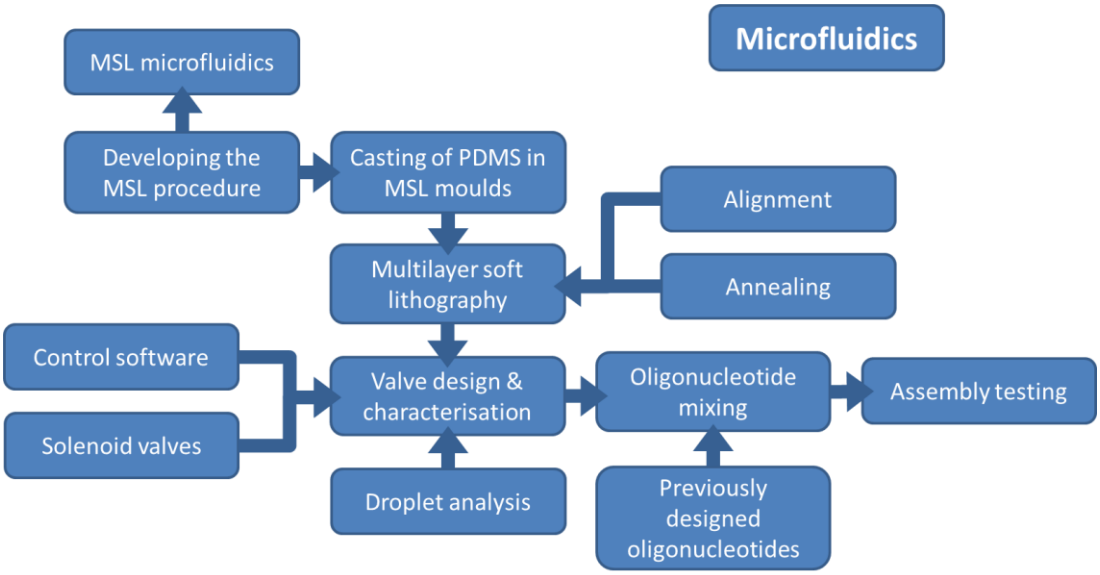


Figure 6.1: Flow chart describing the processes necessary for the fabrication and operation of the microfluidic device. Development of the MSL procedure and its use in the fabrication of moulds for PDMS casts is described in chapter 5. Assembly testing is performed in chapter 8 and 9.

In the previous chapter the manufacture of monolithic and multi-component flow cells was described. The flow cell in section 5.3 had a channel with a cross sectional area of 0.25

mm^2, which was completely enclosed within the part. The flow cell in section 5.4 used the minimum layer thickness of the MSL process to make a chamber 25 micrometres tall, although this chamber was not fully enclosed in the part. All the flow cells described in chapter 5 are static. Once assembled, they are not capable of changing shape. All fluidic control and pumping must be done outside the flow cell.

Here, the design and manufacture of a PDMS microfluidic device is described that is fabricated from multiple layers of PDMS each cast from moulds made by the MSL method. The final PDMS device contains actuatable surfaces and valves that are used to control the flow of several reactants to make complex reaction mixtures. The PDMS microfluidic chip was used to make oligonucleotide mixtures necessary to synthesise two related mutant variant sequences that encode *cis*-regulatory modules (CRMs).

Photolithography and hot embossing were considered as an alternative fabrication methods for the PDMS microfluidic device. The photolithographic process, described in section 3.4, is widely used to pattern moulds for PDMS microfluidic devices. Although photolithography offers higher resolution than can be obtained through ALM, ALM was chosen because of the rapid fabrication times and the prior experience of the investigator (see chapter 5). Hot embossing, described in section 3.7, is widely used to make microfluidic device parts. Although PDMS is not suitable for hot embossing, other easily available materials are, such as PMMA. The microfluidic device required a flexible membrane to be incorporated into the device between multiple aligned layers. PDMS mortar layers offer an obvious solution to this and cannot be incorporated into PMMA devices without clamping of the device. As PDMS is not suitable for hot embossing, this method was not used.

6.1. Droplet microfluidics

Droplet microfluidics is a type of microfluidics in which an aqueous phase (the droplets) are suspended in an oil phase (the carrier flow). The advantages and disadvantages of droplet microfluidics are discussed in chapter 2. Analytes of interest are dissolved in the aqueous phase droplets, the movement of which through the chip can be controlled. Because the droplets do not touch the sides of the PDMS, they are separated by a thin layer of the carrier fluid, thus there is very little risk of contamination.

Alternative microfluidic methods considered include microfluidic spotting using microarray spotting robots and continuous flow devices. Microfluidic spotting allows the placement of sub-µL volumes in specific locations of a glass slide. This method could be used to make place oligonucleotides in specific wells of a multiwall plate to which the DNA assembly solution could be added. This method was not used as spotting of sufficient volumes (~1 µL) without droplets into wells without contamination of the dropping tip was difficult using the available equipment. Continuous flow devices are a viable alternative to droplet systems provided flushing is performed to prevent contamination. The incorporation of a flushing system and the use of droplet microfluidics are equivalently complex in terms of additional liquid handling and on/off chip valving. Droplet microfluidic systems provide more efficient reagent use as the system scales in size as droplets can traverse whole device lengths in the flow of a carrier fluid, which uses less reagents than filling the entire channel, as would be required by continuous flow system. As a result of these considerations, droplet microfluidics was selected as the operating technique for the PDMS microfluidic device.

Droplet microfluidics was employed in an experiment to mix appropriate oligonucleotide sequences for assembly reactions to be performed off chip. This experiment is a proof of principle that the oligonucleotide mixtures can be mixed in

equimolar quantities, without contamination, to produce mixtures capable of assembling into distinct sequences.

A two layer chip was assembled that was capable of performing the mixture experiment. The droplets are produced when a valve from a perpendicularly flowing side channel is opened into a continuously flowing mainline channel. Droplet merging is achieved by moderating valve timing such that each newly created droplet is 'injected' into the main droplet as it flows through the channel. By using valves that sit as close as possible to the main channel as possible, the entire of the injected bolus is carried into the main channel, thus avoiding the possibility of contamination of later droplets by aqueous fluid remaining in the side channel. Figure 6.2 describes this process of merging droplets in flows.

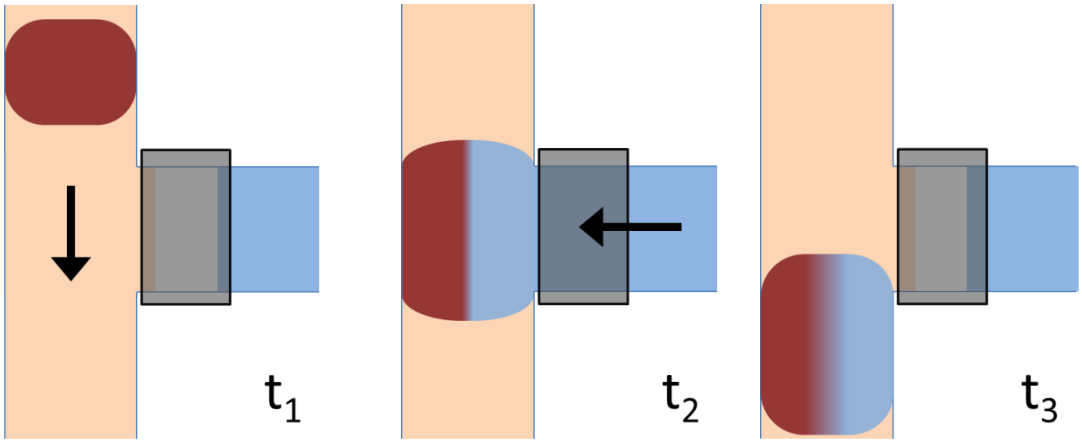


Figure 6.2: Diagram of valve controlled microfluidic droplet merging. Carrier flow (pink) is flowing vertically from top to bottom as indicated in the arrow in the first frame. The valve is indicated by the transparent grey rectangle. At t_1, An aqueous droplet in red is flowing within the carrier flow. At t_2, as the red droplet passes the valve, the valve is opened allowing another aqueous solution to be 'injected' into the red droplet. At t_3, has moved past and the valve is closed.

6.2. Making of PDMS channels using MSL moulds

Simple microfluidic flow cells can be fabricated through the process of multilayer soft lithography [1]. The process is described in figure 6.3. The microchip was comprised of two layers of PDMS bonded together using a PDMS mortar layer. Open channels in the top

surface of one or both of the layers will be closed once bonded to a complementary layer. Further layers can be added to make more complex microfluidic chips.

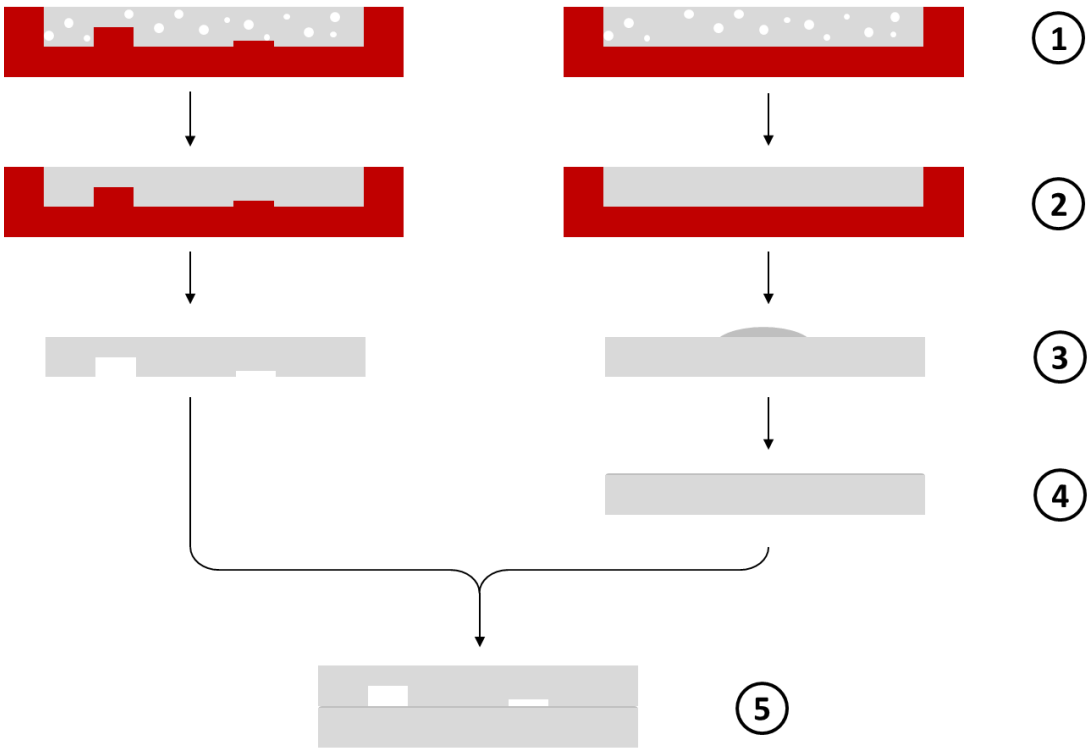


Figure 6.3: Process flow of making a simple PDMS microchip by multilayer soft lithography using MSL moulds. 1) Pour uncured, liquid PDMS into two complementary moulds and degas thoroughly. 2) Cure at 60°C for 1 hour. 3) Remove casts from moulds. 4) Spin coat one side with additional uncured, liquid PDMS and partially cure. 5) align and bond two parts of the PDMS microchip. Uncured PDMS of the mortar layer is shown in a darker shade of grey.

Figure 6.4 shows a top down microscope image of two PDMS layers bonded together by a PDMS mortar layer. The overall chip was fabricated in the manner described in figure 6.3. Producing channels and actuatable membranes within microfluidic chips required the development of methods for the mould, casting, assembly and membrane production as well as testing procedures to determine flow rate through the resulting chips. This chapter describes the characterisation of the resulting chip assemblies at each stage.

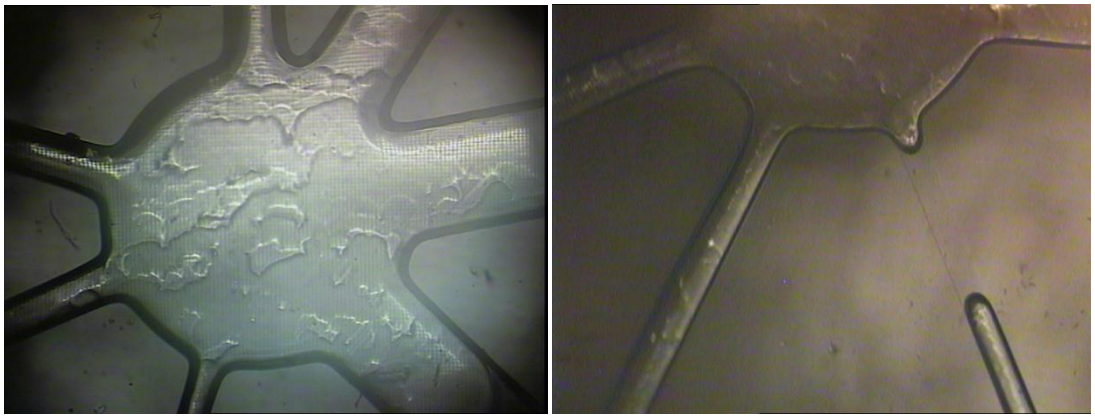


Figure 6.4: Microscope image of PDMS channels test part. Several channels of varying width are arranged radially around a central chamber. The chamber and channels are a single MSL layer thick.

The hatched pattern of the MSL mould surface is reproduced in the PDMS cast in figure 6.4. The regions where the two layers are bonded by the mortar can clearly be seen by the absence of the hatching pattern that is present in unbounded areas. This hatching pattern can also be clearly seen in figure 6.7 (left). The thick, dark line that traces the edges of the channels indicates a curved vertical wall, rather than the straight vertical wall that was designed. Imperfections reminiscent of patched of peeling paint can be seen in the surface of the chamber. These are most likely the result of incomplete cleaning of the mould prior to casting. An SEM image of the cross section through these channels can be seen in figure 6.5.

The PDMS mortar is partially cured before the two parts are aligned and bonded. The partial curing renders the PDMS more viscous but still sufficiently sticky for strong bonding. An insufficiently cured PDMS mortar layer results in inflow of the mortar into the channel features as seen in figure 6.3, right pane. The outline of the original channel can be seen, but the channel is completely blocked. It is not possible to clear a blocked channel without breaking the bonding elsewhere in the part.

Figure 6.4 show two simple, independently cast PDMS microfluidic chips. A significant amount of dust and fibres, black marks and lines in the sealed regions, can be

seen in both chips. The inclusion of fibres, dirt and dust are a result of the fact that the microchips were assembled in a standard lab environment rather than a clean room environment. It was found that fibres, dirt and dust buried in the main body of a layer had no effect on the operation of the device, except when they occurred in portions of the devices that optical measurements where made. Fibres that are on the inner surface of the channels, however, can affect droplets as they traverse the channel. Fibres were often found to be hydrophilic and therefore likely to retain some droplet material as it passes, ready to contaminate the following droplets. Contamination of critical chip areas was not a common occurrence, affecting approximately 1 in 10 finished devices.

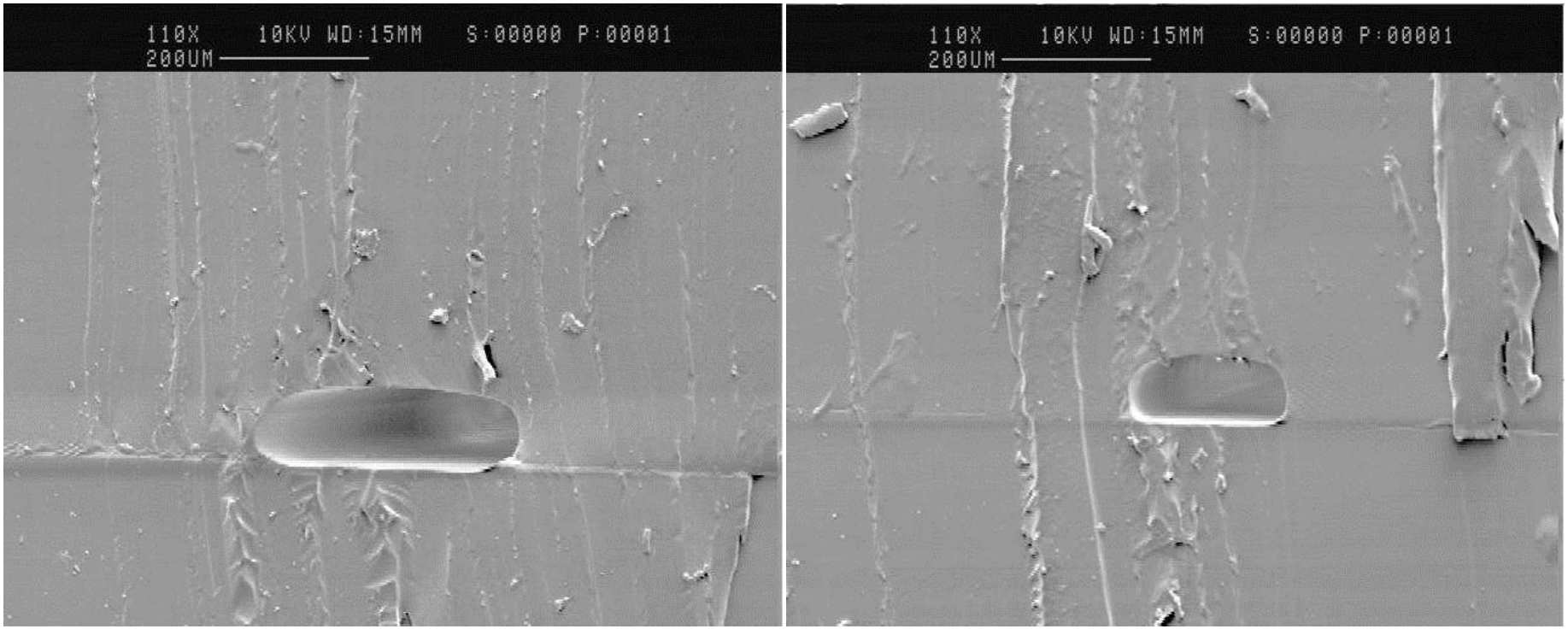


Figure 6.5: Scanning electron microscope (SEM) images of cross sections through channels made by multilayer soft lithography. The cross sections through 300 x 50 micrometre (left) and 200 x 50 micrometre (right) channels are shown. The bonding surface can be seen running across each frame at the bottom of each channel.

6.3. Membranes from the mortar layer

Membranes were made in the PDMS microchips in a two-step process, as described in figure 6.6. A support layer, in this case a silicon wafer, was used to form a thin film of PDMS which is then partially cured. A previously cured layer of PDMS, with channels as appropriate, was then aligned to the partially cured thin film and bonded by curing. Once cured, the PDMS film forms a strong bond to the PDMS part. The film was cut and the part peeled from the silicon support to yield a membrane attached to the PDMS layer. The film now forms a membrane covering any channels in the first layer.

To enclose the membrane, a second PDMS layer must be bonded to the first. A new film of PDMS is spun onto the membrane of the first layer, which acts as a mortar to bond a second pre-cured layer of PDMS to the first.

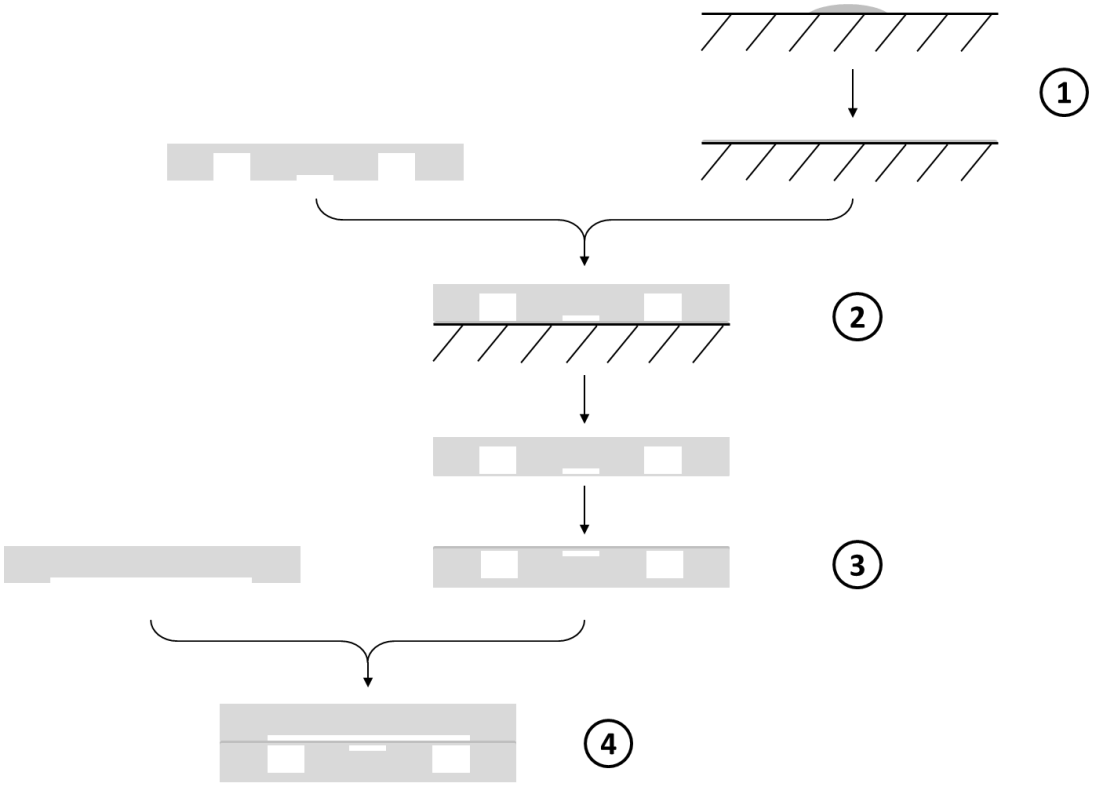


Figure 6.6: Process flow for making PDMS microchip with actuatable PDMS membrane component by multilayer soft lithography. 1) Spin coat silicon wafer disc with uncured, liquid PDMS. 2) PDMS layer is partially cured and pre-cured PDMS layer is aligned. Cure fully and peel off silicon support. 3) Spin coat fresh, uncured PDMS onto the cured PDMS membrane. 4) Partially cure, align and bond to a new pre-cured layer.

6.4. Valve design

Figure 6.6 shows a schematic diagram with a single simple channel in the upper layer and three overlapping channels in the lower layer. Pressurising the middle chamber of the lower layer will seal the channel in the upper layer. In this case and in the rest of the chapter, the upper layer, that contains the channels through which the fluid flows, is termed the fluidic layer, whereas the lower layer, which contains the chambers which are pressurised to seal the channels of the fluidic layer, is termed the pneumatic layer. To make valves in the assembled chip, chambers are arranged so that they overlap but are divided by a membrane. By pressurising the medium in one chamber, the dividing membrane is pushed into the other chamber (see figure 6.7, right panel). Provided the geometry of the membrane and the receiving channel are correct, the receiving channel is closed to flow. The process is reversible upon the removal of the actuating pressure as the membrane relaxes back to the open configuration (see figure 6.7, left panel).

A step feature was produced in the fluidic layer to aid the sealing of the valves. The step can be seen diagrammatic form in figure 6.7 and also in practice in figure 6.8 (unbonded) and 6.9 (bonded). This step is a valve seat for the membrane to seal against when the valve is pressurised. The fluidic layer channels were specified to be 4 layers, around 100 µm, high, whereas the channel through the valve seat region was only one layer, around 25 µm, high. The valve seat was incorporated to ensure rapid sealing of the fluidic channel with the minimum membrane movement.

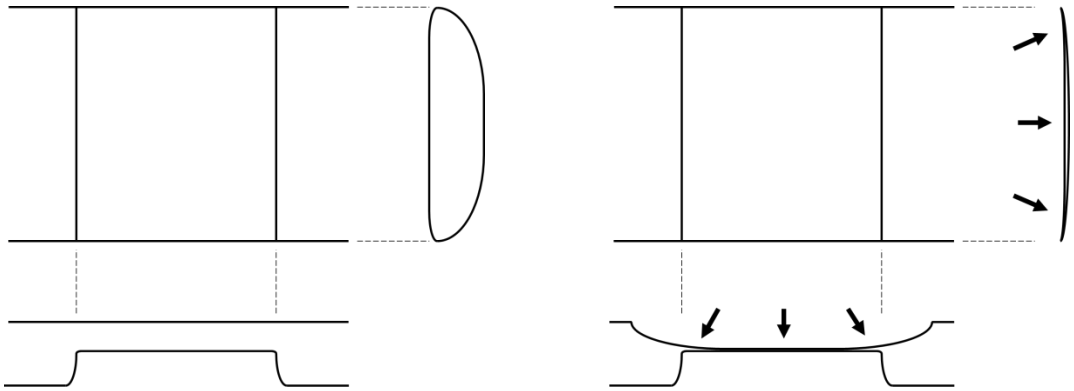


Figure 6.7: Schematic diagram of valve closing of the Quake valves fabricated here. The left panel shows the open valve with a top down view and two orthogonal side views of the channel (SEM images of these two planar views can be seen In figure 6.10. The right panel shows the change in the change in channel cross section as a result of pressure being applied to the chamber above the valve seat. Positive pressure causes the membrane between the fluidic channel and the pneumatic chamber to expand into the fluidic channel, as indicated by the arrows, blocking flow. Diagram not to scale.

6.5. Layer production, alignment and bonding.

Vertical channel heights correspond with those obtained by WYKO measurements of the mould feature sizes presented in chapter 5 (see section 5.1).

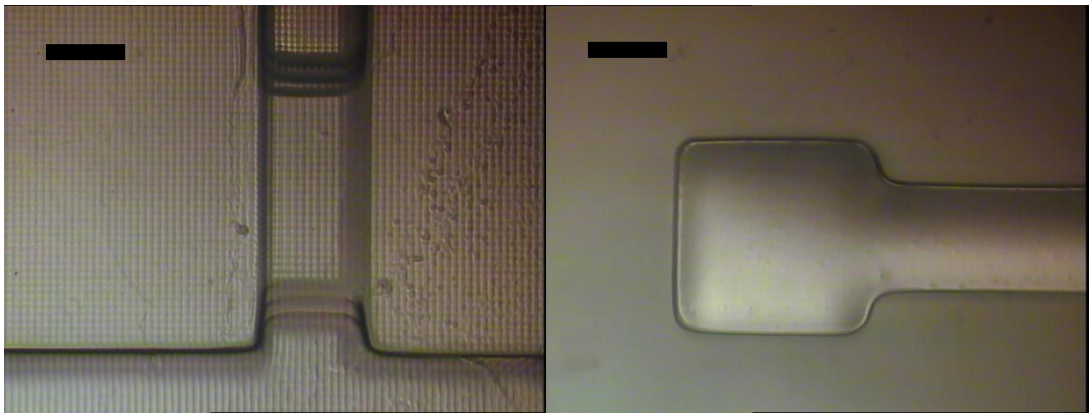


Figure 6.8: Microscope images of PDMS layers prior to sealing against one another. Fluidic channel is seen in the left pane where an inlet channel meets the main channel with a valve seat in centre. Surface of membrane bonded to the pneumatic layer is seen on the right pane. Note the presence of the hatched pattern on the surface of the fluidic layer and the absence of this pattern on the pneumatic layer. Scale bars indicate 200 µm.

The right hand frame of figure 6.8 shows a membrane formed over a chamber in a PDMS layer. The image is focussed on the top of the membrane surface. By focussing the microscope it is possible to see that the membrane is not flat, instead the membrane

appears to sag away from the plane of the top surface of the layer. Spinning of additional liquid PDMS onto the concave membrane surface is likely to result in a thicker spun layer in the region of the membrane. As a result, the valve will not have a uniform cross section, which will affect the sealing of the valve when pressurised.

Alignment of the two layers is carried out by hand using a low magnification dissection microscope. A light source is placed orthogonally to the plane of the lower layer, which is coated with partially cured PDMS. The orthogonal light source highlights the vertical faces within the part and, combined with using a black background surface, maximises the contrast in what is a transparent material. The valve seats and valve membrane were used as marks during alignment.

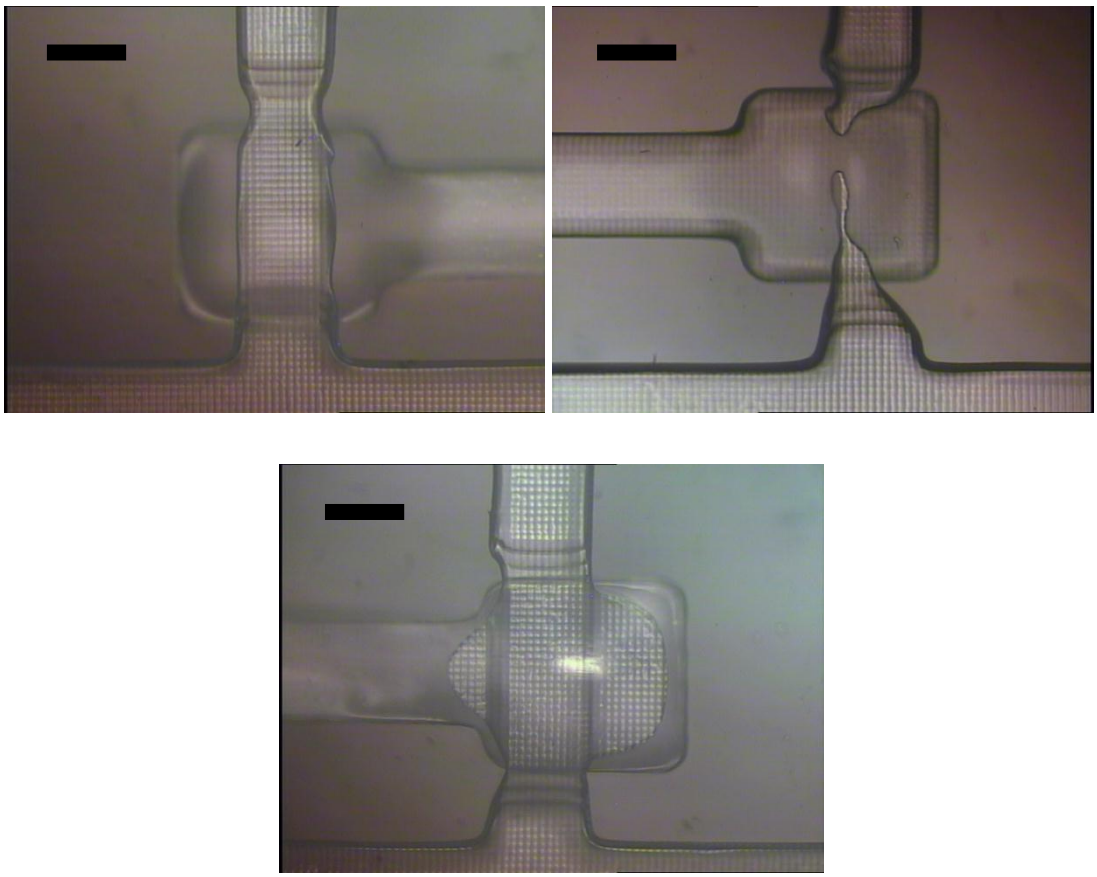


Figure 6.9: Microscope images of two valves made by multilayer soft lithography. The top right pane shows a valve where the PDMS of the mortar layer has ingressed into the fluidic channel permanently sealing it. The bottom pane, by comparison, represents when the membrane layer has not completely sealed to the sides of the channel. Scale bars indicate 200 μm.

Microscope images of valves produced by multilayer soft lithography using PDMS layers cast from MSL moulds can be seen in figure 6.8. The time that the mortar layer is partially cured for is the principle determinant of whether a valve will seal properly (see figure 6.8, top left) or be blocked (see figure 6.9, top right). Through systematic experimentation the optimal partial curing time was found to be 12±1 minutes at 60°C for fresh mixed PDMS. The importance of using fresh mixed PDMS is highlighted by the fact that highly variable results were obtained with PDMS mixed >45 minutes after partial curing. The variation is most likely due to the range of temperatures that the mixed PDMS is exposed to after mixing (the lab temperature is controlled to ±2°C). PDMS tubes were manipulated by hand, heat transfer from which would have accelerated the curing process.

Although an undercured PDMS layer can be easily identified by blocking of channels and valves, overcured PDMS is not as easily identified. Strongly bonded PDMS layers should not be able to be peeled apart easily and the ease with which two layers can be pulled apart is the only identifying trait of overcured PDMS. Devices assembled with overcured PDMS mortar are unable to withstand the pressures necessary to operate the fluidic and pneumatic functions of the device. Perfectly partially cured PDMS can be identified by lightly touching the surface at the periphery of the coated layer with a gloved finger. The partially cured PDMS is tacky enough to stick to the gloved finger and not so liquid as to leave an imprint in the mortar layer.

6.6. Membrane thickness

Generation of a strong, flexible membrane between the two layers of the chip is crucial to the function of the chip overall. Generally, the faster PDMS is spun onto a support surface, the thinner the resulting layer is. The following experiment was performed to determine the relationship between spin speed and layer thickness: PDMS was spun at several speeds (500 – 5000 rpm) onto glass slides, the resulting layers were cured and cut using a scalpel

and the layer thickness determined using the TalySurf as described in section 4.2.2. The graph shown in figure 6.10 provides a summary of the results from this experiment.

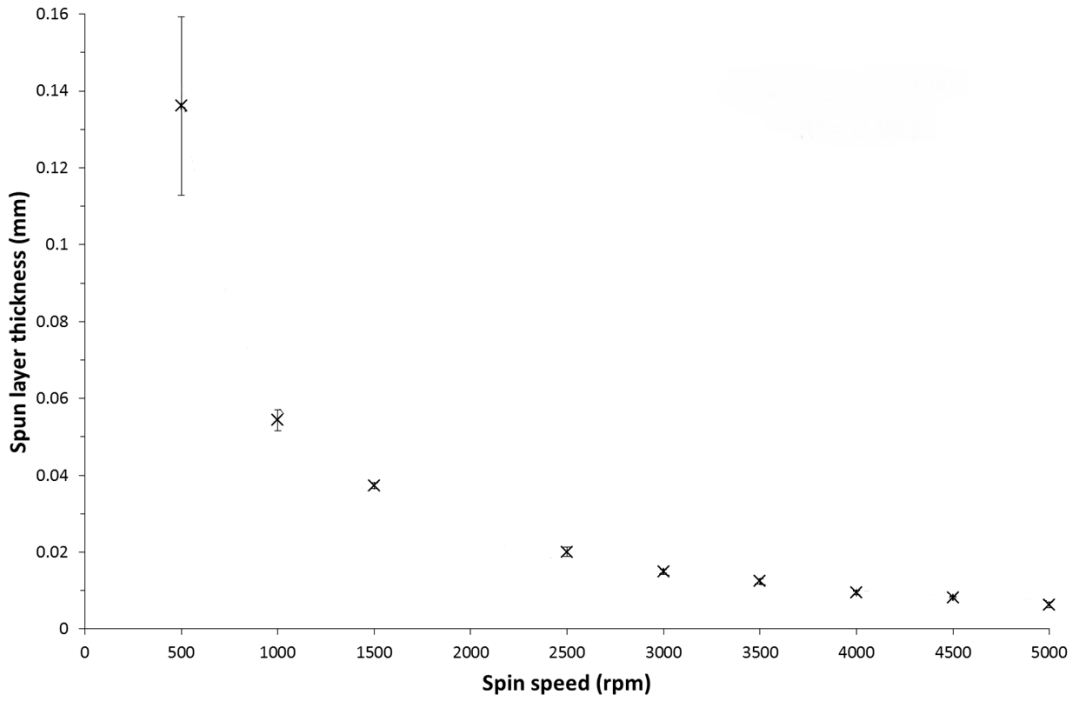


Figure 6.10: Graph of PDMS layer thickness after spinning from 500 to 5000 rpm. Crosses show averages between duplicate spins, each spin is measured in triplicate. Error bars represent one standard deviation of the averages. The black line is a curve, with the equation displayed, fitted to the data.

Interestingly, these data differ significantly from previously published PDMS layer thicknesses [2]. This difference indicates that more variables than just spin speed affect spun layer thickness and illustrates the importance of performing this type of experiment using the specific equipment and conditions available.

The layer thickness measurements were performed on PDMS layers spun onto glass surfaces rather than the silicon and PDMS that are used to produce chips. Several layers were spun onto pre-cured flat PDMS surfaces to determine whether there are significant differences between layer thicknesses on the two surfaces. Because the uncured PDMS layers could not be cut after curing, however, layers were cut before curing. This

delay allowed the spun layers to flow, making the edge indeterminate and the thickness of the resulting layers could not be accurately determined.

As can be seen in the 500 and 1000 RPM data point error bars in figure 6.10, the variation in layer thickness produced by low spin speeds is larger than at higher spin speeds (>1000 RPM). At these speeds, the quantity of PDMS added probably has a significant effect on the layer thickness as the slow speed is insufficient to throw off excess material in the available time. All experiments were performed using a graduated syringe to deposit the liquid PDMS onto the spinning surface. Although care was taken to ensure that 100±10 microlitres of PDMS was applied prior to each spin, this level of variation is clearly sufficient to cause significant variation in the resulting layers at low spin speeds (≤1000 RPM). Whereas at high spin speeds (>1000 RPM), the speed is sufficient to equalise variation applied PDMS volumes with respect to layer thickness.

To directly investigate the layer thickness of the membrane SEM was performed and the results are shown in figure 6.11 and 6.12. By cutting the valves orthogonally to the plane of the membrane allows the membrane thickness to be directly measured. Interestingly, the membrane could also be clearly visualised in other parts of the chip as well (see figure 6.11).

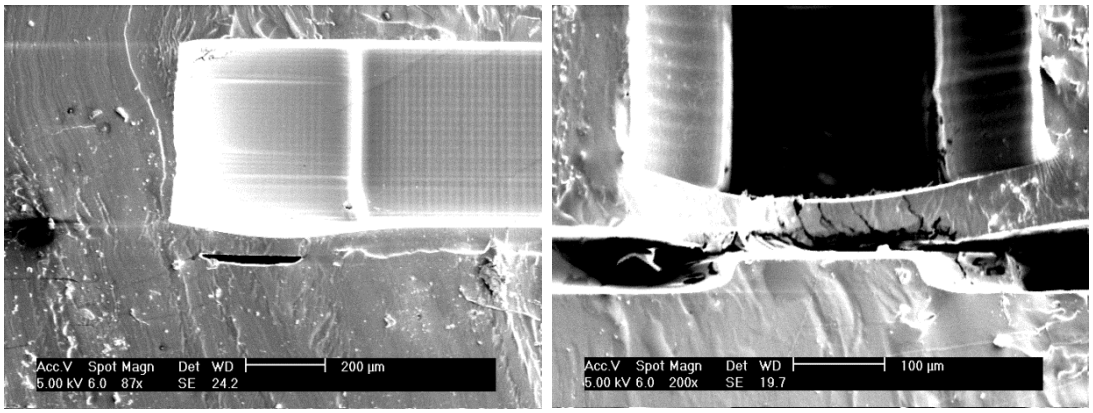


Figure 6.11: SEM images of two orthogonal cross sections through two valves from the same chip assembled by multilayer soft lithography. Scale bars can be seen in the two images. Note the valve seat in the right hand pane, which the membrane is touching.

The membrane of the valve shown in figure 6.11 does not appear to be uniform. During the spinning of the second PDMS layer (mortar layer), the membrane could be seen to sag (experimental observation). It was hypothesised that although the second spinning would result in a flat top surface of PDMS, that the sagged membrane would result in the membrane layer being thicker in the region of the valve than over the rest of the part. This was not considered to be an issue, however, as the top surface of the spun PDMS was expected to be uniform after spinning.

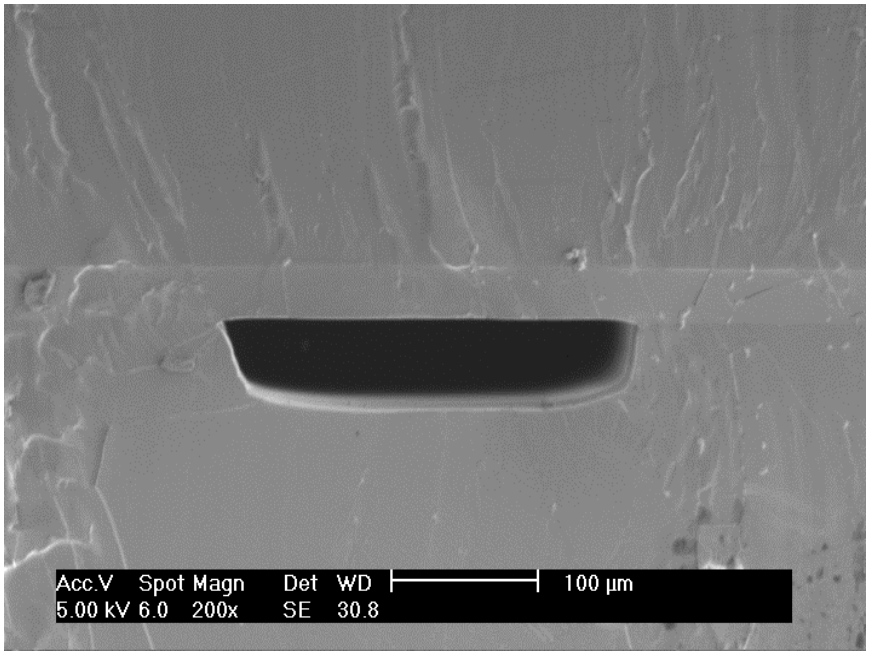


Figure 6.12: SEM image of a channel in a PDMS chip assembled by multilayer soft lithography. The membrane is clearly visible as a band across the top of the channel.

Two independent spins were performed to assemble each chip; the first was performed on silicon at 2500 RPM and the second was performed at the same speed on the reverse of the fully cured PDMS surface of the first spin. Measurements of from these images as well of other valve sections obtained in a similar manner indicate a mean membrane thickness of 50±10 µm. The measured membrane thickness matches the expected thickness of two PDMS layers each spun at 2500 RPM.

6.7. Flow rate through valves at varying pressures

To investigate the relationship between pressure and flow rate through the valves, two experiments were performed. The first varied the pressure applied to the fluidic layer and the second varied the pressure applied to the pneumatic membrane. Figure 6.13 shows the images of valves in the open and closed positions.

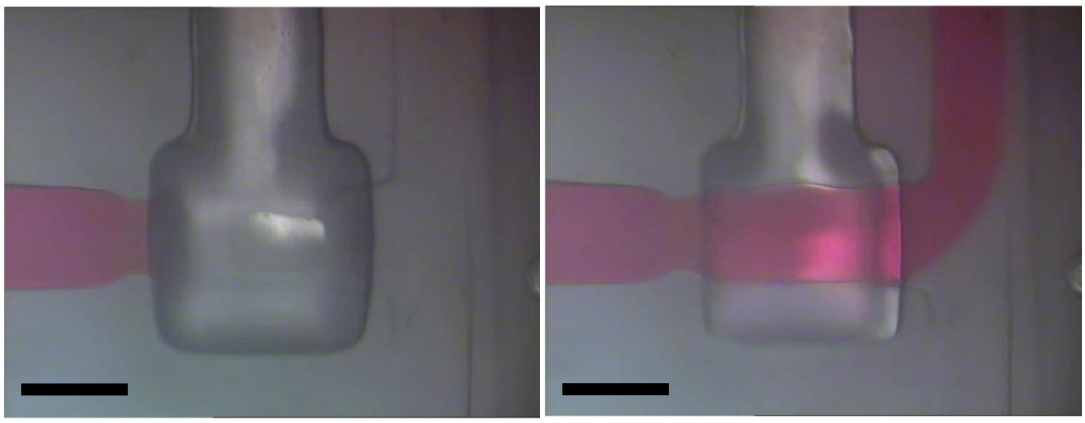


Figure 6.13: Microscope images of a PDMS valve when closed (left) and open (right). Laminar flow of the two fluids (water and water + red food dye) can clearly be seen. Fluid pressure was 34.5 kPa and valve pressure was 137.9 kPa. Black scale bar in both panes represents 250 µm.

In figure 6.13, water carrying red food dye enters from the left, passes through the open valve in the centre of the frame and collides with pure water flowing from the bottom. Both fluids pass out of the outlet towards at the top of the frame. The valve seat cannot be clearly seen in this experiment. When pressurised, figure 6.13, left pane, the chamber on the pneumatic side of the valve expands noticeably. The pressurised valve is ~12% larger than the unpressurised valve which is approximately 500 µm square.

Figure 6.14 shows the through valve flow rate as a function of pressure on the fluidic layer. Regulated air pressure was applied to a reservoir outside of the chip, which was then connected to normal atmospheric pressure via a valve on the microfluidic chip.

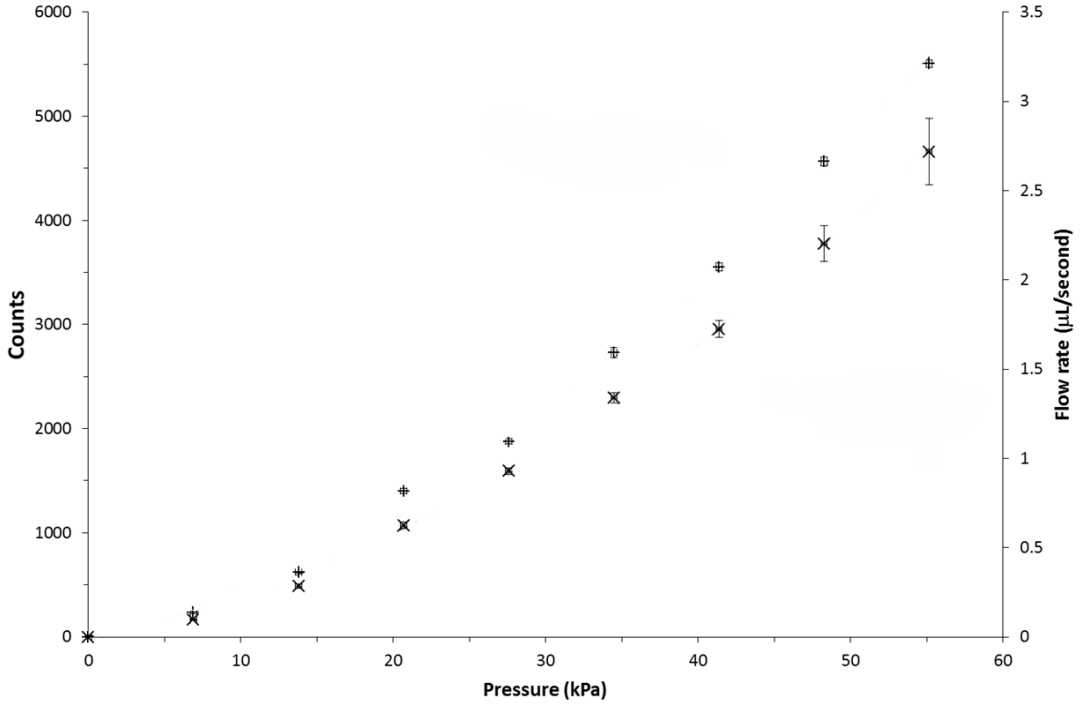


Figure 6.14: Graph of flow rate through a PDMS microvalve over the range of 0 – 55 kPa as measured by two complimentary methods. Measurements using the ruler method are denoted with crosses (x) whilst measurements using a flow meter are denoted with plus signs (+). Vertical error bars represent standard deviations of triplicates whereas horizontal error bars have length 3.45 kPa and are estimated from the pressure gauge.

Interestingly, over the tested range of pressures, the flow rate response is non-linear. This is probably due to expansion of the channels narrowest point in response to the increasing pressure. The larger channel is then able to admit a higher flow rate. Once the PDMS is maximal expanded, the increase in flow rate in response to increasing pressure was expected to become linear. At higher inlet pressures (69-138 kPa) the response from a valve does indeed appear to be linear (data not shown), as expected.

Figure 6.14 indicates how flow rate responds to increasing pressure on the fluidic channel with respect to the outlet. It should be possible, however, to limit the flow rate through the valve by increasing the pressure on the pneumatic side of the channel. Figure 6.15 shows how flow rate through the valve at a fixed pressure of 34.5 kPa changes when various external pressures are applied to the pneumatic channel.

