Development and optimisation of a method for multi-site-directed mutagenesis

Mauro André Mendonça Da Cruz

UMinho | 2019

Universidade do Minho

Escola de Ciências

Mauro André Mendonça Da Cruz

Development and optimisation of a method for multi-site-directed mutagenesis

outubro de 2019

Universidade do Minho

Escola de Ciências

Mauro André Mendonça Da Cruz

Development and optimisation of a novel method for multi-site-directed mutagenesis

Dissertação de Mestrado

Genética Molecular

Trabalho efetuado sob a orientação do

Doutor James Anthony Collins

outubro de 2019

Declaração

Nome: Mauro André Mendonça Da Cruz

Endereço eletrónico: mauro.mendonca@hotmail.de **Telefone**: +351 927 060 881

Número do Cartão de Cidadão: 14221398

Título da dissertação:

Development and optimisation of a novel method for multi-site-directed mutagenesis

Orientador:

Doutor James Anthony Collins

Ano de conclusão: 2019

Designação do Mestrado: Genética Molecular

Atribuição-NãoComercial-SemDerivações CC BY-NC-ND

https://creativecommons.org/licenses/by-nc-nd/4.0/

Universidade do Minho, 31 de outubro de 2019

Assinatura: Mauro Cruz

Acknowledgements

During the year of my thesis I had the opportunity to really experience what science is all about. I learned so much and have come to really appreciate the effort and work that thousands of people do, to advance our understanding in biology. It was also a very challenging year, with multiple hurdles that thankfully I did not have to overcome all by myself.

I first would like to thank Doctor Tony Collins for giving me the opportunity to work on such an interesting topic and for guiding me throughout this whole year. Your support was crucial, and I can't thank you enough for what you taught me and the interesting conversations we had about our work.

Next up I would like to thank Mário Barroca. A smart problem-solving colleague that is only overshadowed by his sense of humour. Your input has helped me a lot in the course of this year and our conversations would always cheer me up.

Ricardo, my lab companion, we would often stay both late to get the results we needed, just talk about random stuff during our downtimes and discuss results. Thank you for being there.

Many more amazing people, including the laboratory technicians and colleagues from other research groups have helped me in one or more ways during this year, and I truly am thankful for having made my life easier, even if in a small way.

Ana, my partner in life and science, you've been an incredible support during my academic carrier. It is thanks to you that I am doing what I am, and your motivational speeches helped me more than once to collect myself and power through everything. A big thanks!

Last but not least, I'd like to thank my family for supporting me throughout these years, especially my grandmother. Your support was incredibly important, and I would not have had the opportunity to pursue my passion for science if it wasn't for you.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo

Desenvolvimento e otimização de um método novo de mutagénese múltipla dirigida

A mutagénese dirigida é um processo de alteração de um ou mais nucleótidos numa molécula de ADN de forma criteriosa e não aleatória. A introdução de múltiplas mutações no ADN é feita frequentemente através da técnica de reação em cadeia da polimerase (PCR) de modo a dar origem a fragmentos que contêm as mutações desejadas. Estes fragmentos podem ser usados para construir um plasmídeo *in vitro* através da adição de misturas enzimáticas baseadas em ligases ou recombinases. Para a construção de plasmídeos *in vivo* é possível utilizar estirpes de clonagem de *Escherichia coli* graças aos seus mecanismos de recombinação genética inerentes, sendo por isso desnecessário a utilização de misturas enzimáticas. Atualmente, protocolos que introduzem várias mutações simultaneamente requerem misturas enzimáticas dispendiosas ou vários passos morosos. Neste trabalho foi desenvolvido e otimizado um novo protocolo com vista a facilitar a introdução de várias mutações simultaneamente de forma rápida e económica.