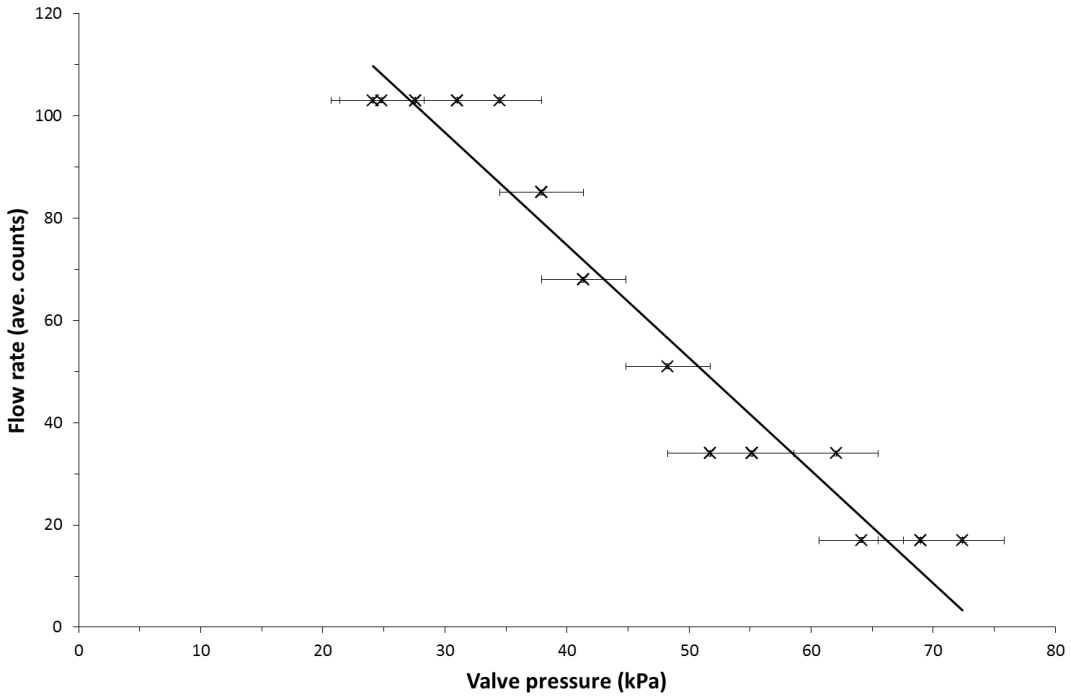


Figure 6.15: Graph of valve pressure from 20.7 to 72.4 kPa against flow rate as measured by flow meter. Crosses represent empirical data and horizontal error bars one each represents uncertainty as to the pressure (±3.44 kPa). Black line is a linear best fit line through the data.

The response curve was expected to be linear through most of the tested range and the data presented in figure 6.15 reflects this expectation. The flow rate through the valves for pressures of <20.7 kPa was constant, indicating that this applied pressure was unable to change the shape of the valve membrane in response to the pressure applied to the fluidic chamber. By using pneumatic chamber pressures of <82.7 kPa the flow rate, with a fluidic channel pressure of 34.4 kPa, can be specified to an order of magnitude from the apparent maximum. As a result of this experiment, the valve pressure was set to be 137.9 kPa in all subsequent experiments so as to ensure that valves were fully sealed when pressurised.

During the valve pressure experiments bubbles were observed to be forming on the valve seats, on the fluid side of the membrane. The source of the bubbles was not, as was expected, either of the two fluidic input lines. Instead the source was the membrane itself: when the membrane is pressurised, air is able to leak through. The thin PDMS

membrane was sufficiently permeable to gasses that air was able to leak through in significant quantities at the pressures used.

Gas permeability of PDMS has been investigated previously and is often regarded as a positive property of the material such as when used for tissue culture[34]. The gas permeation effect has been used to directly control fluid albeit slowly[5–7]. Typically coating with an impermeable layer, such as Parylene C, is employed to reduce gas permeation into channels within PDMS devices. A simple solution, filling the pneumatic lines with water, was found to this issue. By filling the pneumatic line with water, which PDMS is impermeable to, no air is forced through the membrane when the pneumatic line is pressurised. This solution has been implemented previously[8]. PDMS is also permeable to water vapour, meaning that evaporation from channels is an issue for long term experiments.

More generally, bubbles of air could become trapped within the microfluidic device. Bubbles were best prevented by ensuring that the fluidic inlet lines were bubble free. This was achieved by flushing lines with bubbles through the chip until the bubble was removed from the chip. Fluctuating the pressure on the inlet line, by pinching the inlet tubing, was usually sufficient to cause droplets to be moved by the flow of carrier fluid. Occasionally, bubbles would get caught in regions of the channel where flow was low, or alternative paths for carrier fluid flow were available. In these cases, bubbles were brought into the main flow by pressing onto the device and dragging the offending droplet towards the exit channel.

6.8. Microfluidic chip design

To achieve microfluidic mixing of oligonucleotides necessary to assembly members of the CRM mutant library, a microfluidic chip was fabricated that made use of the information described above. The initial design made use of a ‘droplet catcher’ that has been previously

demonstrated[9]. This geometric channel design retarded the flow of a droplet through a channel by providing alternative side paths for the carrier oil to flow around the droplet once the droplet enters the catcher. The subsequent loss of pressure immediately behind the droplet causes the droplet to arrest. Subsequent droplets will then collide with the rear of the arrested droplet, whereupon the side paths are blocked, causing the pressure behind the droplet to increase, forcing the merged droplet from the catcher. Although this catcher was a potentially promising design, it was found that merging droplets in the main channel of the device was more efficient and resulted in the formation of fewer 'satellite droplets'. Satellite droplets are formed when a deformation of a droplet in the flow causes the pinching off of a small portion of the main droplet. Satellite droplets do not flow with the same dynamics as large, channel filling droplets and represent a contamination risk as they may merge with subsequent droplets.

Two moulds for the fluidic and pneumatic layers are shown in figure 6.16 and 6.17. These moulds were built with burn-in settings throughout, 9600 ms exposure time, in order to avoid warping of the part after post-curing. The discovery that incompletely cured R11 moulds inhibit the curing of PDMS casts made in chapter 5, section 5.3.2, informed the decision to flash each mould for 4000-5000 cycles in the UV flasher box.

The channel configuration of the pneumatic layer seen in figure 6.17 (right hand pane) was chosen so as to minimise the space required to fit the valves in. The valve sizes, 0.5 mm^2 was chosen so as to minimise the membrane sagging effect seen in figure 6.8. Small valve seats (1 mm^2) were also chosen so as to minimise the energy required to open the valve.

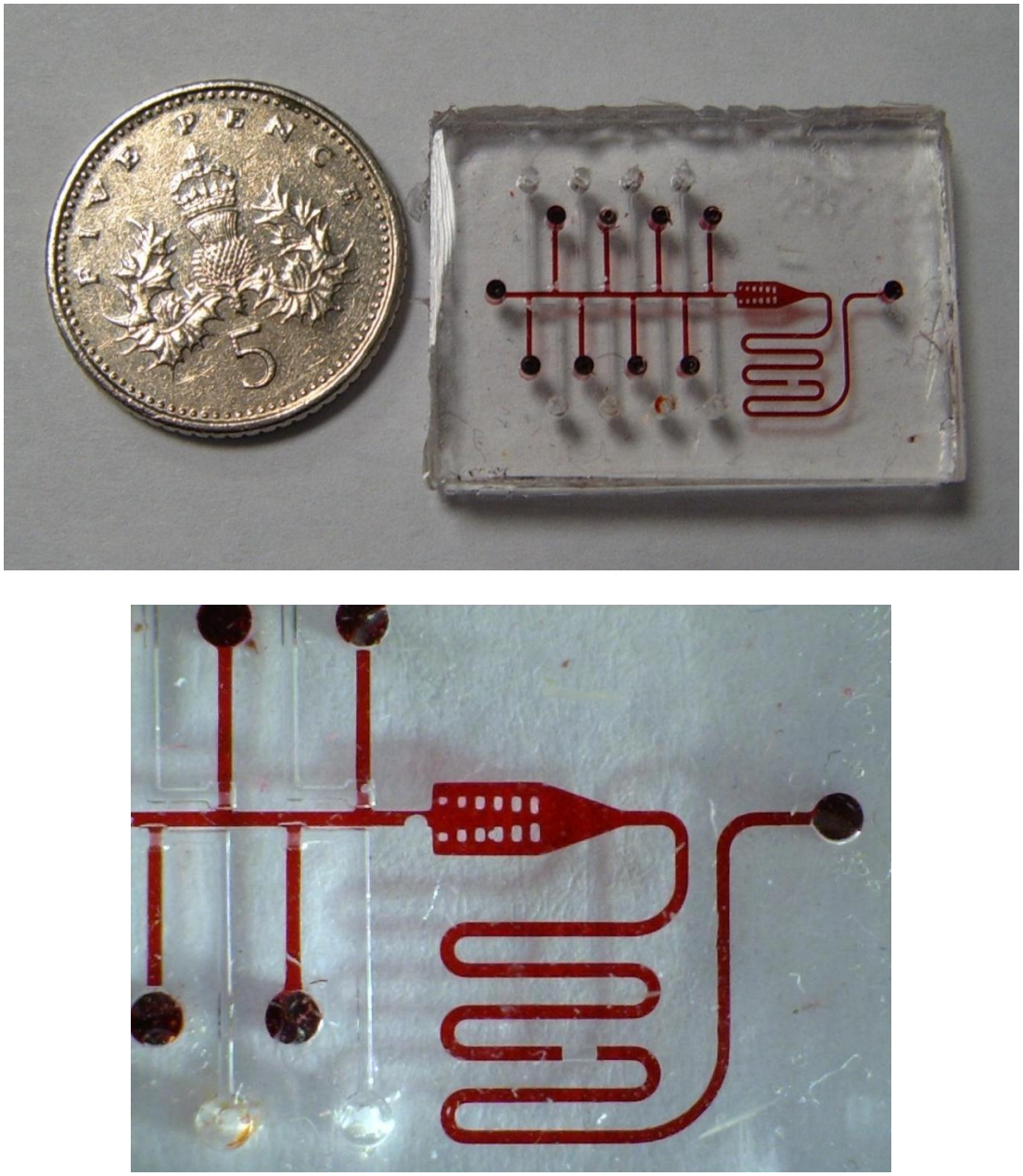


Figure 6.16: Photograph images of a finished whole microfluidic chip (top panel) and detail of the droplet catcher and serpentine (bottom panel). Red food dye is used to fill the fluidic channels, with breaks indicating air bubbles in the system. The pneumatic channels are filled with air and can be distinguished by looking at the detail (bottom panel). A 5 pence piece (Ø 18 mm) can be seen in the top panel for scale.

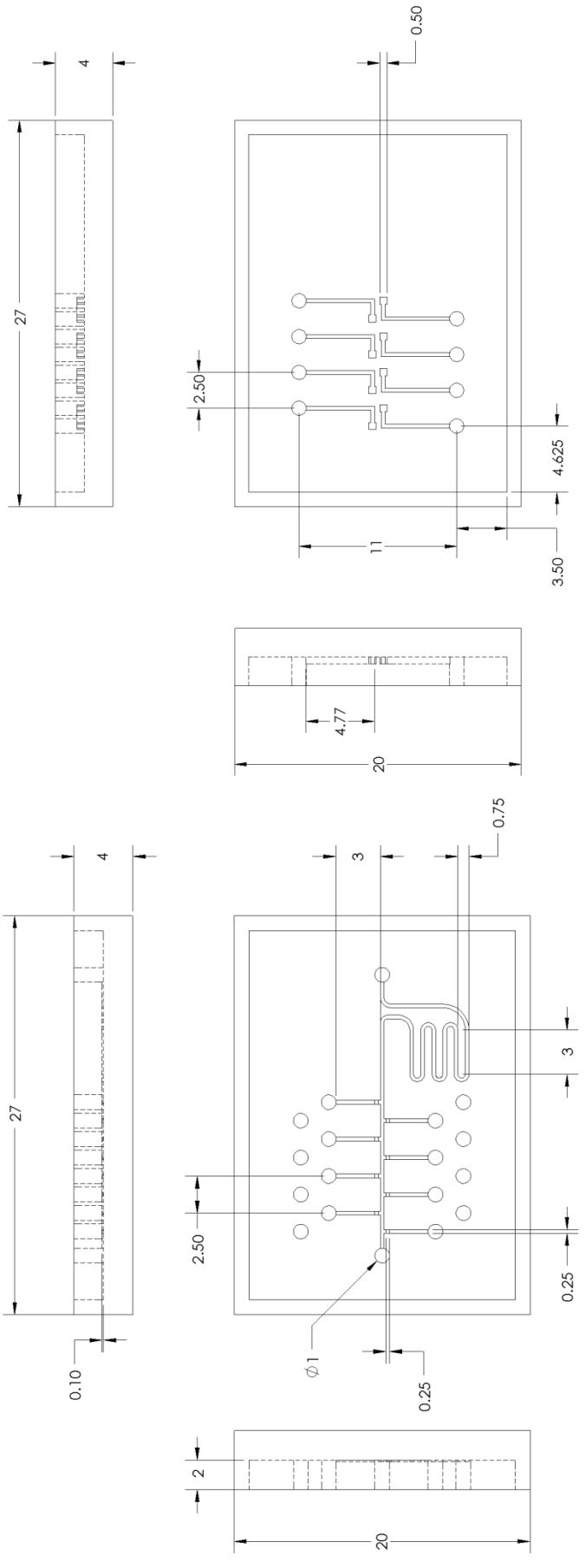


Figure 6.17: Schematic of moulds used in the 8-inlet chip. The left pane shows the mould for the fluidic layer. Note that there are chambers which allow access to the pneumatic layer inlet chambers in this layer. The serpentine is for monitoring droplet mixing after addition. The right hand pane shows the mould for the pneumatic layer. Note the mirror image of the two layers. All measurements are in millimetres.

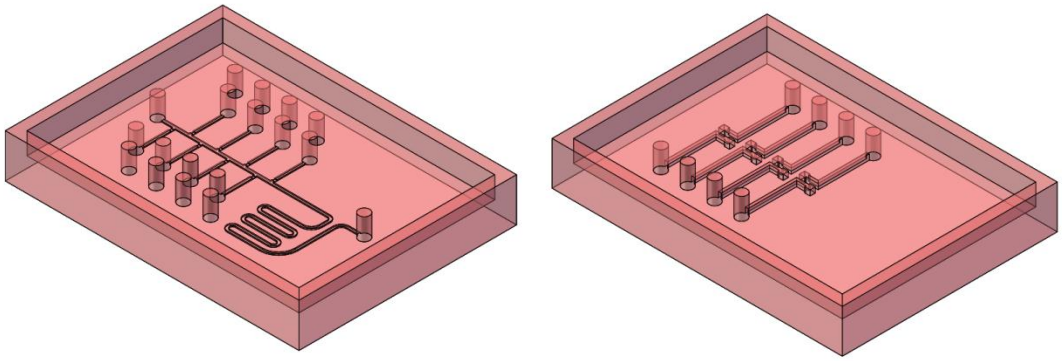


Figure 6.18: Isometric views of 3D CAD models of the moulds for the two layers of the 8-inlet chip. Left pane shows the fluidic layer whilst the right pane shows the pneumatic layer. See related schematic for measurements. Images not to scale.

6.9. Interfacing with the chip

The most common method of interfacing with PDMS microfluidics is by punching a hole through the top surface into the channels below. In these experiments 90° bent needles were used to interface with the chips. Bent needles were used in order to allow access of microscopes and light sources from the top and bottom.

Flat tip needles were used to punch holes through the outside surface of the PDMS layers into chambers in the PDMS layers. A common observation during punching of the holes was that the PDMS would frequently tear, leaving pieces of PDMS inside the chamber able to block the channels. Furthermore, tearing would often cause the inlet port to leak. To seal leaks uncured PDMS daubed around the interface site once needles were inserted. PDMS was either cured for 1-2 hours at 60°C or overnight at room temperature.

Insertion of each needle was time consuming and would often fail: Punching was variable and could yield a hole which would continually leak. Furthermore, the needles were securely held in the PDMS, so accidental movement of the tubing or the needle would easily break the conformal seal. Sealing could not be tested until the daubed PDMS was cured, at which point further PDMS would need to be added. Since PDMS does not stick to

the metal of the needles well, the seals were weak and could be broken with only slight pressure. Overall, issues with chip interfacing and sealing of leaks that would develop was an issue that required solving.

To solve this issue, a part was fabricated by MSL which adapted the bent needles into barbed ends which could be inserted into preformed chambers made during casting. The barbed ends used interference fitting to seal into the inside of the chambers. Bent needles, used to minimise the dead volume between the reservoirs and chip, could be securely glued (with cyanoacrylate superglue) into the MSL adaptor and were then able to resist the torques that would be regularly applied during the manoeuvring of a chip for an experiment. The adaptor could then be pulled out of one chip and fixed into another with ease and a total changeover time is 5-10 minutes without leaks or requiring time consuming post-insertion sealing. Figure 6.18 shows a schematic diagram of the adaptor.

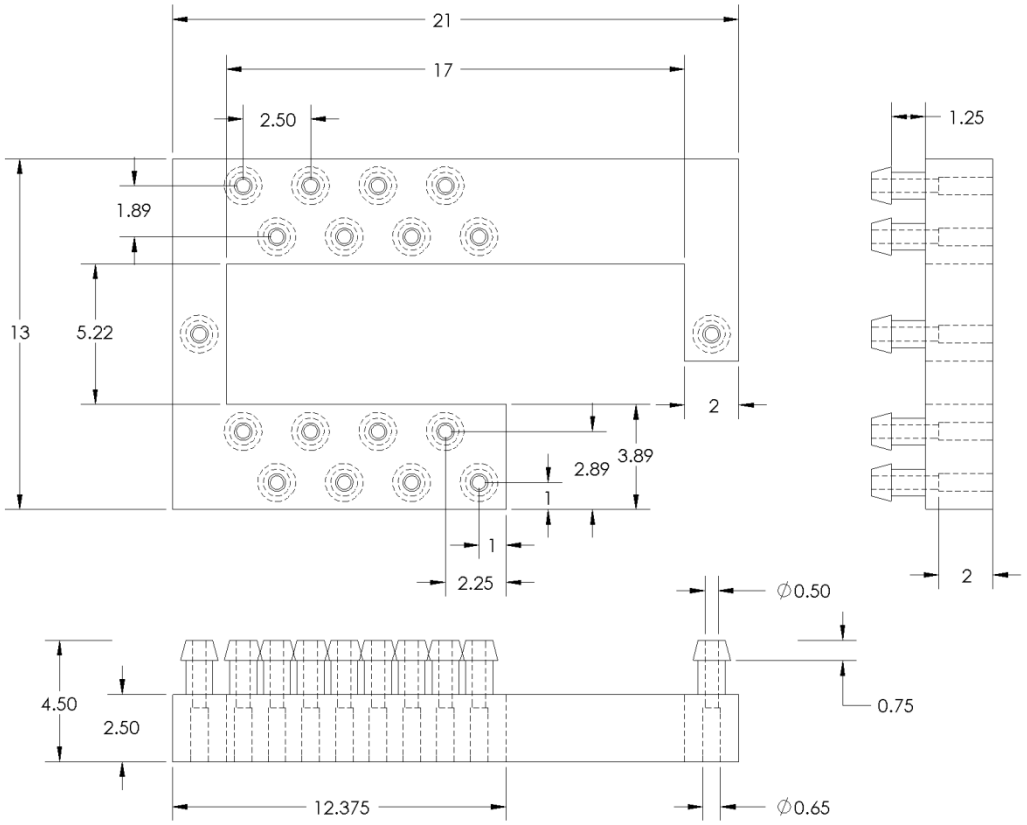


Figure 6.19: Schematic diagram of chip interfacing adaptor. All measurements are in millimetres.

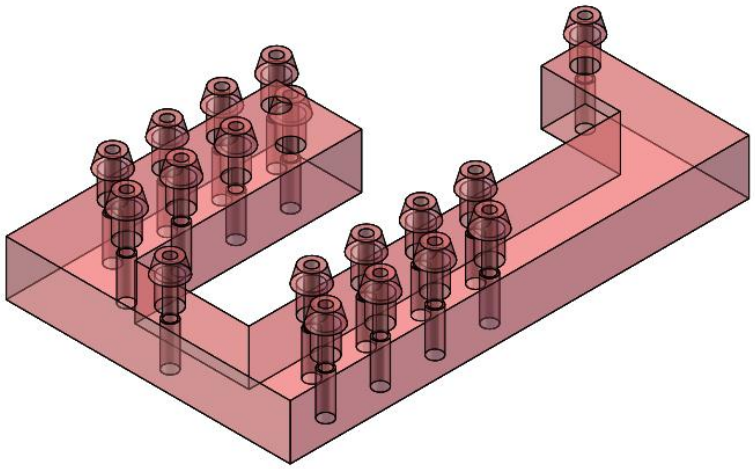


Figure 6.20: Isometric view of 3D CAD model of chip-interfacing adaptor. See related schematic for measurements. Image not to scale.

6.10. Microfluidic setup

Several pneumatic, fluidic and electronic components were necessary for the operation of the microfluidic chip. Figure 6.21 shows a schematic diagram of the tubing that connects the components to the chip. Figure 6.22 shows a picture of the same microfluidic setup. Regulated air pressure is split and applied to each of the fluidic reservoirs. The reservoirs consist of eppendorf tubes with two holes drilled in the lid. 0.5 mm inner diameter (ID) tubing is then pushed through. Whereas the inlet tube is inserted a short distance into the eppendorf, the outlet tube is inserted right to the bottom of the eppendorf.

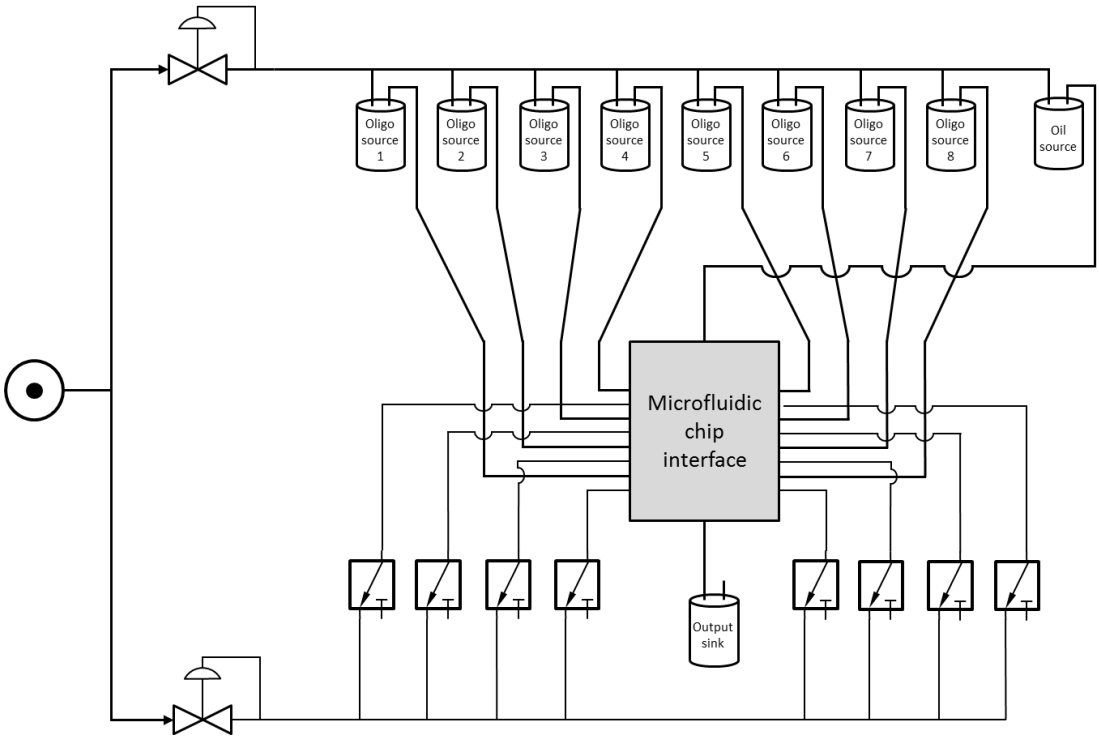


Figure 6.21: Schematic diagram of the arrangement of pneumatic and fluidic tubing connecting oligonucleotide reservoirs to the microfluidic chip. Two regulators control the pressure to the oligonucleotide and oil reservoirs and the solenoid valves, respectively. Solenoid valves control the application of pressure to the on chip pneumatic valves that in turn regulate the flow from each of the oligonucleotide reservoirs through the chip. Finally, the entire of the chip eluent is caught in the output sink.

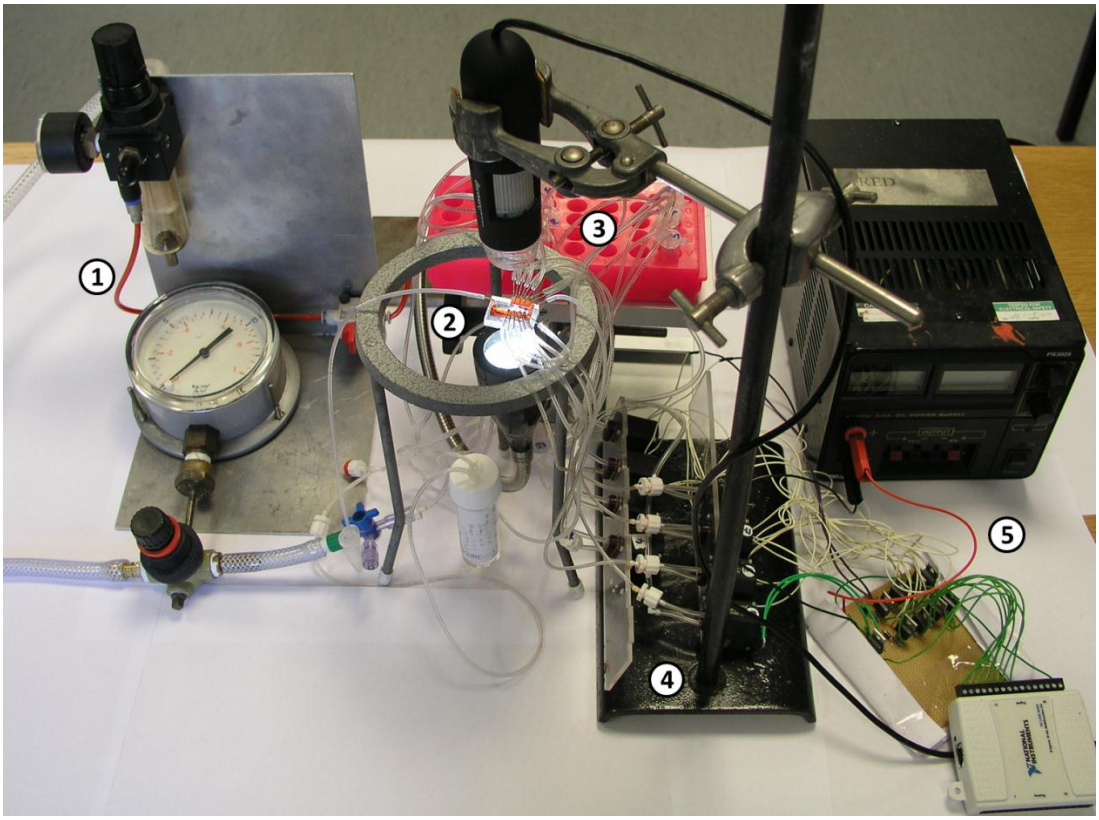


Figure 6.22: Picture of the microfluidic chip setup. Several components can be seen: 1) Pressure regulators. 2) Microfluidic chip, with inserted adaptor, seen between the USB microscope above and LED lamp below. 3) Fluidic sources. 4) Solenoid valves for control of on-chip valves. 5) Electronics for control of valves; consisting of DC power supply, high current circuit and USB-NI-DAQ box. The air pressure source and the computer running LabVIEW (National Instruments version 8.6) are not shown.

6.11. Droplet contamination

Contamination of liquids at droplet forming junctions is a common issue in microfluidic devices. Whilst aqueous droplets do not touch the sides of the device, they will readily form direct contact with other aqueous droplets. Typically surfactants are used to conceal the aqueous nature of droplets from each other, thereby preventing droplet merging. This device requires, however, droplet merging and so the use of surfactants was avoiding. The issue of contamination is demonstrated in figure 6.23.

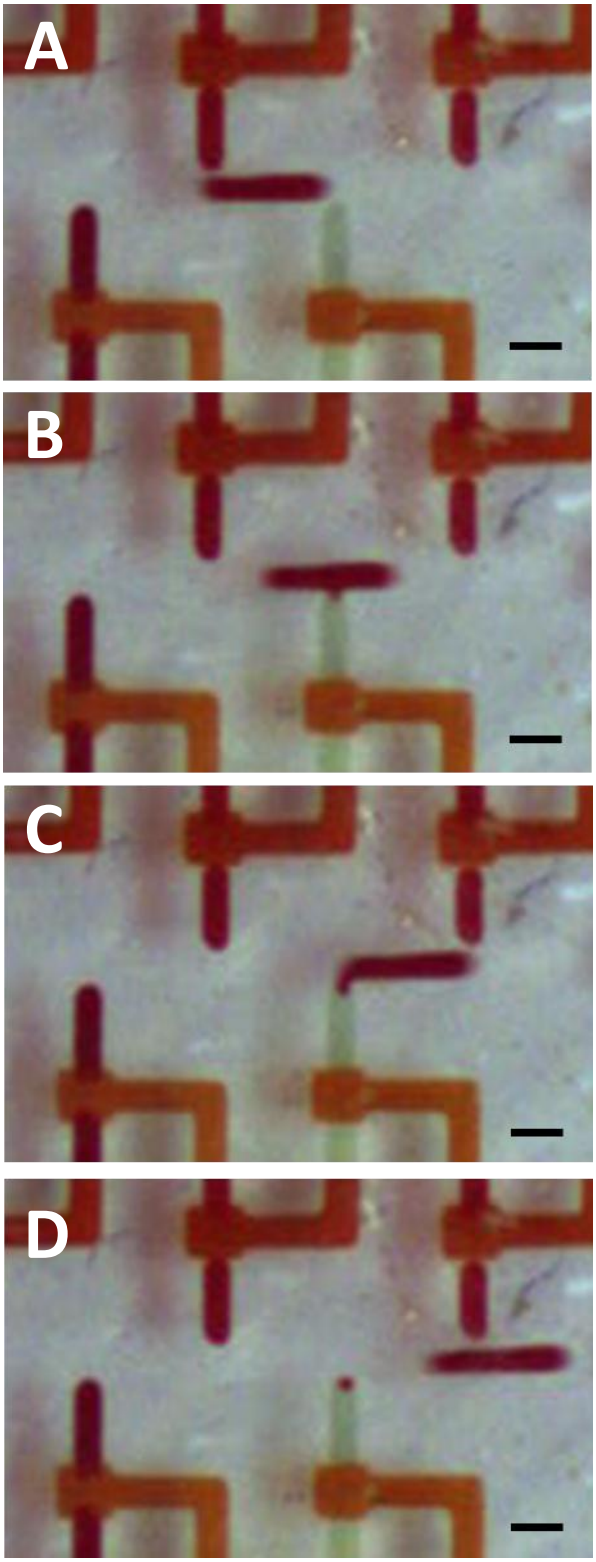


Figure 6.23: Time series of images showing droplet contamination. Each image is one of every four frames of a video. Merging of the droplet containing red dye can be seen with the channel containing green dye in frames B and C. The contamination of the green channel can be seen in frame C. The scale bars are 500 µm in length.

Figure 6.23 shows that droplets can contaminate other aqueous droplets at inlet junctions even when the junction is not actuated. Having valves recessed from the main channel is a common feature of microfluidic devices that include actuateable inlets[5–7]. To prevent this in the microfluidic device used for oligonucleotide mixing, which is highly sensitive to contamination, valves were placed directly beside the main channel. With this arrangement the valve separates a droplet in the main channel from coming into contact with solutions from the side channels.

6.12. Droplet size variability

To measure the variation in droplet sizes an automated droplet counting method was developed. A control program was written that would produce 30 droplets at each specified actuation time for each valve. Another program was written, in MATLAB (see appendixA.1), to read frames off a digital microscope viewing the chip. Drops in the frames were identified and their sizes recorded. In this manner the relationship between actuation time and droplet size could be determined for all the valves on a chip in real time. Figure 6.24 shows the results of one such experiment.

Interestingly, the variation in droplet size was not reduced when using the adaptor, indicating that chip interfacing was not the most significant source of inter-valve flow rate variability as was expected.

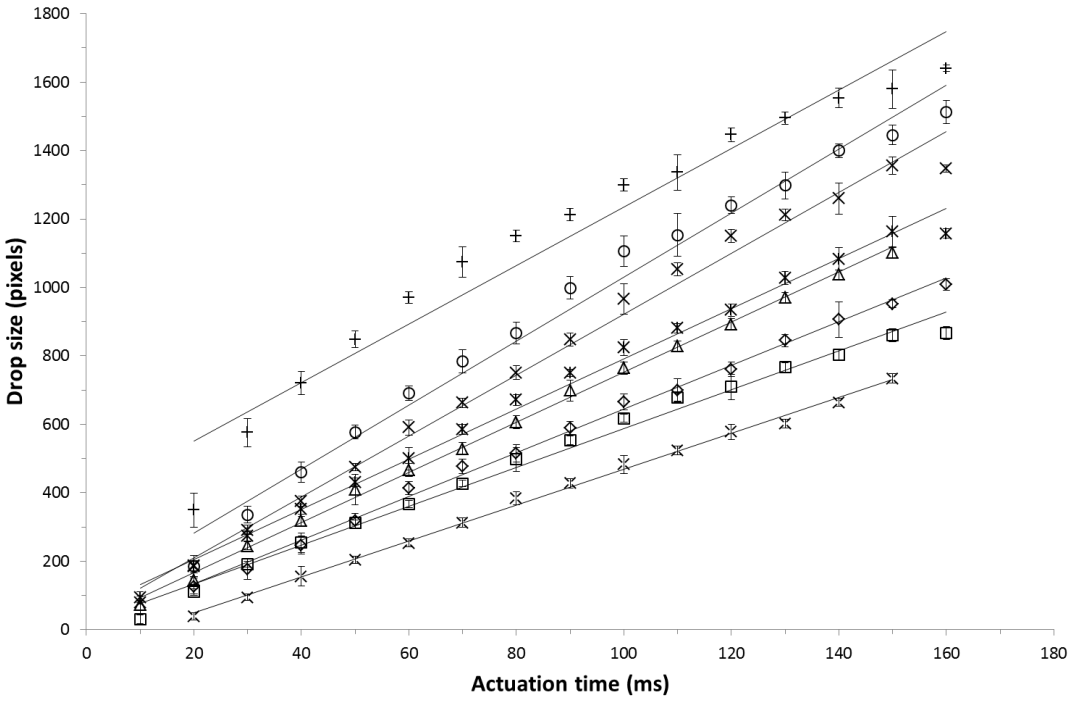


Figure 6.24: Graph of droplet size against actuation time for all 8 valves of a single chip. Each point is represents the average size of up to 30 individual droplets. Vertical error bars represent one standard deviation of the determined droplet sizes. Horizontal error bars represent the variation in valve timing observed (±1 ms).

Figure 6.24 shows how droplet size varies at a specific actuation time for each valve on a single chip. Significant variation is seen between valves. The variation in valve flow rate is probably due to variations between each valve within the chip itself. The alignment of the two layers is probably the greatest source of variation between microfluidic chips. Once layers are aligned, the outer edges of each layer are pressed gently with a pair of forceps to ensure that there is continuous sealing up to the edges of the layers. Placing excessive pressure results in the blocking of channels, particularly at the valves. Placing any pressure, therefore must introduce variation in the thickness of the mortar layer.

The results presented in figure 6.24 indicate that significant differences exist between the valves within a single chip. These differences will almost certainly extend to how the flow rate through each valve, at a fixed fluidic pressure, responds to varying pressure on the pneumatic channel.

Valve opening occurs when the pressure behind the fluid channel overcomes the pressure on the pneumatic side sufficiently to push the valve membrane open. Variations that affect the thickness of the membrane are likely to affect valve opening times and therefore droplet size. The length of the pipe between the on chip Quake valve and the off chip solenoid valve could also have affected the speed at which the valve opens and closes, which in turn will affect the observed flow rates, particularly at fast opening times. The speed of sound in air is approximately 343 m/s. The time for a pressure wave to travel the ~20 cm of tubing between the two valves is approximately 3 ms. Differences in the length of tubing, which were all cut to approximately the same lengths, but not measured, could have contributed to the differences in observed total flow rates. Additionally, the opening rate of the solenoid valves was not examined. It is quite likely that the valves open and close at different speeds or permit different flow rates when open. The available flow meters did not have the millisecond accuracy necessary, however, to test this hypothesis.

The variation in the valve flow rates is controlled for by altering the open time of each valve. By opening each valve for a period inversely proportional to their flow rate, droplets of the same size can be produced. Although this means that each valve does not have the same dynamic range of droplet size, it does allow droplets of the same size to be created.

6.13. ValveControl LabVIEW program design

A program was written in LabVIEW, ValveControl.vi, that could interact with the user through a simple GUI or text file and send instructions to the attached NI-DAQ box. ValveControl has two modes; button and text input. The two modes are essentially exclusive; instructions from a text file will override user button presses when in text input mode. The shortest actuation time required was ~10 ms. It was essential, therefore, that the ValveControl was able to cycle in this length of time or shorter. Furthermore, an

inconsistent cycle length would mean that a valve might remain open too long or not open at all.

For optimum speed, LabVIEW was allowed to determine its own order of action as much as possible. This means that artificial order of action constraints such as sequences or loops were avoided as much as possible. Sequences and loops are frequently used when the programmer must restrict the order of action, as is implicit in a text-based coding environment such as C or MATLAB. LabVIEW, however, is not implicitly restricted in this manner and is, in fact, inherently able to utilise multi-threading to speed up processing times.

6.14. ValveControl GUI

LabVIEW incorporates the writing of a GUI, called a front panel, into the process of writing a program. The valve control GUI is separated into three key sections. The first is the program interface, where the user can specify text input files, reset the timing system, start and stop the simulation. The second in the valve button interface where the user can interact directly with the valves provided a program is not currently running. The valve button interface is useful for initial filling of pneumatic channels, for priming of the fluidic channels and for the removal of bubbles.

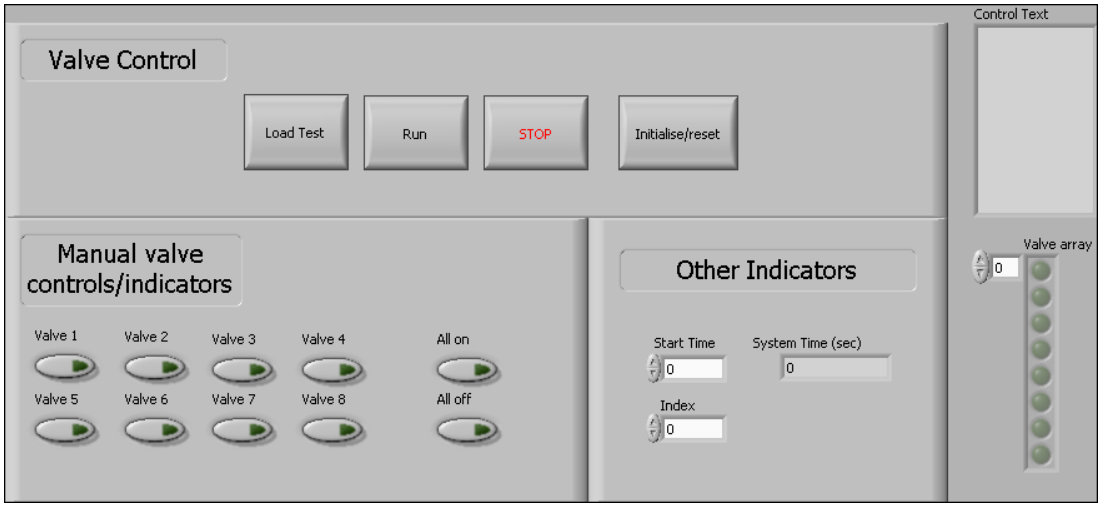


Figure 6.25: Valve Control front panel. This panel allows the direct control of valves when not in 'run' mode via the manual valve control buttons. The current state of the valves is shown in 'valve array'. See text for description of how the GUI functions.

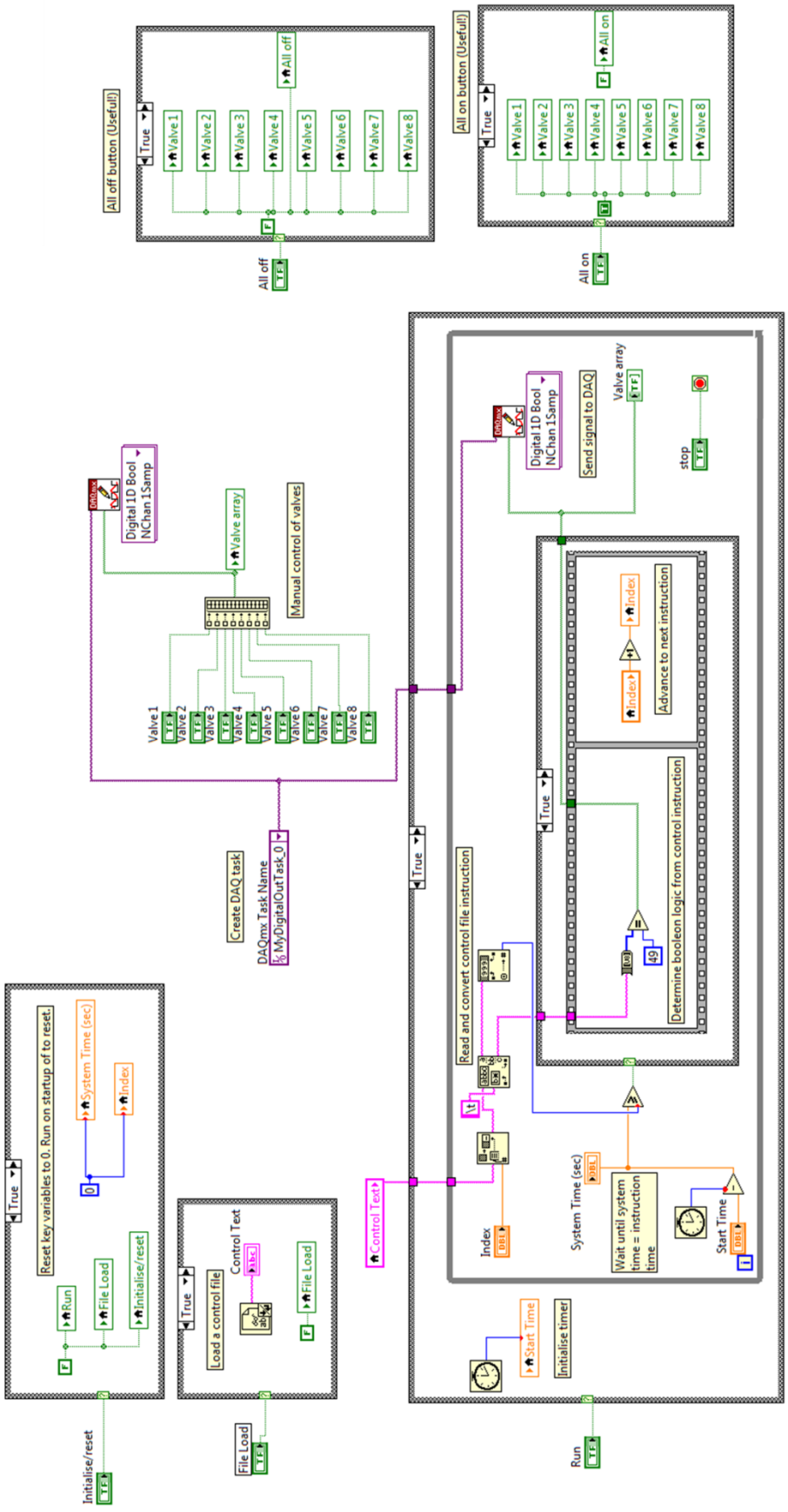


Figure 6.26: Valve Control back panel optimised for speed. Use of logical comparisons of American standard code for information interchange (ascii) characters found in the control text file allows very fast deciphering of instructions which in turn increases cycle time. Flat structure enables LabVIEW to use parallel processing where possible. Valve opening times are within 1 ms of expected using this program. The program is designed to be run in 'continuous' mode.

6.15. ValveControl text input

Text input files followed a simple, tab-delimited table format. Instructions were separated into rows with the time (in milliseconds) of event (relative to pressing 'run') in the first column and the valve state in the second column. When the user wants to specify a text input file, the 'Load file' button is pressed. A dialog asks the user to specify a file, once selected the entire file is read and the end time determined as the largest valve in the first column. ValveControl is then ready to run.

When the user presses the 'run' button, the start time is recorded. The program then repeatedly polls the system clock and compares this with the event time as specified in the user controlled program. When the current time becomes greater than the event time, the event is triggered. The event is encoded in the 8-digit string in the second column of the text input file. Surprisingly, LabVIEW does not implement a function for directly converting of a string of 1's and 0's into a Boolean array. Several methods of reading the string and converting it into a Boolean array suitable for passing to the NI-DAQ assistant were trialled. The fastest method compared each digit to the ascii code for a 0 (48) The results of this comparison are then used to construct a Boolean array which is then passed to NI-DAQ assistant. The program then continues until the current time becomes greater than the end time.

When the program reaches the end of the text input file the valves will remain in the same state as the final instruction indefinitely. Once an inputted instruction set has been finished, however, the valve control buttons become usable again. In order to start the program anew, the user must press 'reset' so that program start time can be reinitialised.

6.16. Additional functionality

Additional functionality was programmed into other versions of the Valve Control program but are not shown here. The user can specify individual valve opening and closing times using a slider. This function allowed modification of valve opening frequency in close to real time; an adjustment is made to a slider for a running valve and a button pressed that applies the change immediately. The sliders were accurate to the nearest millisecond and had a range of 1 – 200 ms.

This function was especially useful when determining delays necessary to achieve droplet merging. The additional computation time required to implement this feature, however, meant that opening times were inconsistent with a variation up to 10 ms on actuation times. Whilst acceptable at lower pressures with longer actuation times, this variation was not acceptable at the higher pressures and actuation times used in the DNA assembly testing experiment.

6.17. DNA assembly testing

A DNA assembly experiment was performed to test the ability of the chip to accurately mix solutions of oligonucleotides. Two sequences were selected for assembly, each sequence required 7 oligonucleotide solutions for assembly to work correctly. Of the 7 oligonucleotide solutions, 6 were common between both of the sequences to be assembled. As a result, all the solutions necessary for the assembly of both sequences could be loaded onto the 8-valve chip simultaneously Because the two sequences are composed nearly identical oligonucleotides incorporation of an incorrect, contaminating oligonucleotide is possible. Contaminated assemblies could then be seen in downstream sequencing. The two sequences are, therefore, highly sensitive to contamination by each other and serve as internal controls.

A total volume of 10 µL of oligonucleotide solution was loaded into each inlet line. The valve actuation times were selected according to the valve flow rate measurement experiment (see figure 6.24) to ensure that the volume of a droplet from each valve was the same. Droplet merging was achieved by timing valve openings so droplets injected directly into previous droplet as the previous droplet travels past the valve.

Merged droplets and carrier fluid were collected from the outlet in 1.5 mL eppendorf tubes, sealed and place on ice at 4°C. To be able to handle the droplets on the bench for the assembly procedure, the output droplets from the procedure were merged in the eppendorf tubes by brief centrifugation. Total volumes of 7 and 4 µL of merged droplet mixtures were obtained from the two assemblies. The obtained volumes were used in two assembly reaction mixtures each totalling 25 µL.

After assembly of the oligonucleotide mixtures using *Taq* ligase, samples of the reactions were taken to be amplified. Amplification was performed exactly as previously described for amplification of normal assembly reactions. Products of the assembly reactions were run on 2% agarose gels to check the successful mixing, ligation and amplification of the two sequences. Figure 6.25 shows the gel of this reaction along with a parallel assembly performed using the normal, bench top method as a control.

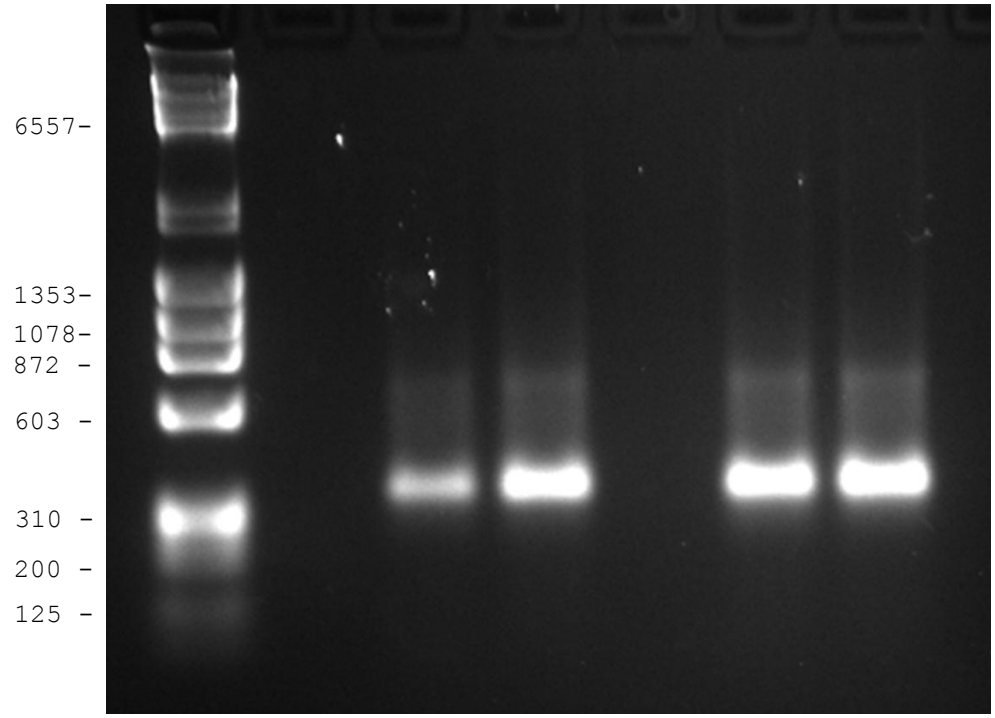


Figure 6.27: Gel of amplification of assembly reaction for two sequences mixed on the microfluidic chip (lanes 3 and 4) or the traditional bench top method (lanes 6 and 7).

As can be seen in figure 6.27 the concentration of the oligonucleotides used in the assembly reaction has an effect on the amount of product produced. Both reactions contained less than the two control reactions, 4 and 7 µL as compared to an equivalent of 9 µL. The fact that the 4 µL band is weaker than the 7 indicates that the concentration of oligonucleotides is limiting. Reactions were performed at between 9 and 16 nM of each oligonucleotide. The chapter on optimisation of CRM assembly conditions (chapter 8) describes the effect of dilution of oligonucleotides in the reactions mixtures. Assembly is successful at 20, but not 2 nM. These results are consistent with these findings.

To measure the mutation rate of sequences assembled through the microfluidic method, the assembled DNA sequences were gel purified and TOPO cloned. The resulting plasmids were then isolated and sent for sequencing. Of a total of 10 plasmids sequenced, representing a total of 295 base pairs of assembled sequence, the number of errors found was 15 the determined error rate was 0.51%. The data used to obtain these values is

shown in table 6.1 This error rate is comparable to the error rate of sequences assembled through the bench top method. Finally, there was no evidence of contamination of either of the two sequences with oligonucleotides from the other sequence. Except for random mutations to sequences were identical to those specified.

Sequence	Insertions	Deletions	Substitutions	Sum(Mutations)	% error
#449 a 000011111	0	1	2	3	1.01%
#449 d 000011111	0	0	1	1	0.34%
#449 c 000011111	0	0	0	0	0%
#449 d 000011111	0	0	1	1	0.34%
#449 e 111100010	0	2	1	3	1.01%
#481 a 111100010	0	2	2	4	1.36%
#481 b 111100010	0	0	0	0	0%
#481 c 111100010	0	0	1	1	0.34%
#481 d 000000100	0	1	1	2	0.68%
#481 e 000000100	0	0	0	0	0%
Total	0	6	9	15	0.51%

Table 6.1: Sequencing data obtained from a set of 10 sequences produced using the PDMS chip mixtures. Sequence numbers are relative to those found in table 8.1. Binary codes represent the sites that are mutated in each sequence. Percentage errors are calculated assuming a total assembled length of 295 base pairs in each sequence. Insertions and deletions are defined and the presence of an unexpected or absence of expected base pair. substitutions are a base pair change. The types of substitutions (transversions or transitions) are not recorded.

6.18. Conclusions

This chapter has demonstrated the feasibility of MSL moulds to make working microfluidic chips with actuatable valves. The time taken to make MSL moulds is shorter than the time required to make moulds through traditional lithographic techniques. In addition, the moulds produced possess features of multiple heights, which would require several steps and might not be feasible by traditional lithographic techniques.

A microfluidic device was designed, made and operated to produce two sequences using half as much oligonucleotide substrate as was previously possible through the use of bench-top methods. The sequences produced were accurate, with an error rate comparable to assembly on the bench-top, and contamination free, as the sequences did not contain any sequences specific to the other. Assessment of the device's performance over time was not assessed as both sequenced were mixed and assembled on the same day.

A proof of concept for the use of a PDMS microfluidic device to make mixtures capable of being assembled correctly has been described. By simply increasing the number of valves at the current density (8 in 10 mm^2) with the 36 valves necessary to make the mixtures to assemble all 512 sequences could be produced using the same method of using MSL moulds described herein.

The final PDMS device is relatively robust; assembly of the fluidic and pneumatic layers with the PDMS mortar provides a strong bond without additional clamping. The weakest part of many microfluidic devices is the inlet/outlet seal, with leaks being caused when the inlet tubing is twisted relative to the device. This issue was reduced by the design and use of the inlet adapter described in section 6.9. The use of the adaptor meant that the device could be handled more roughly and without clamping the inlet tubes relative to the device.

6.19. References

1. Unger, M.A., Chou, H.-P., Thorsen, T., Scherer, A. & Quake, S.R. Monolithic Microfabricated Valves and Pumps by Multilayer Soft Lithography. *Science* **288**, 113-116 (2000).
2. Koschwanez, J.H., Carlson, R.H. & Meldrum, D.R. Thin PDMS Films Using Long Spin Times or Tert-Butyl Alcohol as a Solvent. *PLoS ONE* **4**, 2-6 (2009).
3. Mehta, G. *et al.* Quantitative measurement and control of oxygen levels in microfluidic poly(dimethylsiloxane) bioreactors during cell culture. *Biomedical microdevices* **9**, 123-34 (2007).
4. Regehr, K.J. *et al.* Biological implications of polydimethylsiloxane-based microfluidic cell culture. *Lab on a chip* **9**, 2132-2139 (2009).
5. Eddings, M.A. & Gale, B.K. A PDMS-based gas permeation pump for on-chip fluid handling in microfluidic devices. *Journal of Micromechanics and Microengineering* **16**, 2396-2402 (2006).
6. Johnson, M., Liddiard, G., Eddings, M. & Gale, B. Bubble inclusion and removal using PDMS membrane-based gas permeation for applications in pumping, valving and mixing in microfluidic devices. *Journal of Micromechanics and Microengineering* **19**, 095011 (2009).
7. Randall, G.C. & Doyle, P.S. Permeation-driven flow in poly(dimethylsiloxane) microfluidic devices. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10813-8 (2005).
8. Galas, J.-C., Bartolo, D. & Studer, V. Active connectors for microfluidic drops on demand. *New Journal of Physics* **11**, 075027 (2009).
9. Niu, X., Gulati, S., Edel, J.B. & deMello, A.J. Pillar-induced droplet merging in microfluidic circuits. *Lab on a Chip* **8**, 1837-1841 (2008).

Chapter 7

7. Development of OptiCut

OptiCut is the name of a computer program for the optimisation of oligonucleotide overlap sequences written in MATLAB. This functionality was found to be missing from currently available commercial and free software, as described in section 7.1.1. Therefore, a new program was required to successfully design oligonucleotides that can be assembled into the required CRM sequences. Each sequence is assembled from a set of cognate oligonucleotides that are selected from a larger set of oligonucleotides. Each sequence in the CRM mutant library can be assembled by mixing and ligating a different set of oligonucleotides.

The function of OptiCut is to take a suitable input file and optimise oligonucleotide sequences according to user-specified parameters. The results of the optimisation are then displayed and the output can be produced in a number of different formats. Once obtained, the oligonucleotides can then be combined as specified to produce any of the member sequences. The algorithm has been tested on each of the CRM sequences associated with the project, although only optimisations for CRM-B are presented herein.

OptiCut is designed as a standalone optimisation program with a graphical user interface (GUI) that is installable on Windows® computers. This chapter describes the program, the implemented modules and discusses their effectiveness and efficiency. Finally, further modules are suggested and conclusions from the work drawn. The OptiCut algorithm is presented in appendix A.3.

7.1. Introduction

7.1.1. Definition of the problem

Cis-regulatory modules (CRMs) are sequences which are capable of regulating gene expression of genes on the same chromosome. In this case, the CRM of interest, CRM-B, is a single, short region of DNA (~300bp) in the genome of *Mus musculus*. Previous work has identified that this region is conserved in mammals and more distantly related animals[1]. Furthermore, the ability of this CRM to synergistically regulate the *myod* promoter, in combination with additional novel CRMs and CRMs previously identified in the literature, was previously demonstrated[2,3].

In addition to the CRM itself, consensus binding sites for transcription factors of interest were found previously (see section 2.6). Microarray, bioinformatics and ChIP analysis of the various transcription factor binding sites present in the CRM allowed the prioritisation of several of the binding sites for further analysis. This information and the process by which binding sites were prioritised is described in chapter2.

To investigate the transcription factor interaction landscape that exists within a CRM, specific binding sites can be replaced with null binding sites. A null, or mutant, binding site is defined as a sequence that shows no predicted affinity for the factor predicted to bind there. For this project, null binding sites were generated randomly according to two simple rules: The sequence should show an almost or completely reduced affinity for the specific factor. The sequence changes never changed the overall length of the DNA. Factor affinity for sequences was predicted using the BiFa tool as previously described (see section 2.7). Further rules were that the mutation site should be contiguous and at least four base pairs long. These last rules were implemented to facilitate downstream separation of sequences assembled by Gao assembly (see chapter 8). By comparing reporter expression levels from constructs containing sequences with different

combinations of binding sites, different modes of cooperative regulation can be distinguished, as discussed in chapter 1 (see section 1.5).

A set of sequences which contain all the possible combinations of these mutant sites constitutes a mutant library. Each putative site has an 'on' and 'off' sequence associated with it, the former being the wild-type and the latter a sequence that has hugely reduced or null potential binding to the factor without significantly affecting the binding potential of adjacent or partially overlapping sites. A combinational mutant library for this region consists of every possible combination of 'on' and 'off' sites. In this case, the CRM sequence contains 9 sites which have been prioritised for mutational analysis. The mutant library, therefore, contains 2^9 or 512 sequences in total.

Generation of this library by progressive SDM would involve a large number of reactions (>512) and could not be run in parallel, as the products from one reaction would be used in the next. Furthermore, each step requires reamplification of the previous product, which introduces more random mutations. Sequencing at each step would be necessary to ensure that these mutations do not build up.

Direct synthesis of 512 sequences, approximately 300 bp long for $0.20 per bp would cost in excess of $30,000. Since the sequences are identical except for the mutation sites, there would be a large amount of redundancy expected in such a synthesis. A set of mutually compatible, short, overlapping oligonucleotides could be derived that could be used to assemble the full 512 sequences by DNA assembly. By mixing the right combination of oligonucleotides from this set, any sequence from the library could be assembled.

DNA assembly was chosen as the method by which the library was generated. All the sequences could be generated in parallel without requiring multiple amplification steps progressive SDM would require multiple amplification steps. Chapter 8 compares the two methods of DNA assembly (see section 8.1.2). In this project, the method of ligative

assembly was chosen in preference to the more commonly used PCR-based assembly. Successful ligative DNA assembly requires the optimisation of the sequences of the oligonucleotides that are to be used. This chapter describes the various available methods and their short comings, before describing the software that was developed to perform this optimisation, OptiCut.

7.1.2. Current gene assembly software

Many commercial options exist for researchers requiring synthetic DNA. Several large companies offer gene synthesis services such as Invitrogen GeneArt®[4] and Sigma (GeneOracle)[5]. In addition, several smaller companies offer gene synthesis services: Integrated DNA technologies (IDT)[6], EuroFINS[7], DNA2.0[8], BlueHeron[9], Origene[10], BioMatik[11] and Entelechon[12]. The cost per base for oligonucleotides and synthetic genes have both been decreasing whilst the maximum synthesisable length has been increasing at roughly exponential rates since the mid 90's. Gene synthesis and gene assembly will be discussed primarily in chapter 8.

The many companies that provide synthetic gene services is mirrored by the plethora of optimisation software available for synthetic gene work; Gene Designer[13] , genecomposer[14] , TmPrime[15], Gene2Oligo[16], Assembly PCR oligo maker[17] , DNAWorks[18], Computationally Optimised DNA Assembly (CODA)[19], GeneDesign[20] , OPTIMIZER[21] and gene morphing system (GeMS)[22]. These software principally perform the optimisation of oligonucleotide sequences for PCR or LCR-based assembly reactions. Most will also allow codon optimisation in the case of expressed proteins or reverse translation of a desired protein sequence.

None of the software described above offer the option to optimise multiple sequences in parallel such that the same oligonucleotides can be used in the synthesis of all sequences. Batch sequence processing can be performed by some software, but this is not

for the purpose of parallel, degenerate assembly as described here. By reusing oligonucleotides the whole library can be built using as few as 50 oligonucleotides, from which a set of around 20 are selected to assemble each sequence in the library. In order to maximise the chance of success, the lengths of the overlapping oligonucleotides must be optimised so that they will anneal at a similar temperature suitable for ligating. The following section describes the most widely used method of melting temperature estimation.

7.1.3. Melting temperature estimation

The method of determination of melting temperature is of primary importance to any DNA overlap optimisation program. Generally, the Nearest Neighbour (NN) method using base pair coefficients determined by SantaLucia *et al*.[23] is used for overlaps of <50bp. For overlaps of >50bp the simpler Meinkoth and Wahl algorithm is employed [24]. More recently, optical trap experiments have refined the coefficients used in the NN method still further[25]. In this project the SantaLucia 1998 values were used as these are the most commonly employed values in other software.

Whilst the energy derived from the base pairing of pairs of bases in opposite strands is the primary determinant to oligonucleotide melting temperature. The NN method takes into account the basepair stacking energy of DNA as well as the basepairing energy. In short, a DNA sequence of GGGGGGAAAAAA will have a higher melting point (to its cognate partner) than a sequence of GAGAGAGAGAGA, 41°C and 36°C respectively. Both sequences are the same length and have the same G:A ratio. Based solely on the proportion of G/Cs and A/Ts, therefore, they should have the same melting temperature. However, the extra energy comes from the fact that there is a stacking 'preference' between certain bases adjacent to each other in the chain [26].

2nd base / 1st base	A	C	T	G
A	-7.9 -22.2	-8.4 -22.4	-7.8 -21.0	-7.2 -20.4
C	-8.5 -22.7	-8.0 -19.9	-10.6 -27.2	-7.8 -20.0
T	-8.2 -22.2	-9.8 -24.4	-8.0 -19.9	-8.4 -22.4
G	-7.2 -21.3	-8.2 -22.2	-8.5 -22.7	-7.9 -22.2

Table 7.1: Nearest Neighbour (NN) binding energies for adjacent bases as determined by SantaLucia[23]. The first value in each box is the ΔH (kcal/mol) and the second value is the ΔS (cal/K.mol).

The melting temperature (T_m) of two DNA strands can be determined using equation 7.1,

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln C_T} \qquad \text{Eqn. 7.1}$$

where ΔH° is the change in enthalpy ΔS° is the change in entropy, R is the gas constant (1.987 cal/K.mol) and C_T is the ratio of the concentrations of the two annealing strands. In this case, where strands are not self-complimentary and the two strands are assumed to have equal concentrations, C_T becomes $C_T/4$. The ΔH° and ΔS° are determined from table 7.1 as the sum of the contributions from each pair of nucleotides in the linear sequence.

7.1.4. Competitor identification

Oligonucleotide assembly reactions are sensitive to competition reactions. Each Oligonucleotide has 1 or 2 target oligonucleotides that it is 'meant' to bind with. Each oligonucleotide is, however, capable of binding with any of the others present in solution, including itself. To further complicate the situation, individual oligonucleotides can form hairpin loops within themselves. Heteroduplex binding energy prediction is difficult as the binding energy associated with heterologous bases is not null. Empirical investigation has informed the modification of the binding energy tables used in the NN method[27]. The

predictions developed by this method are not ideal as they are performed in HPLC conditions using acetonitrile as a solvent. Consequently, corrective coefficients are applied to bring the results in line with those observed in aqueous conditions. The method used here was reverse engineered from IDT SciTools[28]. Basically, the method finds regions of homology between two strands and uses the NN method on each region. The final melting temperature is then calculated from the sum of all the homologous regions.

7.2. OptiCut method

A flow chart summarising the optimisation algorithm is shown in figure 7.2. The program optimises the lengths of the overlapping portions between oligonucleotides. The overlaps start and stop at 'cut positions', although first and last cut positions are fixed to the start and the end of each sequence. A set of cut positions and the sequences to be assembled is sufficient to describe the full set of necessary oligonucleotides. The oligonucleotide overlap optimisation program was written in MATLAB with a modular format. The main module receives a FASTA format file containing the combinational mutant library and performs the optimisation procedure. Output consists of an Excel file containing the 5' to 3' base pair sequence of all the necessary oligonucleotides. Alternatively, output can be produced in tab-delineated ASCII format.

Melting temperatures are determined using the SantaLucia coefficients[29] for Nearest Neighbour (NN) method as discussed in the introduction (see section 7.1.3). In brief, the optimisation procedure consists of series of fully defined steps:

1. Determine the temperature of every overlap in the sequence library.
2. Identify the 'hottest' and 'coldest' overlaps.
3. Shift the positions of all the intervening cut sites by one base pair to reduce the length of the 'hot' overlap and increase the length of the 'cold' overlap.
4. Repeat until the end condition is met.

Thus, the 'hot' and 'cold' overlap lengths change whilst the intervening overlap lengths do not change. Although the overlap lengths of the intervening overlaps do not change, their sequences do. Overlaps outside of the 'hot' and 'cold' overlaps are not altered. See figure 7.2 for a flow chart of how the algorithm optimises an oligonucleotide set.

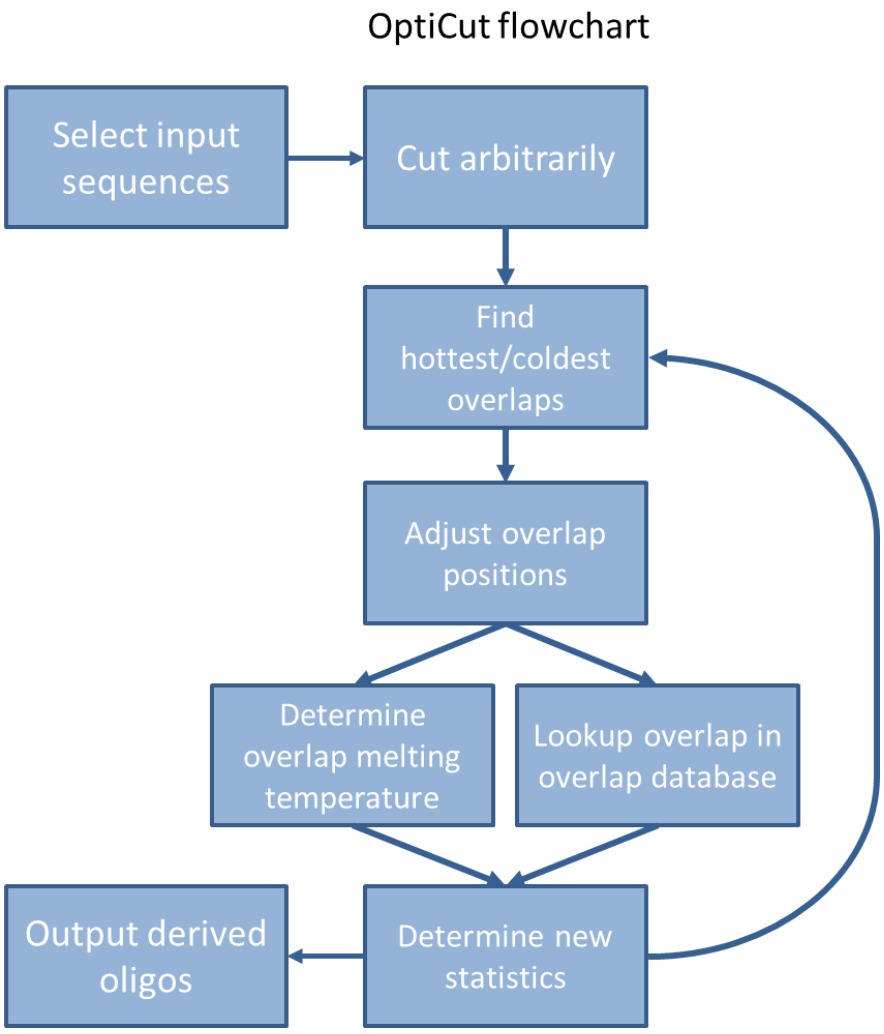


Figure 7.2: Flow chart for the OptiCut program with tranch-shifting employed.

The above procedure is performed on the whole set of library sequences simultaneously. The 'hot' and 'cold' overlaps, therefore represent the average melting temperature of all the sequences within a given sequence length. The cut site position changes are propagated to every sequence in the library. Performance of the algorithm is measured by

the root mean square difference (RMSD) of the melting temperatures of the whole overlap set.

Exhaustive searching is not feasible for anything but very small sets due to the amount of computation required. The following set of equations (eqns 7.2 – 7.4) can be used to determine the approximate time required: The number of overlaps is determined using equation 7.2. The average oligonucleotide length for each overlap is described by equation 7.3. Given that each cut position is allowed to vary from the original by ±2 bp, each cut position can take any of 5 positions. Equation 7.4. describes the total number of position sets that need to be computed.

$$\frac{\#\text{oligonucleotides}}{2} - 1 = \#\text{overlaps} \qquad \text{Eqn. 7.2}$$

$$\frac{\text{Total sequence length}}{\#\text{overlaps}} \times 2 = \text{average overlap length} \qquad \text{Eqn. 7.3}$$

$$\#\text{positions}^{\#\text{overlaps}-1} = \text{Total cut position sets} \qquad \text{Eqn. 7.4}$$

For example: A set of 128 sequences each 100 bp long with 5 oligonucleotides in each strand is to be optimised. A total of 9 overlaps, each with average length 22.2 bp, generate a total of ~400,000 cut position sets. Each cut position set must be applied to each of the 128 sequences in the original set and statistics of the resulting oligonucleotides developed. This process takes a non-trivial amount of time (typically between 0.5 and 1 second). The overall time taken to perform such an exhaustive search is several hours. The exhaustive approach is, therefore, only appropriate for a very limited search range and is incapable of providing a good quality optimisation (as measured by RMSD) of anything but the most homogeneous sequence.

The algorithm will estimate the melting temperature of a sequence and then store the determined melting temperature in a database. If the algorithm then determines that the melting temperature of the same sequence is required it will first check the database to

determine whether the melting temperature has been previously determined. Storing of previous melting temperature results in a database that markedly speeds up the process of optimisation. To test this, a database of 10000 random oligonucleotide sequences of length 22 (representative) was generated (using 'randseq', from the MATLAB bioinformatics toolbox), the melting temperatures assessed and the results stored. This represents the process of reanalysing the sequence every time. Sequences from the database were then searched for 10,000 times within the database to simulate lookup of the sequence instead of reanalysing. The results from this simulation can be seen in figure 7.3 and indicate that lookup was approximately 15x faster than redetermination.

Figure 7.3 indicates that lookup in a database only becomes slower than redetermination when the database is approximately 5,000 members or larger. Therefore, the database lookup method was implemented in preference to a redetermination method. A typical sequence library of 512 sequences, with 10 oligonucleotides per strand, therefore 19 overlaps, would produce a database with maximum size of 9728. Because of the similar nature of the sequences in the library, most of these overlaps would be identical, so the database is likely to be much smaller than this. As the optimisation progresses, and new overlaps are added, it is possible that the library will become > 5000 members in size. At this point a simple switch could be employed where the algorithm determines the melting temperature of new overlaps rather than looking in the database.

Furthermore, experience of the optimisation algorithm indicated that a typical overlap database size is 100-400 members in size. Therefore, no attempt was made to gauge the size of the database and switch between the two possible modes.

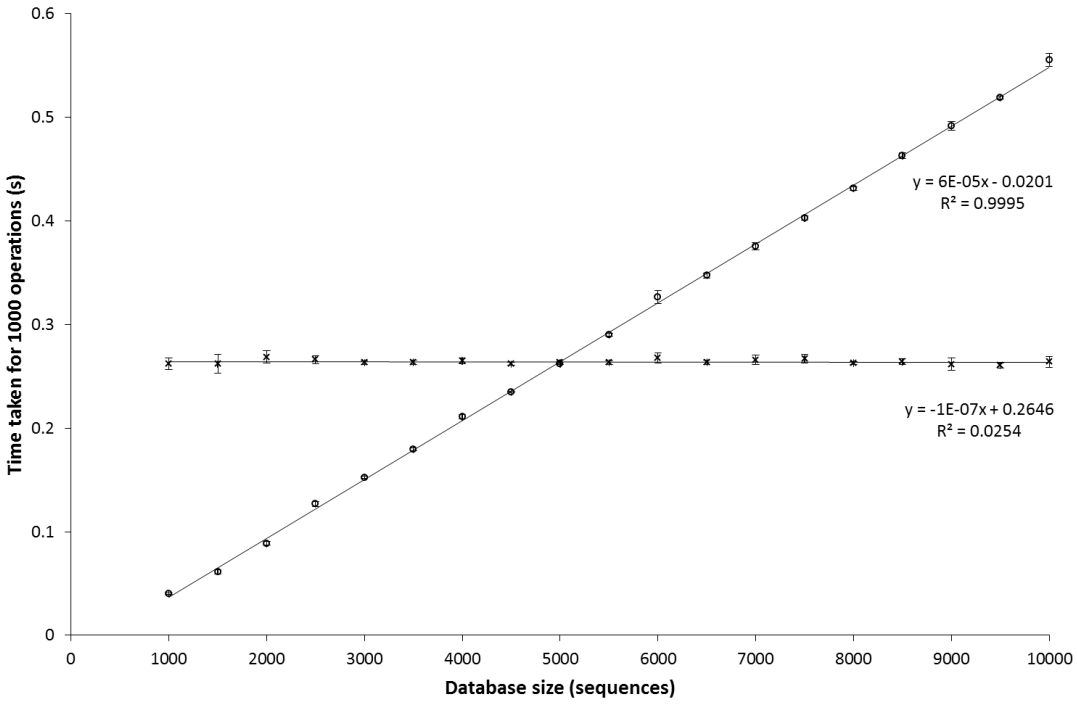


Figure 7.3: Comparison of time taken to determine each melting temperature repeatedly (reanalyse, circles) with time taken to lookup each sequence in the database first (lookup, crosses) for increasing database sizes from 1,000 to 10,000 members. Bars represent the average of 10 repeats of the simulation, which involved 1,000 reanalyses or lookups. Error bars represent one standard deviation of the 10 repeats. Simulation details are explained in the text.

Linear lines of best fit shown in figure 7.3 were fitted using Microsoft Excel 2010. A linear fit was chosen as it seemed to best describe the trend of the line in the region measured.

The initial conditions of the OptiCut algorithm are chosen arbitrarily; such that all the overlaps are of equal length that depends on the number of oligonucleotides per strand. Since the algorithm selects the greatest outliers for optimisation first, the algorithm can handle particularly long or short overlaps by bringing their length closer to the mean in the initial optimisation steps. A formal investigation of the sensitivity of the algorithm to 'badly chosen' initial conditions was not carried out.

7.3. Loop identification

A module was implemented for the identification of loops occurring in the optimisation process. Fully defined optimisation procedures always run the risk of optimising into a local

minimum rather than the true global minimum. In situations where fully exhaustive implementations are impractical, like ours, various strategies can be employed to avoid a local minimum trap. Such strategies include an implementation of a random 'jump' or a brief, restricted 'exhaustive' search. Both these two strategies were implemented as switchable events within the main program. In the event that a loop is found the program either; 1. A random set of cut sites are shifted by a random (<3 bp) amount and the program allowed to continue. Or 2. A brief exhaustive search is performed where the cut sites are shifted by only a few base pairs and every possible temperature determined.

These two methods ensure that the minimum found is sufficiently robust for the reasons described above.

The end state of the optimisation is a loop. The algorithm makes a cut site move in an attempt to reduce the total range of temperature values. The move chosen results in the algorithm making a subsequent move of a cut site which returns the cut sites to their original state. A loop then ensues as the algorithm flips between the two states. Such loops are clearly identifiable on graphs of the RMSD of a given optimisation (see figure 7.3): After the 60^{th} cycle, the RMSD switches between two values for the remainder of the iterations.

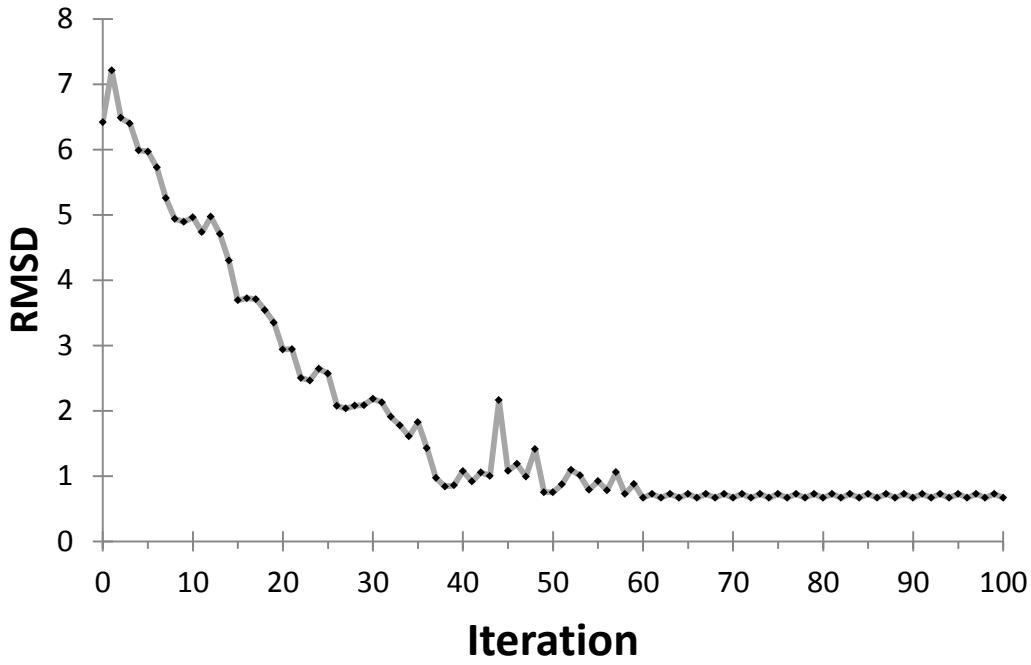


Figure 7.4: Graph of RMSD over iteration of a representative optimisation run. Note the cycle which is reached at around the 60th iteration.

The trace shown in figure 7.4 is representative of a typical optimisation. The trace is obtained by recording the RMSD of the melting temperatures of all the overlaps of the whole sequence set relative to the average overlap melting temperature. A hundred iterations are usually sufficient for the optimisation to reach the cycling state. Most optimisations will reach the cycling state in less than half this number of iterations. The OptiCut GUI includes a checkbox which enables automatic loop detection. Upon detecting a loop, the algorithm will stop optimising and will proceed to display the output. The automatic loop detection check box should be ticked for best performance.

It is possible that the loop-end algorithm might represent only a local minimum rather than the true global minimum. Two methods were implemented to check whether the initial looping minimum could be improved: The first is a simple, short exhaustive search. The second is a temporary cut site locking. In the first, the cut sites are all moved by two base pairs in each direction. The search is exhaustive because every possible combination of positions within this restricted set is considered. In the second, one of the

cut sites that participates in the loop is 'locked', i.e. the algorithm is forced to make the second best move rather than the preferred move. After proceeding for a time, cut sites then became unlocked and the algorithm could then resettle on a new minimum. Locking of sites, however, did not improve the overall effectiveness of the optimisation. A brief exhaustive search was also determined to be inefficient: The optimisation gain was insignificant relative to the amount of computing power necessary to perform it.

7.4. Visualisation of algorithm performance

The cut mutant library can be visualised as a chessboard with a unique sequence in each row and a specific overlap in each column where colour represents melting temperature. Each element of the matrix then represents the melting temperature of the overlap for a given sequence. Graphing the overlap temperature on a colour scale with red representing 'hot' and blue representing 'cold' temperatures allows the quick and clear visualisation of the oligonucleotide set that would compose the mutant library with respect to the heterogeneity of the underlying sequences. Furthermore, side by side comparison of before and after graphs allows easy visualisation of algorithm performance.

Figure 7.5 simultaneously visualises the performance of the algorithm and the level of sequence variation present in the combinational library: The performance can be seen by noting the homogenous colour in the second frame as compared to the first, whilst the variation can be seen in the heterogeneous colour in the first frame. The AT-rich region in the last half of the sequence can be seen in the left panel as a set of columns that appear in blue, indicating lower melting temperatures. The effect on the variation in melting temperature of the overlaps can be seen in the right panel of the same figure. Whilst most of the columns are consistent turquoise/pale green colour, some have significant variation within the column (for example: overlap 14, right hand panel of figure 7.5). The significant difference in the G/C content of the 'on' and 'off' sites in this overlap is the cause of the

colour heterogeneity in a given column. A histogram of the overlap melting temperatures can be seen in figure 7.5. This histogram demonstrates the improvement of the algorithm over arbitrarily assigned cut positions; the starting point of the optimisation procedure.

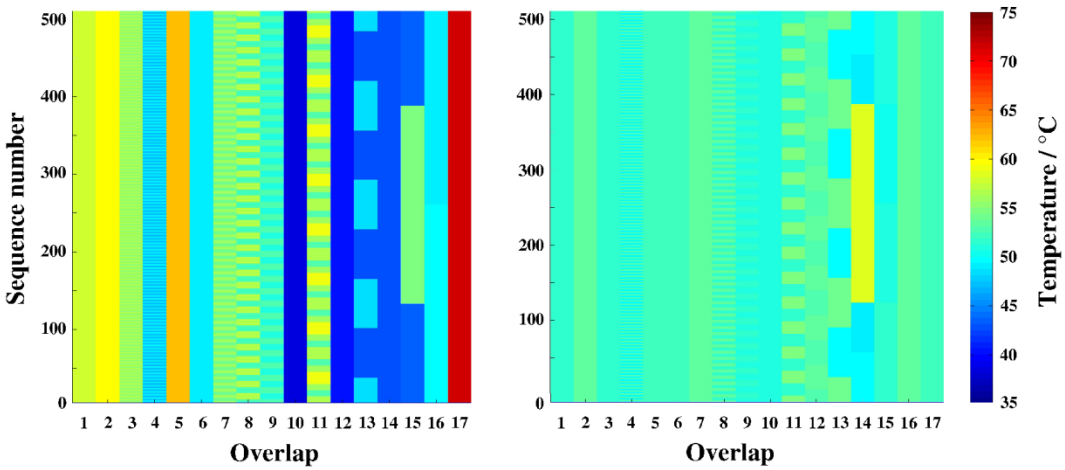


Figure 7.5: Colour histograms visualising performance of algorithm. The left panel is before optimisation, with regularly spaced cut sites. The right panel is after optimisation, with optimised cut sites. Each row represents one sequence, each column one overlap. The colour represents the estimated melting temperature of each overlap.

In figure 7.5, the 15th overlap before optimisation and the 14th overlap after optimisation both exhibit significant temperature variation between sequences. Melting temperature variation between sequences in the same overlap indicates that the mutation sites in this region have a significantly different melting temperature than the wild type sequences. This represents an 'optimisation limit' that is due to the original library sequences. To minimise or avoid this problem, optimisation would have to be able to change the sequence of the mutation sites. Implementation of this would require cross-talk between the optimisation algorithm and the binding site consensus sequence tool.

A more traditional method of visualising algorithm performance is with a histogram of overlap melting temperatures. Histograms represent the frequency with which an individual overlap melting temperature is observed in the whole set. Figure 7.6 shows histogram plots of the same data that can be seen in figure 7.5. An advantage of histogram plots is that the range can be more clearly discerned. The GUI plots only histogram plots.

Like the colour chessboard plot in figure 7.5, the histogram plot in figure 7.6 shows the melting temperature of each overlap in the set. Hence the total area under the histogram is 512 x 17 = 8704 represents the total number of overlaps.

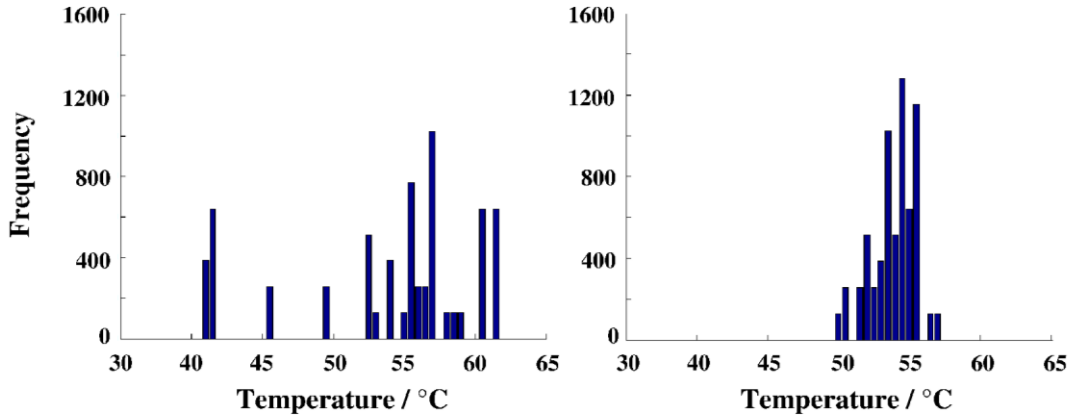


Figure 7.6: Histograms representing performance of the optimisation algorithm. Temperatures of overlaps generated by arbitrarily assigned cut sites are shown in the left panel, whilst the right panel shows optimised overlap positions. Bucket size is 0.1 °C.

The chessboard visualisation makes it clear that the limitations of the optimisation are due to the specific sequences in the library. The difference in melting temperature of the sequences of overlap 14 in figure 7.5 is due to the difference between the mutated and wild type site sequences held therein. Strategies for the mitigation of this limitation are discussed in the further work section of this chapter (see section 7.10). As it stands, the minimum range that the overlap temperatures could ever take is limited by the overlap which possesses the largest mutation site sequence dependant change in melting temperature.

7.5. Competitor identification

A module was written which allows the full set of possible interactions to be analysed, including self-interactions. Hairpin interactions within a single oligonucleotide were not considered. Regulatory modules do, however, often exhibit repeated binding site motifs. As a result, care should be taken where the optimised sequences contain reverse

complimentary sequences that are <1 overlap length apart. To make the algorithm suitable or application where this is the case detection of hairpin loops should be implemented. The module requires a checkbox to be ticked in the GUI for it to be used during optimisation. Once the main optimisation loop is complete, if the checkbox has been ticked, the set of all oligonucleotides is checked for all potential interactions. The chart shown in figure 7.7 shows the output of the competitor identification module. The module will detect intended interactions as well as unintended interactions. Hopefully, the intended interactions will be as strong as or stronger than the unintended interactions.

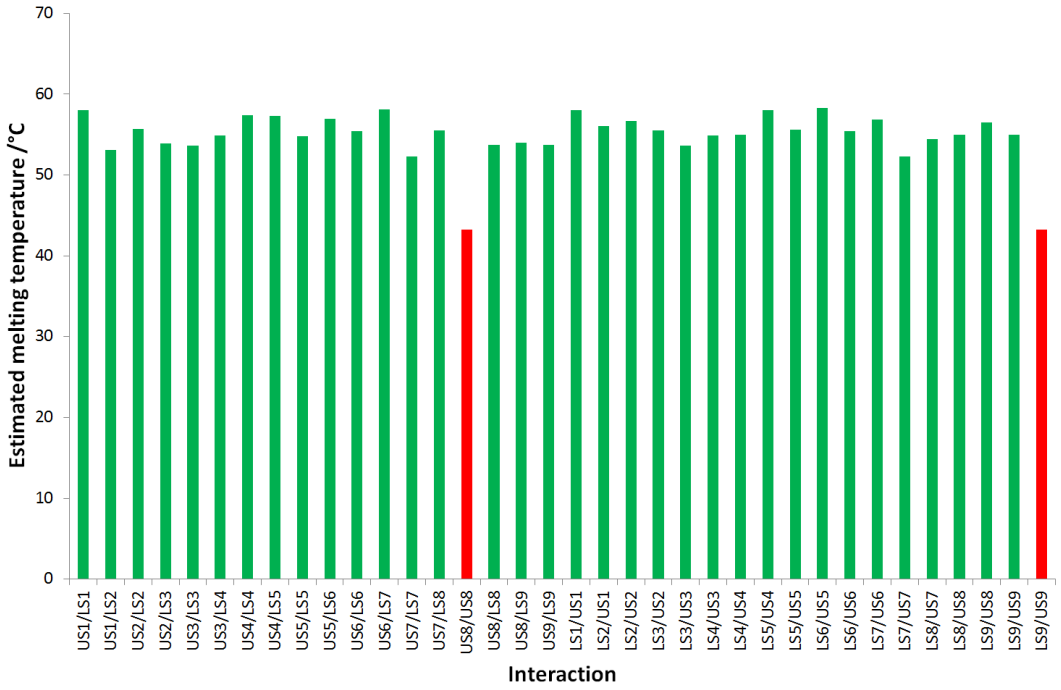


Figure 7.7: Bar chart displaying data obtained by running competitor identification within OptiCut. Intended (green) and unintended (red) interactions for an example set of oligonucleotides necessary to make a single sequence.

To exemplify the output of the competitor identification module, all the feasible interactions of the set of oligonucleotides necessary to build a single sequence is displayed in figure 7.7. The competitor identification module will output 'intended' hits as well as 'unintended' hits. The lower cut off for melting temperature was arbitrarily set at 30°C which means that only interactions of that are estimated at 30°C or more are reported.

As discussed in the introduction (see section 7.1.4) prediction of heteroduplex melting temperature is somewhat imprecise. Therefore, the values here are probably only accurate to the stated value ± 5°C. Even with this uncertainty, the two highest scoring unintended interactions score 10-12°C lower than the intended interactions.

7.6. Efficiency and effectiveness of algorithm

To determine how well the algorithm scales with the size and complexity of the optimisation set, a testing environment was written. The testing environment is described in the methods (see section 4.7.3). In brief, a random sequence generator made sets of sequences with n randomly generated mutation sites of varying, defined length. The algorithm then optimised each of the sets, recording only the time necessary to do so.

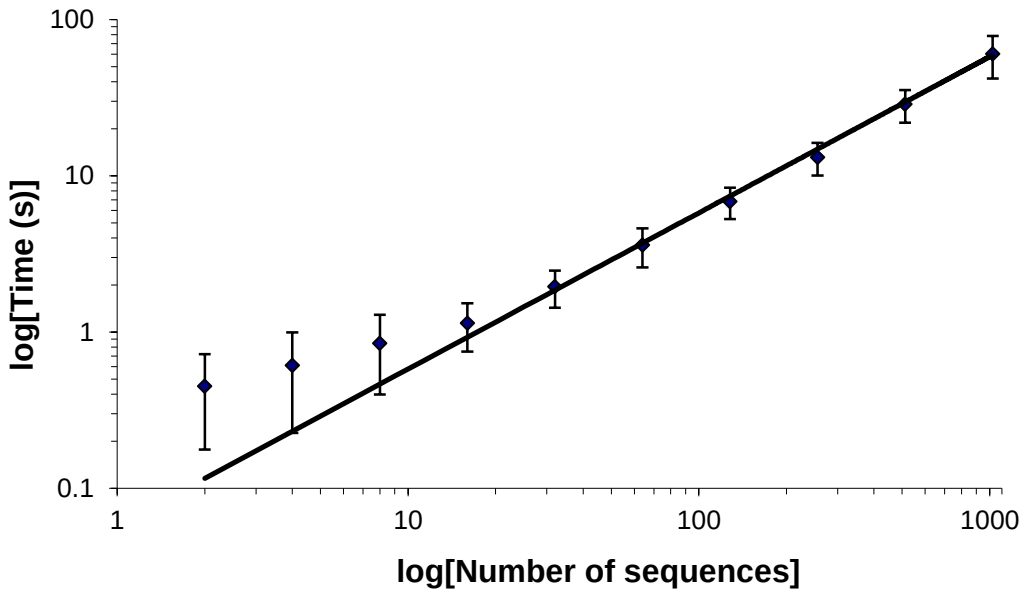


Figure 7.8: Graph of algorithm performance, measured by average total optimisation time required, against number of sequences (blue diamonds). Error bars represent one standard deviation of the optimisation time required. Trendline, shown in black, does not take account of the first three values.

Figure 7.8 shows the how the performance of the algorithm scales with the number of sequences in the optimisation set. Because the algorithm is being applied to a library of all

possible combinations of several mutant sites, the library contains 2^n sequences where n is the number of mutation sites.

As can be seen in figure 7.8, the algorithm scales linearly with the number of sequences in the optimisation set except for optimisations performed on small numbers of sequences. For ≤16 sequences the total optimisation time in generally <1 second. For optimisations of this type, the time taken for the algorithm to initialise the necessary data structures becomes significant with respect to the time spent optimising the sequences. This baseline amount of time required for the algorithm to initialise is probably the main source of the non-linearity observed for sets containing ≤16 sequences.

In contrast to the number of sequences in a given set, the number of iterations is a poor representation of the efficiency of a given optimisation as measured by total time for optimisation minima to be reached. Figure 7.8, 7.9 and 7.10 are representative of the same empirical data.

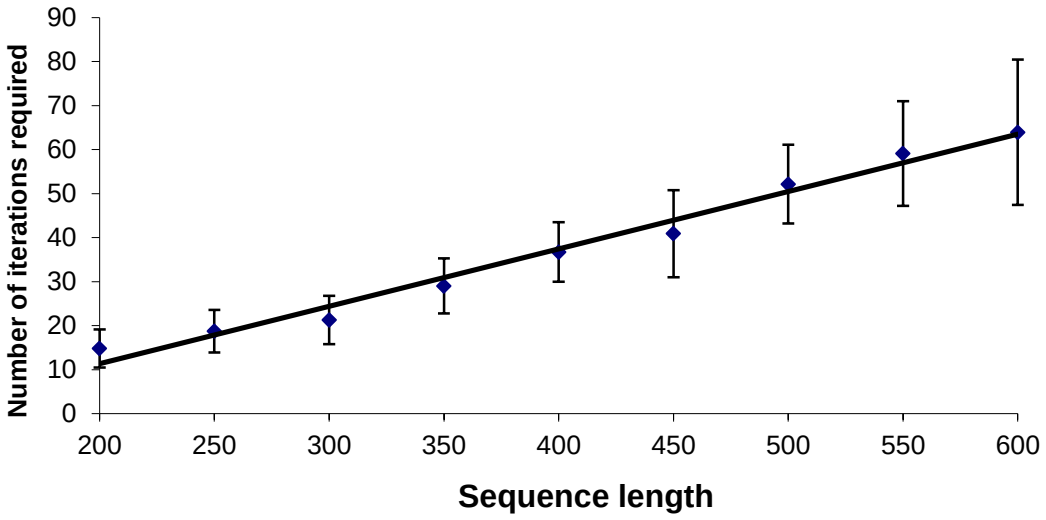


Figure 7.9: Graph of the average number of iterations required before the optimisation minima is reached against length of inputted sequence (blue diamonds). Error bars represent one standard deviation of the values.

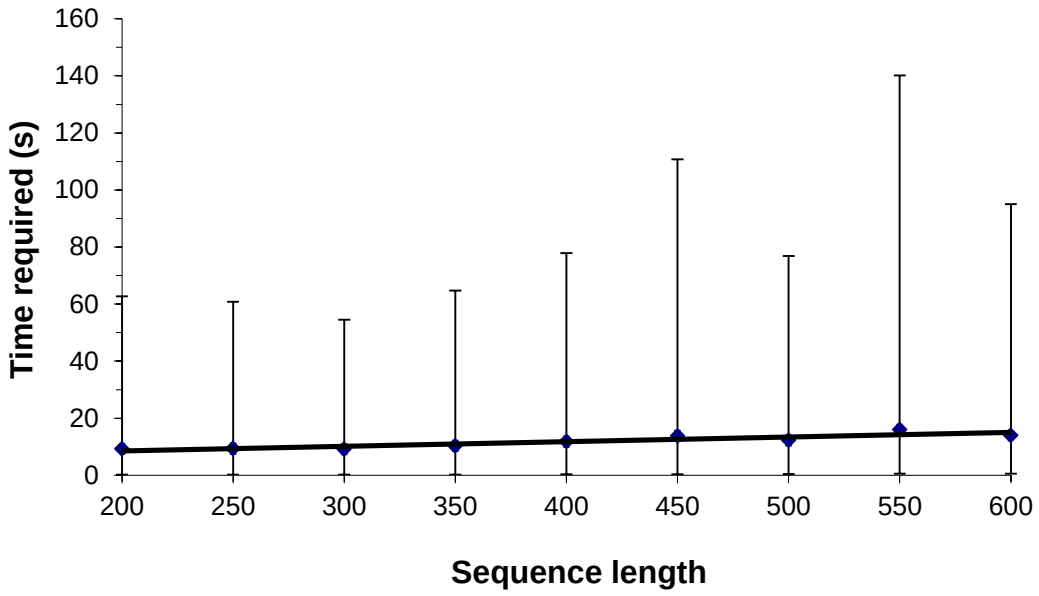


Figure 7.10: Graph of the average time required before the optimisation minima is reached against length of inputted sequence (blue diamonds). Error bars represent one standard deviation of the values.