Estudos iniciais, acompanhados de uma análise da literatura atual, identificaram o tipo de polimerase, competência das células, quantidade de ADN em PCR, extensão de homologias e rácios de fragmento maior para mais pequenos como possíveis fatores importantes para garantir uma elevada eficiência do protocolo de recombinação genética *in vivo* em *E. coli*. O estudo foi feito utilizando um modelo de superfície de resposta, complementado por um *central composite design* para otimizar o processo, minimizar o número de experiências e maximizar informação para o modelo matemático. As variáveis e modelos obtidos foram submetidos a uma ANOVA para avaliar a sua significância estatística. Para todas as experiências, foram usadas estirpes de clonagem de *E. coli*, apresentando uma competência mínima de 10^7. Os resultados mostraram compatibilidade com vários modelos matemáticos, dos quais foi adotado o modelo quadrático ($p = 0.0148$). Todas as variáveis testadas (conteúdo de ADN, homologia, rácio vetor para inserto) foram avaliadas como significativas, porém, não foi identificada interação entre elas. As condições ótimas previstas para recombinação genética foram testadas e confirmadas usando 0.1 ng de ADN, homologias de 52 nucleótidos e um rácio de 1 para 10, para o qual foi obtida uma eficiência de 97%.

O protocolo de mutagénese múltipla dirigida desenvolvido e otimizado neste trabalho oferece eficiências superiores a protocolos atuais, sendo ao mesmo tempo mais económico e simples de efetuar.

Palavras-chave: Engenharia proteica; *In vivo* construção de plasmídeo; Mutagénese; Protocolo; Recombinação homóloga.

Abstract

Development and optimisation of a novel method for multi-site-directed mutagenesis

Multi-site-directed mutagenesis is a powerful tool used to alter several target nucleic acids within a DNA molecule simultaneously. Mutations are generally introduced through a polymerase chain reaction (PCR), producing mutagenic fragments that are assembled into a construct *in vitro* by use of conventional ligase-based or recombination-based enzyme mixtures. Current multi-site-directed mutagenesis protocols are characterised by high costs and/or laborious processes, requiring expensive enzymes and/or multiple steps. Interestingly, many common *Escherichia coli* cloning strains are capable of carrying out *in vivo* recombination, albeit at low frequencies, and were investigated and optimised in the present study for use in a cost-effective multi-site-directed mutagenesis protocol.

Initial studies indicated that the type of polymerase, the cell competence, and process variables such as the PCR template concentration, length of homologous regions and ratio of fragments, to be critical for success of mutagenesis via *in vivo* recombination in *E. coli*. Response surface methodology with central composite design was used to optimise process variables, identify interactions and determine the protocol enabling highest multi-site-directed mutagenesis efficiency. Use of this mathematical and statistical approach enabled experimental design with minimisation of the number of experiments and maximisation of the information obtained, and allowed for data set analysis with development of statistical previsions based on this. Statistical analysis of variables and obtained models was performed via ANOVA. Experimental data showed compatibility with several models, of which a quadratic model with a p value of 0.0148 was adopted. All tested variables: template amount, homology length and vector to insert ratio, were identified as significant, while interactions between these factors were not. Predicted optimal conditions were confirmed by using 0.1 ng template in the PCR reaction, a homology length of 52 base pairs, a large fragment to small fragment ratio of 1:10, and highly competent cells ($\geq 10^7$ cfu/µg), producing an efficiency of 97%.

The hereby developed multi-site-directed mutagenesis protocol requires only three steps (PCR amplification, *Dpn*I digestion and transformation) and represents an easy and simple alternative to current protocols without compromising efficiency.

Key words: Homologous recombination; *In vivo* plasmid assembly; Mutagenesis; *E. coli*; Recombineering.