Figure 7.10 shows that the time required for optimisation depends weakly on overall sequence length. In contrast, figure 7.9 shows that the number of iterations required to optimise each sequence does correlate with sequence length. The observation that the number of iterations is not proportional to the time required was not expected. This observation does indicate, however, that number of iterations before a minimum is reached should not be taken as the sole measure of algorithm performance.

The sequences used to test algorithm performance in this section were generated randomly and therefore are more likely to be homogenous than true, genetic sequence. As a result it is likely that randomly generated sequences will require less optimisation than true, genetic sequences. Although this relationship was not tested, significant performance differences should only occur in repetitive or variable, where one portion is significantly more G/C rich than another, sequences.

The overall performance of the algorithm is characterised by the range of the RMSD of the optimised overlaps. None of the commercially available programs for optimisation of sequences are capable of producing sequence sets that can be used interchangeably in several gene assembly reactions to produce different products by ligative assembly. Optimisation of each member of the mutant sequence library using a program with a batch optimisation function (see section 7.1.2) would produce sequences that have cut sites at different positions in the overall sequence. The oligonucleotides are not, therefore, compatible with each other. The only available optimisation program is the SeqZego program (see section 8.2.1) that was used by the Gao lab. Both programs were used to optimise the same sequence set, a set of 512 sequences of CRM-B, with 6 oligonucleotides per strand. The results are displayed in figure 7.11.

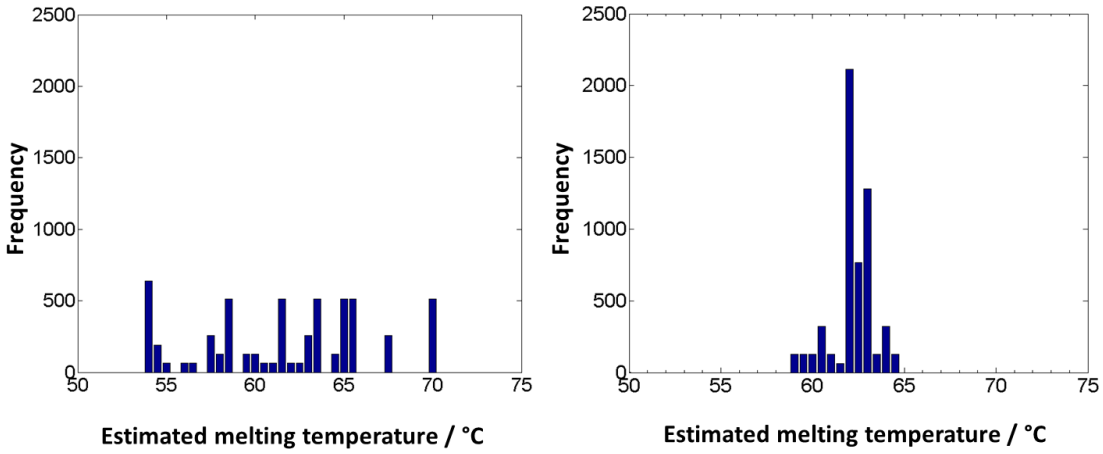


Figure 7.11: Histogram showing comparison of oligonucleotide optimisation by the Gao lab's SeqZego (left) and the OptiCut (right) optimisation algorithms. Optimisations were run on the same sequence sets.

The OptiCut optimised oligonucleotide sets exhibit a significantly smaller range of melting temperatures than the SeqZego optimised oligonucleotide set: 59 – 65°C vs. 54 – 70°C, respectively. Furthermore, the OptiCut optimised melting temperatures appear to have a nearly normal distribution, as compared to the more uniform appearance of the SeqZego optimised melting temperatures.

7.7. Optimisation and cost minimisation

The assembly procedure is based upon selecting a subset of oligonucleotides from a master set, the subset being sufficient to make a given sequence and the master set necessary to make all the sequences. Some oligonucleotides will be used for every sequence whereas some will only be used for a subset of sequences. A given oligonucleotide will be used in 2^{n-a} sequences where 'n' is the total number of mutant sites in the library and 'a' is the number of mutation sites present in a given oligonucleotide. Conversely, in order to be able to build a full library 2^a versions of a given oligonucleotide containing 'a' mutant sites will be required. It is thus in the interest of the experimenter that the number of mutant sites per oligonucleotide be minimised. To minimise the number of oligonucleotides necessary for a complete set, situations where a cut site occurs within a mutation site should be avoided, as this would cause the mutation site to be doubly represented in terms of the number of oligonucleotides necessary.

The OptiCut algorithm can detect when a cut site occurs within a mutation site. In this event, the algorithm will attempt to adjust the cut site position as minimally as possible so as to take the cut site out of the mutation site. This will most likely have an effect on the distribution of overlap melting temperatures and the resulting RMSD. In the event that this option is enabled, both the normal output and the minimised output will be displayed in the output.

7.8. OptiCut Graphical User Interface

The OptiCut algorithm is associated with a GUI that allows users unfamiliar with MATLAB to use the optimisation algorithm. The GUI was written using the MATLAB's GUIDE GUI tool. All GUI screenshots displayed here are taken on Windows 7 with Aero desktop features enabled. The graphical features of the GUI may appear differently on other operating systems. The code for the OptiCut GUI is presented in appendix A.3.1.

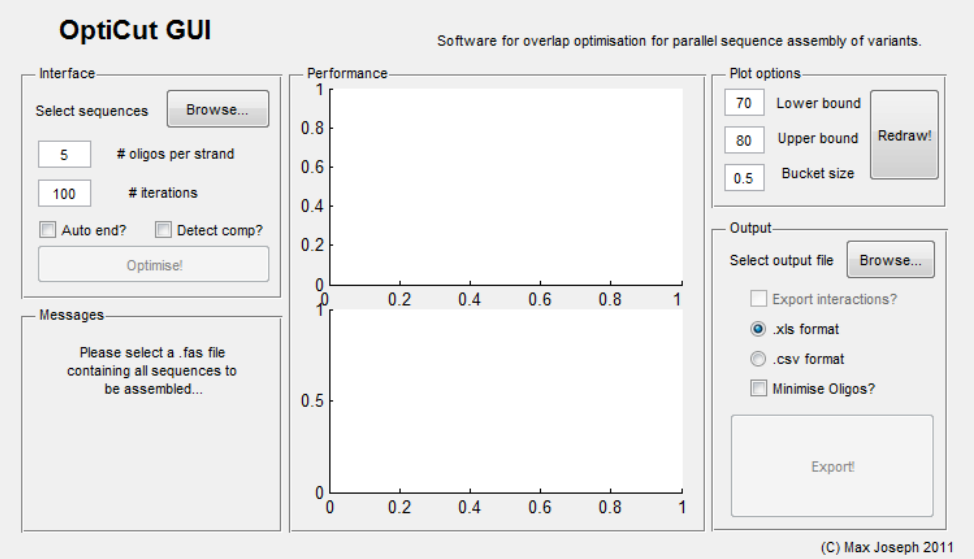


Figure 7.12: Screenshot of the OptiCut GUI on opening. Panels separate the key controls and graphs in the middle show the performance of the algorithm.

The GUI uses panels to separate the controls and displays. Certain controls (such as the Optimise and Export controls in figure 7.12) are disabled and appear greyed out until the necessary information to make the button functional is supplied by the user. Default values are provided where possible but can be modified as required. Button presses will be necessary in order to enact the changes in these user inputs: The Optimise button will have to be pressed for OptiCut to enact a change to the '#oligos' or '#interations' text input boxes.

The 'Browse…' buttons open dialogs for the user to select files appropriately. Whilst the input file 'Browse…' dialog requires the file selected to exist, the export file 'Browse…' dialog can be pointed to a file that does not exist, in which case the appropriate file will be created, or to a file which already exists, in which case the file overwrite dialog will be displayed.

The Messages panel interactively displays messages, suggesting the next step of the optimisation process. It is also used to explain errors that may occur, such as when the cancel button is pressed during the 'Browse…' dialogs.

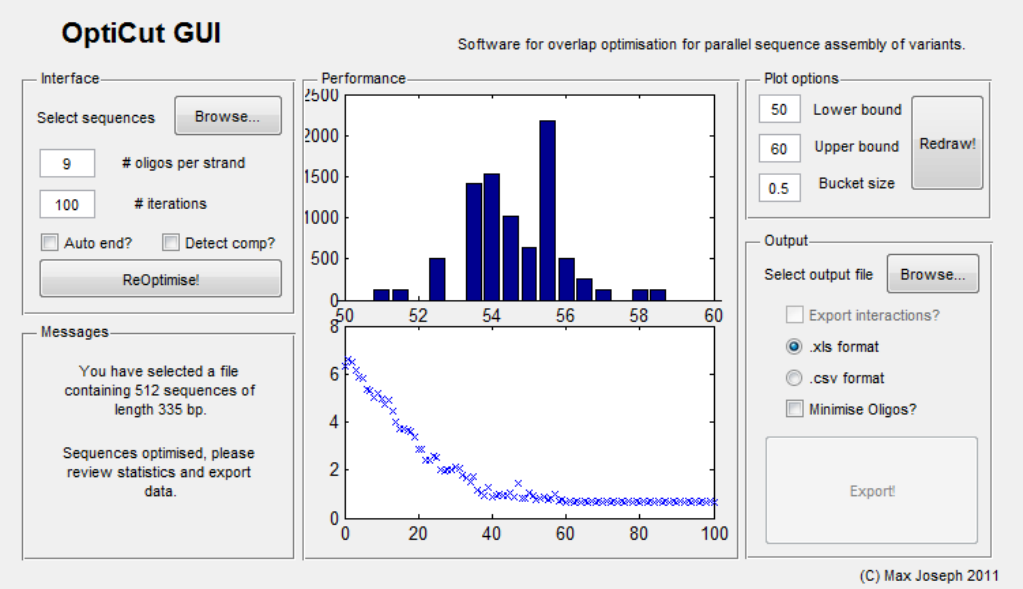


Figure 7.13: Screen shot of the OptiCut GUI after running an optimisation on a sequence set. Note the two plots which display the algorithm performance.

Pressing the 'ReOptimise' button will cause the program to repeat the optimisation procedure, even if no changes to the input have been made. Changes to the input.fas file after pressing Optimise and 'ReOptimise' will not be noticed. The Browse button to select the file must be pressed for the GUI to recognise changes in the input.fas file.

The histogram plot represents the distribution of melting temperatures across the whole overlap set. The histogram plot can be redrawn at any time using the Redraw button in the plot options panel. The Lower bound is the minimum limit of the x-axis whilst the Upper bound is the maximum limit of the x-axis. Entering a values where Lower bound < Upper bound and pressing Redraw will cause the program to throw an error. The plot will not be updated. The error is not fatal, however, and entering appropriate values and pressing Redraw will allow the program to recover.

The output panel allows configuration of the output and export of results of the optimisation algorithm. The 'Export' button is disabled and greyed out until a suitable filename is inputted using the 'Browse…' button in the output panel. Checking the .xls or .csv radio buttons will amend the filenames returned when selecting the filename through

the Output 'Browse…' button. Changing the radio button selection after choosing a filename will not, however, amend the filename suffix, although it will change the filename format. It is also important to note that the .xls button will cause the file to be exported at 2007 .xlsx format. Users who do not have Office 2007 or greater or the compatibility pack for Office 2003 are restricted to using the .csv export option.

7.8.1. Sub functions and installation of the OptiCut GUI

The main GUI function is called OptiCutGUI.m. Running this file on any machine with MATLAB installed will cause the GUI to run. The following functions must also be present: GroupCutINIT.m which contains the optimisation algorithm and GroupCutOutput.m which formats the optimisation algorithm output suitable for printing to a file. GroupCutINIT.m and GroupCutOutput.m are presented in appendix A.3.2 and A.3.3, respectively.

A histogram plotting function used by GroupCutGUI.m was written when the inbuilt MATLAB histogram function was found to lack the necessary flexibility. This histogram plotting function can be found in plothist.m (see appendix A.3.4). Similarly, a novel heteroduplex melting temperature function was written when the MATLAB Bioinformatics Toolbox was found to not possess a function to determine this information. This function was built around the melting temperature algorithm reverse engineered from IDT SciTools[28,30]. The function is called HeteroDimerMeltingTemp.m (see appendix A.3.5). The melting temperature of homologous overlaps was determined by the SantaLucia method, see section 7.1.3 and appendix A.3.6.

The OptiCutGUI has been compiled into an executable that installs the necessary files onto a standard Windows PC. The program needs either a full version of MATLAB or a copy of the MATLAB C runtime (MCR) installed to function. The MCR is essentially a cut down version of MATLAB that contains the minimum necessary for software written in MATLAB to function. The MCR can be used royalty free. Versions of the program with and

without the MCR, the size of each is 414 MB and 1 MB, respectively. The compiled GUI runs identically to the GUI running natively in MATLAB. The compiled GUI cannot be edited, however, whilst the natively running GUI can be edited as required.

7.9. Assembly results

The ultimate test of the optimisation algorithm is the assembly of optimised oligonucleotides. The main discussion of the assembly protocol is discussed in chapter 8. Here a simple demonstration of the algorithm efficacy is presented. An assembly reaction (ligation) was performed as described in the methods (see section 4.4.8.). The result of the ligation reaction was then sampled and the specific sequence amplified by PCR. The product of the PCR was then visualised in a gel which can be seen in figure 7.14.

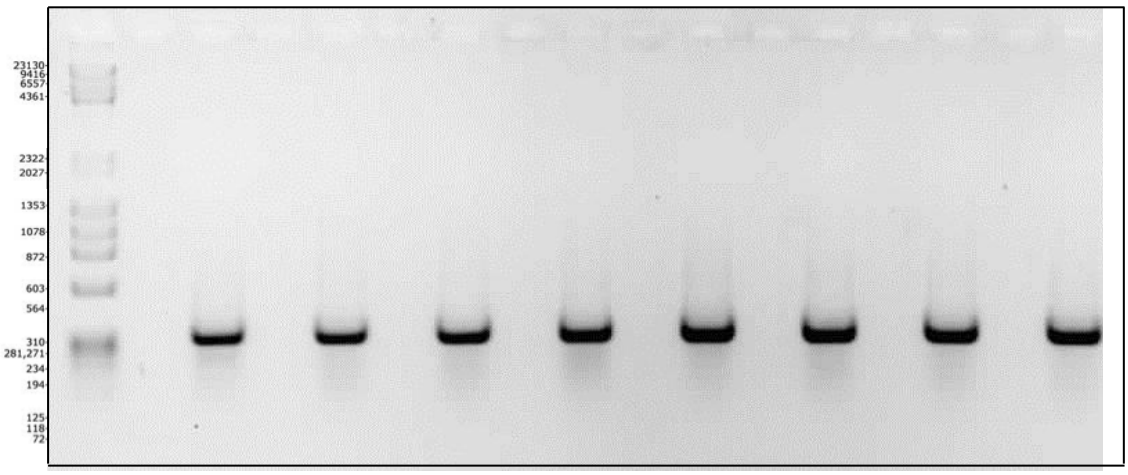


Figure 7.14: Inverted colour agarose gel of 8 of 512 assembled products stained with EtBr. Each lane contains the product of a PCR using a sample from *Taq* ligase assembled oligonucleotides designed by the OptiCut program.

Single, clear bands slightly larger than the 310 marker band were seen in the gel of the PCR products (see figure 7.14). The presence of the specific band of the right weight indicates that the algorithm is working as intended and is successfully generating overlap sequences that can be used to assemble the target sequence (as indicated by sequencing of the assemblies). This result contrasts markedly with many gene assembly results which

produce many product bands. As these alternative bands represent competing products that reduce the overall efficiency of the assembly of the desired product, a single band is significantly better. A minor, fuzzy band with a smaller molecular weight can also be seen on the gel. This band is probably partial amplification of incompletely assembled products of the ligation reaction and also the primers used in the PCR.

To determine the error rate of the assembly reaction and ensuing amplification steps, PCR products were TOPO cloned and colonies were Sanger sequenced. A total of 42 colonies from 19 separate TOPO cloning procedures were sequenced (for breakdown of sequencing results see table 8.2). The average error rate in all sequenced constructs was 0.42%. This compares favourably with other gene assembly methods[19,31–33]. Incorporation of mutant oligonucleotides (heteroduplex binding events) is thought to be the major contributing factor to incorrect products, above the contribution of erroneous incorporations by DNA polymerase[34]. However, previous analysis of gene assembly by PCR has indicated that mutations in overlap regions is roughly half that in single strand regions[19].

7.10. Further work

Currently, Opticut does not perform codon optimisation, but could be extended to do. Codon optimisation is typically employed when sequences are to be expressed as protein; particularly when a protein is from one species and is to be expressed in another. Reverse translation of a desired protein sequence in a given species could also be implemented. In the case discussed herein, the DNA sequences themselves are the active elements, and are not meant to be expressed as protein. As such a codon optimisation module was not implemented. Change of codons could also be employed to minimise the mutation-site dependant melting temperature variation as seen in columns in figure 7.5, although any

codon changes would have to be made with regards to the efficiency with which that codon was translated in the target organism.

The most commonly used method for gene assembly, particularly of long genes, involves PCR using overhanging primers. As discussed in the introduction (see section 7.1.) this is not suitable for the application used here. OptiCut could be modified to perform optimisations on oligonucleotides designed for a PCR-based assembly process.

Here, the OptiCut program is applied to a set of sequences that are all largely identical. The OptiCut program could equally be applied, as efficiently, to sequences with more significant differences. The current program, however, will not accept inputs that have sequences with non-uniform lengths. Modification of the initial overlap positioning, handling of variable numbers of overlaps per sequence would be required. The main algorithm would be largely unchanged apart from looking up overlap positions on a per overlap basis rather than using the same positions for every sequence.

The competitor identification module currently only considered interactions between oligonucleotides and not self-interactions within the same nucleotides. The module could be developed further to include consideration of these loops. Thus giving the user more confidence that the oligonucleotide set selected would function as intended.

The major weakness of this assembly method is the mutation-site sequence dependant melting temperature variation, as mentioned in section 7.4. This variation cannot be completely avoided as the mutation site must be, by definition, different to the wild-type site. The differences could be minimised however and an additional module could be developed for the suggestion of mutation sites. On one level, the module could suggest mutation sequences that result in minimal change in melting temperature over the site for the user to check with the BiFa tool for suitability. On another level, the module could communicate with the BiFa tool or the TRANSFAC database itself to determine

whether the mutation removed the target site without affecting neighbouring sites or introducing additional sites.

7.11. Conclusions

This chapter describes the OptiCut algorithm and associated GUI which allows the user to optimise oligonucleotide sequences for ligative assembly reactions. Currently available sequence optimisation software does not provide oligonucleotide sequences that can be used in the assembly of several related DNA sequences. The OptiCut software described herein can be used to define a set of sequences that can be used to assembly a library of related sequences. The key conclusions of this chapter are:

- The OptiCut program (described in section 7.2.) can be used to optimise sequences for a library containing hundreds or thousands of individual sequences rapid and efficiently.
- OptiCut has an easy-to-use and clear GUI and option for .xlsx or .csv outputs (described in section 7.8).
- Optimisation scales linearly with number of sequences (see section 7.6).
- The algorithm can be run on any Windows desktop computer (provided either MATLAB or MCR are installed).
- Proof of concept has been obtained through a successful assembly of OptiCut-optimised oligonucleotides (see section 7.9).

The OptiCut program is small, cheap in terms of system resources and can be installed and run on large proportion of computers. These facts should facilitate the programs use by researchers investigating problems by the use of similar mutant libraries.

In its current form, the goodness of the optimisation result is dependent on the variation present in the input sequence and the number of oligonucleotides per strand. The

OptiCut algorithm is not currently setup for optimisation of overlaps for assembly by, for instance, polymerase-based assembly.

7.12. References

1. Atchley, W. R., Fitcht, W. M. & Bronner-frasert, M. Molecular evolution of the MyoD family of transcription factors. **91**, 11522–11526 (1994).
2. Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* **132**, 2685–2695 (2005).
3. Ten Broek, R. W., Grefte, S. & Von den Hoff, J. W. Regulatory factors and cell populations involved in skeletal muscle regeneration. *Journal of cellular physiology* **224**, 7–16 (2010).
4. GeneArt, I. Gene synthesis by GeneArt. (2011).at <http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cloning/gene-synthesis.html?CID=fl-genesynthesis>
5. Sigma-Aldrich Sigma Gene Synthesis (GeneOracle). (2011).at <http://www.sigmaaldrich.com/life-science/custom-oligos/custom-dna/product-highlights/gene-synthesis.html.>
6. Technologies, I. D. IDT Custom Gene Synthesis. (2011).at <http://eu.idtdna.com/catalog/customgenesyn/page1.aspx>
7. Operon, E. M. Eurofins Gene Synthesis. (2011).at <http://www.eurofinsdna.com/products-services/gene-synthesis.html.>
8. DNA2.0 DNA2.0 Gene synthesis. (2011).at <https://www.dna20.com/index.php?pageID=17>
9. BlueHeron Blue Heron: The Gene Synthesis Company. (2011).
10. OriGene OriGene gene synthesis. (2011).at <http://www.origene.com/gene-synthesis/>
11. BioMatik BioMatik custom gene synthesis. (2011).at <http://www.biomatik.com/SiteContent/Gene-Synthesis.aspx>
12. Entelechon Entelechon perfectGene. (2011).at <http://www.entelechon.com/synthetic-products-2/gene-synthesis/>
13. Villalobos, A., Ness, J., Gustafsson, C., Minshull, J. & Govindarajan, S. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7**, 285 (2006).
14. Lorimer, D., Raymond, A., Walchli, J., Mixon, M., Barrow, A., Wallace, E., Grice, R., Burgin, A. & Stewart, L. Gene composer: database software for protein construct design, codon engineering, and gene synthesis. *BMC biotechnology* **9**, 36 (2009).
15. Bode, M., Khor, S., Ye, H., Li, M.-H. & Ying, J. Y. TmPrime: fast, flexible oligonucleotide design software for gene synthesis . *Nucleic Acids Research* **37** , W214–W221 (2009).
16. Rouillard, J.-M., Lee, W., Truan, G., Gao, X., Zhou, X. & Gulari, E. Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Research* **32**, W176–180 (2004).
17. Rydzanicz, R., Zhao, X. S. & Johnson, P. E. Assembly PCR oligo maker: a tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production. *Nucleic Acids Research* **33**, W521–5 (2005).
18. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis . *Nucleic Acids Research* **30** , e43–e43 (2002).

19. Wassman, C. D., Hatfield, G. W. & Lathrop, R. H. Computationally optimised DNA Assembly of synthetic genes. *International Journal of Bioinformatics Research Applications* **4**, 324–336 (2009).

20. Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D. & Bader, J. S. GeneDesign 3.0 is an updated synthetic biology toolkit. *Nucleic acids research* **38**, 2603–6 (2010).

21. Puigbò, P., Guzmán, E., Romeu, A. & Garcia-Vallvé, S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research* **35**, W126–31 (2007).

22. Jayaraj, S., Reid, R. & Santi, D. V. GeMS: an advanced software package for designing synthetic genes. *Nucleic Acids Res* **33**, 3011–3016 (2005).

23. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1460–5 (1998).

24. Meinkoth, J. & Wahl, G. Hybridization of nucleic acids immobilized on solid supports. *Analytical biochemistry* **138**, 267–84 (1984).

25. Huguet, J. M., Bizarro, C. V., Forns, N., Smith, S. B., Bustamante, C. & Ritort, F. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15431–6 (2010).

26. Protozanova, E., Yakovchuk, P. & Frank-Kamenetskii, M. D. Stacked-unstacked equilibrium at the nick site of DNA. *Journal of molecular biology* **342**, 775–85 (2004).

27. Jones, A. C., Austin, J., Hansen, N., Hoogendoorn, B., Oefner, P. J., Cheadle, J. P. & O'Donovan, M. C. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. *Clinical chemistry* **45**, 1133–40 (1999).

28. Owczarzy, R. & Behlke, M. Calculation of Tm for Oligonucleotide Duplexes. *Integrated DNA Technologies* 1–4 (2005).

29. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1460–5 (1998).

30. Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey, J. a, McQuisten, K. a, Almabrazi, H. G., Pedersen, K. F., Lin, Y., Garretson, J., McEntaggart, N. O., Sailor, C. a, Dawson, R. B. & Peek, A. S. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic acids research* **36**, W163–9 (2008).

31. Stemmer, W. P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 10747–51 (1994).

32. Chalmers, F. M. & Curnow, K. M. Scaling up the ligase chain reaction-based approach to gene synthesis. *BioTechniques* **30**, 249–52 (2001).

33. Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X. & Church, G. M. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).

34. Borovkov, A. Y., Loskutov, A. V., Robida, M. D., Day, K. M., Cano, J. A., Le Olson, T., Patel, H., Brown, K., Hunter, P. D. & Sykes, K. F. High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides. *Nucleic Acids Research* gkq677 (2010).

Chapter 8

8. CRM library assembly

The CRM mutant library was created by DNA assembly. DNA assembly was employed so that all the variants in the combinatorial mutant library of CRM-B could be assembled rapidly and in parallel. Two gene assembly methods were used here. The first, Gao synthesis[1], involves the simultaneous synthesis of all the oligonucleotides required for every sequence on a microfluidic chip. The sequences are then removed from the chip for simultaneous, one pot assembly. The second, OptiCut assembly, was employed due to difficulties in obtaining full length sequences from the Gao assembly. In OptiCut assembly, the sequences are assembled separately in parallel using multiwall plates. These successful assemblies were then. The Gao assembly was only partially successful even after significant optimisation. As a result, the OptiCut assembly process was developed so as to obtain successfully assembled sequences for testing in chapter 9. In contrast to the Gao assembly process, the OptiCut process uses separately synthesised oligonucleotides that are optimised using the OptiCut sequence optimisation algorithm described in chapter 7. A key step of the assembly process described in this chapter, the accurate and equimolar mixing of oligonucleotides prior to assembly, was performed in the microfluidic device described in chapter 6.

This chapter firstly describes the optimisation of the Gao assembly and the analysis of the partial success, in section 8.2. Secondly, the OptiCut assembly process is described in section 8.3. Finally, the possible further work is discussed in section 8.4 and conclusions are drawn in section 8.5. The flow chart depicted in figure 8.1 represents this.

The assembly process was first optimised on the bench top and thereafter the same process was used to the assembly of mixtures made using the microfluidic device described in chapter 6.

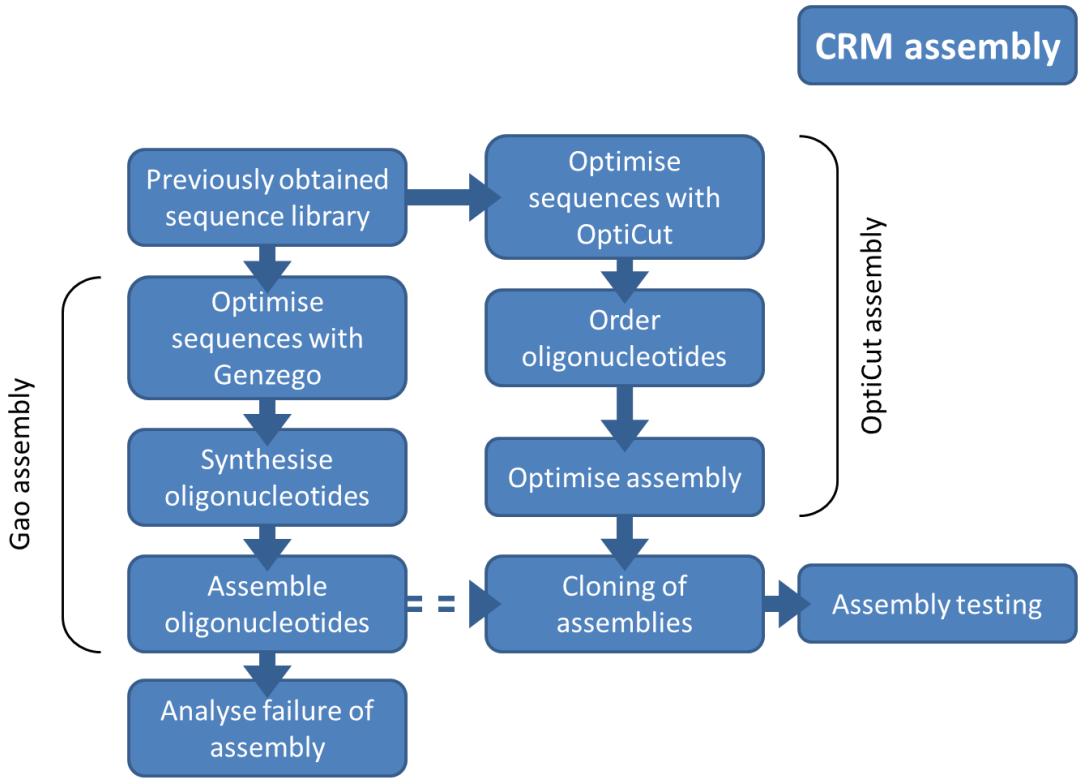


Figure 8.1: Flow chart outlining the two attempted CRM library assembly processes. Initially, the Gao assembly path was expected to yield assemblies for cloning and testing. Due to the failure of the Gao assembly (as denoted by the dashed arrow), the OptiCut assembly process was employed and the failure of the Gao assembly process was analysed.

8.1. Introduction to gene assembly

8.1.1. Current gene assembly methods

The first assembly of a synthetic double stranded DNA molecule was reported more than 40 years ago. In 1970 Agarwal *et al.* assembled *in vitro* a 77 base pair sequence that encoded a yeast alanine transfer RNA[2] [19]. The first synthesis of a gene more than 1 kb in length was performed by Ferretti *et al.*[34]. Most recent development of gene synthesis in the last 5 years has focused on synthesising longer constructs. A milestone was reached

recently when an entire genome was synthesised[5,6] with implications for the development of so-called 'synthetic life'.

Several methods exist for the synthesis of large numbers of genes in parallel[7,8]. Other gene synthesis methods which were investigated for their suitability are TopDown[9], TBIO [10], overlap extension PCR[11–14] and dual-asymmetric PCR[13] . Various gene assembly methods have been reviewed previously[15]. The ligase chain reaction (LCR) method has been used successfully in gene assembly[16]. The advantages of LCR based gene are principally the minimal use of polymerase compared to PCR based approaches. Genes assembled by the LCR method are regarded to be of higher quality, or possessing fewer errors, than PCR based methods. LCR based approaches require more DNA template ad every base of the final gene sequence is synthesised twice. In PCR based approaches, most bases are synthesised once with only base pairs in the overlap regions being synthesised twice.

8.1.2. DNA assembly of the CRM

At 300 base pairs the CRM is a relatively short sequence of DNA for an assembly reaction. Furthermore, since the mutation sites are not scattered uniformly along the sequence, instead clustering towards one end, the density of sites is high. The most frequently used DNA assembly method is PCR-based: Long primers of 50-100 base pairs are designed with short regions of homology 15-20 base pairs long between them. A PCR then fills in the gaps between the primers. See figure 8.2 (right panel) for a diagram of the PCR-based gene assembly method. The oligonucleotides overlap with each other and leave gaps between each oligonucleotide on each strand. The overlapping portions of the oligonucleotides then act as primers for a DNA polymerase to fill in, making a contiguous sequence. The CRM could be fully described by 3-6 PCR primers of this type. However, due to the concentration of sites, there would be 5-6 sites on just one of the primers. In order to describe the whole mutant library, therefore, 2^5-2^6 (32-64) versions of a single primer would be necessary. If a

mutation site occurred within an overlap region then the variants of both primers would be affected. Finally, analysis of the final sequence from previous PCR assemblies have indicated the rate of error in single strand regions is higher than double strand regions[17]. Due to the high density of sites and the requirement for high sequence fidelity, the ligative assembly approach was selected as the preferred method by which to generate the library.

The oligonucleotides used for ligative assembly, shown in figure 8.2 (left panel), differ from those used for PCR-based assembly in that there are no gaps between oligonucleotides on the same strand[18]. Instead, both strands are contiguous, which removes the requirement for a polymerase step to fill in the gaps. A DNA ligase can be used to ligate pairs of oligonucleotides together. A high temperature ligase, such as *Taq* ligase, can be used. Correctly annealed overlapping oligonucleotides will anneal at a higher temperature than improperly annealed oligonucleotides that are only partially complementary. *Taq* ligase, active at 40 – 60°C, will actively ligate at a higher temperature than other ligases, such as T4 DNA ligase. Therefore, by using *Taq* ligase at a high temperature it is less likely that incorrect assemblies will be produced.

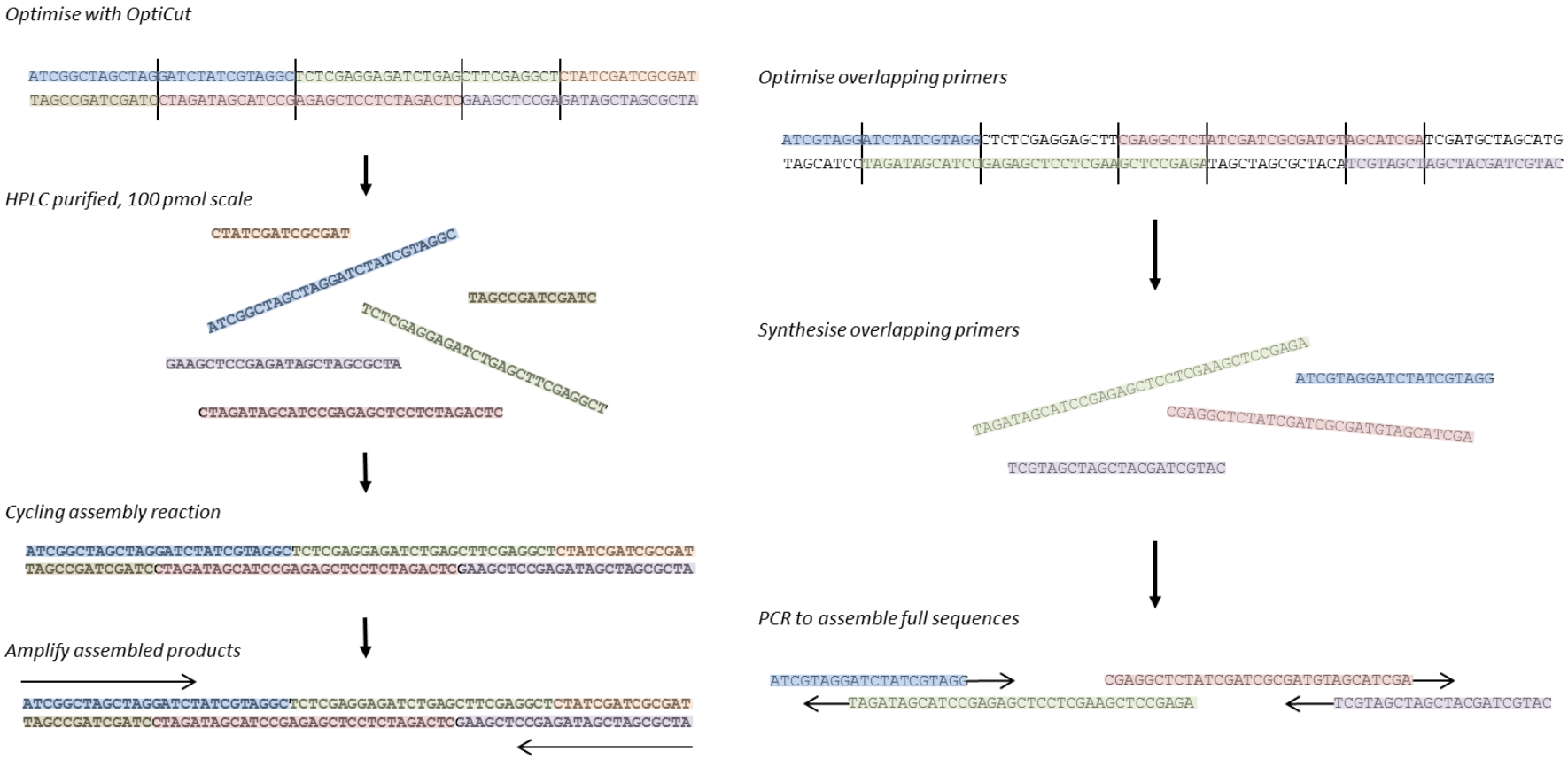


Figure 8.2: Diagram comparing ligative (left) with PCR-based (right) gene assembly. Ligative assembly uses oligonucleotides that cover both strands in their entirety whereas PCR-based assembly only cover both strands in overlapping regions, the rest being covered in one strand or the other. The sequences in this diagram are representative only, longer overlaps are usually used (25 – 50 base pairs).

8.1.3. Literature examples of library assembly

Libraries of mutants are typically used for directed evolution studies that seek to optimise sequence-function relationships without *a priori* knowledge of mechanism. There are several methods of creating libraries, such as error-prone PCR, mutator strains, oligonucleotides or recombination, that are suited to different applications[19].

Mutant libraries could be created by progressive site directed mutagenesis (SDM) of plasmids. SDM involves amplification of a sequence with primers that contain a mismatch in their binding site. The product of the reaction will contain the sequence of the primer rather than of the template, thus introducing a mutation. Due to the annealing limitations of the primer, SDM is typically limited to altering ≤3 bp of contiguous sequence. Furthermore, an amplification step is required for every additional mutation required. As each amplification involves the risk mutation, it is likely that mutations would be added during the production of a whole combinatorial library of sequences by SDM. The use of SDM is generally limited to the modification of enzyme properties[20–22]. Such directed evolution studies will also make use of random SDM and error prone PCR [23,24]. Previous work utilising SDM includes modifying properties of GFP[25].

Recombinant libraries of various genes have been produced in the past:

- Bacterial lipases recombined by a primerless PCR method[26].
- Recombinant human collagen variants produced by modular recombination of gene-coding fragments[27].
- Recombination of 15 subtilisin gene variants through a spiked overlap extension PCR method followed by clonal selection[28].
- Synthetic antibodies produced without DNA amplification, although gel purification and clonal selection were employed[29].

None of these methods are suitable for the library in this case due to the constraints of working with a regulatory module in a high-throughput, parallel manner. The requirement that the sequence outside of the mutations sites remain unaltered means that recombination based upon restriction enzymes is not suitable. The high density of the sites precludes the use of a PCR-based approach.

Libraries described in this chapter represent an advance from other recently developed methods, such as systematic evolution of ligands by exponential enrichment (SELEX)[30]. The SELEX method makes use of barcode tags, randomised sequences and Illumina sequencing to identify the sequences which transcription factors bind to in mammalian cell lines. Although the application of our method is similar to that of SELEX, the sequence length (14 bp in SELEX) is small compared to the sequence length by our method and randomly generated. The study of libraries composed from larger sequences is important in order to elucidate the interactions between transcription factors that may have binding sites separated by 100s of base pairs. Random mutations could be incorporated through the application of oligonucleotides containing randomised sequence in specific base pair positions. 4^n variants would be produced, where n is the number of bases in the varying region. Successful base pairing could then be achieved by generating complementary oligonucleotides with random sequence in the relevant positions. Downstream analysis could then be employed to identify specific mutants after a selective assay step.

8.1.4. Downstream separation

The Gao assembly is a one-pot assembly where the oligonucleotides necessary to assemble every sequence in the library is present in the reaction mixture. In previous examples of one-pot assembly the full length sequences possess unique primers[7] or are easily separable on the basis of length. In this case, the number of variants to be synthesised (several

hundreds) means that separation by primer variation prohibitive. Furthermore, the sequences were all of the same length, which makes separation on sequence length impossible. Clonal selection is a viable solution to separation of the variants, but would require many more samplings from the set than unique members of the set. Furthermore, the likelihood that a given sample would contain a mutation means that several members would have to be sampled several times.

To estimate the number of samplings that would be necessary to be reasonably certain of selecting every member of the set, bootstrapping was performed. This bootstrapping was performed using a script written in MATLAB (see A.2). From the bootstrapped results, the probability distribution function (PDF) and the cumulative distribution function (CDF) can be plotted.

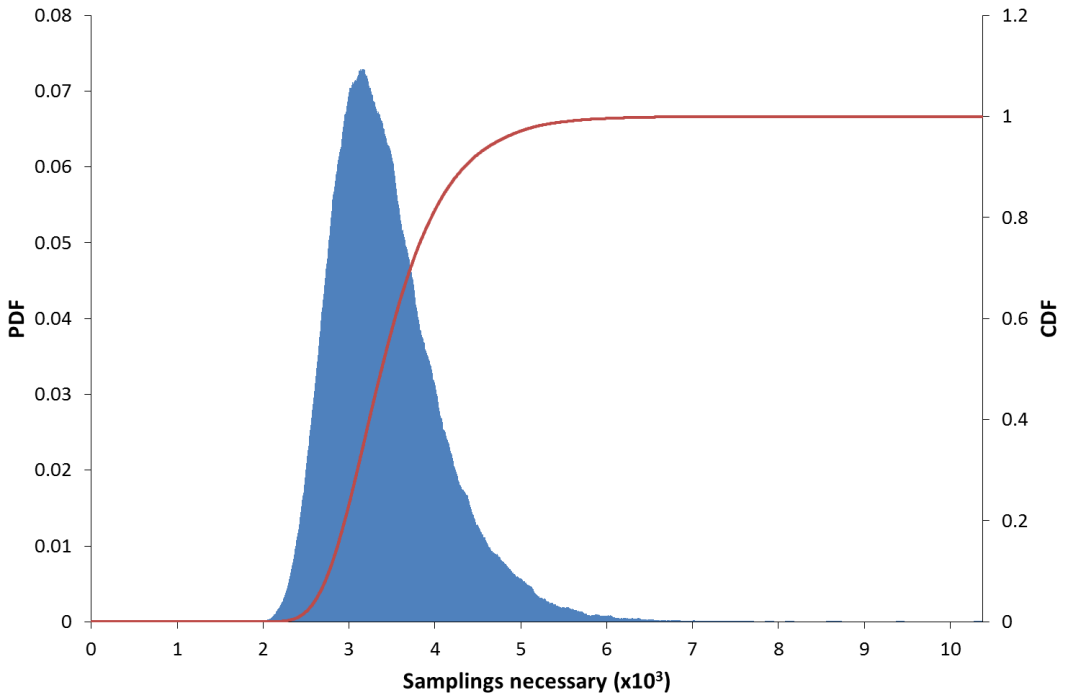


Figure 8.3: Bootstrapped PDF (solid blue histogram) and CDF (red line) of number of samplings required to select every member of a set of 512 sequences at least once based purely upon random selection. Note that the CDF approaches but never reaches 1 and that the most likely number of samplings necessary is equal to the mode of the PDF.

In figure 8.3, the PDF and CDF the represents the likelihood that, having sampled a specific number of times, the whole population would have been isolated. The CDF is asymptotic to a probability of 1: i.e. It is necessary to sample an infinite number of times to be *certain* of having sampled the whole population. Using the CDF, the number of samplings required to be 95% certain of having sampled the population is determined to be 4725. Random mutations in a sequence mean that each sequence is likely to need to be synthesised more than once. A representative error rate and the overall sequence length are required to estimate the effect on the sampling described.

8.2. Optimisation of the Gao assembly protocol

The Gao assembly process was developed by a group from the University of Houston led by Xiaolin Gao and is described in section 4.4.3. Briefly, the Gao assembly process consists of several distinct steps: The oligonucleotide sequences are designed and synthesised. The synthesised oligonucleotides are cleaved from the chip and amplified by ligator PCR. The generic oligonucleotide primers are cleaved off. Finally, the assembly of the full length sequence is attempted from the amplified constituent oligonucleotides. See figure 4.4 for a summary of this process.

8.2.1. Gao assembly sequence optimisation

The Gao lab used several pieces of software to produce the oligonucleotide sequences prior to synthesis. The primary software was called SeqZego, written by Dr. Rafal Debrek. At the time of writing, this software was unpublished and not freely available. The primary function of SeqZego was to produce overlapping oligonucleotides. A range for the number of oligonucleotides and their lengths could be selected. Selecting incompatible values, however, produced an error. By exploring the additional functions of the SeqZego program it was noted that checking the “Tm optimisation” box would result in truncated oligonucleotides being presented in the output. Since such oligonucleotides would not be

able to produce successful assembled sequences, the Tm optimisation offered by SeqZego was not utilised. The cut positions and overlaps were, therefore defined arbitrarily. The wide variation of overlap melting temperatures that resulted from this sequence approach was not considered not be an issue by members of the Gao lab.

The generic primer sequences that are added to each of the oligonucleotides can also be added by SeqZego. Once the oligonucleotide sequences were determined the output was written to files readable by Microsoft Excel. A script was written in MATLAB to confirm whether these oligonucleotides were indeed capable of forming correct sequences. The script tried to reconstruct each of the input sequences from the oligonucleotides present in the output. This script would produce a descriptive error whenever a sequence could not be assembled.

8.2.2. Oligonucleotide mixture amplification

The first step in the Gao assembly procedure is the one-pot amplification of all the oligonucleotides cleaved from the synthesis chip. The cleavage of the synthesis chip was performed as described in section 4.4.4. Once cleaved, all the synthesised oligonucleotides from the synthesis chip are present in 10 – 20 µL of deionised water. Each oligonucleotide is thought to be present at the femtomolar level. To obtain sufficient oligonucleotides for the assembly procedure, the cleaved oligonucleotide mixture is amplified by PCR. The synthesis work was performed at the University of Houston, US. The cleaved oligonucleotides were then assembled, amplified and assessed at the University of Warwick, UK.

Temperature during in the primer annealing step of the PCR has an effect on the amount of PCR product produced. To test this relationship a gradient PCR was prepared where the annealing temperature was varied between 45 and 55°C.

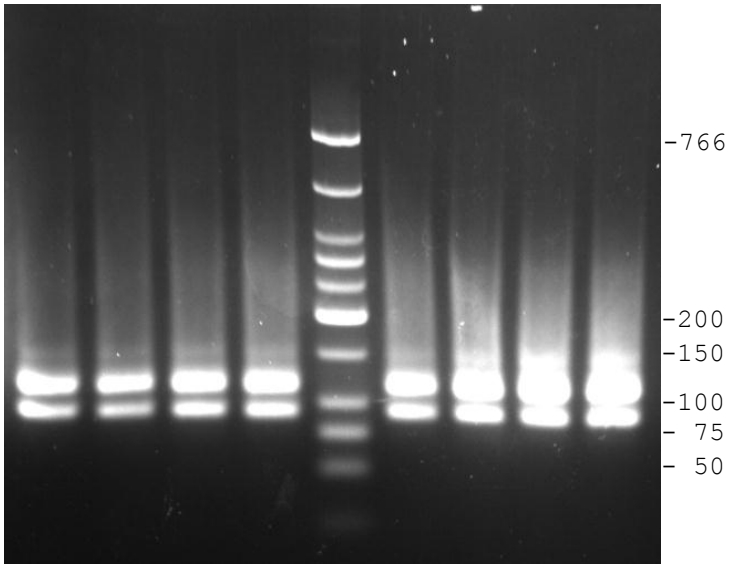


Figure 8.4: Effect of changing annealing temperature on oligonucleotide amplification. The annealing temperatures at 45, 47, 48 and 49°C in lanes 1-4 and 51, 52, 53 and 55°C in lanes 6-9.

The oligonucleotide primer melting temperature is estimated to be 50°C and effectiveness of the PCR when the temperature of the annealing step was increased over 50°C. Figure 8.4 shows that the temperature of the annealing step does not reduce the amount of oligonucleotide amplified at each temperature. The other synthesised CRM yielded similar results to the results shown here. The two expected band sizes are 123 and 84 base pairs. There is a weak band visible that is nearest in size to the 150 base pair marker. This band was unexpected and was not part of the synthesis. Furthermore, there is a significant high molecular weight smear in each lane that could indicate a non-specific amplification.

8.2.3. Oligonucleotide mixture digest

The first step of the Gao assembly process once the cleaved chip oligonucleotides have been amplified is to digest off the short oligonucleotide primer sequences that were used in the amplification of the oligonucleotides. The oligonucleotide primers sequences contained a recognition site for a type IIS restriction enzyme, in this case MlyI. The recognition site was placed so that the cut site would excise the complete oligonucleotide primer sequence.

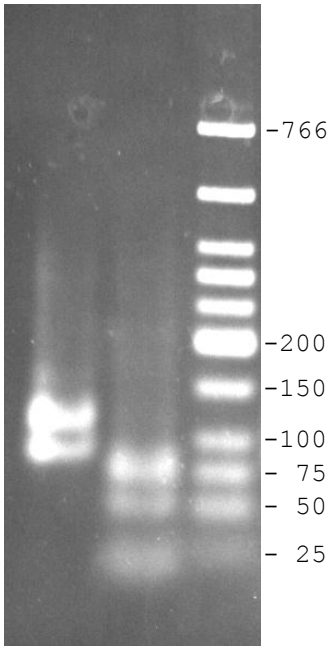


Figure 8.5: Gel of digested oligonucleotide cleavage mixtures from the Gao synthesis chips for two CRMs. Lane 1 contains a sample of the undigested template DNA whereas lane 2 contains a sample of the template DNA digested with MlyI.

The oligonucleotides from the amplification PCR were digested according to the protocol described in section 4.4.5. Figure 8.5 shows the oligonucleotide mixture before and after digestion for two amplified oligonucleotide cleavage products. The successful reaction can be seen by comparing lane 1 and 2: Before digestion with MlyI there are two higher molecular weight bands which correspond to the long and short possible oligonucleotide sequences (see figure 8.5). After digestion, the two higher molecular weight bands have disappeared and been replaced by two lower molecular weight bands. Additionally, a new very low molecular weight band can now be seen in the digested lane. This new band corresponds to the primers sequences that have been cut off from the larger oligonucleotide sequences. The expected band sizes are 123 and 84 before digest which convert to bands of 77 and 53 base pairs respectively after digest. The cleaved oligonucleotide primer sequence is 15 base pairs in length.

8.2.4. Gao Assembly PCR

The oligonucleotides are ready to undergo ligative assembly once the oligonucleotide primers have been fully digested. Using a restriction enzyme that leaves only blunt ends, such as MlyI, and a ligase that only ligates overlapping or nicked strands, such as *Taq* ligase, means that removal of the cut oligonucleotide primers from the reaction mixture before ligation is unnecessary. The ligation of the digested oligonucleotide mixtures was carried out as described in the methods (see section 4.4.6). Once the ligation reaction was complete, an aliquot was taken and a PCR amplification along the full length sequence.

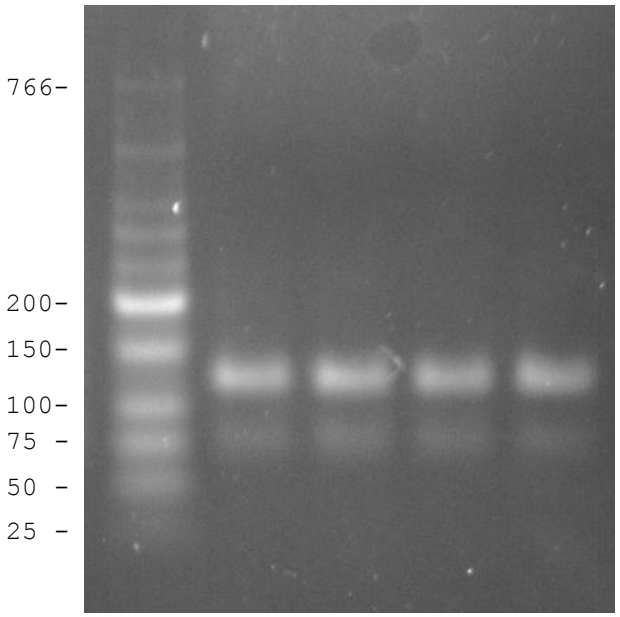


Figure 8.6: Gel of amplified assembly reaction products. The basic assembly reaction protocol was followed. A gradient annealing temperature step was performed in the amplification. Lane 1: Ladder. Lane 2-5: 45, 50, 55, 60°C annealing temperature.

Figure 8.6 shows gel which exemplifies the output from the Gao assembly PCR after ligation of the digested oligonucleotides. There are no bands of the expected size (335 base pairs) in the gel shown on figure 8.6. It does not appear that the assembly of the oligonucleotides into full length sequences was successful. The bands that are visible, however, are of different size than the bands present in the cleaved oligonucleotide mixture that was the template for the amplification. Indicating that some form of assembly other than what was

intended has occurred. Furthermore this assembly was, at least partially, able to be amplified in the ensuing PCR. The band sizes are approximately the same as the undigested oligonucleotide cleavage mixture, as seen in lane 1 of figure 8.5, indicating that the failure could have been simply due to the wrong substrate being used. Repeated experiments, however, with the freshly digested oligonucleotide cleavage mixture yielded similar results.

No separation of the digested ends from the digested oligonucleotide cleavage mixture could result in the simple religation of the ends in the assembly reaction. As *Taq* ligase only ligates blunt ended fragments, however, this is unlikely. Incompletely synthesised strands could result in strands that do not possess the binding site for the restriction enzymes, but do possess sufficient overhang for the binding of ends digested from fully synthesised strands. It is possible that these ends were able to catalyse PCR on the incompletely synthesised strands, resulting in a re-emergence of the complete, undigested-like sequence after the reaction. This process, however, would need to be the dominant reaction in order to amplify exponentially. Since the first step after the cleavage of the chip is to amplify the chip oligomix, the vast majority of the strands are likely to complete as they result from the amplification reaction, not the chip synthesis which is likely to be incomplete. This side reaction is, therefore, unlikely to have taken place. The alternative explanation, that the major products of the reaction are simply coincidentally approximately the same size as other sequences used previously, becomes the most likely explanation.

8.2.5. Changing concentration of oligomix

The failure of several attempted assembly reactions, as described in section 8.2.4, implied that the ligative or PCR steps were at fault. In the following section, the process whereby the Gao assembly ligation reaction conditions were optimised is described. During this process, the key parameters that were varied are the concentration of the oligonucleotides

in the assembly reaction, the number and type of temperature cycling that the ligation mixture undergoes. The results shown here are a selection from the multiple assembly and amplification reactions that were performed.

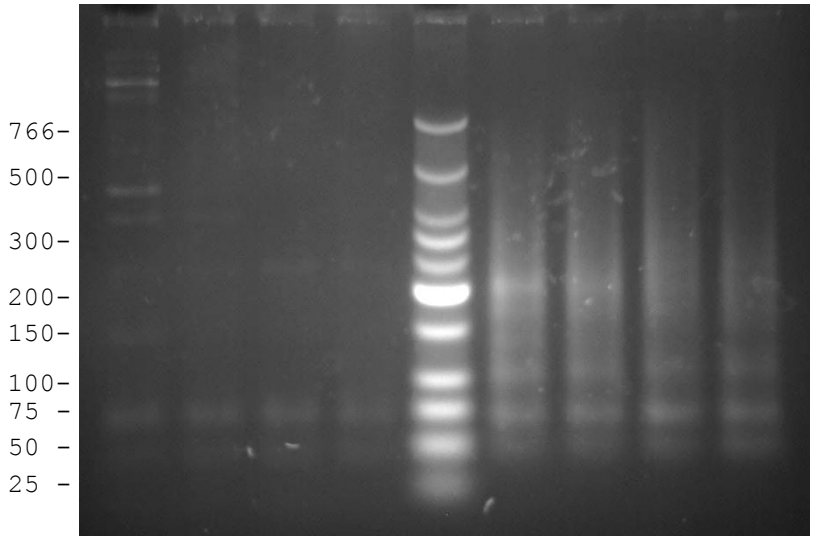


Figure 8.7: Comparison of oligonucleotide concentration on the Gao assembly reaction. The four left hand lanes are all at a high substrate concentration whilst the four right hand lanes are all at a low substrate concentration.

Figure 8.7 compares high and low concentrations of the oligonucleotide mixture template. As described in section 4.4.6, the high concentration is approximately 4 µg template, whereas the low concentration is approximately 0.5 µg template. Interestingly, a lower oligonucleotide concentration results in the presence of more bands in each lane. It is possible that each band is amplified from a partially ligated template that was not fully ligated. The amplification uses primers specific to sequences on either end of the full length assembled sequence. Amplification by these primers, therefore, indicates that the oligonucleotides in the middle of the assembly are not selective as there exist several ways of getting from the forward to the reverse primers. It is possible that the primers were ar fault. The primers were obtained from the Gao lab and the same primers were used throughout. The oligonucleotide amplification primers, however, worked consistently (see section 8.2.3) and were synthesised using the same method.

The presence of more bands in the lower substrate concentration conditions could be due to the lack of inhibition of the ligation reaction by truncation mutants or otherwise unligatable oligonucleotides. Despite the presence of more bands, no strong bands are observed. The bands can be seen in lane 1 (high concentration) is probably due to a contamination of the PCR mixture in this lane.

8.2.6. Changing the conditions of the post-assembly PCR

After initial difficulty with the assembly, the paper by Smith *et al.*[39] was used as a source of an assembly method, see section 4.4.7 for description of method. The effect of the use of an alternative assembly method can be seen in figure 8.8.

The assembly reaction mixture was put through a series of temperature cycles during the ligation step. At the start, the solution is heated to a temperature sufficient to melt all DNA double strands in the reaction mixture. The mixture is then slowly cooled and, ideally, each oligonucleotide would bind to its complementary partner(s). A portion of the oligonucleotides, however, will bind to partially complementary or mutated partners that are inevitably present in the reaction. By cooling the mixture slowly, and using a high temperature ligase, the most complementary oligonucleotides anneal and care ligated first. The improperly pairs oligonucleotides are then given successive opportunities to bind to their intended partners by cycling the reaction mixture from a high melting temperature (95°C), through a slow cool (over ~15 minutes) to a baseline temperature (~40°C). The probability that mutated oligonucleotides are incorporated into the assembled sequence is reduced by maintaining a relatively high baseline temperature, above which these oligonucleotides will not significantly anneal.

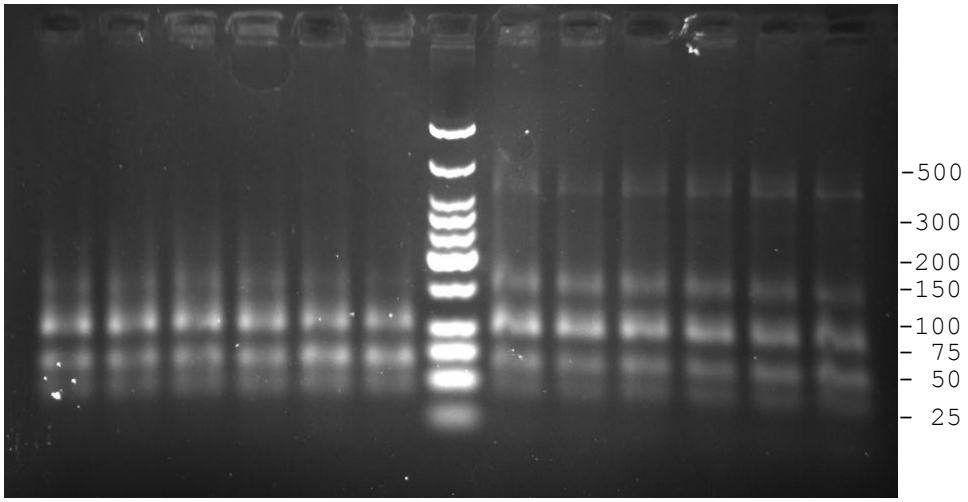


Figure 8.8: Gel of amplified assembly reaction for two CRMs. The temperature cycling used in the protocol was modified from the Gao assembly cycling protocol in this experiment. Additionally a gradient was applied during the PCR reaction: The annealing temperature of lanes 1-6 and 8-13 were 45, 48, 50, 55, 60, 65°C respectively. The lanes on the left side of the ladder were run with a high oligonucleotide concentration whilst the lanes on the right hand side of the marker were run with a low oligonucleotide concentration.

A range of primer annealing temperatures was employed in the reaction shown in figure 8.8 to whether the issue was with the PCR step rather than the assembly step, which had been modified thus far. It was expected that as the melting temperature increases, the primers would not be able to anneal to the template and the amount of product would decrease. As can be seen in figure 8.8, however, there are bands present in every lane. This observation would seem to indicate that the PCR using the full assembly primers is having no effect on the products and that the bands are present from prior to the amplification or that some component of the template mixture is able to prime even at temperatures >65°C. Several high molecular weight bands can be seen in figure 8.8 indicating that the assembly has been at least partially successful. None of these bands, however, are of the expected size of 335 base pairs. Multiple, spurious bands are regularly reported in this type of parallel, one-pot gene assembly[35].

To identify what had been amplified several of these bands were excised, purified and cloned into TOPO vector. The resulting plasmids were sequenced. The resulting

sequencing traces were uniformly unreadable. It is likely that the band is comprised of several distinct sequences that would make sequencing accurately impossible.

8.2.7. Use of alternative primers

The use of full length sequence primers necessitates that sufficient copies (10^5) of the full length sequence were successfully ligated. The sequencing data from the bands that could be synthesised indicated that a portion of the internal sequence had been assembled correctly. Furthermore, the primers used for the full length sequence amplification step were obtained from the Gao lab. These primers were of an artificial sequence and so there was no positive control to ensure that the primers themselves were working and able to amplify sequences that possessed the appropriate sequences. The assembled sequence, however, was natural and so a set of primers was obtained to an internal portion of the sequence.

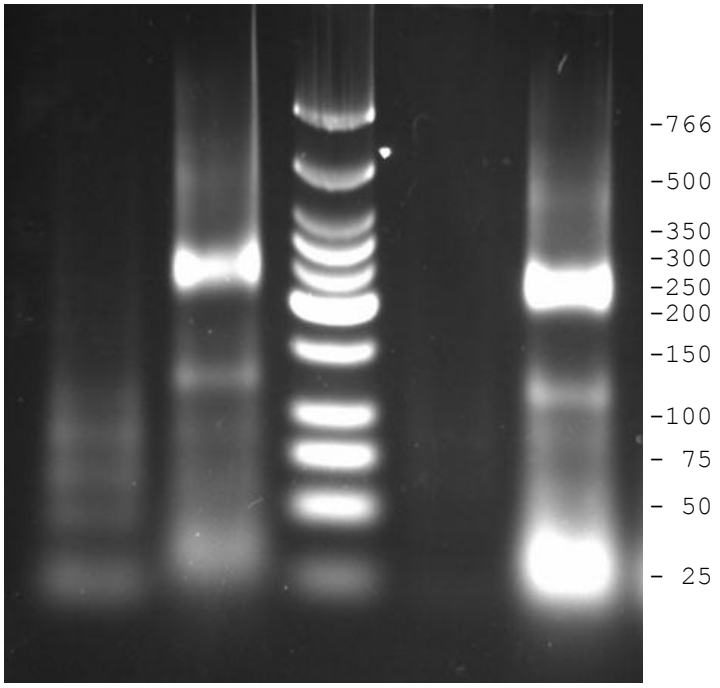


Figure 8.9: Gel of amplification products comparing old and new primers. Unassembled (lane 1 and 4) and assembled (lane 2 and 5) template using high (left hand pair) and low (right hand pair) template concentrations.

The set of primers used in the amplifications shown in figure 8.9 are specific to internal sequences within the CRM sequence being amplified. Clear, strong bands can be observed at the expected product size of 239 base pairs. This presence of this single strong band indicates that the assembly was at least partially successful; the middle section of the sequence had been ligated and was able to be amplified. Weaker bands can also be seen at lower molecular weights as well as significant smears at both higher and lower molecular weights.

To determine which of the ends of the assembled sequence had failed to ligate and/or which if the original primers were at fault a PCR was performed using different combinations of the primers. For simplicity, the original full length primers were called P1f and P1r and the new set of primers was called P2f and P2r. Four possible combinations of the two pairs of primers were used to amplify a typical ligation reaction mixture. The gel of the PCR products is shown in figure 8.10.

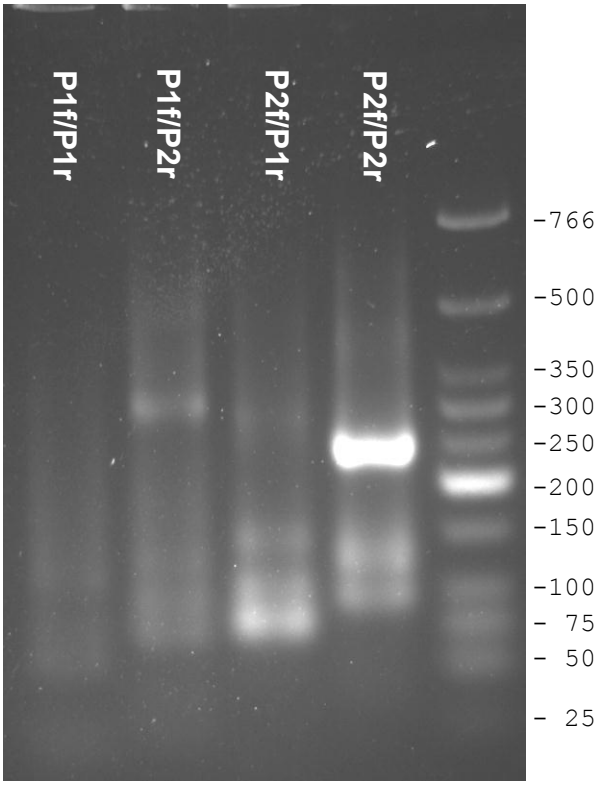


Figure 8.10: Amplification of Gao ligation products using internal/external primer combinations. Note the single, strong band in lane 4.

There is a single strong band present in lane 4 of figure 8.10. This band corresponds to the internal portion of the CRM which is being successfully amplified by the new primer pair. Additional weaker bands are visable which could be indicative of a competative binding interaction between two oligonculeotides present in the oligonucleotide mixture. Furthermore, there are weak bands in each of lanes 2 and 3. These weaker bands correspond to the expected sizes of the corresponding products: P1f/P1r, 335 base pairs. P1f/P2r, 296 base pairs. P2f/P1r, 278 base pairs. P2f/P2r, 239 base pairs.

This result appears to indicate that it was the fragment amplification primer was at fault during the preceeding optimisation process. It is possible that the primers provided by the Gao lab were unsuitable for the amplification of the full length sequence. The presence of weak bands in lanes 2 and 3 of figure 8.10 indicate that these primers are capable of producing a small amount of product of the appropriate size. To test this, new versions of

the Gao-sunthesised primers could have been ordered and used in the amplification of the full length assembly. Due to constraints of time, however, this was not performed.

The amplification product (strong band in lane 4) of the shorter primer pairs (239 bp) was isolated by gel purification and TOPO cloned. The resulting plasmid was then sequenced with TOPO specific primers. An example trace can be seen in figure 8.11.

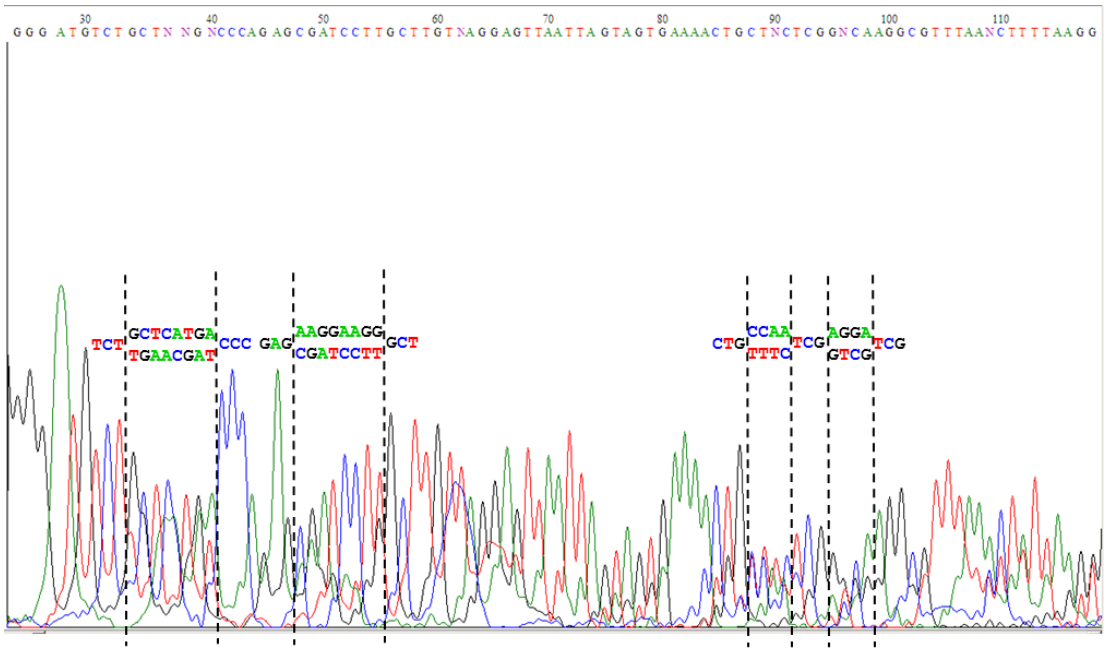


Figure 8.11: Sequence traces indicating partially successful assemblies of the Gao oligonucleotides. Four mutation sites are indicated by the overlay. The two alternative sequences (wild-type and mutated) are shown for the mutation sites.

Whilst the sequencing quality was generally poor, the sequence trace shown in figure 8.11 was obtained. Several mutation sites can be identified in the sequence trace. Generally the two highest traces correspond to the two expected bases at a given position of the mutation. Outside of the identified mutation sites the trace is generally clearer with single strong traces being observed. However, there are cases outside of the mutation sites where there are significantly high secondary traces. It is difficult to draw any conclusions from the available sequences beyond the fact that a portion of the several of the variants has been successfully assembled and amplified.

8.2.8. Reasons for failure of the Gao assembly

During the oligonucleotide sequence selection step, before any oligonucleotide selection has taken place, the Gao lab use several programs to attempt to determine whether there are likely to be any significnat competative binding events. Although the program did identify likely interactions, the melting point of these interactions signifncantly below the intended interactions. It is possible that an unidentified binding interaction, such as a hairpin loop, could have occurred that sequestered the oligonucleotides from the ends of the sequence away from the ligating strand, thus preventing the full length sequence from successfully assembling. Interactions of this type could be avoided by assessing the potential of oligonucletoides to form hairpin loops during optimisation. The method of competitor identification, described in section 7.5, could be modified for this task. Once identified, the lengths of oligonucleotides with a significant chance of forming hairpin loops could be changed such that cut sites appear within the looping region.

The synthesis chip contained 4096 wells in a roughly square configuration. The light that deblocks a given well is focussed in the centre of the chip, meaning that the light becomes progressively less focussed towards the periphery of the chip. Members of the Gao lab recommended that all the sequences to be synthesised should be placed as close to the middle of the chip as possible, avoiding the perphery, for this reason. The synthesis chip did, however, possess quality control sequences in the periphery wells to which a tagged probe was bound to ensure that synthesis had occurred successfully even in this peripheral wells. The quality control sequences bound their probes, indicating that synthesis was successful right up to the periphery of the chip.

The quality of the oligonucleotides produced by the Gao synthesis method is not quantified. No purification was performed on the chip oligonucleotides prior to amplification and assembly. It is possible that the presence of truncation mutants inhibited

the Gao assembly. Oligonucleotides synthesised close to the periphery of the chip are more susceptible to mutation as light used to direct the synthesis reaction is less focussed than in the centre of the chip. As a result, correct full length examples of some sequences will not be present in equimolar concentrations as others.

In order to synthesise the 512 member library, two strands of each sequence are required. Each strand was specified to be split into 6 oligonucleotides, which implies a total of 6144 oligonucleotides. Since the chip only contained a maximum of 4096 wells, some oligonucleotides, the oligonucleotides that occur in every sequence were not repeatedly synthesised. The resulting oligonucleotide mixture, therefore, did not contain each a molar ratio of each oligonucleotide equal to frequency with which that oligonucleotide would be required for the assembly of every sequence. This lack of equimolarity in the original oligonucleotide mixtures is likely to have had a detrimental affect on the success of the assembly reaction.

The quality of the SeqZego optimisation that the oligonucleotide sequences were obtained from was quite poor. Figure 7.11 (left) shows a histogram of the melting temperature of all the overlaps in the oligonucleotide set. The melting temperatures range from 54 to 70°C with a fairly uniform distribution indicating that the oligonculeotides will assemble across a wide range of temperatures. This lack of optimisation could mean that the higher melting temperature sequences are either annealed non-selectively at lower temperatures leading to improper or truncated sequences being produced.

8.3. Opticut assembly

Due to the difficulty in obtaining full length sequence products from the Gao gene assembly method an alternative gene assembly method was developed. Fully optimised individual oligonucleotides were to be bought from commercial sources. The advantages of taking this change in approach were as follows:

1. Oligonucleotides would be of known quantity, quality and concentration, allowing precise equimolar mixing of the assembly components.
2. Oligonucleotide overlap sequences could be optimised using OptiCut and therefore exhibit a smaller range of melting temperatures.
3. The OptiCut assembly was multipot, requiring no downstream separation or sampling.

Buying oligonucleotides of known purity meant that all the necessary oligonucleotides were present in the assembly solution at a sufficient concentration was known with a high degree of confidence relative to the Gao method. Furthermore, because the oligonucleotides used in the assembly reaction did not need to be amplified prior to use there is less chance of mutations being introduced as a result of the PCR amplification. This reduced need for amplification was expected to be observed in the error rate of the final sequence.

8.3.1. OptiCut optimised oligonucleotide assembly

The assembly of the OptiCut optimised oligonucleotides was comparatively much simpler than the Gao assembly. A simple slow cool with several cycles was employed for the ligation reaction as described in the methods section 4.4.7. Once complete, an aliquot of the assembly reaction was sampled amplified using specific primers to the full length assembly. The gel of the product of several such reactions can be seen in figure 8.12.

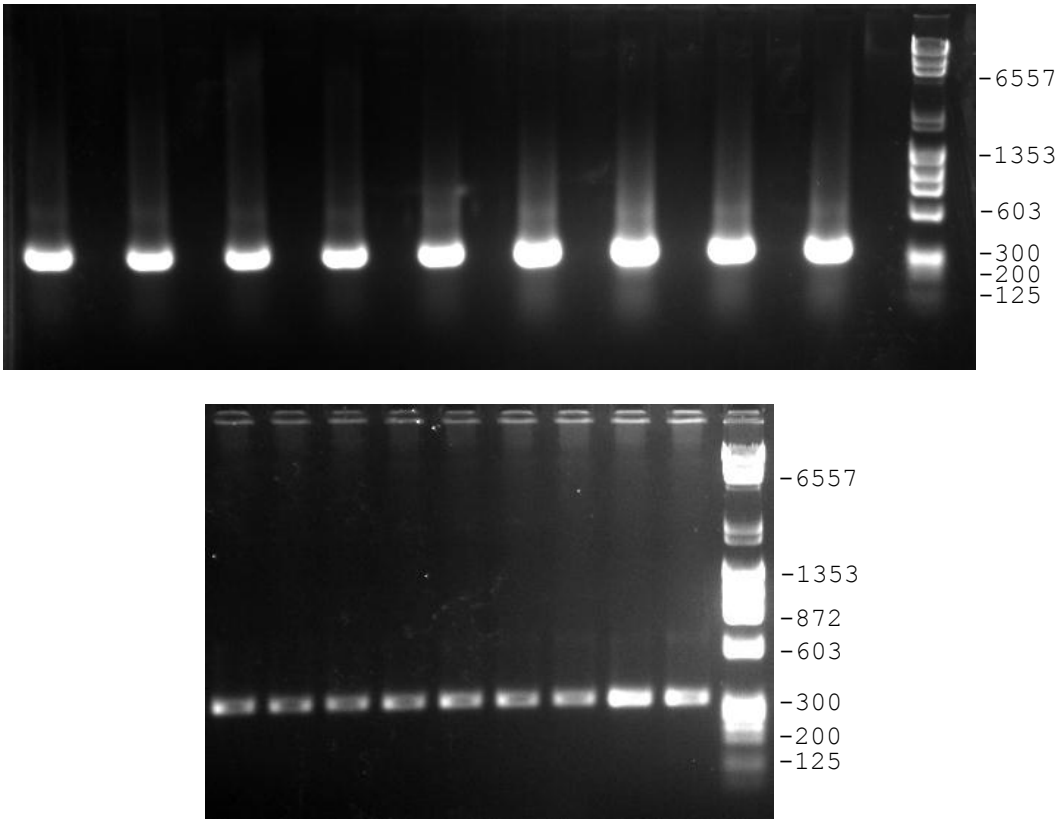


Figure 8.12: Gel showing successful amplification of full length oligonucleotide assemblies (top pane) and efficiency of the purification protocol for the same (bottom pane).

In figure 8.12, each band in the top pane is the whole of a PCR volume for a unique sequence from the combinatorial mutant library whereas each band in the bottom pane is an aliquot of the gel purification of the same band. A single strong band at the expected product size of 327 base pairs is observed. An additional, significantly weaker band can also be seen in each lane. Care was taken during the gel purification to avoid excising the weak band with the strong band. Interestingly, despite this care the weak band still reappeared in the purification check gel shown in the bottom pane of figure 8.12.

8.3.2. Use of high fidelity DNA polymerase

The DNA polymerase used in the amplification PCR was *Taq* Gold. This polymerase is for use with a wide variety of amplicons and boasts a low level of false positive results. Phusion DNA polymerase is a high fidelity polymerase which is specifically for long amplifications or when accuracy is particularly important. The high fidelity is due to the proof reading

capability of the Phusion DNA polymerase, if an incorrect base is added to the nascent strand then the polymerase process is halted allowing the incorrect base to be excised and the replication process to begin again. The manufacturer of Phusion DNA polymerase claim an error rate of the 1 per 4.4 x 10^7 base pairs.

To test whether the Phusion DNA polymerase could be used to amplify the ligated oligonucleotide mixtures a direct comparison was performed. Two oligonucleotide mixtures were prepared and ligated. A sample of each was then amplified by either *Taq* Gold or Phusion DNA polymerase. Other than the buffers and the polymerases the conditions of each pair of amplifications was identical.

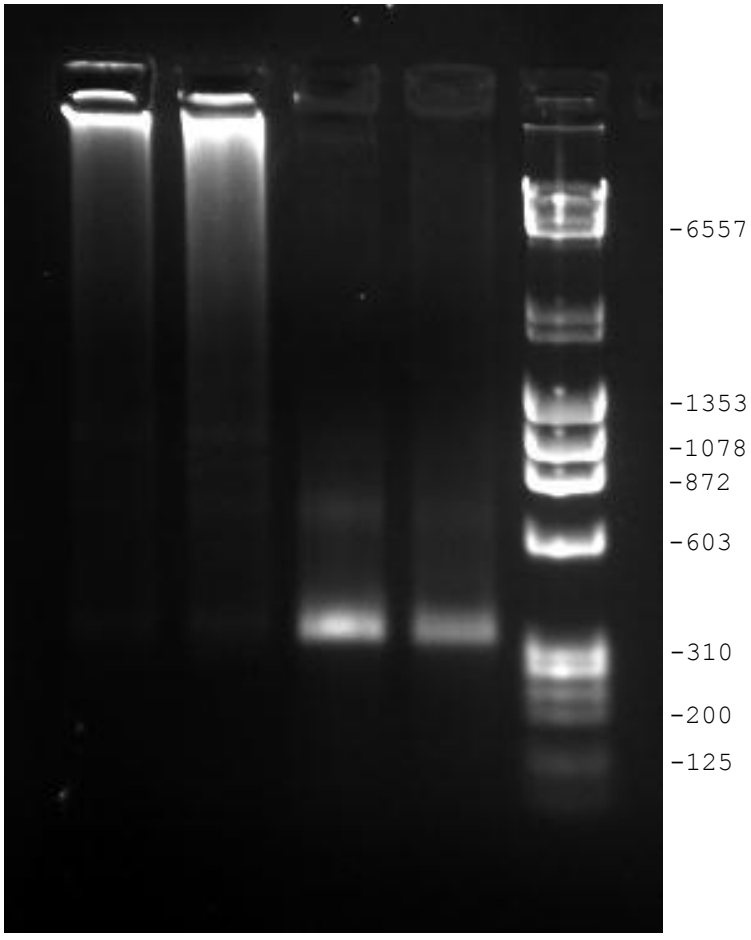


Figure 8.13: Gel of amplified assembly products. Comparison of polymerase types: Lanes 1 and 2 were amplified using high fidelity phusion polymerase whereas lanes 3 and 4 were amplified using *Taq* Gold polymerase.

Figure 8.13 shows that high levels of the final product were only produced by the *Taq* Gold polymerase. Weak bands can also be seen in the Phusion DNA polymerase lanes, but the majority of the DNA appears to be in the high molecular weight smear towards the top of the gel.

8.3.3. Demonstration of the necessity of the ligation step

The necessity of the ligase step is clearly indicated in figure 8.14. An oligonucleotide mixture was prepared and split into two. One part received the *Taq* ligase whilst the other received no enzyme. Both mixtures were then put through the ligation procedure and samples taken for amplification. It is entirely possible that an effect like that observed with TBIO[10] and other PCR based DNA assembly methods might be taking place. The DNA may not be fully assembling the ligase step and instead being assembled by the PCR step.

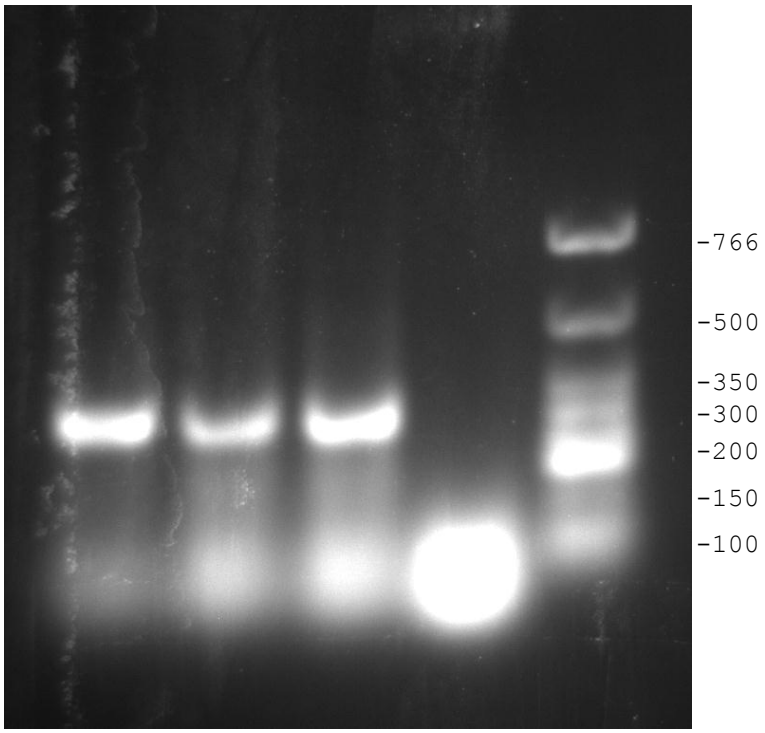


Figure 8.14: Gel of products from the amplification PCR. Lanes 1-3 had ligase added to the assembly reaction whereas lane 4 did not. All other treatments were the same.

Figure 8.14 clearly shows the necessity of the ligase in the two step assembly. Without the ligase the specific, high molecular weight band is not observed. Furthermore, the fuzzy, low

molecular weight band of the remaining assembly oligonucleotides and PCR primers is much stronger than in the other lanes (comparison of lane 4 and lanes 1-3 of figure 8.14). The strength of this band is due to the overlapping presence of unassembled oligonucleotides and PCR primers.

8.3.4. Assembly of whole CRM-B library

The entire of the 512 sequence CRM-B library was assembled using the OptiCut optimised oligonucleotide set. The assembly was carried out in 96-well qPCR plates as described in section 4.4.10. Samples of stored plates, sealed and cooled to 4°C, have been used as template for successful ligation reactions 4 months after the initial assembly and could be viable beyond this time frame. This could be due to the fact that the assembly is a ligase- rather than polymerase dependant reaction. High fidelity DNA polymerases have a 3' to 5' exonuclease proofreading activity that could result in degradation of the assembled sequences during storage. *Taq* ligase, by contrast, does not have any exonuclease activity and sequences can, therefore, be expected to have a longer lifetime in storage without purification.

Each assembled member of the combinatorial was assigned a global identifying number from 1 to 512. In addition, each sequence was associated with a string of 1's and 0's that identified which sites were mutated. For instance, '#67 110001101' denotes sequence number 67, which possesses mutated binding sequences at site positions 1, 2, 6, 7 and 9 and wild type binding sequences at site positions 3, 4, 5 and 8. Table 8.1 shows every member of the CRM-B mutant library that has been assembled to date (all 512 members) as they occur in the qPCR plates in which they were assembled. The order of each sequence in the library shown in table 8.1 is determined by the order in which that sequence's binary code occurs in a 9-bit reflected binary code (also known as 'Gray code'[40]

or a ‘Hamiltonian cycle’). This order was produced as a result of the MATLAB function ‘ff2n’ that was used to generate the library sequences.

Plate1

	1	2	3	4	5	6	7	8	9	10	11	12
A	#1 000000001	#2 100000001	#3 110000001	#4 010000001	#5 011000001	#6 111000001	#7 101000001	#8 001000001	#9 001100001	#10 101100001	#11 111100001	#12 011100001
B	#13 010100001	#14 110100001	#15 100100001	#16 000100001	#17 000110001	#18 100110001	#19 110110001	#20 010110001	#21 011110001	#22 111110001	#23 101110001	#24 001110001
C	#25 001010001	#26 101010001	#27 111010001	#28 011010001	#29 010010001	#30 110010001	#31 100010001	#32 000010001	#33 000011001	#34 100011001	#35 110011001	#36 010011001
D	#37 011011001	#38 111011001	#39 101011001	#40 001011001	#41 001111001	#42 101111001	#43 111111001	#44 011111001	#45 010111001	#46 110111001	#47 100111001	#48 000111001
E	#49 000101001	#50 100101001	#51 110101001	#52 010101001	#53 011101001	#54 111101001	#55 101101001	#56 001101001	#57 001001001	#58 101001001	#59 111001001	#60 011001001
F	#61 010001001	#62 110001001	#63 100001001	#64 000001001	#65 000011001	#66 100011001	#67 110011001	#68 010011001	#69 011001101	#70 111001101	#71 101001101	#72 001001101
G	#73 001101101	#74 101101101	#75 111101101	#76 011101101	#77 010101101	#78 110101101	#79 100101101	#80 000101101	#81 000111101	#82 100111101	#83 110111101	#84 010111101
H	#85 011111101	#86 111111101	#87 101111101	#88 001111101	#89 001011101	#90 101011101	#91 111011101	#92 011011101	#93 010011101	#94 110011101	#95 100011101	#96 000011101

Plate2

	1	2	3	4	5	6	7	8	9	10	11	12
A	#97 000010101	#98 100010101	#99 110010101	#100 010010101	#101 011010101	#102 111010101	#103 101010101	#104 001010101	#105 001110101	#106 101110101	#107 111110101	#108 011110101
B	#109 010110101	#110 110110101	#111 100110101	#112 000110101	#113 000100101	#114 100100101	#115 110100101	#116 010100101	#117 011100101	#118 111100101	#119 101100101	#120 001100101
C	#121 001000101	#122 101000101	#123 111000101	#124 011000101	#125 010000101	#126 110000101	#127 100000101	#128 000000101	#129 000000111	#130 100000111	#131 110000111	#132 010000111
D	#133 011000111	#134 111000111	#135 101000111	#136 001000111	#137 001100111	#138 101100111	#139 111100111	#140 011100111	#141 010100111	#142 110100111	#143 100100111	#144 000100111
E	#145 000110111	#146 100110111	#147 110110111	#148 010110111	#149 011110111	#150 111110111	#151 101110111	#152 001110111	#153 001010111	#154 101010111	#155 111010111	#156 011010111
F	#157 010010111	#158 110010111	#159 100010111	#160 000010111	#161 000011111	#162 100011111	#163 110011111	#164 010011111	#165 011011111	#166 111011111	#167 101011111	#168 001011111
G	#169 001111111	#170 101111111	#171 111111111	#172 011111111	#173 010111111	#174 110111111	#175 100111111	#176 000111111	#177 000101111	#178 100101111	#179 110101111	#180 010101111
H	#181 011101111	#182 111101111	#183 101101111	#184 001101111	#185 001001111	#186 101001111	#187 111001111	#188 011001111	#189 010001111	#190 110001111	#191 100001111	#192 000001111

Continued overleaf

Plate3

	1	2	3	4	5	6	7	8	9	10	11	12
A	#193 000001011	#194 100001011	#195 110001011	#196 010001011	#197 011001011	#198 111001011	#199 101001011	#200 001001011	#201 001101011	#202 101101011	#203 111101011	#204 011101011
B	#205 010101011	#206 110101011	#207 100101011	#208 000101011	#209 000111011	#210 100111011	#211 110111011	#212 010111011	#213 011111011	#214 111111011	#215 101111011	#216 001111011
C	#217 001011011	#218 101011011	#219 111011011	#220 011011011	#221 010011011	#222 110011011	#223 100011011	#224 000011011	#225 000010011	#226 100010011	#227 110010011	#228 010010011
D	#229 011010011	#230 111010011	#231 101010011	#232 001010011	#233 001110011	#234 101110011	#235 111110011	#236 011110011	#237 010110011	#238 110110011	#239 100110011	#240 000110011
E	#241 000100011	#242 100100011	#243 110100011	#244 010100011	#245 011100011	#246 111100011	#247 101100011	#248 001100011	#249 001000011	#250 101000011	#251 111000011	#252 011000011
F	#253 010000011	#254 110000011	#255 100000011	#256 000000011	#257 000000010	#258 100000010	#259 110000010	#260 010000010	#261 011000010	#262 111000010	#263 101000010	#264 001000010
G	#265 001100010	#266 101100010	#267 111100010	#268 011100010	#269 010100010	#270 110100010	#271 100100010	#272 000100010	#273 000110010	#274 100110010	#275 110110010	#276 010110010
H	#277 011110010	#278 111110010	#279 101110010	#280 001110010	#281 001010010	#282 101010010	#283 111010010	#284 011010010	#285 010010010	#286 110010010	#287 100010010	#288 000010010

Plate4

	1	2	3	4	5	6	7	8	9	10	11	12
A	#289 000011010	#290 100011010	#291 110011010	#292 010011010	#293 011011010	#294 111011010	#295 101011010	#296 001011010	#297 001111010	#298 101111010	#299 111111010	#300 011111010
B	#301 010111010	#302 110111010	#303 100111010	#304 000111010	#305 000101010	#306 100101010	#307 110101010	#308 010101010	#309 011101010	#310 111101010	#311 101101010	#312 001101010
C	#313 001001010	#314 101001010	#315 111001010	#316 011001010	#317 010001010	#318 110001010	#319 100001010	#320 000001010	#321 000011110	#322 100001110	#323 110001110	#324 010001110
D	#325 011001110	#326 111001110	#327 101001110	#328 001001110	#329 001101110	#330 101101110	#331 111101110	#332 011101110	#333 010101110	#334 110101110	#335 100101110	#336 000101110
E	#337 000111110	#338 100111110	#339 110111110	#340 010111110	#341 011111110	#342 111111110	#343 101111110	#344 001111110	#345 001011110	#346 101011110	#347 111011110	#348 011011110
F	#349 010011110	#350 110011110	#351 100011110	#352 000011110	#353 000010110	#354 100010110	#355 110010110	#356 010010110	#357 011010110	#358 111010110	#359 101010110	#360 001010110
G	#361 001110110	#362 101110110	#363 111110110	#364 011110110	#365 010110110	#366 110110110	#367 100110110	#368 000110110	#369 000100110	#370 100100110	#371 110100110	#372 010100110
H	#373 011100110	#374 111100110	#375 101100110	#376 001100110	#377 001000110	#378 101000110	#379 111000110	#380 011000110	#381 010000110	#382 110000110	#383 100000110	#384 000000110

Continued overleaf

Plate5	1	2	3	4	5	6	7	8	9	10	11	12
A	#385 000000100	#386 100000100	#387 110000100	#388 010000100	#389 011000100	#390 111000100	#391 101000100	#392 001000100	#393 001100100	#394 101100100	#395 111100100	#396 011100100
B	#397 010100100	#398 110100100	#399 100100100	#400 000100100	#401 000110100	#402 100110100	#403 110110100	#404 010110100	#405 011110100	#406 111110100	#407 101110100	#408 001110100
C	#409 001010100	#410 101010100	#411 111010100	#412 011010100	#413 010010100	#414 110010100	#415 100010100	#416 000010100	#417 000011100	#418 100011100	#419 110011100	#420 010011100
D	#421 011011100	#422 111011100	#423 101011100	#424 001011100	#425 001111100	#426 101111100	#427 111111100	#428 011111100	#429 010111100	#430 110111100	#431 100111100	#432 000111100
E	#433 000101100	#434 100101100	#435 110101100	#436 010101100	#437 011101100	#438 111101100	#439 101101100	#440 001101100	#441 001001100	#442 101001100	#443 111001100	#444 011001100
F	#445 010001100	#446 110001100	#447 100001100	#448 000001100	#449 000001000	#450 100001000	#451 110001000	#452 010001000	#453 011001000	#454 111001000	#455 101001000	#456 001001000
G	#457 001101000	#458 101101000	#459 111101000	#460 011101000	#461 010101000	#462 110101000	#463 100101000	#464 000101000	#465 000111000	#466 100111000	#467 110111000	#468 010111000
H	#469 011111000	#470 111111000	#471 101111000	#472 001111000	#473 001011000	#474 101011000	#475 111011000	#476 011011000	#477 010011000	#478 110011000	#479 100011000	#480 000011000

Plate6	1	2	3	4	5	6	7	8	9	10	11	12
A	#481 000010000	#482 100010000	#483 110010000	#484 010010000	#485 011010000	#486 111010000	#487 101010000	#488 001010000	#489 001110000	#490 101110000	#491 111110000	#492 011110000
B	#493 010110000	#494 110110000	#495 100110000	#496 000110000	#497 000100000	#498 100100000	#499 110100000	#500 010100000	#501 011100000	#502 111100000	#503 101100000	#504 001100000
C	#505 001000000	#506 101000000	#507 111000000	#508 011000000	#509 010000000	#510 110000000	#511 100000000	#512 000000000				
D												
E												
F												
G												
H												

Table 8.1: Table of all members of the CRM-B mutant library assembled by OptiCut-optimised ligative oligonucleotide assembly. The members of the library are arranged by their occurrence on the 96-well qPCR plates in which they were assembled. The binary code associated with each well denotes the presence of a mutated (1) or wild type (0) sequence present at each site in the sequence.

8.3.5. Sequencing of amplified assemblies

The full length sequences were cloned into TOPO vector as described in the methods, see section 4.3.7, for sequencing and downstream use. Sequencing reactions were performed using a primer on the TOPO vector backbone point towards the cloning site. The results from these initial sequencing reactions were not successful (see figure 8.15). Initially high quality sequence quickly gave way to very low peak heights which continued for the result of the sequencing reaction. The sharpness of the drop in quality of sequence was indicative of the presence of a secondary structure in the DNA such as a hairpin loop. A hairpin loop would be difficult for the polymerase to bypass and would result in a significant number of prematurely terminated products.

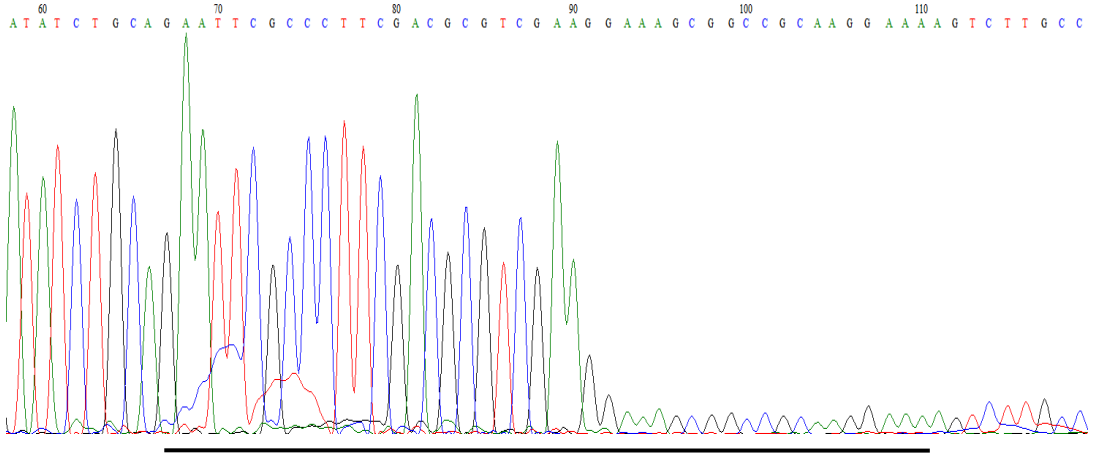


Figure 8.15: Sequence trace of CRM cloned into TOPO vector without addition of any chaotropic agents. The black bar denotes a 43 base pair region of homology. Note the sharp drop in sequence quality within the denoted region.

Chaotropic agents such as DMSO[41] or betaine[42] are frequently added to PCR and sequencing reactions in order to reduce secondary structures present in the DNA. DMSO was added as standard to all PCR reactions performed during the assembly of the full length sequences. Neither DMSO or beteine agent is not usually added to sequencing reactions due to the fact that DMSO increases the rate of degradation of the tubing used in the sequencers. Whilst the addition of DMSO did not significantly affect the sequencing

reaction, the addition of beteine did marginally improve the signal after the signal drop (data not shown). Even with the improved signal:noise ratio the low quality of the sequence made determination of error rate during the assembly process difficult to determine.