Table of Contents

Acknowledgements iii

Resumo v

Abstract vi

Table of Contents vii

List of Abbreviations and Acronyms ix

List of Figures x

List of Tables xii

1. Introduction 1

1.1 Protein Engineering 1

1.2 Single-Site-Directed Mutagenesis 2

1.3 Multi-Site-Directed Mutagenesis 5

1.4 Fragment Assembly in Multi-Site-Directed Mutagenesis 7

1.5 Recombination 8

1.5.1 Recombination Types 9

1.5.2 RecA-Dependent Recombination 15

1.5.3 RecA-Independent Recombination 20

1.6 Exogenous Recombination Pathways in *E. coli*: the Red & RecET Systems 21

1.7 Response Surface Design (RSM) 23

1.8 Central Composite Design (CCD) 24

2. Rational and Objectives 26

3. Materials and Methods 27

3.1 Primer Design 27

3.2 PCR 27

3.3 Gel Electrophoresis 28

3.4 Enzymatic Digestions 29

3.4.1 Plasmid Confirmation 29

3.4.2 Template Removal 29

3.5 *E. coli* Strains 30

3.6 Competent Cells 30

3.6.1 Modified Dagert and Ehrlich Method 30

3.6.2 Modified Inoue et al. Method 31

3.6.3 In-house Magnesium Chloride and Calcium Chloride Method 31

3.6.4 Rubidium Chloride Method 31

3.7 Transformation 32
3.8 Bacterial Cell Storage: Glycerol Cell Stocks 32
3.9 Plasmid Extraction 33
3.10 DNA Quantification 33
3.11 Sequencing 33
3.12 Central Composite Design 34
4. Results 35
4.1 Method Development 36
4.1.1 PCR Conditions 38
4.1.2 Template Concentration 39
4.1.3 False Positives: *DpnI* Digestion 42
4.1.4 DNA Polymerase 42
4.1.5 Transformation Volume 43
4.1.6 Fragment Ratio 44
4.1.7 Primer Size 44
4.2 NZYProof DNA Polymerase 45
4.3 RSM-CCD for Protocol Optimisation 47
4.3.1 Output Response and Input Factors 47
4.3.2 Experiment Design and Analysis 47
4.3.3 Experiment Run 48
4.3.4 Model Fit, Diagnose and Interpretation 51
4.3.5 Model Confirmation 52
5. Discussion 54
5.1 Possible Epigenetic Interference by DNA Polymerases 54
5.2 Microhomology-Mediated Recombination and Fragment Bias 55
5.3 Identification of Crucial Factors for Mutagenesis Efficiency 58
5.4 *recA1 E. coli* Strains as Hosts for Recombination *In Vivo* 59
5.5 Protocol Mutagenesis Efficiency 61
6. Conclusions and Future Perspectives 63
7. References 64
Annexe I – Map of Construct pET22b-pXyl 77
Annexe II – Map of Cloning Vector pUC18 78
Annexe III – Primers 79
Annexe IV – PCR Conditions used with Phusion® High-Fidelity Polymerase 80
Annexe V – PCR Conditions used with NZYProof Polymerase 81

List of Abbreviations and Acronyms

A – Adenine

amp – Ampicillin

bp – Base pair

C – Cytosine

CCC – Central composite circumscribed

CCD – Central composite design

CCF – Central composite face-centred

CCI – Central composite inscribed

cfu – Colony forming unit

Chi – Crossover hotspot instigator

D-loop – Displacement loop

DNA – Deoxyribonucleic acid

ds – Double-stranded

Fwd – Forward

G – Guanine

g – Relative centrifuge force

h – Hour

iPCR – Inverse polymerase chain reaction

kb – Kilobases

LB – Lysogeny broth

nt – nucleotides

OD – Optical density

PCNA – Proliferating cell nuclear antigen

PCR – Polymerase chain reaction

pXyl – Xylanase from *Pseudoalteromonas haloplanktis* (Uniprot: Q8RJN8)