8.3.6. Palindromic regions present in original sequences

The initial failure of the sequencing reaction with the full length sequences with traces indicative of secondary structures being present in the DNA led to a close investigation of the original sequence. A long stretch of sequence at either ends of the full sequence were identified as being homologous. Furthermore, when cloned into TOPO vector, the region of homology was increased due to palindromic sequences present in the TOPO vector backbone. The result of this homology was that a structure reminiscent of a Holliday junction[43] is very likely.

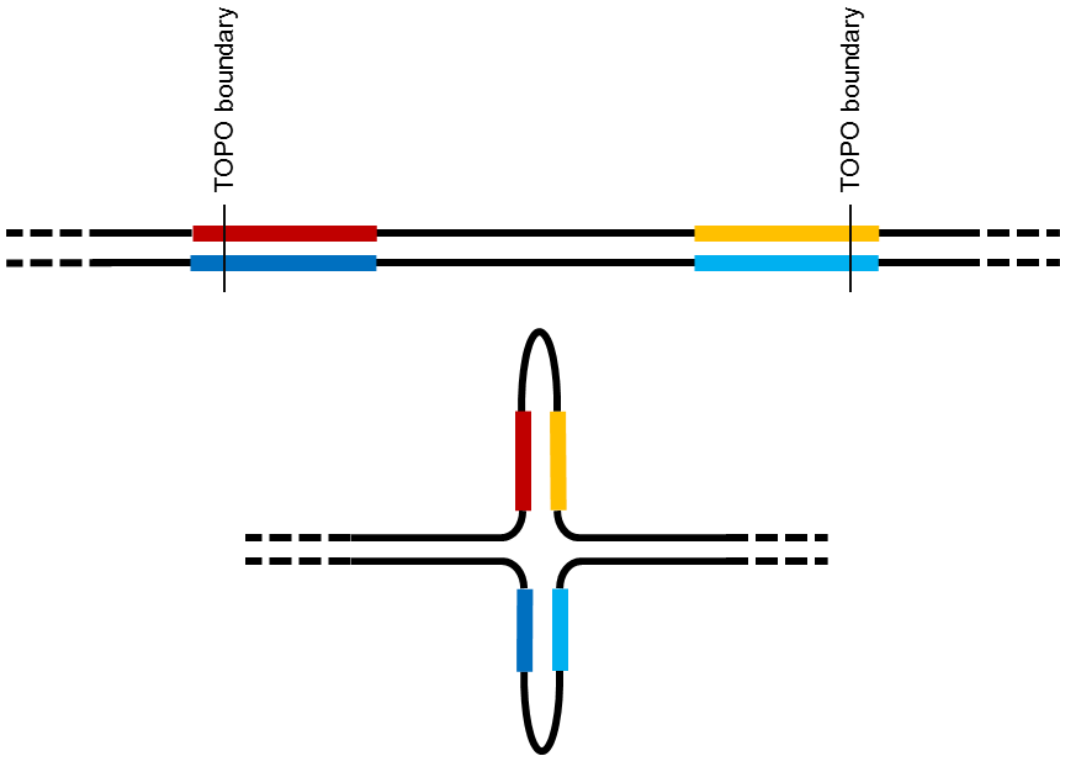


Figure 8.16: Diagram of secondary structure that could form in the CRM constructs that inhibit the sequencing reaction. The TOPO boundaries denote the start and finish of the cloned CRM sequence. The rest of the plasmid DNA, depicted by dashed lines, is omitted for clarity. Diagram not to scale.

The total length of the homologous region is 43 base pairs and has an estimated melting temperature of 73°C. The secondary structure represented in figure 8.16 is likely to be the cause of the sequencing failure observed. To test the hypothesis, the majority of the homologous region was removed by recombinant PCR of unamplified, ligated assembly with a new pair of primers. Sequencing of this product yielded high quality sequence traces from which the error rate can be accurately determined.

8.3.7. Error rate

All the possible sequences in the library were assembled and amplified in parallel using a 96 well format. 5 1/3rd 96 well plates were necessary to amplify all the sequences. Of this library 20 distinct sequences were cloned and a total of 42 sequence traces were obtained sequencing a total of 12390 bases of assembled sequence. The error rate in the sequenced DNA was found to be 4.2 per kb or 0.42%. The data used to determine this error rate is displayed in table 8.2. An error rate of 0.42% implies that each sequence would have to be sampled an average of 1.377 times in order to be confident that at least one copy of correct sequence had been obtained. This value compares favourably with other reported error rate values[17,35,44,45]. Methods of improving this error rate are discussed in the following section 8.4.

Figure 8.17 shows an example alignment of 6 sequences generated in order to obtain sequence #505 001000000. Figure 8.18 shows the relative distribution of these mutation sites across all aligned sequences. Interestingly, this distribution appears to be skewed to the first position of the overlap. It should be noted that in this analysis the normalised positions 0 and 1 do not represent the 5' or 3' end of the synthesised sequences specifically. Rather 0 is the first base downstream from a cut site and 1 is the first base upstream of a cut site.

Sequence	Insertions	Deletions	Substitutions	Sum(Mutations)	% error
#161 a 000011111	0	0	4	4	1.36%
#161 b 000011111	0	1	4	5	1.69%
#161 c 000011111	0	0	3	3	1.01%
#161 d 000011111	0	1	2	3	1.01%
#257 a 111100010	0	2	0	2	0.68%
#257 b 111100010	0	1	0	1	0.34%
#257 c 111100010	0	0	0	0	0%
#257 d 111100010	0	3	1	4	1.36%
#385 a 000000100	0	0	0	0	0%
#385 b 000000100	0	1	0	1	0.34%
#385 c 000000100	0	0	0	0	0%
#385 d 000000100	0	1	1	2	0.68%
#449 a 000001000	0	0	0	0	0%
#449 b 000001000	0	2	3	5	1.69%
#449 c 000001000	0	1	0	1	0.34%
#449 d 000001000	0	1	3	4	1.36%
#481 a 000001000	0	0	1	1	0.34%
#481 b 000001000	0	0	0	0	0%
#481 c 000001000	0	0	0	0	0%
#481 d 000001000	0	0	0	0	0%
#481 e 000001000	0	0	0	0	0%
#482 100001000	0	0	2	2	0.68%
#483 110001000	0	0	0	0	0%
#484 010001000	0	0	0	0	0%
#486 111001000	0	1	0	1	0.34%
#487 101001000	0	0	1	1	0.34%
#488 001001000	0	0	0	0	0%
#497 a 000100000	0	1	2	3	1.01%
#497 b 000100000	0	0	2	2	0.68%
#497 c 000100000	0	0	1	1	0.34%
#497 d 000100000	0	0	0	0	0%
#505 a 001000000	0	0	0	0	0%
#505 b 001000000	0	0	2	2	0.64%
#505 c 001000000	0	0	0	0	0%
#505 d 001000000	0	0	2	2	0.64%
#505 e 001000000	0	0	0	0	0%
#509 a 010000000	0	0	0	0	0%
#509 b 010000000	0	1	1	2	0.64%
#509 c 010000000	0	0	0	0	0%
#511 a 100000000	0	0	2	2	0.64%
#511 b 100000000	0	0	0	0	0%
#512 a 000000000	0	3	1	4	1.36%
#512 b 000000000	0	0	0	0	0%
Total	**0**	**20**	**38**	**52**	**0.42%**

Table 8.2: Sequencing data obtained from a set of 42 sequences produced by the bench-top OptiCut sequence assembly process. Sequence numbers are relative to those found in table 8.1. Binary codes represent the sites that are mutated in each sequence. Percentage errors are calculated assuming a total assembled length of 295 base pairs in each sequence. Insertions and deletions are defined and the presence of an unexpected or absence of expected base pair at a specific positions. Substitutions are a base pair change at a specific position. The types of substitutions (transversions or transitions) are not recorded.

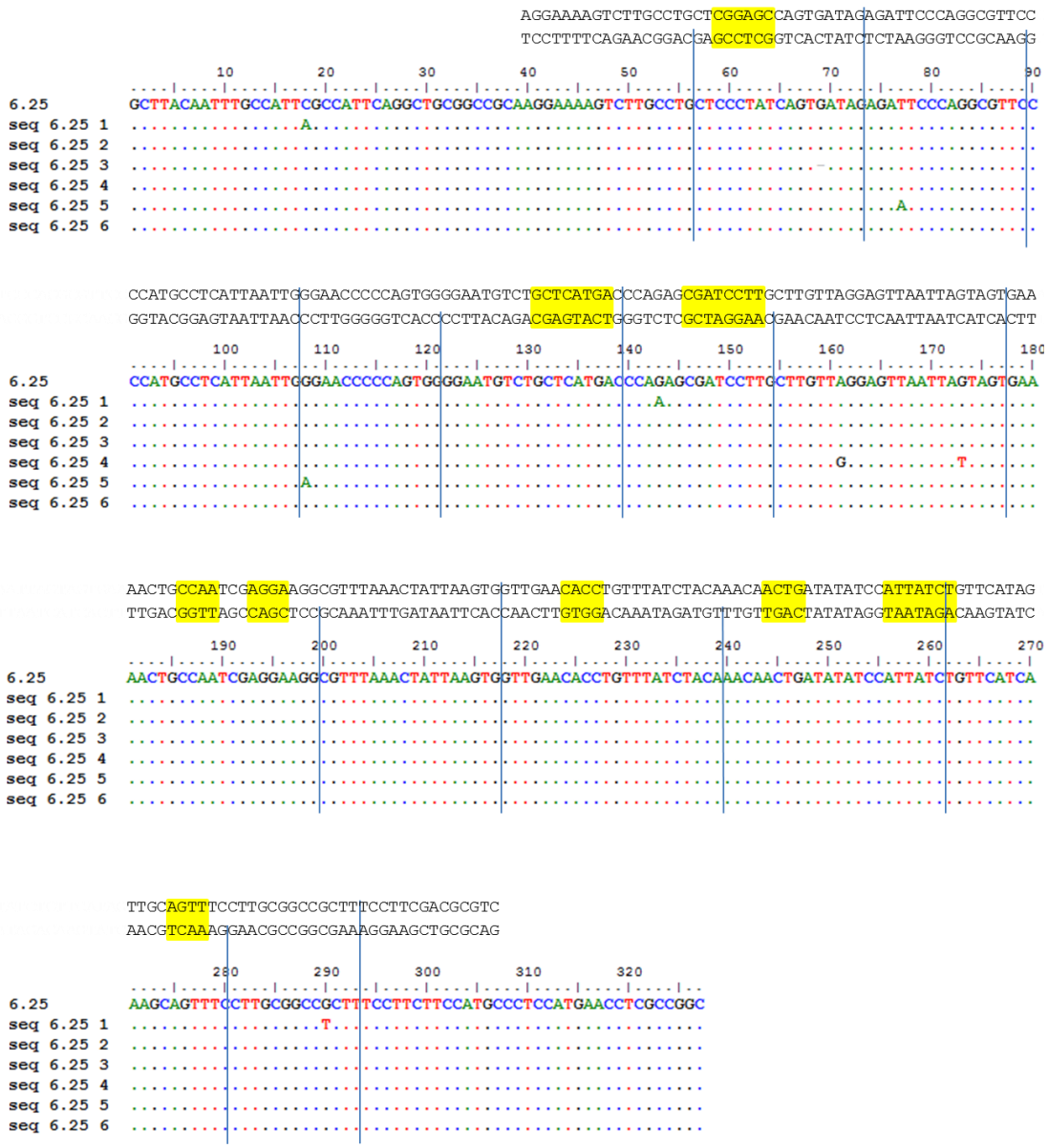


Figure 8.17: Aligned sequencing data obtained from four sequencing reactions of sequence #505 001000000. The top sequence of the alignment is the expected sequence (6.25) and the sequenced sequences (seq 6.25 1-6). The assembly oligonucleotide start and stop points (cut sites) are indicated by the blue vertical lines dropping from the double strand sequence above each row of the alignment. Cut sites in the top strand are represented by bars reaching up to the top strand, conversely cut sites in the bottom strand are represented by bars reaching to the bottom strand only. Intended mutation sites are marked in yellow. Unintended mutations are visible in the alignment sequences.

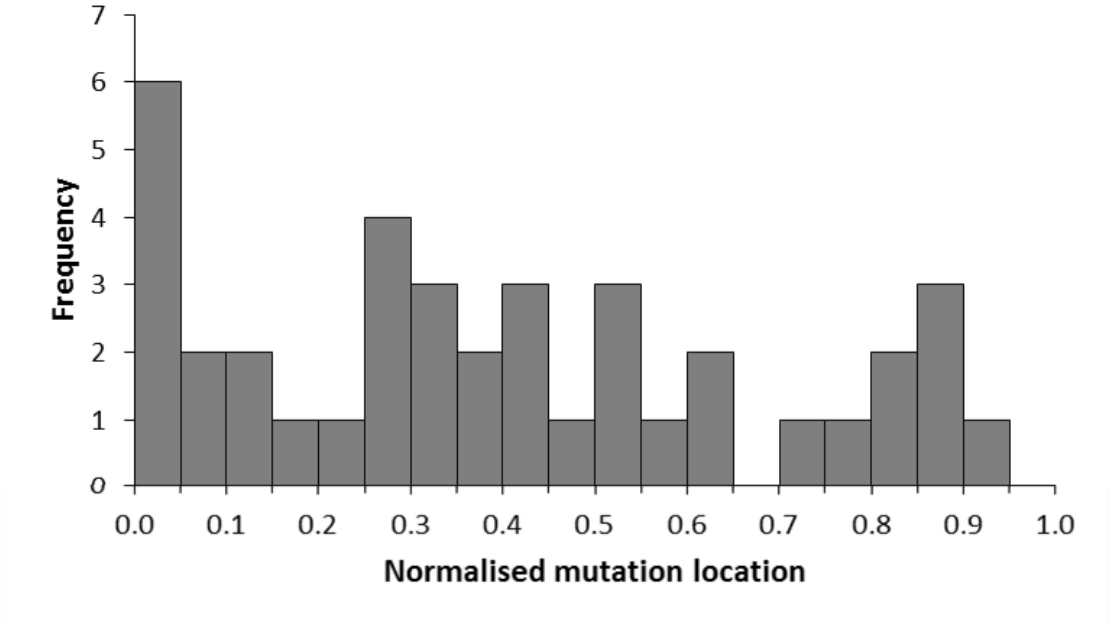


Figure 8.18: Histogram of normalised mutation location showing the distribution of relative positions of mutations within each overlap in sequenced data. Histogram bars are each 0.05 units wide.

It is impossible to know with this analysis whether the source of the mutation was the oligonucleotide of the strand which possesses the cut site or the strand which does not. The apparent skew of the normalised mutation location distribution shown in figure 8.18 does appear to indicate that the source of some mutations is the oligonucleotide substrate and/or the assembly process. This is because if all the mutations were due purely to errors occurring during amplification then the distribution seen in figure 8.18 would be uniform. Unfortunately, this sample is not large enough for the apparent skew in the normalised mutation location distribution to be significant.

8.4. Further work

In this chapter, the feasibility of the creation of the library using bench-top methods is demonstrated. Unfortunately there are two primary downstream bottlenecks. The first bottleneck is the cloning of each PCR product into its target vector. To alleviate this bottleneck a high throughput cloning strategy is required that would allow cloning of each

PCR product of the into the target vector. Ideally a rapid cloning method, such as TOPO cloning, would be employed with a recipient vector.

The second bottleneck is gene expression testing of each plasmid. Currently the effect of each CRM is determined by measurement of the level of green fluorescent protein (GFP) in transient transfections in C2C12 cells by flow cytometry. This work is described in detail in chapter 4. Ideally the measurement of transcription from the gene of interest would be performed in live cells in real time.

A single pot assembly of the oligonucleotides could possible by combining all ~42 oligonucleotides in a single ligation reaction. Some oligonucleotides will be expected to participate in multiple sequences, whereas others will only participate in a small number of sequences and the molarity of each oligonucleotide should reflect this. For reasons of sampling (see section 8.1.4), however, this is not an efficient approach to take.

Enzymes such as MutS digest DNA with mismatched base pairs. A band shift assay can then be employed to separate sequences containing mismatches from correct sequences[46]. It is possible, however, for this system not to catch mutations as two strands that have complementary mutations will not exhibit mismatching. Two identically mutated strands are extremely unlikely, however. A commercially available combination of enzymes called ErrASE (Novici Biotech, US) allows the cutting of mismatched sequences before amplification to remove these sequences from the population[7].

Further analysis of the mutations could be performed by high throughput sequencing of a mixture of assemblies. A higher number of samples, >1000 sequences could allow the deconvolution of the two possible sources of mutation; the oligonucleotide substrate and/or assembly process and the amplification process.

8.5. Conclusions

This chapter details the attempts to generate a combinatorial mutant library of DNA sequences using two related gene assembly approaches. The main conclusions that are drawn from this work are as follows:

- One-pot assembly of the library using the Gao method was only partially successful.
- The multi-pot assembly of the library using OptiCut optimised oligonucleotides was successful.
- The error rate determined by traditional Sanger sequencing was 0.42 %.
- Secondary structures, present as a result of the input sequence, whilst they did not prevent the sequence from successfully assembling, did make sequencing difficult.

Unfortunately, the Gao assembly was not successful despite the attempts at optimising the ligation reaction and PCR conditions. Had the Gao assembly been successful, however, significant work would have remained in sampling the one pot assembly for mutants of interest (see section 8.1.4). By converting to a multi-pot assembly the full length sequences required could be assembled and amplified on demand.

A key advantage of the gene assembly process used here to make mutational libraries over methods such as site directed mutagenesis (SDM) is that the range of types of mutations is increased. SDM is limited to making changes to only a few basepairs at a time. The gene assembly process described here can create sequences with multiple mutation sites in the assembled sequence in a single reaction.

The entire mutant library is now assembled, amplified and purified as linear DNA and is ready to be cloned into any appropriate vector. The sampling of this mutant library and regulation of gene expression experiments are discussed in chapter 9.

8.6. References

1. Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X. & Church, G. M. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
2. Agarwal, K. L., Buchi, H., Caruthers, M. H., Gupta, N., Khorana, H. G., Kleppe, K., Kumar, A., Ohtsuka, E., Rajbhandary, U. L., Van de Sande, J. H., Sgaramella, V., Weber, H. & Yamada, T. Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature* **227**, 27–34 (1970).
3. Nambiar, K. P., Stackhouse, J., Stauffer, D. M., Kennedy, W. P., Eldredge, J. K. & Benner, S. A. Total synthesis and cloning of a gene coding for the ribonuclease S protein. *Science* **223**, 1299–1301 (1984).
4. Ferretti, L., Karnik, S. S., Khorana, H. G., Nassal, M. & Oprian, D. D. Total synthesis of a gene for bovine rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 599–603 (1986).
5. Gibson, D. G., Benders, G. A., Andrews-pfannkoch, C., Denisova, E. A., Baden-tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Iii, C. A. H. & Smith, H. O. Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science* **319**, 1215–1220 (2008).
6. Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Iii, C. A. H. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 12–16 (2009).
7. Kosuri, S., Eroshenko, N., Leproust, E. M., Super, M., Way, J., Li, J. B. & Church, G. M. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nature biotechnology* **28**, (2010).
8. Kong, D. S., Carr, P. A., Chen, L., Zhang, S. & Jacobson, J. M. Parallel gene synthesis in a microfluidic device. *Nucleic acids research* **35**, e61 (2007).
9. Ye, H., Huang, M. C., Li, M.-H. & Ying, J. Y. Experimental analysis of gene assembly with TopDown one-step real-time gene synthesis. *Nucleic acids research* **37**, e51 (2009).
10. Gao, X., Yo, P., Keith, A., Ragan, T. J. & Harris, T. K. Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Research* **31**, e143 (2003).
11. Bryksin, A. V. & Matsumura, I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* **48**, 463–5 (2010).
12. Dong, B., Mao, R., Li, B., Liu, Q., Xu, P. & Li, G. An improved method of gene synthesis based on DNA works software and overlap extension PCR. *Mol Biotechnol* **37**, 195–200 (2007).
13. Sandhu, G. S., Aleff, R. A. & Kline, B. C. Dual asymmetric PCR: one-step construction of synthetic genes. *BioTechniques* **12**, 14–6 (1992).
14. Gordeeva, T. L., Borschevskaya, L. N. & Sineoky, S. P. Improved PCR-based gene synthesis method and its application to the Citrobacter freundii phytase gene codon modification. *Journal of microbiological methods* **81**, 147–52 (2010).
15. Czar, M. J., Anderson, J. C., Bader, J. S. & Peccoud, J. Gene synthesis demystified. *Trends in biotechnology* **27**, 63–72 (2009).
16. Au, L., Yang, F., Yang, W., Lo, S. & Kao, C. Gene Synthesis by a LCR-Based Approach : High-Level Production of Leptin-L54 Using Synthetic Gene in Escherichia coli. *Biochemical and Biophysical Research Communications* **203**, 200–203 (1998).
17. Wassman, C. D., Hatfield, G. W. & Lathrop, R. H. Computationally optimised DNA Assembly of synthetic genes. *International Journal of Bioinformatics Research Applications* **4**, 324–336 (2009).

18. Xiong, A., Peng, R., Zhuang, J., Liu, J., Gao, F., Chen, J., Cheng, Z. & Yao, Q. Non-polymerase-cycling-assembly-based chemical gene synthesis : Strategies , methods , and progress. *Biotechnology Advances* **26**, 121 – 134 (2008).

19. Wang, T., Zhu, H., Ma, X., Zhang, T., Ma, Y. & Wei, D. Methods for Mutant Library Construction Casting a Wider Net. *Molecular Biotechnology* **34**, (2006).

20. Farinas, E. T., Bulter, T. & Arnold, F. H. Directed enzyme evolution. *Current opinion in biotechnology* **12**, 545–51 (2001).

21. Otten, L. G. & Quax, W. J. Directed evolution: selecting today's biocatalysts. *Biomolecular engineering* **22**, 1–9 (2005).

22. Shivange, A. V., Marienhagen, J., Mundhada, H., Schenk, A. & Schwaneberg, U. Advances in generating functional diversity for directed protein evolution. *Current opinion in chemical biology* **13**, 19–25 (2009).

23. Fujii, R., Kitaoka, M. & Hayashi, K. Error-prone rolling circle amplification: the simplest random mutagenesis protocol. *Nature protocols* **1**, 2493–7 (2006).

24. Stephens, D. E., Singh, S. & Permaul, K. Error-prone PCR of a fungal xylanase for improvement of its alkaline and thermal stability. *FEMS microbiology letters* **293**, 42–7 (2009).

25. Sawano, a & Miyawaki, a Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic acids research* **28**, E78 (2000).

26. Zha, D., Eipper, A. & Reetz, M. T. Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. *Chembiochem : a European journal of chemical biology* **4**, 34–9 (2003).

27. Chan, S. W. P., Hung, S.-P., Raman, S. K., Hatfield, G. W., Lathrop, R. H., Da Silva, N. a & Wang, S.-W. Recombinant human collagen and biomimetic variants using a de novo gene optimized for modular assembly. *Biomacromolecules* **11**, 1460–9 (2010).

28. Ness, J. E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T. V., Govindarajan, S., Mundorff, E. C. & Minshull, J. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nature biotechnology* **20**, 1251–5 (2002).

29. Ge, X., Mazor, Y., Hunicke-Smith, S. P., Ellington, A. D. & Georgiou, G. Rapid construction and characterization of synthetic antibody libraries without DNA amplification. *Biotechnology and bioengineering* **106**, 347–357 (2010).

30. Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J. M., Yan, J., Sillanpää, M. J., Bonke, M., Palin, K., Talukder, S., Hughes, T. R., Luscombe, N. M., Ukkonen, E. & Taipale, J. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Research* **20**, 861–873 (2010).

31. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites - A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* **22**, 1859–1862 (1981).

32. Dorman, M. a., Noble, S. a., McBride, L. J. & Caruthers, M. H. Synthesis of oligodeoxynucleotides and oligodeoxynucleotide analogs using phosphoramidite intermediates. *Tetrahedron* **40**, 95–102 (1984).

33. Srivannavit, O., Gulari, M., Gularia, E., LeProust, E., Pellois, J. P., Gao, X. & Zhou, X. Design and fabrication of microwell array chips for a solution-based, photogenerated acid-catalyzed parallel oligonuclotide DNA synthesis. *Sensors and Actuators A: Physical* **116**, 150–160 (2004).

34. Srivannavit, O., Gulari, M., Hua, Z., Gao, X., Zhou, X., Hong, A., Zhou, T. & Gulari, E. Microfluidic Reactor Array Device for Massively Parallel In-situ Synthesis of Oligonucleotides. *Sensors and actuators. B, Chemical* **140**, 473–481 (2009).

35. Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X. & Church, G. M. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).

36. Zhou, X., Cai, S., Hong, A., You, Q., Yu, P., Sheng, N., Srivannavit, O., Muranjan, S., Rouillard, J. M., Xia, Y., Zhang, X., Xiang, Q., Ganesh, R., Zhu, Q., Matejko, A., Gulari, E. & Gao, X. Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Methods* **32**, 5409–5417 (2004).

37. Gao, X., Gulari, E. & Zhou, X. In Situ Synthesis of Oligonucleotide Microarrays. *Biopolymers* **73**, 579–596 (2004).

38. Szybalski, W., Kim, S. C., Hasan, N. & Podhajskab, A. J. Class-IIS restriction enzymes - a review. *Gene* **100**, 13–26 (1991).

39. Smith, H. O., Hutchison III, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly : phi X174 bacteriophage from synthetic oligonucleotides. *PNAS* **100**, 15440–15445 (2003).

40. Gray, F. Pulse Code Communication. (1953).

41. Chakrabarti, R. & Schutt, C. E. The enhancement of PCR amplification by low molecular-weight sulfones. *Gene* **274**, 293–8 (2001).

42. Henke, W., Herdel, K., Jung, K., Schnorr, D. & Loening, S. a Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic acids research* **25**, 3957–8 (1997).

43. Holliday, R. A mechanism for gene conversion in fungi. *Genetical research* **5**, 285–307 (1964).

44. Chalmers, F. M. & Curnow, K. M. Scaling up the ligase chain reaction-based approach to gene synthesis. *BioTechniques* **30**, 249–52 (2001).

45. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).

46. Carr, P. a, Park, J. S., Lee, Y.-J., Yu, T., Zhang, S. & Jacobson, J. M. Protein-mediated error correction for de novo DNA synthesis. *Nucleic acids research* **32**, e162 (2004).

Chapter 9

9. Analysis of CRM position-effect and mutant constructs

As described in chapter 1, there are two models of CRM action, the billboard and enhanceosome models (see section 1.5). The previous investigation of this CRM system has not indicated which of these models best describes the mechanisms by which these previously identified CRMs achieve the regulatory synergy observed (see section 2.5). Furthermore, the plasmid into which the previously identified CRMs have been cloned is an artificial construct that does not accurate represent the environment that the CRMs function in. As a result, it is possible that the synergistic effects observed are a product of the artificial environment rather than representative of processes which occur *in vivo*. To investigate outstanding questions, the position and orientation of one CRM with respect to the others was modified in three CRM contexts and the effect on reporter expression measured by flow cytometry (see section 9.2).

In addition to the above, this chapter describes the measurement of the reporter expression changes induced by the mutation of single sites within CRM-B in one CRM context (see section 9.3). The mutation sites were identified based on the information presented in chapter 2 (see section 2.6.4). Furthermore, by cloning the mutant CRM sequences in both the forward and backwards directions, the orientation sensitivity was measured. These experiments represent the sampling of a small portion of the mutant library generated in chapter 8.

9.1. Description of model system and analysis

Transient transfection assays were used to investigate the activity of different CRM constructs by quantification of expression of a reporter gene. This section will introduce the

different methods of model systems used in this analysis, with reference to their power and limitations as well alternative methods where applicable.

9.1.1. C2C12 cells

The model system used in this study was the C2C12 mouse muscle myoblast cell line. This cell line is capable of reproducing the major steps of muscle differentiation and is a well-established model system for studying myogenesis. C2C12 cells were first isolated from dystrophic mouse muscle[1]. From these isolates a heterogeneous cell population was derived. By serial passaging a population of cells was obtained that was capable of proliferation and differentiation in cell culture conditions. This population is thought to be derived from the muscle satellite cells: A population of adult stem cells responsible for repairing damage to muscles sustained during adulthood[2]. Satellite cells are described in more detail in section 2.4.

The C2C12 cell line differentiates in appropriate (serum-starved, confluent) conditions into contractile myotubules that produce characteristic muscle proteins. The cell line exhibits some plasticity; application of bone morphogenetic protein-2 (BMP-2) shifting differentiation towards osteoblastic pathway[3]. Similarly, inhibiting myogenic differentiation shifts C2C12 differentiation towards the adipocytic pathway[4]. In their undifferentiated state, C2C12 cells are characterised as small, mononucleate and fast replicating. During myogenic differentiation myoblasts stop replicating and fuse into multinucleate myotubes. This differentiation process can be induced by cell-cell contact and removal of fibroblast growth factor (FGF) from the medium. There is evidence that during differentiation, a subpopulation of C2C12 cells remain in an undifferentiated state expressing low levels of MyoD and Myf5[5], analogous to the specification of the satellite cell population during development and adult regenerative myogenesis.

C2C12 cells are tetraploid[6] and therefore it is impossible to insert DNA by homologous recombination from a suitable vector such as a bacterial artificial chromosome (BAC). An alternative mouse cell line used for myogenic differentiation studies is the 10T1/2 cell line. This line, however, requires treatment with demethylating agents before myogenic differentiation will take place[7]. Primary cells can be obtained from samples of mouse muscle. The size of such samples and the fact that primary cells can only be maintained for a limited number of population doublings mean that significant numbers of cells, as would be required for flow cytometry analysis, is difficult to obtain.

9.1.2. Transient transfection to study reporter gene expression

Plasmids encoding genes of interest or marker genes can be inserted into eukaryotic cells by the process of transient transfection. Plasmids are double stranded, circular and small, usually less than several thousand base pairs in length. Cultured cells will take up plasmids in solution and transport them to the nucleus, where factors present therein can initiate expression of a reporter gene. A common reporter gene is GFP, the fluorescence of which is proportional to the of GFP as measured by antibody staining[8]. Transfection additives such as Lipofectamine 2000™ can be included in order to increase transfection efficiency[9]. Lipofectamine operates by lipofection, where liposomes that are capable of merging with the cell membrane are used to encapsulate the double stranded plasmid DNA[10]. This method of transfection is widely used and effective. Plasmids are degraded rapidly by cells, however, so the lifetime of a plasmid in a cell is short. If the appropriate sequences are present on the plasmid and factors present in the cell then genes on the plasmids can be expressed during the plasmid's lifetime in the cell (48-72 hours).

Plasmids exist separately from the host DNA and do not usually integrate into the host genome. This fact combined with the lack of structural sequences on the plasmid DNA mean that there is no epigenetic modification of the plasmid DNA. Epigenetic modification

can cause a section of DNA sequence to condense from open, accessible euchromatin to closed, inaccessible heterochromatin[11]. Chromatin condensation and decondensation events, an important mechanism of gene regulation, will not be represented by transient gene expression experiments involving plasmids used here.

Plasmid copy number is an important determinant of resulting gene expression: More copies of a plasmid will mean more of the resulting protein product[12]. Furthermore, many thousands of copies of a given plasmid will enter a cell during a transfection, meaning that gene dosing, competition between plasmids for available factors, can affect the obtained results[13]. Ideally, all cells will receive the same number of plasmids, making the variation in resulting measurements dependant solely upon plasmid sequences. In reality, variation in plasmid copy number will account for some of the variation in the resulting measurements. Plasmid transfection strategy is, therefore, important to control, to ensure that the variation due to transfection efficiency is minimised.

Transient gene expression studies offer a way of rapidly determining, roughly, the sequence-response relationship of a given system[14]. Plasmids can be easily manipulated in bacteria to contain sequences of interest in specific arrangements. Manipulation of plasmids is performed using the toolkit of molecular biology including, but not limited to restriction digest, ligation, recombination and mutagenesis. Once transient transfection studies in cell lines indicate a relationship, more controlled studies can be initiated.

9.1.3. Studying reporter gene expression using stable transfectants

Stable transfection of cells is achieved when a DNA sequence, usually from a transfection, is inserted into the genome of the host cell. One mechanism that is frequently used is borrowed from retroviruses. Specific homologous sequences in the host and the infecting plasmid undergo recombination due to the co-transfection of a plasmid encoding a recombinase gene. The most widely used example of this is the Cre-Lox system[15]. The

insertion of recognition sequences for recombination is done randomly which can result in position-effect variegation, as the sites where the sequence might insert have different packaging environments. Stably incorporated sequences will, however, be copied through future generations, although there is a risk that the inserted sequence will be randomly excised during replication. Stable transfections using recombination offer a method for generating single copy, stable insertions of genes from plasmids into specific sites in the host genome.

The gold standard of genetic analysis is usually a stably transfected whole mouse model. This approach is not feasible with a combinatorial experiment such as this one potentially involving hundreds of variants due to the length of time required to generate the constructs necessary and then breed the mice to generate the necessary animals.

9.2. Testing of position and orientation effects of CRM-B

The looping hypothesis suggests that CRM-B is brought close enough to the promoter through protein-protein bridges for the regulatory factors bound to the CRM to act on the formation of the DNA polymerase holoenzyme. Although the plasmid is not a perfect analogy of the chromosome, it is possible to test certain aspects of the looping hypothesis in a plasmid situation. For instance, if CRM-B does form a loop to bring itself close to the promoter, it should be able to do so from any position on the plasmid or orientation of that position.

For consistency, the differentiation time point chosen for all CRM expression studies was the same in all experiments and the same as had been used in previous work (Dr. H. Crutzen, P. Downton). The C2C12 cells in individual wells were transfected with the appropriate plasmid 24 hours before, when still proliferating, being transferred to a differentiation medium (serum starved) when confluent. The cells then remained in the

differentiation medium for a further 20 hours before the cells were fixed. See methods section 4.5 for full details of the transfection, differentiation and fixing protocols.

9.2.1. Plasmid construct design

The vector backbone for all expression studies was consistent across all experiments. The basic plasmid, the pGL plasmid, which contains an ampicillin resistance gene, an origin of replication, had been modified previously (H. Crutzen) to include the *myod* core promoter, a venus GFP (vGFP) encoding sequence (with 3 nuclear localisation sequences on the 3' end) and a SV40 poly adenosine (pA) sequence. A total of 16 combinations were cloned which had every possible combinations of CRMs, including none at all. The relative expression values of each of the constructs can be seen in figure 9.3. The basic backbone can be seen in figure 9.4. Combinations of the four CRMs; A, B, C and CER were cloned into the multiple cloning site (MCS) upstream of the *myod* promoter by H. Crutzen. More recently P. Downton repeated cloned the DRR, another region known to regulate *myod*, into the same constructs to generate a total of 32 constructs. The relative expression values from all 32 of these constructs can be seen in figure 2.4.

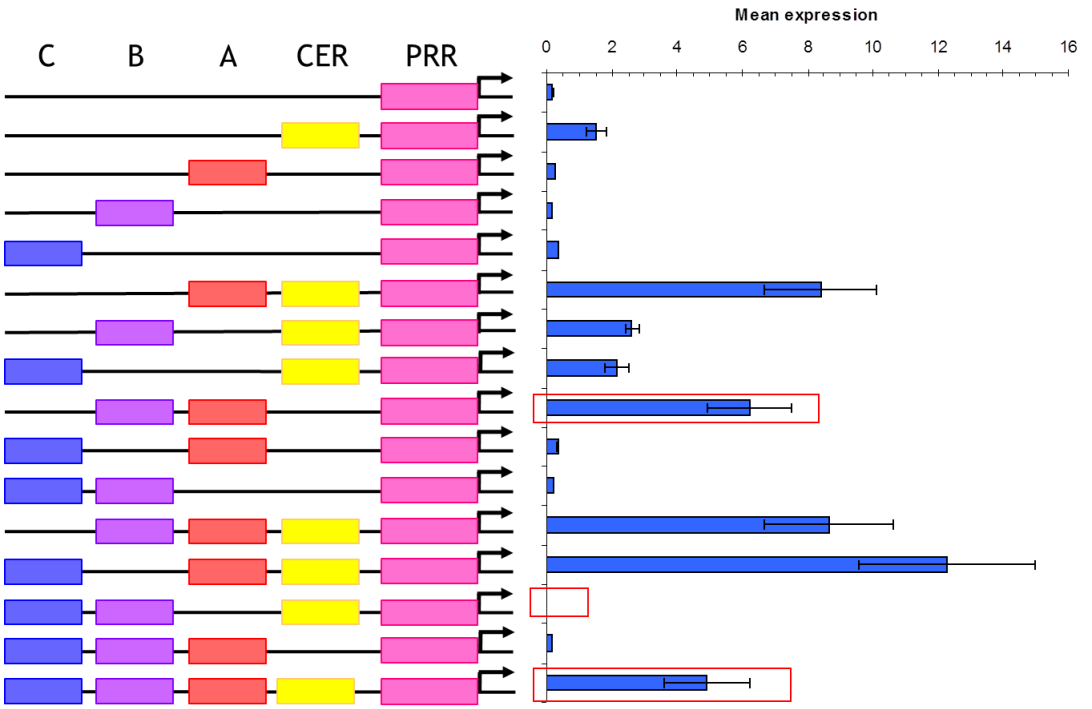


Figure 9.1: Previous work on combinatorial CRM constructs (H. Crutzen). Three constructs, highlighted in red, were used to test the position/orientation effects of CRM-B. Note that the scheme in the left pane is not to scale; there is no 'blank' sequence replacing absent CRMs.

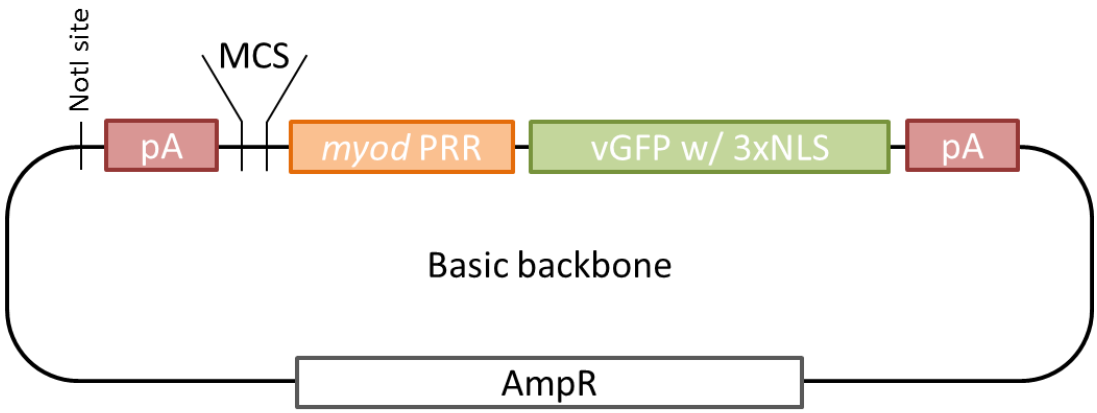


Figure 9.2: Schematic diagram of the plasmid vector backbone used in all expression studies. The NotI restriction enzyme binding site is marked, pA refers to polyadenlyation sequence, MCS to the multiple cloning site, *myod* PRR to the core promoter of mouse *myod* gene, vGFP w/ 3xNLS to the venusGFP encoding sequence with 3 nuclear localisation sequences attached and AmpR to the amplicillin resistance gene. NB the diagram is not to scale.

Figure 9.2 shows the basic plasmid backbone common to all previous experiments with the CRMs and the *myod* PRR, performed by H. Crutzen and P. Downton. The backbone consists

of an amplicilin resistance gene used for bacterial selection, the *myod* PRR and the venus GFP (vGFP) reporter flanked by two poly adenosine (pA) sites. In addition to these modules, an MCS is present upstream of the *myod* PRR and an additional, unique NotI site upstream of the 5' pA site. In previous experiments, the various CRM combinations were cloned into the MCS.

Figure 9.3 shows the CRM combinations cloned into the MCS to generate three constructs produced by H. Crutzen. Note that the order and orientation of the CRMs in these constructs does not change; CRM-C is upstream of CRM-B is upstream of CRM-A is upstream of the CER. The DRR, not represented in this figure, would be cloned downstream of the CER.

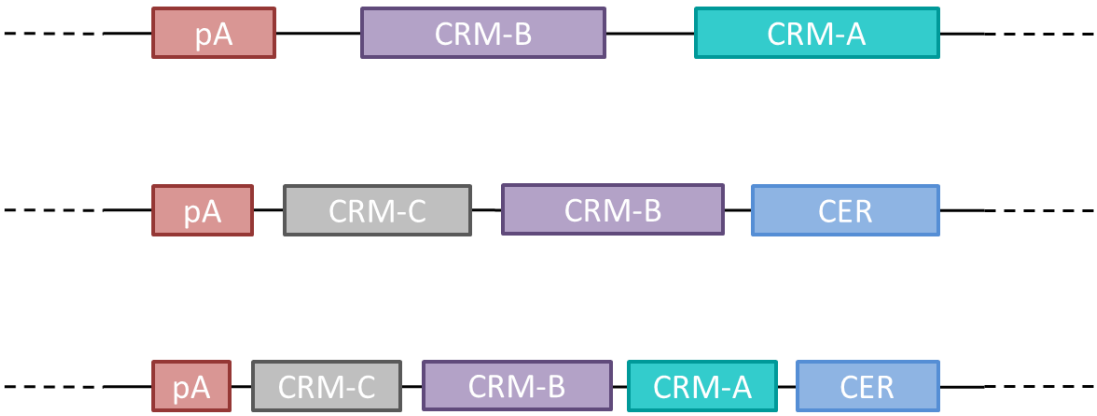


Figure 9.3: Schematic diagram of arrangement of CRMs in the control plasmids used as controls in the position/orientation experiments. The dotted lines indicate where the region joins the rest of the plasmid. The inserts shown here replace the region between the NotI site and the MCS of the backbone vector. NB diagram not to scale.

The three constructs were chosen based on work by H. Crutzen, shown in figure 9.3. CRM-B appears to have a large synergistic regulatory effect when combined with CRM-A, comparing the expression level of A-PRR and B-PRR with A-B-PRR. Conversely, CRM-B appears to have a negative effect on the overall expression level in two other cases; comparing C-CER-PRR with C-B-CER-PRR and C-A-CER-PRR with C-B-A-CER-PRR. These results appear to indicate that CRM-B is capable of exerting context-dependant effects on

the overall level of gene expression. It should be noted that the results obtained by H. Crutzen were not entirely consistent with the results obtained by P. Downton; the A-B-PRR synergistic effect is lost. In this case, the experiments performed by P. Downton are thought to be more reliable as they involve more repeats, were performed on a more accurate flow cytometer and did not use plasmids that were identified as defective. The choice of plasmids for testing was made on the basis of the information available at the time, which was the data obtained by H. Crutzen.

To investigate the how the position and order of the CRMs might affect their synergistic effects on *myod* activity, CRM-B was cloned into the NotI site upstream of the 5' pA so that it was out of position with respect to the other CRM cloned into the MCS. Three constructs, seen in figure 9.4, were prepared for comparison to the three constructs seen in figure 9.3.

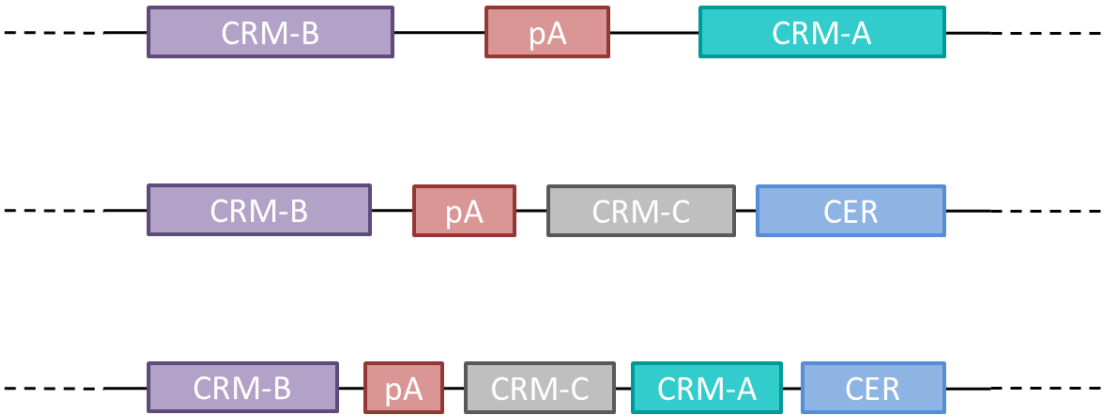


Figure 9.4: Schematic diagram of arrangement of CRMs in the test plasmids used for position/orientation experiments. The dotted lines indicate where the region joins the rest of the plasmid. The inserts shown here replace the region between the NotI site and the MCS of the backbone vector. NB diagram not to scale.

In addition to cloning CRM-B out of position, CRM-B was also cloned into the NotI site in both the forward and backward orientation. A total of six test plasmids were produced; a pair for each of the arrangements shown in figure 9.4. The expression of these test plasmids was compared to the three control plasmids, as shown in figure 9.3. All plasmid

manipulation was performed according to the techniques described in section 4.3 of the methods.

9.2.2. Expression analysis results

Figure 9.5 shows the normalised results of the experiment to test position and orientation effects in CRM-B. To emphasise the fact that CRM-B has been taken out of its normal order with respect to the promoter and the other CRMs, the construct labels on figure 9.5 represent the order of CRMs in each construct from 5' to 3'.

Figure 9.5 shows that there is clear segregation of the three CRM combinations. The yellow bars in figure 9.5 show the relative expression of the A-B construct with the A-IR-B construct, that contains the intervening region between A and B, and the A-B constructs with CRM-B cloned into the NotI site in either the forward or reverse conformation. Whilst there is an increase in the relative expression by including the IR, there is no change when B is cloned into the NotI site than when B is cloned into the normal site. B was not cloned into the NotI site in the A-IR-B construct.

The light orange bars in figure 9.7 show the relative expression of the CER-B-C construct and the CER-B-C constructs with CRM-B cloned into the NotI site in either the forward or reverse conformation. There is no significant difference between these three constructs.

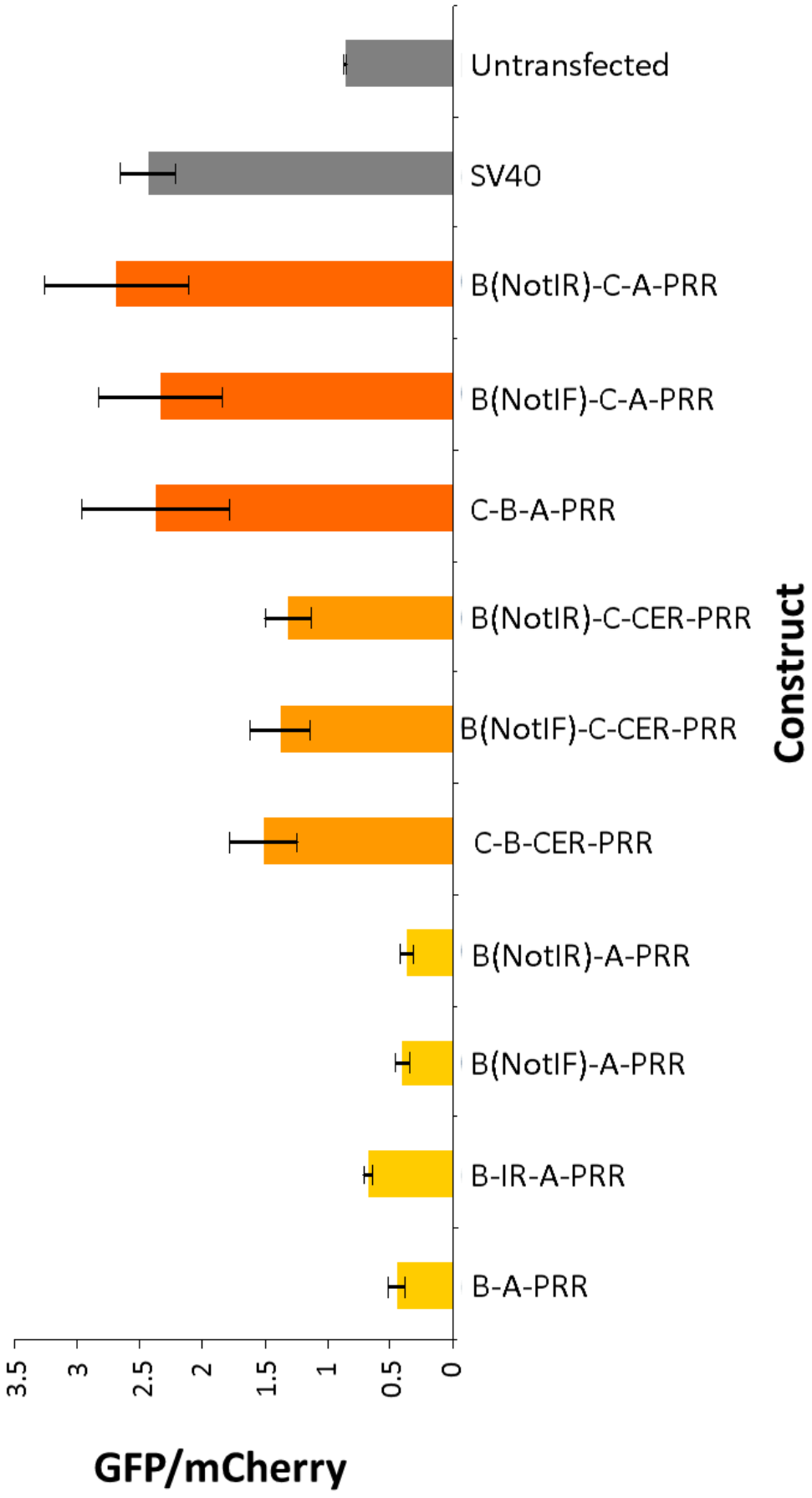


Figure 9.5: Comparison of expression of constructs containing three combinations of CRMs. The effect of cloning CRM B into the NotI site in either the forward (NotIF) or reverse (NotIR) conformations can be seen. Expression values are shown relative to mCherry internal control. SV40 and untransfected cell controls are also shown.