RSM – Response surface model

Rev – Reverse

rpm – Rotations per minute

SDSA – Synthesis-dependent strand annealing

SOB – Super optimal broth

SOC – Super optimal broth with catabolite repression

ss – Single-stranded

SSB – Single-strand binding protein

T – Thymine

T° – Temperature

Tm – Melting temperature

U – Unit

2FI – Two-factor interaction

°C – Degrees Celsius

List of Figures

Figure 1. Overview of the overlap extension method for site-directed mutagenesis.................................3

Figure 2. Illustration of iPCR primer layouts and representation of the various steps involved in iPCR with the commercial site-directed mutagenesis kit QuickChangeTM (Agilent Technologies)............................4

Figure 3. Overview of the QuikChange™ Kit multi-site-directed mutagenesis kit from Agilent Technologies..6

Figure 4. Multi-site-directed mutagenesis method based on *in vitro* homologous recombination..............7

Figure 5. Broken fork repair model..10

Figure 6. Double-strand break repair model...12

Figure 7. Gap-filling recombinational repair model...14

Figure 8. The RecBCD pathway...16

Figure 9. The RecFOR pathway model..17

Figure 10. Mechanism of resection of the three types of dsDNA ends by helicase RecQ and nuclease RecJ...18

Figure 11. Envisioned RecA-loading mechanism by SbcB15..19

Figure 12. Exogenous recombination systems in *E. coli*..22

Figure 13. Central composite design space..25

Figure 14. Flowchart of the developed multi-site-directed mutagenesis protocol by *in vivo* recombineering in *E. coli*...37

Figure 15. Amplification of F1, F2 and F3..38

Figure 16. Effect of PCR template concentration on recombination efficiency and number of transformants..39

Figure 17. Molecular weight analysis of pET22b-pXyl mutants...40

Figure 18. pET22b-pXyl truncations..41

Figure 19. Transformation efficiency of varying volumes...44

Figure 20. Amplification of fragments F4, F5 and F6..45

Figure 21. High molecular DNA entities produced by NZYProof...46

Figure 22. Template preparation, F4, F5 and F6 fragment amplification and optimization with Phusion...49

Figure 23. Surface response model for recombination efficiency...53

Figure 24. Mechanism of production of single and double mutants..57

List of Tables

Table 1. *E. coli* strains used in the present work..30

Table 2. Analysis of insert sequences...40

Table 3. Transformation associated recombination results of Y43E, Y315R, T403R mutants..................43

Table 4. Central composite design matrix...47

Table 5. Evaluation of the designed experiment...48

Table 6. Results corresponding to the RSM-CCD experiment..50

Table 7. Model summary statistics for recombination efficiency...51

1. Introduction

1.1 Protein Engineering

Mutagenesis – the act of altering nucleotides within a DNA or RNA sequence – is a fundamental tool commonly used in molecular biology and protein science today. Various types of nucleotide alterations can be carried out, substitution(s), insertion(s) and deletion(s), so as to alter the target sequence and its properties. It is central to the investigation and engineering of the structure and function of DNA, RNA and proteins and has been successfully used in the tailoring of these molecules for specific functions and applications.

Protein engineering is a powerful technique making use of mutagenesis approaches wherein alteration of the protein gene sequence is used to impact the amino acid content and hence thereby alter the protein characteristics (Zoller, 1991). The main approaches used in modern protein engineering are based on site-directed mutagenesis, in which specific amino acid changes are made, and random mutagenesis, in which random changes are introduced in the protein sequence (Clore et al., 2011; Turanli-Yildiz et al., 2012; Wong et al., 2006). Both approaches have various advantages and limitations and thus the approach chosen should be suited to the particular objectives and available tools.

Random mutagenesis, or directed evolution, does not require extensive knowledge of the studied protein and is based on the preparation of large libraries of randomly mutated sequences and the screening of these for specific desired characteristics e.g. enhanced stability, increased activity, etc. (Clore et al., 2011; Forloni et al., 2018; Romero and Arnold, 2009). While a number of different kits and techniques for the efficient construction of large random mutant libraries do exist, screening for the desired mutants is much more technically challenging and requires development of high through-put low cost screening methods to identify the few desired mutants in the thousands (sometimes hundreds of thousands) of mutants prepared. Furthermore, the mechanisms by which mutations are introduced may not be entirely unbiased, and dependent on the method used. To give an example, the use of *Taq* DNA polymerase, to introduce mutations during an error-prone polymerase chain reaction (PCR), results in a strong tendency for transition substitutions (purines substituted by purines, A ⇆ G; pyrimidines substituted by pyrimidines, C ⇆ T), thereby leading to a less randomly mutated and possibly smaller acquired library of protein mutants (Wong et al., 2006). Another downside of the random mutagenesis approach is the accumulation of non-relevant mutations which interfere with identification of the effective mutation(s) (Robinson et al., 2009; Wong et al., 2006).

Site-directed mutagenesis is often also called “rational design” because the alterations introduced in the target protein do not occur randomly, but rather are targeted, predetermined changes designed by the scientist and based upon acquired information (Antikainen and Martin, 2005; Arnold, 1993). Site-directed mutagenesis is, in practice, relatively simple to carry out, but success with this approach, in identifying the correct mutations to introduce, is based on an in-depth knowledge of the protein being studied and of protein design principles and factors governing protein structure and function (Antikainen and Martin, 2005). In particular, knowledge of the protein structure, function, activity, stability and physicochemical characteristics as well as information on closely related homologs are key to success with this approach. Site directed mutagenesis can involve the introduction of single alterations (single-site-directed mutagenesis), which is commonly used in fundamental studies investigating the function of specific amino acids, or the introduction of numerous alterations (multi-site-directed mutagenesis) which is often required in engineering a specific property in a protein. A large variety of *in vitro* and *in vivo* techniques and commercial kits have been developed for each of these approaches, with difference being observed in the costs, time, labour requirements and efficiencies of each and the most important of these will be discussed below.

1.2 Single-Site-Directed Mutagenesis

There are a number of methods described for performing single-site-directed mutagenesis, but presently the most commonly used *in vitro* methods are based on either overlap extension PCR or inverse PCR (iPCR) (Antikainen and Martin, 2005; Ho et al., 1989; Silva et al., 2017). The overlap extension method requires multiple rounds of PCR (see Figure 1 for details), consisting of two separate initial PCRs, each one with a different pair of primers, followed by a third PCR. Each pair of primers, a & b and c & d, has one primer with the intended mutation (primer b and c), and together produce fragment AB and CD, respectively, which partially overlap. In some cases, unwanted PCR products can appear while amplifying segments AB and CD. In those cases, gel purification of desired amplicons is recommended. The partial overlap at the mutation site enables the synthesis of a whole double-stranded DNA molecule with the intended mutation in a subsequent PCR. The final linear PCR product is then cloned into the appropriate plasmid by restriction enzyme- and DNA-ligase-dependent cloning (Ho et al., 1989).