The dark orange bars in figure 9.5 show the relative expression of CER-A-B-C construct and the CER-A-B-C constructs with CRM-B cloned into the NotI site in either the forward or reverse conformations. Again, there is no significant difference between any of these three constructs.

These results support the assertion that the previously identified sequences are indeed acting as CRMs. Their presence, independent of their relative position or orientation, in constructs containing the target *myod* promoter is sufficient for their regulatory effect to be observed. These observations indicate that the CRMs are acting as independent entities that cooperate to achieve regulation of the *myod* promoter.

According to the enhanceosome model of CRM function, relative spatial arrangement of CRMs is important their function. Thus, it would be expected that, by changing the position and orientation of a CRM, that the ability of that CRM to function as part of an enhanceosome would be impaired. The obtained results do not indicate that this is the case. The billboard model of CRM function implies that it is the presence of a factor, not its relative position, that affects function. Whilst positioning a factor further away from its site of interaction might affect the likelihood of interaction, orientation is unlikely to have an effect. The results shown here indicate that changing the position and the orientation of the CRM did not significantly alter the ability of the CRM to regulate *myod* activity. These results are, therefore, consistent with both the billboard and enhanceosome models of CRM function.

9.3. Mutational analysis of CRM activity

9.3.1. Selection of sites within CRM-B for mutational analysis

To investigate the regulation of *myod*, a library of combinatorial mutants of a CRM, CRM B, were prepared. CRM B was chosen because it is the smallest of the CRMs, contained

several binding sites for relevant transcription factors and previous work by H. Crutzen and P. Downton, see figure 2.2, indicated that the addition of CRM B to the C-A-CER-PRR construct resulted in a significant upregulation of promoter activity. It was hoped, therefore, that mutation of the CRM B sequence would prevent this enhancing effect and would be readily observable for flow cytometry analysis. Furthermore, a 'fuller' construct that contains most of the identified regulatory regions of *myod* is most likely to reflect the physiological situation, where all CRMs are available.

A total of 65 sites were found in CRM B using the Binding Factor (BiFa) tool developed by Dr. J. Reid. Briefly, the BiFa tool searches a given sequence for matches to the consensus binding sequences of all known DNA binding proteins found in the TransFac database[16]. The consensus binding sequence for a given transcription factor may only weakly depend on some base pairs within its binding site, as defined by the relative strength of the base pairs in the transcription factor's binding motif. Furthermore, the database often over represents transcription factors, with a factor having several valid consensus binding sites. As a result of these two facts, the BiFa tool returns many hits for the same site as shown in figure 2.7. To reduce the number of hits to only the most promising, the list of potential binding sites is limited to those sites that bind factors that have been shown to be expressed in differentiating C2C12 cells by microarray analysis (previous work by H. Crutzen). Further analysis of the binding sites is performed by assigning certain sites to signalling pathways known to be important to the regulation of *myod* specifically or to cellular differentiation in general. By combining this information with data from previous experiments, described in section 2.6, several sites were selected for mutation in CRM B.

The following 7 sites in CRM B were selected for mutational analysis (in order from 5' to 3'): AP1, Ets, NFY, Ets, FOXO, Ebox and Lef1. Two additional sites were regarded as significantly interesting, a serum response factor (SRF) binding site that overlaps both the

AP1 and first Ets site and an Ebox site that overlaps with the FOXO site. These additional sites could not be mutated according to the used mutation strategy without affecting the other mutant sites as any changes that obliterated one binding site also affected the binding score of the other. See figure 2.6 for a diagram of the sites on CRM-B.

As discussed in the background chapter (see section 2.6), the factors that bind to these predicted sites are known to affect regulation of the *myod* promoter. Each of the selected sites will now be briefly discussed in turn with regard to why the site was selected.

The AP1 and first Ets site partially overlap. Ets is known to coregulate the binding of AP1 to target promoters and so these two sites were of particular interest. The SRF binding site that was not selected for mutation overlaps with both the AP1 and first Ets sites. SRF is known to regulate the *myod* promoter through the DRR and, therefore, could also regulate the *myod* promoter through other means[17]. This fact elevated the importance of all three sites as this could be a site of potential competitive binding interactions.

The NFY site is known to bind Runx2 which, as discussed in section 2.6, is known to regulate the osteoblastic differentiation pathway[18] that C2C12 cells are known to be able to follow[3]. The second Ets site is adjacent to the NFY site, which could indicate that the Ets site is a site of coregulation of Runx2 binding.

The Ebox site, which is known to bind MyoD, is adjacent to a FOXO site, which is able to bind members of the FOXO family. The MyoD protein is known to autoregulate the *myod* gene[19] and members of the FOXO family are known to regulate differentiation in several cell types[20,21] including myogenesis[22] and is a known mediator of insulin signalling in cells[23]. An additional Ebox overlapped with the FOXO site and could not be mutated without affecting the FOXO site.

Enhanceosomes require specific arrangements of transcription factors in order to function[24]. The final binding site selected for mutation was capable of binding the

architectural protein Lef1. Lef1 is capable of bending DNA when binding to its specific binding site[25] and could therefore be important in the formation of a specific enhanceosome structure that could be responsible for the regulation of *myod*. Furthermore, Lef1 mediate the Wnt signalling response for some genes[26] and has been implicated in playing a role in somitogenesis[27].

In additional to the investigatory mutations, an artificial binding site was added to the CRM B sequence. This artificial binding site was specific to a bacterial DNA binding protein that is able force DNA to loop. In this case, the binding protein, itself a dimer, forms a tetramer with another dimer bound to another identical binding site. The non-covalent interaction forces the DNA to loop, the idea being to be able to force the interaction of two CRMs by coexpressing the specific bacterial DNA binding protein. This binding site was not used in this study, but is available for use in subsequent studies.

9.3.2. Selection of several members of the mutant library for further analysis

Several mutant sequences from the previously assembled mutant library (see chapter 8) were selected for testing in a CRM-plasmid context. Due to the limited amount of time available, only a limited set of sequences could be tested. Each sequence in the selected set contained a single mutation at a different target site whilst the rest remained as wild-type sequence. The highly expressing plasmid from the previous experiment (figure 9.5, dark orange) was selected as the vector for this experiment. The CRM combination for this plasmid had the largest CRM-B dependant effect of all the plasmids measured (see figure 9.6): When B is removed the plasmid expresses at a low level and when B is added the plasmid expresses highly. It was hoped that by choosing this vector system a full range of B activities will be observable.

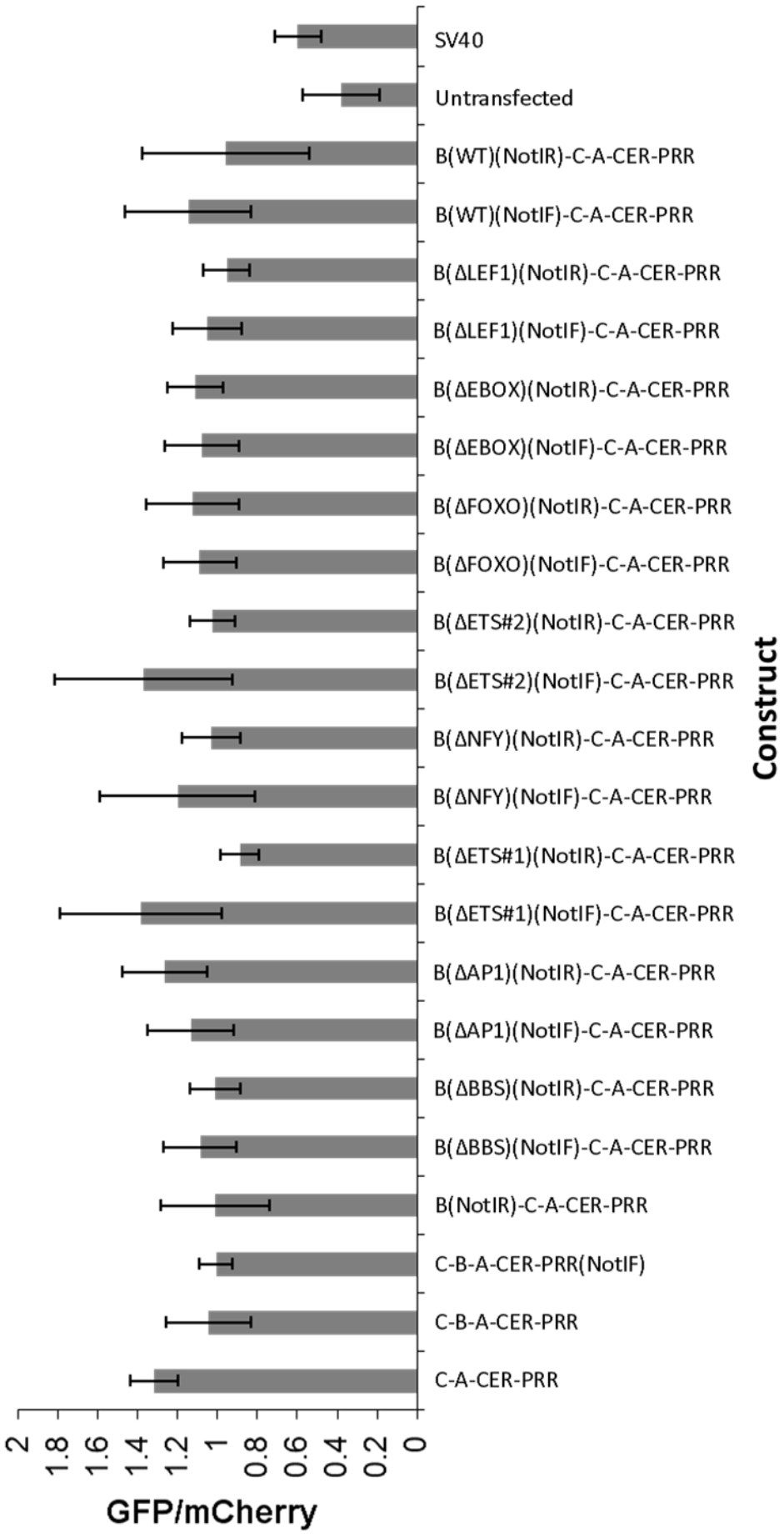


Figure 9.6: Comparison of CRM-B mutant expression. Each mutant is cloned into the NotI site in either the forward or reverse conformations and the expression of each is compared to the expression of the wild-type and the normal CRM-B. Mutation of a site is represented with a Δ[site name]. Bars represent the average GFP expression normalised against a mCherry control of >10000 cells obtained from four identical repeats, with duplicates in each repeat. Error bars represent the mean ± 1 standard deviation.

Figure 9.6 shows the relative expression of several versions of mutated CRM-B sequences in either the forward or reverse conformations in the context of other CRMs. All the tested combinations expressed highly, above the positive control. The results here (bars 2-4) confirm the results from the position orientation experiment, which was performed with the same plasmids. The CER-A-C plasmid was expected to only express at a low level, however, and here it can be seen expressing at a higher level than the CER-A-B-C plasmid (compare bars 1 and 2). All the plasmids tested expressed higher than untransfected cells, indicating that the GFP-containing plasmids were having detectable level of expression over background autofluorescence.

Although there appear to be some minor differences, such as with CER-A-C(ΔETS#1)(NotIR) these differences are not regarded as significant. It appears from the data presented in figure 9.6 that single site mutations of CRM-B in the context of CER-A-C do not have a significant effect on overall expression.

Since the results presented here are the average expression of >10000 cells, differences in expression between subpopulations of cells might be missed. Gene regulation is a stochastic event. There is some evidence that gene expression occurs in bursts of high activity followed by periods of relative silence[28,29]. Regulation of noise in gene expression, where one cell or population of cells expresses highly and another expresses at a lower level is an important aspect of gene regulation[30,31]. Processes such as quorum sensing in bacteria, where group behaviours are coordianated, have been shown to involve 'noisy' gene regulation[32]. The results shown in figure 9.6 could indicate a role for the mutated sites in inducing 'noisy' gene expression. Some constructs, particularly B(ΔETS#1)(NotIF)-C-A-CER-PRR, B(ΔNFY)(NotIF)-C-A-CER-PRR, B(ΔETS#2)(NotIF)-C-A-CER-PRR, B(WT)(NotIF)-C-A-CER-PRR and B(WT)(NotIR)-C-A-CER-PRR, all exhibit a high level of noise their average expression value between replicates, but not between duplicates. In contrast, some constructs, particularly C-B-A-CER-PRR, B(ΔETS#1)(NotIR)-C-A-CER-PRR and

B(ΔLEF1)(NotIR)-C-A-CER-PRR exhibit much lower noise levels, matching the noise level of the SV40 positive control. Interestingly, the ETS#2 and the NFY site bind to overlapping sites on CRM-B, as seen in chapter 3, figure 2.6. This information could indicate that this site has a role in the regulation of noisy gene expression from the *myod* promoter.

It is possible that the regulatory effect of the mutations was not evident due to the model that was chosen. C2C12 cells are tetraploid (see section 9.1.1) and therefore could be unable to reproduce the appropriate regulatory environment in which the mutations could have significant effects. It is possible that significant differences in expression caused by the mutations are not seen at this differentiation time point.

It should be noted that, based on the results from P. Downton (see section 2.5.1, figure 2.5), the level of expression of C-A-CER-PRR was expected to be much lower than the level of C-B-A-CER-PRR. This was not observed in the data presented in figure 9.6, comparing the first two columns. The results presented here are the average of four identical repeats of the same experiment, with duplicate wells in each repeat. Furthermore, the expression level of each construct is in excess of that observed with the positive control SV40 promoter, which is consistently observed in both of P. Downton's results. The confidence in these results is, therefore, high.

9.3.3. Mutation analysis of A-CER construct

Mutation of single sites within CRM B in the context of the CRMs CER, A and C did not have a significant effect on overall gene expression. The decision to use the C-B-A-CER-PRR/C-A-CER-PRR system for the combinatorial mutant library investigation was made on the basis of the results obtained by H. Crutzen (summarised in figure 9.1) which indicated a strong CRM B-dependant downregulation in expression levels. The later results from P. Downton, made with updated plasmids that corrected an expression issue with some of the earlier plasmids and using a flow cytometer calibrated with fluorescent beads, changed this

indication. Instead of a B-dependant downregulation, a B-dependant upregulation was observed that was not as strong as the previous downregulatory effect. For these reasons, the mutational analysis of CRM A in the context of A-CER-PRR was performed by P. Downton at the University of Warwick. Several sites within CRM A were individually mutated by a site directed mutagenesis (SDM) approach. Figure 9.7 shows the results from this individual site mutation experiment. As can be clearly seen in figure 9.7, the mutation of individual sites in CRM A within the context of A-CER-PRR has a significant effect on the level of overall normalised gene expression. This data, obtained by P. Downton, indicates that single site mutation of a CRM within the context of other CRMs is able, in contrast to the results presented in section 9.3, to have a significant, measureable effect on relative gene expression.

It is possible that whilst the effect of the single site mutations within the context of a large collection of CRMs is lost within the morass of completing signals coming from each CRM, the effect of single site mutations can be seen within the context of only a pair of CRMs.

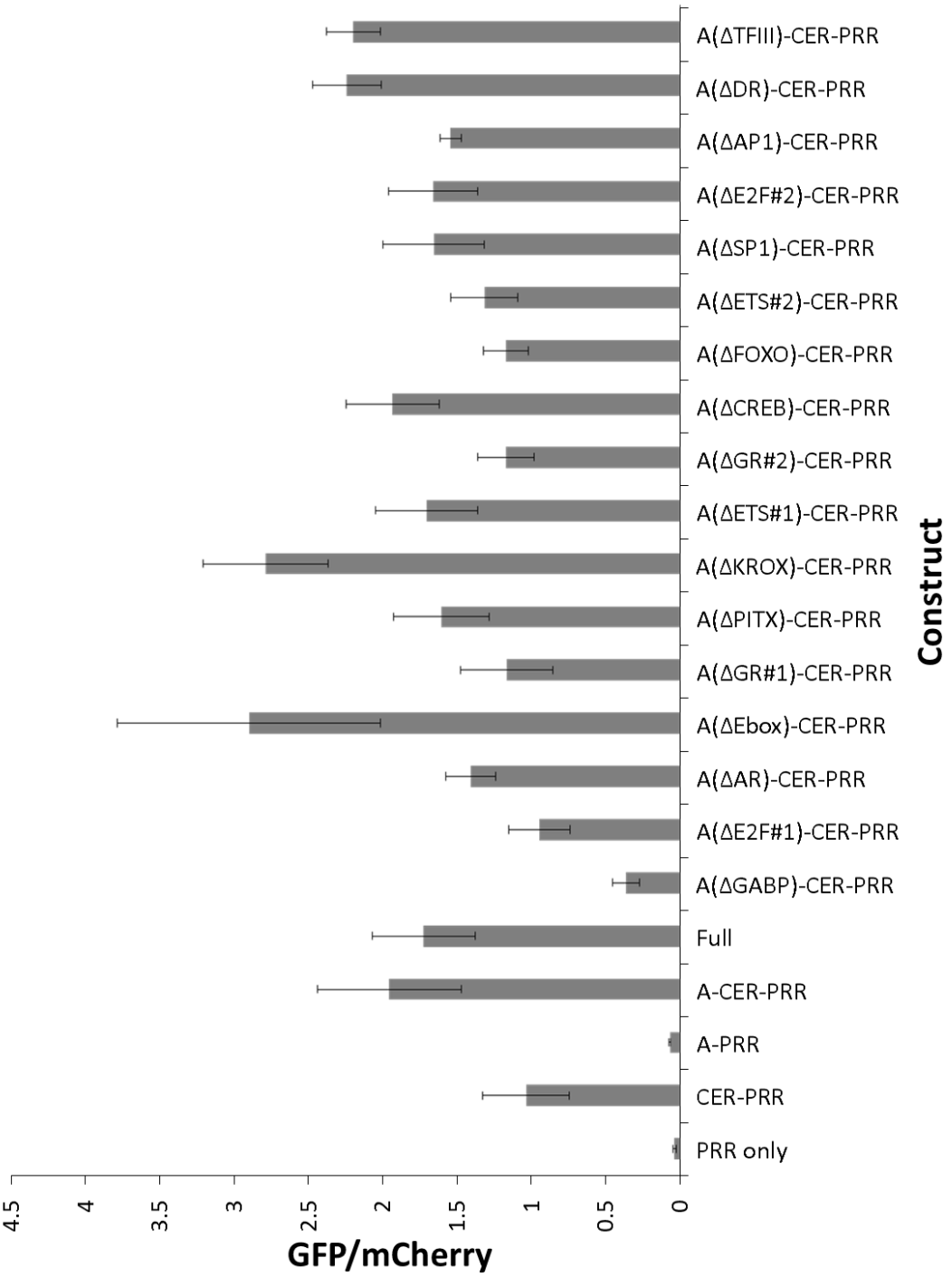


Figure 9.7: Graph of relative normalised expression values for several individually mutated sites in CRM A expressed within the context of A-CER-PRR. Bars represent averages of two repeats whilst error bars represent one standard deviation of these values. 'Full' refers to the full construct that contains all CRMs. NB Data obtained by P. Downton.

9.4. Further work

Not all the highly interesting sites could be mutated according to the used mutation strategy. The mutation strategy, contiguous mutations ≥4 bp long that obliterate the

selected site without affecting the binding scores of the neighbouring or overlapping sites, was determined in response to the downstream separation strategy that was to be used for the separation of specific mutants from the one-pot Gao assembly. Although the Gao assembly was not used (see chapter 8), the mutation sites sequences remained the same. It is possible that by using another mutation strategy, the additional highly interesting sites described in section 9.3 could be analysed.

Although the analysis of the CRM B mutants in the context of C-B-A-CER-PRR showed that none of the sites necessary for a high level of expression it is possible that the CRM B mutant library could be used in another CRM context. In the case of successful identification of sites contributing to the regulation of the *myod* promoter two further avenues of investigation are available: The binding of the predicted factor to the CRM could be confirmed by ChIP and the factor responsible for the affect could be identified by reproduction of the effect in the presence of the wild type site when the specific factor is knocked down using RNA interference (RNAi).

The current combinatorial mutant library replaces bioinformatically identified transcription factor binding sites with null sequences (as described in section 2.7). Several similar investigations could be envisaged from this starting point. The entire sequence except for the identified binding sites could be replaced with null sequence and specific sites combinatorially reintroduced in order to develop a full understanding of the sites necessary for various functions of the CRM. Additionally, the enhanceosome model of CRM action involves the interaction of CRMs away from the core transcriptional machinery, artificially driving the CRMs to interact using unique bacterial binding sites could elucidate whether sites, or combinations thereof, are involved in CRM-CRM interaction verses CRM activity. To this end, a bacterial binding site was introduced to each of the sequences, but due to constraints of time these sites were not employed.

9.5. Conclusions

This chapter presents evidence that the position and orientation of CRMs in the context of transiently expressed plasmids does not affect how those plasmids contribute to the regulation of *myod1* promoter also on the plasmid. This is important evidence in support of the looping hypothesis and against the billboard hypothesis.

Because mutation of single sites in CRM-B did not indicate any significant effects, mutation of several sites simultaneously should be performed. For instance, mutation of all the Ebox sites should be performed. The flexibility of the assembly process allows additional mutation sites beyond the original set to be incorporated in to the library subsequently. For instance, although ChIP data indicated that NF-κB was not bound to the relevant site in the developmental context, the NF-κB site could be mutated by replacing the relevant oligonucleotides with appropriately designed oligonucleotides.

In the event that significant differences were found in response to deleting a binding site within a given CRM, the putative hypothesis would be that the predicted binding factor is then unable to bind. This hypothesis could be tested by using RNAi against this factor, a similar expression level in response to knock down of the factor using the wild-type sequence as when using the deleted sequence would support the hypothesis. Furthermore, ChIP and/or DNaseI footprinting could be performed to confirm that the specific factor or other factors were bound to the CRM.

9.6. References

1. Yaffe, D. & Saxel, O. A Myogenic Cell Line with Altered Serum Requirements for Differentiation. *Differentiation* **7**, 159-166 (1977).
2. Buckingham, M. Skeletal muscle progenitor cells and the role of Pax genes. *Comptes rendus biologies* **330**, 530-3 (2007).
3. Katagiri, T. *et al.* Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage. *The Journal of cell biology* **127**, 1755-66 (1994).
4. Yeow, K., Phillips, B., Dani, C., Cabane, C. & Zoubir, E. Inhibition of myogenesis enables adipogenic trans-differentiation in the C2C12 myogenic cell line. *FEBS Letters* **506**, 157-162 (2001).
5. Yoshida, N., Yoshida, S., Koishi, K., Masuda, K. & Nabeshima, Y. Cell heterogeneity upon myogenic differentiation: down-regulation of MyoD and Myf-5 generates “reserve cells”. *Journal of cell science* **111 (Pt 6**, 769-79 (1998).
6. Chang, B.H., Smith, L., Huang, J. & Thayer, M. Chromosomes with delayed replication timing lead to checkpoint activation, delayed recruitment of Aurora B and chromosome instability. *Oncogene* **26**, 1852-61 (2007).
7. Lassar, A.B., Paterson, B.M. & Weintraub, H. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* **47**, 649-56 (1986).
8. Subramanian, S. & Srienc, F. Quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. *Journal of biotechnology* **49**, 137-51 (1996).
9. Dodds, E., Dunckley, M.G., Naujoks, K., Michaelis, U. & Dickson, G. Lipofection of cultured mouse muscle cells : a direct comparison of Lipofectamine and DOSPER. *Gene Therapy* **5**, 542-551 (1998).
10. Felgner, P.L. *et al.* Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 7413-7 (1987).
11. Jiang, C. & Pugh, B.F. Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews. Genetics* **10**, 161-72 (2009).
12. Soboleski, M.R., Oaks, J. & Halford, W.P. Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **19**, 440-2 (2005).
13. Mercola, M., Goverman, J., Mirell, C. & Calame, K. Immunoglobulin Heavy-Chain Enhancer Requires One or More Tissue-Specific Factors. *Science* **227**, 266-270 (1985).
14. Carey, M. & Smale, S.T. *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*. 640pp (Cold Spring Harbor Laboratory Press: 2000).
15. Sauer, B. Functional Expression of the cre-lox Site-Specific Recombination System in the Yeast Saccharomyces cerevisiae. *Molecular and Cellular Biology* **7**, 2087-96 (1987).
16. Wingender, E. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Briefings in bioinformatics* **9**, 326-32 (2008).
17. L’honore, A. *et al.* Identification of a New Hybrid Serum Response Factor and Myocyte Enhancer Factor 2-binding Element in MyoD Enhancer Required for MyoD Expression during Myogenesis. *Molecular Biology of the Cell* **18**, 1992-2001 (2007).
18. Lian, J.B. *et al.* Networks and hubs for the transcriptional control of osteoblastogenesis. *Reviews in endocrine & metabolic disorders* **7**, 1-16 (2006).
19. Thayer, M.J. *et al.* Positive autoregulation of the myogenic determination gene MyoD1. *Cell* **58**, 241-8 (1989).
20. Tuteja, G. & Kaestner, K.H. SnapShot: forkhead transcription factors I. *Cell* **130**, 1160 (2007).
21. Tuteja, G. & Kaestner, K.H. Forkhead transcription factors II. *Cell* **131**, 192 (2007).
22. Kitamura, T. *et al.* A Foxo/Notch pathway controls myogenic differentiation and fiber type specification. *The Journal of Clinical Investigation* **117**, 2477-2485 (2007).

23. Taniguchi, C.M., Emanuelli, B. & Kahn, C.R. Critical nodes in signalling pathways: insights into insulin action. *Nature reviews. Molecular cell biology* **7**, 85-96 (2006).
24. Panne, D. The enhanceosome. *Current opinion in structural biology* **18**, 236-42 (2008).
25. Love, J.J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791-795 (1995).
26. Eastman, Q. & Grosschedl, R. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Current opinion in cell biology* **11**, 233-40 (1999).
27. Galceran, J., Sustmann, C., Hsu, S.-C., Folberth, S. & Grosschedl, R. LEF1-mediated regulation of Delta-like1 links Wnt and Notch signaling in somitogenesis. *Genes & development* **18**, 2718-23 (2004).
28. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183-6 (2002).
29. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216-26 (2008).
30. Paulsson, J. & Ehrenberg, M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Physical review letters* **84**, 5447-50 (2000).
31. Pedraza, J.M. & van Oudenaarden, A. Noise propagation in gene networks. *Science (New York, N.Y.)* **307**, 1965-9 (2005).
32. Teng, S.-W. *et al.* Measurement of the copy number of the master quorum-sensing regulator of a bacterial cell. *Biophysical journal* **98**, 2024-31 (2010).

Chapter 10

10. Conclusions

Gene expression in higher animals is regulated, in part, by complex interactions between CRMs that facilitate the formation and release of RNAPII from target promoters. Internal and external information is integrated at CRMs by means of competitive and cooperative binding of transcription factors to specific sites. Transcription factors bound to CRMs can enhance or silence target genes by direct interaction with the core transcriptional machinery or indirectly by affecting chromatin conformation or recruitment of coactivators. The way the expression of a gene changes in response to stimuli is, therefore, hardwired into the CRMs by means of the arrangement of the different transcription factor binding sites present therein. Current methods for investigating the mechanisms by which transcription factors cooperate or compete to facilitate proper CRM function are not sufficiently combinatorial to address the huge potential interaction space of a system that can contain several CRMs each with several transcription factor binding sites. Chapter 1 describes the mechanisms, as they are currently understood, underpinning CRM interactions that bring about regulation of genes in a variety of contexts. This chapter then concludes that a combinatorial, holistic approach is required to efficiently understand the mechanisms behind CRM function in higher animals. It is likely that through further study of these mechanisms functional motifs will emerge that can then be identified in other genes.

This thesis describes the development of novel microfluidic, computational and molecular biology techniques for the generation of a combinatorial mutant library (all possible combinations of 9 binary sites; 512 sequences). This mutant library consists of variants of a CRM sequence that has previously been shown to, in cooperation with other

previously identified CRMs, synergistically regulate the *myod* promoter. The current understanding of the regulation of *myod* and how the expression of this gene relates to muscle specification in the developing embryo and adult muscle is described in chapter 2. The interactions synergistic effects of three other recently identified CRMs, in combination with two CRMs that have been known about for some time, are also described in chapter 2.

10.1. Microfluidics for biological and chemical applications

Microfluidic devices, particularly droplet microfluidics, fabrication techniques, control methods and are reviewed in chapter 3. This review served as the basis for the design of the PDMS microfluidic device described in chapter 6. The EnvisionTec Perfactory mini MSL process was used, in chapter 5, to produce novel microfluidic devices both directly and indirectly, by casting of PDMS in MSL moulds. These devices were then applied, in collaboration with others, to different experimental situations, including microbial biofilm culture and electrochemistry. In addition to making monolithic, inactive devices, moulds made by the MSL process were used to make patterned PDMS layers that were subsequently assembled into microfluidic devices by multilayer soft lithography, as described in chapter 6.

Two flow cells were fabricated with chambers a single or a few layers thick (see section 5.4). One of these flow cells were used to improve upon currently available radial flow cells and the second was used to make measurements of low concentrations of a specific solute down to the nM level. Both these flow cells and an additional flow cell, that was used for microbial biofilm culture (see section 5.3.1), utilised the MSL procedure's ability to make complicated internal geometries in one process to make monolithic devices with appropriately placed channels. The flow of fluids through the radial flow cell and the flow cell used for biofilm culture were analysed by COMSOL fluid dynamics modelling. This

modelling indicated that the flow cell design was appropriate for the functions required (see sections 5.3.1 and 5.4).

A final flow cell was assembled from multiple parts, each fabricated by the MSL procedure as well as quartz-quality glass (see section 5.5). This final flow cell overcame the fact that the material that MSL parts are made from is not sufficiently optically clear for some applications. This flow cell was then applied to make highly sensitive UV-absorbance measurements.

Chapter 6 describes the assembly of a microfluidic device for the contamination-free mixing of oligonucleotides for the assembly of a CRM mutant library. The device was fabricated using moulds made from MSL to create patterned layers of PDMS that were assembled by multilayer soft lithography (see section 6.2 and 6.3). The device made use of pneumatic 'Quake valves' to control the fluid flow necessary to make droplets of the correct size with the correct timings (see sections 6.4, 6.7 and 6.12). The device is applied to the mixing of the oligonucleotides that were subsequently assembled and amplified off chip. The results of this experiment indicated that the device was capable of making the mixtures in the right stoichiometry without contamination (see section 6.17). Furthermore, sequencing of the assembled sequences indicated that the error rate was comparable to that achieved by bench-top assembly.

By overcoming the issue of improper casting of PDMS on MSL parts, this thesis has shown it is possible to fabricate PDMS layers suitable for multilayer soft lithography (see section 5.3.3). Furthermore, since MSL is a rapid, maskless technique that can produce a finished cast in 2-3 hours, layers and finished devices can be produced quickly; 24 hours from design to device.

A system of valves was set up for the independent control of the 8 on-chip microvalves (see section 6.10). The design of the valve, described herein (see section 6.4),

shows it is possible to achieve simultaneous and contamination-free merging of droplets. These valves controlled the formation of droplets from each of a set of 8 reservoirs that were merged to make the appropriate oligonucleotide mixtures in the microfluidic device. To fabricate these valves, a PDMS membrane was incorporated using a PDMS mortar procedure optimised herein (see section 6.5). Reviews of the literature had indicated that sealing the device in this manner resulted in the strongest seals possible (see section 3.10).

To use the resulting microfluidic device to make mixtures suitable for assembly of CRM mutant sequences, the operation of the device was characterised. The flow rate through the valves was quantified using a real-time droplet measuring system based on a USB microscope (see section 6.12). Software and systems necessary for the control of the fluid flow and monitoring of the droplet output were developed (see sections 6.13 – 6.16). These systems allowed the control of the valve opening signal to occur with a 1 ms time resolution, which is less than the solenoid valve opening time as estimated by the manufacturer.

The mixing of 8 reagents through the use of droplet microfluidics using a PDMS device composed of layers patterned using moulds made by MSL has not been reported in the literature. In this project, feasibility of using the device to generate mixtures suitable for mutant CRM assembly is demonstrated. It is possible apply the current device to other situations requiring precise addition and mixing of multiple reagents in droplets.

10.2. Optimisation of oligonucleotide overlap sequences

Currently available software was investigated (see section 7.1.2) but found not to be suitable for the optimisation of sequences that could be assembled into a CRM mutant library (see section 7.1.1). As a result, novel software, named OptiCut was developed for the optimisation of DNA sequences necessary for the assembly of the CRM mutant library in chapter 8. The software described in chapter 7 is written in MATLAB and produces a

minimal set of optimal oligonucleotides. Subsets of this set can then be mixed and assembled into members of a combinatorial mutant library for a given CRM.

The efficiency of the optimisation algorithm is investigated and the time taken for optimisation was found to be proportional to the number of sequences in the library rather than sequence length. The efficiency of the algorithm was assessed by comparison to the optimisation algorithm used by the Gao lab and the OptiCut algorithm was found to significantly reduce the spread of oligonucleotide melting temperatures (see section 7.6). Various modules were implemented for the OptiCut program; identification of competitors can be performed to check that the desired sequences will be produced by the assembly process and reduction of costs by minor sequence adjustment (see sections 7.5 and 7.6 respectively). An OptiCut GUI is developed, which allows non-experienced users to optimise oligonucleotide sequences inputted via the widely used FASTA format (see section 7.8). Oligonucleotide sequences can then be outputted in a variety of formats for downstream use. Finally, the effectiveness of the optimisation process is demonstrated by the assembly of a set of optimised sequences using bench-top methods.

10.3. Optimisation of CRM assembly

The assembly of the CRM combinatorial mutant library was attempted using two, related gene assembly processes. Both methods used a form of ligative assembly that had been shown to be successful in the past[1–4]. The first method, termed 'Gao synthesis', used sets of oligonucleotides that were synthesised in one step and simultaneously assembled in one pot after initial amplification (see sections 4.4.2 and 4.4.3). Despite a significant optimisation process, described in chapter 8, this method did not yield successful assemblies (see section 8.2). The Gao assembly process did yield a partial product, however, that matched in sequence with what was expected (see section 8.2.7). It is possible that a full length could have been obtained by changing the primers for the full

length amplification and further optimisation. The generally poor quality of the sequencing results (see section 8.2.7) and the potential difficulty with sampling individual sequences from the one-pot assembly (see section 4.4.2) indicated that this approach was not likely to yield sufficient results within the available time and therefore this approach was abandoned.

Another assembly method was developed to generate the sequences for the combinatorial mutant library due to the failure of the Gao assembly method (see section 8.3). The new method avoided used individual, commercially synthesised, purified and phosphorylated oligonucleotides whose quality could be ensured. By using individual oligonucleotides, the sequences could be assembled independently (multi-pot assembly), avoiding problems with competitive binding interactions. Furthermore, the assembled sequences would then not need any downstream purification. To ensure the oligonucleotides sequences were optimal an optimisation algorithm, OptiCut, was developed, which is described in chapter 7.

Finally, the successful assembly of the CRM mutant library using methods optimised on the bench-top for the assembly of the OptiCut-optimised sequences is demonstrated in chapter 8. The error rate of the assembly was determined to be 0.42% (see section 8.3.7).

The assembly process was a successful proof of concept of both the OptiCut optimisation strategy and the assembly methods that were developed as part of the Gao assembly optimisation. This approach could be applied to other systems to generate mutant libraries that are not feasible with traditional techniques such as SDM or gene ablation.

10.4. Investigation of CRM position/orientation and CRM mutation analysis

In the final results chapter, chapter 9, the involvement of a model CRM (CRM-B) in the regulation of *myod* in the context of other CRMs cloned into a model plasmid was investigated. Previous work investigating the regulatory interactions of the CRMs and the *myod* promoter had always involved the CRMs being cloned in the same relative positions and orientations. To test whether the position or orientation of the CRMs affects the regulatory contribution of a given CRM, CRM B was cloned upstream of other CRM elements and in either the forward and reverse confirmations (see section 9.3). The results of this experiment indicated that the position and orientation of the CRMs had no effect on the expression from the *myod* promoter.

To begin to investigate the mechanisms through which the CRMs interact to bring about regulation of the *myod* promoter the combinatorial mutant library of CRM-B produced in chapter 8 was used. By cloning the CRM B sequence in the context of other CRMs in the model plasmid, changes in expression levels when compared to the wild type CRM-B sequence could be assigned to specific sites within CRM-B (see section 9.3.2). The mutation of specific sites of CRM B did not result in significant changes in *myod* promoter activity. Interestingly, however, single site mutation of another CRM within the context of a different set of CRMs, work by P. Downton, did result in significant changes in expression levels, indicating that the mutation of sites within individual CRMs can be measured (see section 9.3.1).

10.5. Future work

The microfluidic devices developed during this thesis have shown that MSL can be used to make useful devices either monolithically or by multilayer soft lithography. There is significant possibility for the fabrication of bespoke microfluidic devices for a variety of

applications in chemistry and biology. The key items of further work with regards to the PDMS microfluidic device are listed below and then each is discussed in the following text.

- Integration of additional valves (~30) for the control of all necessary oligonucleotide solutions.
- Integration of a multiplexer to control a larger (>30) number of valves.
- Incorporation of heating elements so that the process of gene assembly can occur on chip.

With the current device, the oligonucleotides necessary for only two members of the mutant library can be assembled in a single run without switching reservoirs. Ideally, sufficient valves could be incorporated onto the device that all the necessary solutions to enable the mixing of the oligonucleotides necessary to assembled any of the sequences from the library (~30 solutions). The type of layer annealing process used here, partially cured mortar, results in the strongest interlayer bonding (see sections 3.10, 6.2 and 6.3) and could therefore support a relatively high density of valves (up to 1 valve per mm^2). Packaging of a sufficient number of valves to enable individual control of each of the solutions required to make the full library is, therefore, possible.

An even larger number of reagents (>30) could be mixed by either operating multiple chips in parallel or increasing the number of available valves. Operating multiple chips is problematic because it requires the transfer of droplets between devices in a reliable manner. Similarly, the number of valves that can be placed on a device is limited by the space available. A multiplexer can enable the control a larger number of valves using less than 1 off-chip valve per on-chip valve[5]. The area required to interface with an on-chip valve is larger than the area of the valve. To make maximum use of the available space, therefore, a multiplexer could be incorporated that would minimise the interfaces

necessary to control sufficient valves. This approach could compromise the fidelity of the valve operation, however, as every line is indirectly connected each other.

By incorporating heating elements and suitable reaction chambers, the entire process of gene assembly to be miniaturised to fit on a microfluidic chip similar to that demonstrated here. Due to the fact that PDMS is relatively permeable to water vapour, the internal or external surfaces of the device would have to be coated with a layer impermeable to water vapour to prevent evaporation affecting the reaction (see sections 2.12 and 6.7). Increasing the proportion of the workflow that occurs in the microfluidic device would improve the efficiency of the process of synthetic gene assembly.

Further stages of the process can be incorporated onto the microfluidic device; addition of an appropriately prepared, linearised vector could enable one step assembly and cloning into the target vector. Modification of the linearised vector with topoisomerase at each end, similar to the TOPO® TA cloning® kit from Invitrogen Ltd., could be used to maximise cloning efficiency. Once circularised, this vector could be used to transform *E. coli* for growing up[6]. Once grown, the amplified plasmid could then be isolated from the culture on the microfluidic device[7]. In a following step, this plasmid DNA could be used to transform cells of a suitable type[8]. Given the appropriate intermediate quality control steps or *post hoc* analysis, a miniaturisation of the whole process can be envisaged that would enable the rapid discovery of CRM interaction motifs in a variety of cellular situations.

The OptiCut software can be readily applied to the optimisation of libraries of sequences for applications besides investigation of regulatory modules, such as directed evolution. The following list summarises the key items of further work.

- Additional modules could be readily implemented for the OptiCut programs for the optimisation of oligonucleotides from PCR-based gene assembly or codon optimisation for protein expression.

- A module that generates appropriate mutant sequences at sites selected by the user, possibly by interacting with the BiFa tool, could improve optimisation. This would be achieved by reducing or avoiding the optimisation limit caused by sequence differences between the mutant and wild-type versions of the sequence.
- At this stage the competitive binding module only indicates that potential competitive binding interactions may occur. This module could be made significantly more useful if it suggested changes to the optimised sequences that would minimise or avoid these competitive interactions.
- The core of the optimisation algorithm could be applied to one of the commercially available sequence optimisation software.

Although the testing of the CRM B mutant library indicated that there was no measureable effect of mutating the individual sites predicted to be present on CRM B, significant effects were observed following the mutation of individual sites on CRM A. By applying the same library generation procedure instead to CRM A, the combinatorial interactions that facilitate the expression levels observed in this latter context can be examined.

In this project the effect of each of the DNA constructs was assessed by transient transfection performed on large numbers of cells that were, at the appropriate time point, fixed and analysed by flow cytometry. Although flow cytometry is able to provide simultaneous expression data from a large number of cells in a short amount of time, the process gives just a snapshot in time. Ideally, the effect of each construct could be monitored through time so changes in gene expression could be mapped in individual cells through time. This could be achieved by using real-time fluorescence microscopy that, with appropriate cell tracking software, could be used to obtain expression data in hundreds of cells over time. Furthermore, by multiplexing using multi-well plates, hundreds of constructs could be assessed simultaneously. The relatively long lived, slow to mature GFP

protein can obfuscate meaningful through time expression data, however. An alternative reporter system could be used instead to avoid this problem.

In summary, this project has addressed a bottleneck in the investigation of CRM function. That is, the generation of a set of sequences that constitutes a library of combinatorial mutants; where each sequence possesses a unique combination of transcription factor binding sites. The prioritisation process, assembly methods and optimisation algorithms presented herein can be applied to any genetic system where combinatorial investigation can be usefully applied. A droplet microfluidic device was presented that can be applied to any situation where rapid, accurate and contamination free mixtures of up to 8 constituents are required. This system was successfully applied herein to the generation of oligonucleotide mixtures necessary for the assembly of members of a combinatorial mutant library.

10.6. References

1. Chalmers, F.M. & Curnow, K.M. Scaling up the ligase chain reaction-based approach to gene synthesis. *BioTechniques* **30**, 249-52 (2001).
2. Horspool, D.R., Coope, R.J. & Holt, R. a Efficient assembly of very short oligonucleotides using T4 DNA Ligase. *BMC research notes* **3**, 291 (2010).
3. Au, L.-chun, Yang, F.-yuan, Yang, W.-jung, Lo, S.-hwa & Kao, C.-feng Gene Synthesis by a LCR-Based Approach : High-Level Production of Leptin-L54 Using Synthetic Gene in Escherichia coli. *Biochemical and Biophysical Research Communications* **203**, 200-203 (1998).
4. Zhou, X. *et al.* Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Methods* **32**, 5409-5417 (2004).
5. Hua, Z. *et al.* A versatile microreactor platform featuring a chemical-resistant microvalve array for addressable multiplex syntheses and assays. *Journal of Micromechanics and Microengineering* **16**, 1433-1443 (2006).
6. Nagamine, K. *et al.* On-Chip Transformation of Bacteria accomplished for the first time using a microbial array. *Analytical biochemistry* **77**, 4278-4281 (2005).
7. Northrup, V. a, Backhouse, C.J. & Glerum, D.M. Development of a microfluidic chip-based plasmid miniprep. *Analytical biochemistry* **402**, 185-90 (2010).
8. Denoual, M. *et al.* Vacuum casting to manufacture a plastic biochip for highly parallel cell transfection. *Measurement Science and Technology* **17**, 3134-3140 (2006).

Appendix A

A. Appendix A: MATLAB code

This appendix contains a compilation of MATLAB code written during the course of this thesis. Given the appropriate input files, these programs should be able to run as written here.

A.1. Video reading code

This code was written to enable obtaining and analysing the real time images from a USB microscope. These images were then used to determine the size of droplets produced in the microfluidic chip described in chapter 6.

```
% function [DropDataStore,StatStore] = DropAnalQuery
% optional function call code.
%program to analyse video and determine drop sizes

%% User input information
[fileName,filePath] = uigetfile('C:\Documents and Settings\msrgbh\My
Documents\MATLAB\Image analysis\Drop anal\*.avi');
if fileName == 0, disp('No file selected... exiting...'); return;
end
videoObj = mmreader([filePath fileName]); % read the file. Some
codecs might not work and return an error here.

% writeName = [fileName(1:end-4) ' 70ms '];
nFrames = videoObj.NumberOfFrames;

% create the video windows with the appropriate properties
hVideo1 = video.VideoPlayer('WindowCaption', 'Original Video');
hVideo1.WindowPosition(1) = round(0.4*hVideo1.WindowPosition(1)) ;
hVideo1.WindowPosition(2) = round(1.5*(hVideo1.WindowPosition(2))) ;
hVideo1.WindowPosition([4 3]) = [200 200];

hVideo2 = video.VideoPlayer('WindowCaption', 'Motion Vector');
hVideo2.WindowPosition(1) = hVideo1.WindowPosition(1) + 350;
hVideo2.WindowPosition(2) =round(1.5* hVideo2.WindowPosition(2));
hVideo2.WindowPosition([4 3]) = [200 200];

hVideo3 = video.VideoPlayer('WindowCaption', 'bbox');
hVideo3.WindowPosition(1) = hVideo2.WindowPosition(1) + 350;
hVideo3.WindowPosition(2) = round(1.5*(hVideo3.WindowPosition(2))) ;
hVideo3.WindowPosition([4 3]) = [200 200];

h = waitbar(0,'Processing frames...');
```

```
% These values will change with every video.
LOIrow = 425;
LOIcolLeft = 155;
LOIcolRight = 181;
IOIlength = 300;

% periodicity of the droplets, number of frames. Necessary for the
% appropriate statistics
UpperPeriodLimit = 26;
LowerPeriodLimit = 19;

% setup data stores
NumDropsDetected = 0;
Step = 1;
valve = 1;
DropSizes = zeros(1,30);
DropSizeVector = zeros(1,11);
DropErrors = zeros(1,11);
figure(1);
N = 0;
StatStore = {'Frame index' 'Blob size'};
%% begin loop for image analysis.
DropDataStore = zeros(8,11,2);
for i = 1:nFrames
    image = read(videoObj,i);
%         IOI = image(LOIrow-
5:LOIrow+IOIlength,LOIcolLeft:LOIcolRight,:); % vertical droplets
    IOI = image(LOIrow-IOIlength:LOIrow+5,LOIcolLeft:LOIcolRight,:);
% horizontal droplets
    bwIOI = ~im2bw(IOI,0.45); %change to black and white. value
might change depending on the colour/brightness

    LOI = bwIOI(end-7,1:end);

    % commented code below is for use when changing the IOI
parameters
    % above.
%     image(LOIrow,LOIcolLeft:LOIcolRight,:) = 255 -
image(LOIrow,LOIcolLeft:LOIcolRight,:);
    %     figure(1) = imshow(image);

    %     if isempty(OldLOI), OldLOI = NewLOI; end
    %     ChangeInLOI = OldLOI - NewLOI;
    %     TotChangeInLOI = sum(ChangeInLOI(:));

%     subplot(2,2,[1 3]), imshow(image);
%     subplot(3,2,[1 3]), imshow(image);
%     subplot(3,2,2), imshow(IOI);
%     subplot(3,2,4), imshow(bwIOI);

    % update the frames
    step(hVideo1, image);
    step(hVideo2, IOI);
    step(hVideo3, bwIOI);

    [B,L,N] = bwboundaries(bwIOI,8,'noholes');
```

```
    % ignore any droplets that touch the sides. Record the sizes of
the
    % droplets in the statstore
    if N == 1 && sum(bwIOI(:,1)) == 0 && sum(bwIOI(:,end)) == 0 &&
sum(bwIOI(1,:)) == 0 && sum(bwIOI(end,:)) == 0
        %             subplot(3,2,[1 3]), imshow(image);
        %             subplot(3,2,2), imshow(IOI);
        %             subplot(3,2,4), imshow(bwIOI);
        DropSize = sum(bwIOI(:));
        StatStore(end+1,:) = {i DropSize};
    elseif N > 1
        for j = 1:length(B)
            logical1 = B{j}(:,1) > 1 & B{j}(:,1) < length(L(:,1));
            logical2 = B{j}(:,2) > 1 & B{j}(:,2) < length(L(1,:));
            if  sum(logical1) == length(logical1) && sum(logical2)
== length(logical2)
                BlobSize = sum(sum(L==j));
                if BlobSize > 5;
                    DropSize = BlobSize;
                    StatStore(end+1,:) = {i DropSize};
                end
            end
        end
    end

    % increment the counter and adjust the waitbar
    DropCount = length(StatStore(:,1)) - 1;
    waitbar(i / nFrames,h,['Processing frames... ' int2str(i) ' of '
int2str(nFrames) '. Drop count = ' int2str(DropCount) '. N = '
int2str(N)]);

end
%% Develop statistics
figure(3)
StatMatrix = cell2mat(StatStore(2:end,:));
LowBound = 30;
HighBound = 400;
for i = 1:120
    vector = StatMatrix(StatMatrix(2:end,1) > LowBound &
StatMatrix(2:end,1) <= HighBound,2);
    stdDropSize(i) = std(vector);
    vector(vector < stdDropSize(i) || vector > stdDropSize(i)) = [];
    DropSize(i) = mean(StatMatrix(StatMatrix(2:end,1) > LowBound &
StatMatrix(2:end,1) <= HighBound,2));
    stdDropSize(i) = std(vector);
    LowBound = HighBound;
    HighBound = HighBound + 400;
end
errorbar(DropSize,stdDropSize);
```

A.2. Probability bootstrapping

This code was used to determine the number of samplings would likely be required to have sampled every member of a mixture of assembled sequences. This code was used to obtain the data presented in section 8.1.7.

```
%512 balls.
Limit = 100000;
h = waitbar(0);
Store = zeros(Limit,1);
for j = 1:Limit
    AllSelected = 0;
    Population = zeros(512,1);
    Samples = 0;
    while ~AllSelected
        Selection = ceil(rand(1)*512);
        Population(Selection) = 1;
        if all(Population)
            AllSelected = 1;
            break
        end
        Samples = Samples + 1;
    end
    Store(j) = Samples;
    waitbar(j/Limit,h);
end

%%
Data = zeros(max(Store),1);
for j = 1:max(Store)
    Data(j) = sum(Store==j);
end

%%
CDF = zeros(length(Data),1);
for j = 1:length(Data)
    CDF(j) = sum(Data(1:j));
end
```

A.3. OptiCut program

The OptiCut optimisation algorithm was packaged within a GUI so that users who were not experienced in MATLAB could easily use the program. The following code is sufficient to generate the GUI, run the optimisation, generate the results graphs and format the outputs. This additional code is provided in the following sections. The OptiCut algorithm is presented in chapter 7. See section 7.8.1 for details of the additional components.

A.3.1. OptiCut GUI

```
function varargout = GroupCutGUI(varargin)
% GROUPCUTGUI M-file for GroupCutGUI.fig
%      GROUPCUTGUI, by itself, creates a new GROUPCUTGUI or raises
the existing
%      singleton*.
%
%      H = GROUPCUTGUI returns the handle to a new GROUPCUTGUI or
the handle to
```

```
%      the existing singleton*.
%
%      GROUPCUTGUI('CALLBACK',hObject,eventData,handles,...) calls
the local
%      function named CALLBACK in GROUPCUTGUI.M with the given input
arguments.
%
%      GROUPCUTGUI('Property','Value',...) creates a new GROUPCUTGUI
or raises the
%      existing singleton*.  Starting from the left, property value
pairs are
%      applied to the GUI before GroupCutGUI_OpeningFcn gets called.
An
%      unrecognized property name or invalid value makes property
application
%      stop.  All inputs are passed to GroupCutGUI_OpeningFcn via
varargin.
%
%      *See GUI Options on GUIDE's Tools menu.  Choose "GUI allows
only one
%      instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help GroupCutGUI

% Last Modified by GUIDE v2.5 12-Oct-2011 10:01:59

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @GroupCutGUI_OpeningFcn, ...
                   'gui_OutputFcn',  @GroupCutGUI_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT


% --- Executes just before GroupCutGUI is made visible.
function GroupCutGUI_OpeningFcn(hObject, eventdata, handles,
varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to GroupCutGUI (see VARARGIN)

% Choose default command line output for GroupCutGUI
handles.output = hObject;
```

```
% Update handles structure
guidata(hObject, handles);
% UIWAIT makes GroupCutGUI wait for user response (see UIRESUME)
% uiwait(handles.figure1);


% --- Outputs from this function are returned to the command line.
function varargout = GroupCutGUI_OutputFcn(hObject, eventdata,
handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
varargout{1} = handles.output;


% --- Executes on button press in Browse1_pushbutton.
function Browse1_pushbutton_Callback(hObject, eventdata, handles)
% hObject    handle to Browse1_pushbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
[FileName Path] = uigetfile([pwd '.fas'],'Select a .fas file
containing all sequences...');
if isequal(FileName,0) || isequal(Path,0)
    set(handles.Message2_text,'String','User pressed cancel when
selecting input filename');
else
    FullFileName = [Path '\' FileName];

setappdata(handles.Browse1_pushbutton,'FullFileName',FullFileName);
    [FASTAHeader, FASTASequence] = fastaread(FullFileName);

setappdata(handles.Browse1_pushbutton,'FASTAHeader',FASTAHeader);

setappdata(handles.Browse1_pushbutton,'FASTASequence',FASTASequence)
;
    NumSeq = length(FASTASequence);
    LengthSeq = length((FASTASequence{1}));
    MessageString = ['You have selected a file containing
',num2str(NumSeq),' sequences of length ',num2str(LengthSeq),'
bp.'];
    set(handles.Message1_text,'String',MessageString);
    MaxNumOligos = floor(LengthSeq/3);
    MessageString = ['Enter an integer value for # oligos between 3
and ' num2str(MaxNumOligos) ' and a number of iterations (100 is
suggested).'];
    set(handles.Message2_text,'String',MessageString);
    set(handles.Optimise_pushbutton,'Enable','on');
end


% --- Executes on button press in Optimise_pushbutton.
function Optimise_pushbutton_Callback(hObject, eventdata, handles)
% hObject    handle to Optimise_pushbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
%init
set(hObject,'Enable','off');
```

```
set(hObject,'String','Optimising...');
% refresh(handles.Optimise_pushbutton);
set(handles.Message2_text,'String','Please wait whilst sequences are
optimising...');
NumIterations_inputtext_Callback(handles.NumIterations_inputtext,
eventdata, handles)
OligosPerStrand_inputtext_Callback(handles.OligosPerStrand_inputtext
, eventdata, handles)

NumLig = getappdata(handles.OligosPerStrand_inputtext,'NumOligos');
FASTAHeader = getappdata(handles.Browse1_pushbutton,'FASTAHeader');
FASTASequence =
getappdata(handles.Browse1_pushbutton,'FASTASequence');
NumIterations =
getappdata(handles.NumIterations_inputtext,'NumIterations');

DetectEndpointFlag = get(handles.DetectEndpoint_checkbox,'Value');
CompIDFlag = get(handles.DetectCompetitors_checkbox,'Value');

LengthSeq = length((FASTASequence{1}));
MaxNumOligos = floor(LengthSeq/3);
if NumLig > 2 && NumLig < MaxNumOligos
    [RMSDVector,TmMatrix,UniqueSeqCell,CutSites,HitsStore] =
GroupCutINIT(FASTAHeader,FASTASequence,NumLig,NumIterations,DetectEn
dpointFlag,CompIDFlag);

    set(handles.RMSD_axes,'YLim',[0 ceil(RMSDVector(1)/10)*10]);
    plot(handles.RMSD_axes,0:1:length(RMSDVector)-1,RMSDVector,'x');

    BucketSize = 0.5;
    LowerBound = floor(min(TmMatrix(:))/10)*10;
    UpperBound = ceil(max(TmMatrix(:))/10)*10;
    histstore = plothist(TmMatrix,(UpperBound-LowerBound)*2+1);

bar(handles.Histogram_axes,LowerBound:BucketSize:UpperBound,histstor
e);
    set(handles.Histogram_axes,'XLim',[LowerBound UpperBound]);

    set(handles.HistLow_text,'String',int2str(LowerBound));
    set(handles.HistHigh_text,'String',int2str(UpperBound));

    setappdata(hObject,'SeqCell',UniqueSeqCell);
    setappdata(hObject,'RMSD',RMSDVector);
    setappdata(hObject,'TmMatrix',TmMatrix);
    setappdata(hObject,'CutSites',CutSites);
    setappdata(hObject,'HitsStore',HitsStore);
else
    MessageString = ['POTENTIAL ERROR DETECTED: You have selected
the number of oligos per strand to be ' int2str(NumLig) '. Enter an
integer value for # oligos between 3 and ' num2str(MaxNumOligos) '
and a number of iterations (100 is suggested).'];
    set(handles.Message2_text,'String',MessageString);
    return
end
set(hObject,'Enable','on');
set(hObject,'String','ReOptimise!');
% set(handles.Export_pushbutton,'Enable','on');
set(handles.Message2_text,'String','Sequences optimised, please
review statistics and export data.');
```

```
%cycle

%final


function OligosPerStrand_inputtext_Callback(hObject, eventdata,
handles)
% hObject    handle to OligosPerStrand_inputtext (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of
OligosPerStrand_inputtext as text
%        str2double(get(hObject,'String')) returns contents of
OligosPerStrand_inputtext as a double
NumOligosStr = get(hObject,'String');
NumOligos = str2double(NumOligosStr);
setappdata(hObject,'NumOligos',NumOligos);

% --- Executes during object creation, after setting all properties.
function OligosPerStrand_inputtext_CreateFcn(hObject, eventdata,
handles)
% hObject    handle to OligosPerStrand_inputtext (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end



function NumIterations_inputtext_Callback(hObject, eventdata,
handles)
% hObject    handle to NumIterations_inputtext (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of
NumIterations_inputtext as text
%        str2double(get(hObject,'String')) returns contents of
NumIterations_inputtext as a double
NumIterationsStr = get(hObject,'String');
NumIterations = str2double(NumIterationsStr);
setappdata(hObject,'NumIterations',NumIterations);
set(handles.RMSD_axes,'XLim',[0 NumIterations]);
% set(handles.RMSD_axes,'XTick',0:1:NumIterations);
% XTickLabelCell = cell(1,NumIterations+1);
% for i = 0:20:NumIterations
%     XTickLabelCell(i+1:i+20) = {int2str(i) '' '' '' '' '' '' '' ''
'' int2str(i+10) '' '' '' '' '' '' '' '' ''};
% end
% set(handles.RMSD_axes,'XTickLabel',XTickLabelCell);

% --- Executes during object creation, after setting all properties.
```

```
function NumIterations_inputtext_CreateFcn(hObject, eventdata,
handles)
% hObject    handle to NumIterations_inputtext (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes during object creation, after setting all properties.
function Message1_text_CreateFcn(hObject, eventdata, handles)
% hObject    handle to Message1_text (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called


% --- Executes during object creation, after setting all properties.
function Histogram_axes_CreateFcn(hObject, eventdata, handles)
% hObject    handle to Histogram_axes (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: place code in OpeningFcn to populate Histogram_axes


% --- Executes on button press in xls_radiobutton.
function xls_radiobutton_Callback(hObject, eventdata, handles)
% hObject    handle to xls_radiobutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of xls_radiobutton
set(handles.csv_radiobutton,'Value',0);

% --- Executes on button press in csv_radiobutton.
function csv_radiobutton_Callback(hObject, eventdata, handles)
% hObject    handle to csv_radiobutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of csv_radiobutton
set(handles.xls_radiobutton,'Value',0);

% --- Executes on button press in Export_pushbutton.
function Export_pushbutton_Callback(hObject, eventdata, handles)
% hObject    handle to Export_pushbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
OutputCell = getappdata(handles.Optimise_pushbutton,'SeqCell');
if get(handles.MinimiseOligos_checkbox,'Value')
```

```
    Sequences =
getappdata(handles.Browse1_pushbutton,'FASTASequence');
    Headers = getappdata(handles.Browse1_pushbutton,'FASTAHeader');
    CutSites = getappdata(handles.Optimise_pushbutton,'CutSites');
    [NumSites, SitePositions] = FindSiteSet(Sequences);
    ModifiedCutSites = CutSites;
    for j = 1:length(CutSites)
        CutSite = CutSites(j);
        for k = 1:NumSites
            MutSite = SitePositions{k};
            if CutSite >= MutSite(1) && CutSite <= MutSite(2);
                if CutSite-MutSite(1) < MutSite(2)-CutSite;
                    %cutsite is closer to start of mutsite
                    ModifiedCutSites(j) = MutSite(1)-1;
                elseif CutSite-MutSite(1) == MutSite(2)-CutSite
                    %cutsite is in the middle of the mutsite
                else
                    %cutsire is closer to the end of the mutsite
                    ModifiedCutSites(j) = MutSite(2)+1;
                end
            end
        end
    end

    if isequal(ModifiedCutSites,CutSites)
        %no modifications have been found
    else
        %modifications have been found
        [uniqueseqcell] =
GroupCutOutput(ModifiedCutSites,Sequences,Headers);
        uniqueseqcell(1,1:3) = {'Sequence' 'Participating sequences'
'Name'};

        OutputCell(1,1:3) = {'Sequence' 'Participating sequences'
'Name'};
        OutputCell(1:length(uniqueseqcell),5:7) = uniqueseqcell;
    end
else
    ModifiedCutSites = 0;
    %user has selected no modification, change ouput
    OutputCell = getappdata(handles.Optimise_pushbutton,'SeqCell');
    OutputCell(1,1:3) = {'Sequence' 'Participating sequences'
'Name'};
end

ExportFileName =
getappdata(handles.ExportBrowse_pushbutton,'ExportFileName');
OutputXLS = get(handles.xls_radiobutton,'Value');
if OutputXLS
    xlswrite(ExportFileName,OutputCell);
else
    csvwrite(ExportFileName,OutputCell);
end
Message = ['The file ' ExportFileName ' has been written
successfully.'];

HitsStore = getappdata(handles.Optimise_pushbutton,'HitsStore');
if get(handles.ExportInteractions_checkbox,'Value')
    InteractionsFilename = 'InteractionsData.xlsx';
    xlswrite(InteractionsFilename,HitsStore);
```

```
end
set(handles.Message2_text,'String',Message);

% --- Executes on button press in Stats_checkbox.
function Stats_checkbox_Callback(hObject, eventdata, handles)
% hObject    handle to Stats_checkbox (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of Stats_checkbox


% --- Executes on button press in MinimiseOligos_checkbox.
function MinimiseOligos_checkbox_Callback(hObject, eventdata,
handles)
% hObject    handle to MinimiseOligos_checkbox (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of
MinimiseOligos_checkbox


% --- Executes on button press in ExportBrowse_pushbutton.
function ExportBrowse_pushbutton_Callback(hObject, eventdata,
handles)
% hObject    handle to ExportBrowse_pushbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
OutputXLS = get(handles.xls_radiobutton,'Value');
if OutputXLS
    suffix = '.xlsx';
else
    suffix = '.csv';
end

[FileName Path] = uiputfile([pwd suffix],'Select a .fas file
containing all sequences...','OptiCut_output.xlsx');
if isequal(FileName,0) || isequal(Path,0)
    set(handles.Message2_text,'String','User pressed cancel when
selecting export filename');
else
    FullFileName = [Path '\' FileName];
    setappdata(hObject,'ExportFileName',FullFileName);
    set(handles.Export_pushbutton,'Enable','on');
    set(handles.ExportInteractions_checkbox,'Enable','on');
end

% --- Executes during object creation, after setting all properties.
function Optimise_pushbutton_CreateFcn(hObject, eventdata, handles)
% hObject    handle to Optimise_pushbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called
set(hObject,'Enable','off');


% --- Executes during object creation, after setting all properties.
function Export_pushbutton_CreateFcn(hObject, eventdata, handles)
```

```
% hObject    handle to Export_pushbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called
set(hObject,'Enable','off');



function HistLow_text_Callback(hObject, eventdata, handles)
% hObject    handle to HistLow_text (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of HistLow_text as
text
%        str2double(get(hObject,'String')) returns contents of
HistLow_text as a double


% --- Executes during object creation, after setting all properties.
function HistLow_text_CreateFcn(hObject, eventdata, handles)
% hObject    handle to HistLow_text (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end



function HistHigh_text_Callback(hObject, eventdata, handles)
% hObject    handle to HistHigh_text (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of HistHigh_text as
text
%        str2double(get(hObject,'String')) returns contents of
HistHigh_text as a double


% --- Executes during object creation, after setting all properties.
function HistHigh_text_CreateFcn(hObject, eventdata, handles)
% hObject    handle to HistHigh_text (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
```

```
end



function BucketSize_text_Callback(hObject, eventdata, handles)
% hObject    handle to BucketSize_text (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of BucketSize_text
as text
%        str2double(get(hObject,'String')) returns contents of
BucketSize_text as a double


% --- Executes during object creation, after setting all properties.
function BucketSize_text_CreateFcn(hObject, eventdata, handles)
% hObject    handle to BucketSize_text (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end


% --- Executes on button press in Redraw_pushbutton.
function Redraw_pushbutton_Callback(hObject, eventdata, handles)
% hObject    handle to Redraw_pushbutton (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
LowerBound = str2double(get(handles.HistLow_text,'String'));
UpperBound = str2double(get(handles.HistHigh_text,'String'));
BucketSize = str2double(get(handles.BucketSize_text,'String'));
TmMatrix = getappdata(handles.Optimise_pushbutton,'TmMatrix');

histstore = plothist(LowerBound,UpperBound,BucketSize,TmMatrix);
bar(handles.Histogram_axes,LowerBound:BucketSize:UpperBound,histstor
e);
set(handles.Histogram_axes,'XLim',[LowerBound UpperBound]);


% --- Executes on button press in DetectEndpoint_checkbox.
function DetectEndpoint_checkbox_Callback(hObject, eventdata,
handles)
% hObject    handle to DetectEndpoint_checkbox (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of
DetectEndpoint_checkbox
if get(hObject,'Value')
    set(handles.NumIterations_inputtext,'Enable','off');
else
    set(handles.NumIterations_inputtext,'Enable','on');
```

```
end


% --- Executes on button press in DetectCompetitors_checkbox.
function DetectCompetitors_checkbox_Callback(hObject, eventdata,
handles)
% hObject    handle to DetectCompetitors_checkbox (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of
DetectCompetitors_checkbox


% --- Executes on button press in ExportInteractions_checkbox.
function ExportInteractions_checkbox_Callback(hObject, eventdata,
handles)
% hObject    handle to ExportInteractions_checkbox (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of
ExportInteractions_checkbox


% --- Executes during object creation, after setting all properties.
function ExportInteractions_checkbox_CreateFcn(hObject, eventdata,
handles)
% hObject    handle to ExportInteractions_checkbox (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns
called
set(hObject,'Enable','off');
```

A.3.2. GroupCutINIT

This function is the core of the optimisation process. It performs the iterative sequence changes and produces the final output and suitable statistics.

```
function
[rmsdvector,bestTmmatrix,uniqueseqcell,bestcutposvector,HitsStore] =
GroupCutINIT(FASTAHeader,FASTASequence,numlig,iterations,DetectEndpo
intFlag,CompIDFlag)
% Program for cutting groups of sequenes in a manner to minimise the
number
% of different oligos in the final design.
%clear
%tic
%inputfilename = 'C:\Documents and Settings\moac\My Documents\PhD\UK
assembly\Opticut\GroupCut\remo11-allcombinations-9-MUT.fas';
%numlig = 9;
numhalflig = numlig*2-1;
%iterations = 100;
%cyccheckdelay = 20;
```

```
%fprintf('%s','Loading sequences... ');
% [headercell sequences] = fastaread(inputfilename);
%fprintf('%s\n','done.');
numseq = length(FASTAHeader);
lenseq = length(FASTASequence{1});
esthalfliglen = round(lenseq/numhalflig); %guess a solution
cutposvector = [1
esthalfliglen:esthalfliglen:esthalfliglen*numhalflig-1 lenseq];
originalcutposvector = cutposvector;

uniqueseqholder = cell(1,4);
cutposstore = zeros(numhalflig+1,iterations+1);
Tmstore = zeros(numseq,numhalflig,iterations+1);
rmsdvector = zeros(1,iterations+1);

cutposstore(:,1) = cutposvector;

%fprintf('%s','Making first guestimate... ');
for j = 1:numseq
    sequence = FASTASequence{j}; %load in the new sequence
    Tmvector = zeros(1,numhalflig); %reset store of Tm's
    for i = 1:numhalflig
        start = cutposvector(i);
        stop = cutposvector(i+1);
        seq = sequence(start:stop); %find the segment section
        logical = strcmp(seq,uniqueseqholder(:,1)); %check to see if
this segment has been determined previously
        if sum(logical) > 0 %if so, use the previously determined
value
            [a b] = find(logical==1);
            Tmvector(i) = uniqueseqholder{a,2};
        else %if not, record the pertinant points about the novel
overlap and determine the new overlap.
            %             uniqueseqholder(end+1,1:2) = [start stop];
            uniqueseqholder{end+1,1} = seq;
            Tm = TmNNSanta98(seq);
            Tmvector(i) = Tm;
            uniqueseqholder{end,2} = Tmvector(i);
            uniqueseqholder(end,3:4) = {start stop};
        end
        start = stop + 1;
    end
    Tmstore(j,:,1) = Tmvector; %record the store of Tm's
end
diff = Tmstore(:,:,1) - mean(mean(Tmstore(:,:,1)));
diffvector(1:numhalflig) = mean(diff(:,1:end));
sqdiffvector = diffvector.^2;
rmsd = sqrt(mean(sqdiffvector));
rmsdvector(1) = rmsd;
%fprintf('%s\n','done.');

h = waitbar(0,'Running optimisation...');
%fprintf('%s','Performing iterations... ');
for k = 1:iterations
    %look at stats, find the positions which need changing
    [minval mindiffpos] = min(diffvector);
    [maxval maxdiffpos] = max(diffvector);
    % generate director
    director = zeros(1,numhalflig+1);
```

```
    if mindiffpos < maxdiffpos %the smallest temp is to the left of
the largest, shift all the positions to the right by one
        director(mindiffpos+1:maxdiffpos) = 1;
    else %the smallest temp is to the right of the largest, shift
all the positions to the left by one
        director(maxdiffpos+1:mindiffpos) = -1;
    end
    %do the changes
    cutposvector = cutposvector + director;
    %write to store
    cutposstore(:,k+1)  = cutposvector;
    %do the stats on the new set of positions.
    for j = 1:numseq
        sequence = FASTASequence{j}; %load in the new sequence
        Tmvector = zeros(1,numhalflig); %reset store of Tm's
        for i = 1:numhalflig
            start = cutposvector(i);
            stop = cutposvector(i+1);
            seq = sequence(start:stop); %find the segment section
            logical = strcmp(seq,uniqueseqholder(:,1)); %check to
see if this segment has been determined previously
            if sum(logical) > 0 %if so, use the previously
determined value
                [a b] = find(logical==1);
                Tmvector(i) = uniqueseqholder{a,2};
            else %if not, record the pertinant points about the
novel overlap and determine the new overlap.
                %                 uniqueseqholder(end+1,1:2) =
[start stop];
                uniqueseqholder{end+1,1} = seq;
                Tm = TmNNSanta98(seq);
                Tmvector(i) = Tm;
                uniqueseqholder{end,2} = Tmvector(i);
                uniqueseqholder(end,3:4) = {start stop};
            end
            start = stop + 1;
        end
        Tmstore(j,:,k+1) = Tmvector; %record the store of Tm's
    end
    diff = Tmstore(:,:,k+1) - mean(mean(Tmstore(:,:,k+1)));
    Tmvector(1:numhalflig) = mean(Tmstore(:,1:end,k+1));
    diffvector(1:numhalflig) = mean(diff(:,1:end));
    sqdiffvector = diffvector.^2;
    rmsd = sqrt(mean(sqdiffvector));
    rmsdvector(k+1) = rmsd;

    %     fprintf('%f\t',cutposvector);fprintf('\n');
    %     fprintf('%f\t',diffvector);fprintf('\n')
    %     fprintf('%f\t',Tmvector);fprintf('\n')

    if k > 4 && DetectEndpointFlag
        if rmsdvector(k-3:k-2)==rmsdvector(k-1:k)
            %loop identified
            rmsdvector(k+1:end) = [];
            break
        end
    end
    waitbar(k/iterations);
end
%fprintf('%s\n','done.');
```

```
% find the best solution and stick with it
[minrmsd iternum] = min(rmsdvector);
bestcutposvector(1,:) = cutposstore(:,iternum);
bestTmmatrix = Tmstore(:,:,iternum);

UScutposvector = [bestcutposvector(1:2:end) bestcutposvector(end)];
LScutposvector = [bestcutposvector(1) bestcutposvector(2:2:end)];

%find the number of unique oligos in this set
uniqueseqcell = cell(1,3);
uniqueligcount = cell(numlig*2,4);

%fprintf('%s','Developing output data... ');
for i = 1:numlig*2
    uniqueligcount{i,2} = 0;
    if i <= numlig
        cutposvector = UScutposvector;
        index = i;
        start = UScutposvector(index);
        stop = UScutposvector(index+1);
        prefix = 'US';
    else
        cutposvector = LScutposvector;
        index = i - numlig;
        start = LScutposvector(index);
        stop = LScutposvector(index+1);
        prefix = 'LS';
    end
    uniqueligcount{i,4} = stop-start;
    for j = 1:numseq
        sequence = FASTASequence{j};
        if strcmp(prefix,'US')
            seq = sequence(start:stop);
        else
            seq = seqrcomplement(sequence(start:stop));
        end
        logical = strcmp(uniqueseqcell,seq);
        if sum(logical(:,1)) < 1
            uniqueseqcell{end+1,1} = seq;
            uniqueseqcell{end,2} = 1;
            uniqueligcount{i,1} = [prefix,int2str(index)];
            uniqueligcount{i,2} = uniqueligcount{i,2} + 1;
            uniqueseqcell{end,3} = [FASTAHeader{1,j} '_' prefix '_'
int2str(index)];
        elseif sum(logical(:,1)) == 1
            uniqueligcount{i,3} = uniqueligcount{i,3} + 1;
            [x y] = max(logical(:,1));
            uniqueseqcell{y,2} = uniqueseqcell{y,2} + 1;
        else
            %            fprintf('%s\n','Something terrible has
occured. Error #1');
        end
    end
end
%fprintf('%s\n','done.');

if CompIDFlag
    waitbar(0,h,'Running competitor identification...');
    HitCounter = 0;
    CompCounter = 0;
```

```
    ComparisonStore = cell(1,3);
    for j = 1:numseq
        sequence = FASTASequence{j};
        header = FASTAHeader{j};
        OligoList = cell(numlig*2,2);
        for k = 1:numlig*2
            if k <= numlig
                cutposvector = UScutposvector;
                index = k;
                start = UScutposvector(index);
                stop = UScutposvector(index+1);
                prefix = 'US';
            else
                cutposvector = LScutposvector;
                index = k - numlig;
                start = LScutposvector(index);
                stop = LScutposvector(index+1);
                prefix = 'LS';
            end

            OligoList{k,1} = sequence(start:stop);
            OligoList{k,2} = [FASTAHeader{1,j} '_' prefix '_'
int2str(index)];
        end

        for l = 1:numlig*2
            PrimaryOligo = OligoList{l,1};
            PrimaryName = OligoList{l,2};
            index = [];
            for m = 1:numlig*2
                if l <= numlig && m <= numlig
                    % US comp with US
                    SecondaryOligo = OligoList{m,1};
                elseif l <= numlig && m > numlig
                    % US comp with LS,
                    SecondaryOligo = seqrcomplement(OligoList{m,1});
                elseif l > numlig && m <= numlig
                    % LS comp with US
                    SecondaryOligo = seqrcomplement(OligoList{m,1});
                elseif l > numlig && m > numlig
                    % LS comp with LS
                    SecondaryOligo = OligoList{m,1};
                end
%                 SecondaryOligo = OligoList{m,1};
                SecondaryName = OligoList{m,2};
                PrimFind =
strcmp(PrimaryOligo,ComparisonStore(:,1));
                SecoFind =
strcmp(SecondaryOligo,ComparisonStore(:,2));
                CompFind = PrimFind + SecoFind;
                index = find(CompFind==2);
                if isempty(index)
                    CompCounter = CompCounter + 1;
                    [comparison,longerid] =
HeterodimerMeltingTemp(PrimaryOligo,SecondaryOligo);
                    ComparisonStore(CompCounter,:) = {PrimaryOligo
SecondaryOligo comparison};
                else
                    comparison = ComparisonStore{index,3};
                end
                Tms = cell2mat(comparison(:,4));
```

```
                Hits = Tms > 30;
                if ~all(Hits==0)
                    HitsIndicies = find(Hits);
                    for o = 1:sum(Hits)
                        HitCounter = HitCounter + 1;
                        Tm = Tms(HitsIndicies(o));
                        HitsStore(HitCounter,:) =
{PrimaryName,SecondaryName,Tm};
%                           ['Possible competative binding event
detected between ' PrimaryName ' and ' SecondaryName '.']
                    end
                end
            end
        end
        j;
        waitbar(j/numseq);
    end
else
    HitsStore = [];
end
close(h);
```

A.3.3. GroupCutOutput

This code is used to format the output data into either .xls or .csv formats.

```
function [uniqueseqcell] =
GroupCutOutput(cutposvector,FASTASequence,FASTAHeader)
%function for generating output for GroupCut
numlig = length(cutposvector)/2;
numseq = length(FASTASequence);

UScutposvector = [cutposvector(1:2:end) cutposvector(end)];
LScutposvector = [cutposvector(1) cutposvector(2:2:end)];

uniqueseqcell = cell(1,3);
uniqueligcount = cell(numlig*2,4);
%fprintf('%s','Developing output data... ');
for i = 1:numlig*2
    uniqueligcount{i,2} = 0;
    if i <= numlig
        cutposvector = UScutposvector;
        index = i;
        start = UScutposvector(index);
        stop = UScutposvector(index+1);
        prefix = 'US';
    else
        cutposvector = LScutposvector;
        index = i - numlig;
        start = LScutposvector(index);
        stop = LScutposvector(index+1);
        prefix = 'LS';
    end
    uniqueligcount{i,4} = stop-start;
    for j = 1:numseq
        sequence = FASTASequence{j};
        if strcmp(prefix,'US')
            seq = sequence(start:stop);
```

```
        else
            seq = seqrcomplement(sequence(start:stop));
        end
        logical = strcmp(uniqueseqcell,seq);
        if sum(logical(:,1)) < 1
            uniqueseqcell{end+1,1} = seq;
            uniqueseqcell{end,2} = 1;
            uniqueligcount{i,1} = [prefix,int2str(index)];
            uniqueligcount{i,2} = uniqueligcount{i,2} + 1;
            uniqueseqcell{end,3} = [FASTAHeader{1,j} '_' prefix '_'
int2str(index)];
        elseif sum(logical(:,1)) == 1
            uniqueligcount{i,3} = uniqueligcount{i,3} + 1;
            [x y] = max(logical(:,1));
            uniqueseqcell{y,2} = uniqueseqcell{y,2} + 1;
        else
%            fprintf('%s\n','Something terrible has occured. Error
#1');
        end
    end
end
%fprintf('%s\n','done.');
```

A.3.4. PlotHist

This simple function is used in preference to the hist.m function in built into MATLAB as the in built function does not offer sufficient control over the histogram.

```
function output = plothist(data,buckets)
%plothist(data,buckets)
%%
MinVal = min(data(:));
MaxVal = max(data(:));
BucketSize = (MaxVal-MinVal)/buckets;


output = zeros(buckets,1);

for i = 1:buckets
    output(i) = sum(data(MinVal+((i-
1)*BucketSize)<=data&data<=MinVal+(i*BucketSize)));
end
%%
% heatmap(:,1) = Tmstore(:,1,1);
% heatmap(:,19) = Tmstore(:,end,1);
% heatmap(:,2:18) = Tmstore(:,:,1);
% heatmap = bestTmmatrix;
% heatmap = Tmstore(:,:,1);

% heatmap(1:end,end+1) = min(Tmstore(:));
% heatmap(1:end,end+1) = max(Tmstore(:));
% imagesc(heatmap)
```

A.3.5. HeterodimerMeltingTemp

This function estimates the melting temperature of all possible binding configurations of two oligonucleotides. This information is then used to determine whether sequences in the set might undergo competitive binding interactions. This process is described in section 7.1.4.

```
function [comparisonstore,longerid] =
HeterodimerMeltingTemp(sequence1,sequence2)

% sequence1 = 'CGACGCGTCGAAGGAAAGCGGCCGCA';
% sequence2 = 'CTTCGACGCGTCG';

sequencearray = {sequence1 sequence2};

%column1 = dH, column2 = dS. Both in calories.
lookuptable = {'AA/TT'  -7.900  -22.2;
    'AG/CT' -7.800  -21.0;
    'AT/AT' -7.200  -20.4;
    'AC/GT' -8.400  -22.4;
    'GA/TC' -8.200  -22.2;
    'GG/CC' -8.000  -19.9;
    'GC/GC' -9.800  -24.4;
    'TA/TA' -7.200  -21.3;
    'TG/CA' -8.500  -22.7;
    'CG/CG' -10.600 -27.2;
    'Terminal A-T base pair'    2.300   4.1;
    'Terminal G-C base pair'    0.100   -2.8};

%constants, change at your peril.
R = 1.9872; %Universal gas constant
primerconc = 50e-6;
saltconc = 0.05;
ambienttemperature = 25;

%determine the type of comparison we're dealing with, make changes
to the
%identity of sequences so that the comparison can be performed.
%lenstate1: the max length of a seq2 underhang of seq1
%lenstate2: the max # times that seq2 can fit within seq1 ie if
%lenseq1=lenseq2, lenstate2 = 1.
if length(sequence1) > length(sequence2) % seq1 is the longer, all
is well
    longerid = 1; %these ids tell the calling function whether a
switch has taken place.
    sequence2 = seqrcomplement(sequence2); %complement cos we
compare characters, not basepairing (cheaty). reverse cos its
obvious. 5'-3'
    maxlength = length(sequence1);
    minlength = length(sequence2);
    lenstate1 = minlength - 1;
    lenstate2 = maxlength - minlength + 1;
elseif length(sequence1) < length(sequence2) % seq2 is the longer,
this much be changed
```

```
    longerid = 0;
    sequence1 = sequencearray{2};
    sequence2 = seqrcomplement(sequencearray{1}); %seq2 is always
rcomplemented, even if it is actually sequence 1. confusing, huh?
    maxlength = length(sequence1);
    minlength = length(sequence2);
    lenstate1 = minlength - 1;
    lenstate2 = maxlength - minlength + 1;
elseif length(sequence1) == length(sequence2) % the seqs are the
same length. This is acceptable.
    longerid = -1;
    sequence2 = seqrcomplement(sequence2);
    maxlength = length(sequence1);
    minlength = length(sequence2);
    lenstate1 = minlength - 1;
    lenstate2 = 1;
else
    fprintf('%s\n','Error during relative sequence length
determination. This is terminal.');
end

numcomparison = length(sequence1) + length(sequence2) - 1;
comparisonstore = cell(1,3); %stores all the comparisons for a
particualr sequence.

for i = 1:numcomparison
    %1. find which state we are in so that we can appropriately
generate overlaps. ie seq2 underhanging seq1 etc.
    if i <= lenstate1 %seq2 underhangs seq1
        overlap1 = sequence1(1:i);
        overlap2 = sequence2(end-i+1:end);
    elseif i > lenstate1 && i <= lenstate2 + lenstate1 %seq2 is
encompassed in seq1
        overlap1 = sequence1(i-lenstate1:i);
        overlap2 = sequence2;
    elseif i > lenstate2 + lenstate1 %seq2 underhangs seq1 at the
other end.
        overlap1 = sequence1(i-lenstate1:end);
        overlap2 = sequence2(1:end-(i-lenstate1-lenstate2));
    else
        fprintf('%s\n','Error during overlap generation. #1');
    end

    %2. Find the bits of the given overlap match
    lenoverlap1 = length(overlap1);
    lenoverlap2 = length(overlap2);
    matchflag = 0; %this probably doesnt need to be here.
    if lenoverlap1 ~= lenoverlap2
        fprintf('%s\n','Error during overlap generation. #2');
    else
        rowdex = 1; %for the comparisonstore
        matcharray = cell(rowdex,1); %reset the matcharray, this is
the short term store in which all matching bits of sequence are
recorded.
        matchsequence = ''; %reset the matchsequence, to be inserted
into the match array.
        logical = overlap1 == overlap2; %compare the overlap
strings.
        j = 1;
        while j <= lenoverlap1 %go through the overlap
            matchflag = 0;
```

```
            while logical(j) && j <= lenoverlap1 %whilst the
basepairs int he overlap match...
                matchflag = 1; %turn on the match indicator
                matchsequence = [matchsequence overlap1(j)]; %add
the relevent base to the matchsequence
                j = j + 1;
                if j > lenoverlap1
                    break
                end
            end %stop when the matched bit ends...
            if matchflag %we found a matching bit, and its end. we
now record the match
                matcharray{rowdex,1} = matchsequence;
                matchsequence = '';
                rowdex = rowdex + 1;
                matchflag = 0;
            else
                j = j + 1;%no match found, move on.
            end
        end
        %3. now we should have a matcharray of all matching bits, do
maths.
        if ~isempty(matcharray{1,1});
            nummatchsegment = length(matcharray);

            for j = 1:nummatchsegment
                sumdH = 0;
                sumdS = 0;
                sequence = matcharray{j,1};
                numbase = length(sequence);
                for k = 1:numbase
                    if k == 1 || k == numbase
                        base = sequence(k);
                        if strcmp(base,'A') || strcmp(base,'T');
                            sumdH = sumdH + lookuptable{11,2};
                            sumdS = sumdS + lookuptable{11,3};
                        elseif strcmp(base,'G') || strcmp(base,'C');
                            sumdH = sumdH + lookuptable{12,2};
                            sumdS = sumdS + lookuptable{12,3};
                        else
                            fprintf('%s\n',['Unrecognised base
(',base,') recieved at position ',int2str(i),' in inputted
sequence.']);
                        end
                    else
                        basestack = sequence(k:k+1);
                        if strcmp(basestack,'AA') ||
strcmp(basestack,'TT');
                            sumdH = sumdH + lookuptable{1,2};
                            sumdS = sumdS + lookuptable{1,3};
                        elseif strcmp(basestack,'AG') ||
strcmp(basestack,'CT');
                            sumdH = sumdH + lookuptable{2,2};
                            sumdS = sumdS + lookuptable{2,3};
                        elseif strcmp(basestack,'AT') ||
strcmp(basestack,'AT');
                            sumdH = sumdH + lookuptable{3,2};
                            sumdS = sumdS + lookuptable{3,3};
                        elseif strcmp(basestack,'AC') ||
strcmp(basestack,'GT');
                            sumdH = sumdH + lookuptable{4,2};
```

```
                            sumdS = sumdS + lookuptable{4,3};
                        elseif strcmp(basestack,'GA') ||
strcmp(basestack,'TC');
                            sumdH = sumdH + lookuptable{5,2};
                            sumdS = sumdS + lookuptable{5,3};
                        elseif strcmp(basestack,'GG') ||
strcmp(basestack,'CC');
                            sumdH = sumdH + lookuptable{6,2};
                            sumdS = sumdS + lookuptable{6,3};
                        elseif strcmp(basestack,'GC') ||
strcmp(basestack,'GC');
                            sumdH = sumdH + lookuptable{7,2};
                            sumdS = sumdS + lookuptable{7,3};
                        elseif strcmp(basestack,'TA') ||
strcmp(basestack,'TA');
                            sumdH = sumdH + lookuptable{8,2};
                            sumdS = sumdS + lookuptable{8,3};
                        elseif strcmp(basestack,'TG') ||
strcmp(basestack,'CA');
                            sumdH = sumdH + lookuptable{9,2};
                            sumdS = sumdS + lookuptable{9,3};
                        elseif strcmp(basestack,'CG') ||
strcmp(basestack,'CG');
                            sumdH = sumdH + lookuptable{10,2};
                            sumdS = sumdS + lookuptable{10,3};
                        end
                    end
                end
                % get the relevant data. this is really fast.
                temperature = (sumdH*1000)/(sumdS +
R*log(primerconc)) + (16.6*log10(saltconc)) - 273.15;
                dG = sumdH - (ambienttemperature +
273.15)*(sumdS/1000);
            end
            %4. record the data in the important bit of the store.
            comparisonstore(i,1:5) = {overlap1 overlap2
nummatchsegment temperature dG};
        else
            comparisonstore(i,1:5) = {overlap1 overlap2 0 0 0};
%note the zero values here. NaNs are NOT appropriate. becareful
becuase in a lot of these comparison stores, 0 will be the highest
value of the set.
        end
    end
end
```

A.3.6. TmNNSanta98

This code is an optimised version of the appropriate parts of oligoprop.m that is provided as part of the MATLAB bioinformatics toolbol. This function is used to estimate the melting temperature of overlaps during the optimisation program.

```
function tm = TmNNSanta98(seq)

% seq = 'ATCGCTTAGCTCGCGGATT';
```

```
seq_length = length(seq);

numSeq = double(nt2int(seq));
baseNum = [sum(numSeq == 1) sum(numSeq == 2) sum(numSeq == 3)
sum(numSeq == 4) sum(numSeq == 15)];

temp = 25;            % temperature in Celsius
salt = 0.05;          % salt concentration in moles per liter (M)
primerConc= 50e-6;    % concentration of primers in mole per liter
(M)

if (sum(baseNum)<14)
    basic = 2 * (baseNum(1) + baseNum(4)) + 4 * (baseNum(2) +
baseNum(3)); % TM BASIC [9]
    saltadj = basic - 16.6 * log10(0.05) + 16.6 * log10(salt); % TM
SALT ADJUSTED [9]
else
    basic = 64.9 + (41 * ((baseNum(3) + baseNum(2) - 16.4) /
sum(baseNum))); %TM BASIC [1],[9]
    saltadj = 100.5 + (41 * ((baseNum(3) + baseNum(2))/
sum(baseNum))) - (820/sum(baseNum)) + (16.6 * log10(salt)); %TM SALT
ADJUSTED [9]
end
%%
Sant98_H = [-7.9,-8.4,-7.8,-7.2,;-8.5,-8,-10.6,-7.8,;-8.2,-9.8,-8,-
8.4,;-7.2,-8.2,-8.5,-7.9,];
Sant98_S = [-22.2,-22.4,-21,-20.4,;-22.7,-19.9,-27.2,-21,;-22.2,-
24.4,-19.9,-22.4,;-21.3,-22.2,-22.7,-22.2,];
ind = sub2ind([4 4],numSeq(1:seq_length-1),numSeq(2:seq_length));
NN = [sum(Sant98_H(ind)),sum(Sant98_S(ind))];

% initiation with terminal  5'
if(numSeq(1) == 2 || numSeq(1) == 3)
    NN(1,:) = NN(1,:) + [0.1 -2.8];
elseif(numSeq(1) == 1 || numSeq(1) == 4)
    NN(1,:) = NN(1,:) + [2.3 4.1];
    % NN(2,1) = NN(2,1) + 0.4;
end

% initiation with terminal  3'
if(numSeq(end) == 2 || numSeq(end) == 3)
    NN(1,:) = NN(1,:) + [0.1 -2.8];
elseif(numSeq(end) == 1 || numSeq(end) == 4)
    NN(1,:) = NN(1,:) + [2.3 4.1];
end

NNdelta=zeros(4,2);
%%

b = 4;

% [NN, NNdelta] = near_neigh(numSeq, length(numSeq), selfCompFlag,
nFlag);
tm = (NN(:,1) * 1000 ./ (NN(:,2) + (1.9872 * log(primerConc./b)))) +
(16.6 * log10(salt)) - 273.15; %TM NEAREST NEIGHBOR
```

Appendix B

B. Appendix B: Papers

This appendix contains a compilation of the papers written during and as a result of this PhD thesis as of October 2012. Each paper is briefly described and my contribution to the work is outlined.

B.1. Continuous-channel flow linear dichroism

Authors: Xi Cheng, Maxim B. Joseph, James A. Covington, Timothy R. Dafforn, Matthew R. Hicks and Alison Rodger.

Published: Analytical Methods, 2012, **4**, 3169-3173

Summary: This paper compared three flow systems for the measurement fast kinetic reactions (<600 ms) using linear dichroism (LD). The minimum dead time reported here was 25 ms, using < 100 µL sample volume per time point. The three flow systems used were: A quartz FC-20 cuvette in a SFM-300 stopped flow device. A µ-slide III$^{3\ in\ 1}$ channel cell and a custom made single channel quartz window cell with T-mixer (Q1x1). The custom made cell was designed and fabricated during this project and is described in detail in section 5.5.

The work completed as part of this thesis that contributed to the paper was the design, construction and assembly of the custom Q1x1 flow cell. The flow cell was designed in SolidWorks and fabricated in three pieces using the EnvisionTec Perfactory Mini MSL machine. The pieces were then assembled with two pieces of quartz glass (5x5 mm and 10x10 mm) to produce the final device. A T-mixer was then employed to mix solutions ahead of the measuring cell.

B.2. Dissolution Kinetics of Polycrystalline Calcium Sulfate-Based Materials: Influence of Chemical Modification

Authors: Robin D. Fisher, Michael M. Mbogoro, Mike E. Snowden, Maxim B. Joseph, James A. Covington, Pat R. Unwin, Richard I. Walton.

Published: ACS applied materials & interfaces, 2011, **3**, 3528-3537.

Summary: This work experimentally determined the dissolution kinetics of crystals of gypsum and another related $CaSO_4$-containing crystallite in the presence of several 'humid creep inhibitor' chemicals. The experimental work was carried out by Robin Fisher and Michael Mbogoro. The work was also validated a 2D model of the flow cell, created by Mike Snowden, that gave a theoretical framework to the experimental work.

The design and manufacture of the flow cell was carried out by Maxim Joseph using the EnvisionTec Perfactory Mini system and is described in section 5.4. This work required a radial flow cell with as close to true radial flow as possible. Simulations performed by Maxim Joseph and Mike Snowden indicated that the flow in contemporary commercially available devices was unlikely to be sufficiently radial. The radial symmetry of the outlets from the radial flow chamber and the incorporation of an inlet channel to the centre of the radial flow chamber were critical to the success of this device.

B.3. Ultrasensitive Detection of Dopamine Using a Carbon Nanotube Network Microfluidic Flow Electrode

Authors: Siriwat Sansuk, Eleni Bitziou, Maxim B. Joseph, James A. Covington, Martyn G. Boutelle, Patrick R. Unwin and Julie V. Macpherson

Published: Analytical Chemistry, *in press.*

Summary: This work considered the detection of dopamine using a carbon nanotube network in a flow injection microfluidic setup. This work was able to accurately and repeatedly measure the concentration of dopamine in solutions to the 5 pM in sample volumes of 50 µL. The fabrication and operation of the carbon nanotube electrode and flow injection setup was performed by Siriwat Sansuk and Eleni Bitziou.

The design and manufacture of the flow cell was carried out by Maxim Joseph using the EnvisionTec Perfactory Mini system, as described in section 5.4. To maximise the detection level and minimise the sample volume, a thin layer flow cell was required. Design of the flow chamber, including testing of the z-resolution of the Perfactory system, and the inlet and outlet geometries to incorporate the necessary reference and counter electrodes, was critical to the success of this device.

B.4. Insights into "fermentonomics": evaluation of volatile organic compounds (VOCs) in human disease using an electronic "e-nose"

Authors: Ramesh P. Arasaradnam, N. Quraishi, I. Kyrou, Chuka U. Nwokolo, Maxim Joseph, S. Kumar, Karna D. Bardhan, James A. Covington.

Published: Journal of Medical Engineering & Technology, 2011, **35**(2), 87-91.

Summary: This work is the first in a series of papers validating the use of an 'e-nose' as a technique for the identification and diagnosis human disease. Stool samples from normal, healthy individuals and from individuals diagnosed with several inflammatory bowel diseases were compared using the e-nose and as mass spectroscopic approach. Both techniques were able to cluster the individuals according to the disease states. Ramesh Arasaradnam, N. Quraishi, I. Kyrou and Chuka Nwokolo assessed patients, took and handled samples as well as assessing the samples using the e-nose designed by James Covington. Maxim Joseph assessed the samples by mass spectroscopy using a head space sampler and analysed the data using principle component analysis. This work is not described in the present thesis.