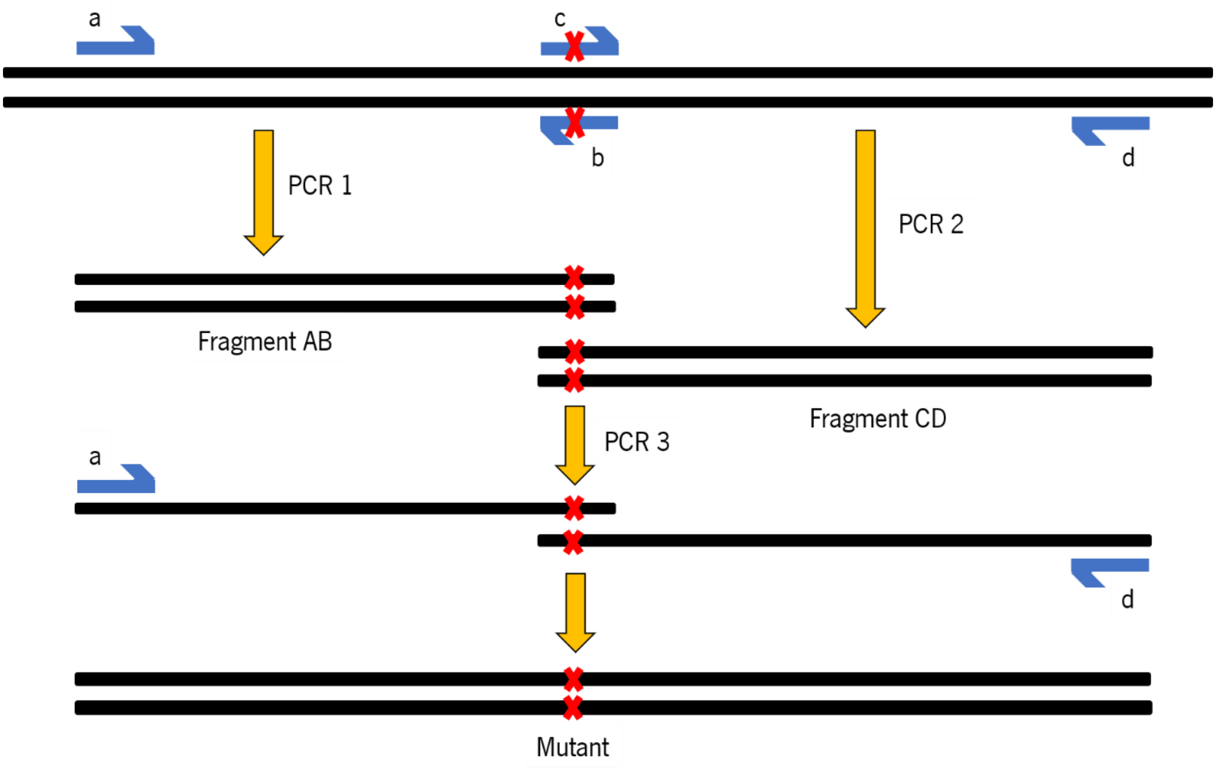


Figure 1 | Overview of the overlap extension method for site-directed mutagenesis. Two pairs of primers, each containing one primer with the desired mutation (shown as red crosses in primers b and c), are required. Primers a & b and c & d are used in two separate PCRs to amplify their corresponding segments, AB and CD. Since the produced fragments in both PCRs have the mutated segment in common, it serves as a "primer" in the third PCR and a complete double-stranded DNA molecule with the desired mutation can be synthesized. To amplify the complete segment, primers a and d can be reutilised.

In iPCR, primers facing "outwards," away from each other (see Figure 2A), are used, in contrast to a typical PCR wherein "in-facing" flanking primers are employed. It is simpler than the overlap extension method and numerous variations of the method, making use of various types of primer combinations, have been described (see Figure 2A). Non-overlapping, partially overlapping or fully overlapping primers can be used, giving rise to PCR products which are then circularised and transformed to the host. iPCR using non-overlapping primers is a multi-step process involving mutant introduction by PCR, *Dpn*I digestion of methylated template DNA, circularisation by phosphorylation and ligation and transformation to a suitable host (*Escherichia coli*) for propagation. It is the basis of numerous commercial mutagenesis kits, e.g. the Phusion (Thermo Scientific), Q5 (New England Biolabs), and KOD-Plus (Toyobo) site-directed mutagenesis kits (Silva et al., 2017). iPCR using fully overlapping primers makes use of a simpler protocol and forms part of the commonly used QuickChange™ Site-Directed Mutagenesis Kit (Agilent Technologies)

(Figure 2B). Here, both primers contain the intended mutation, and PCR gives rise to mutated DNA strands which are treated with the endonuclease *Dpn*I, and directly transformed to competent cells (Bauer et al., 1995a). The exact mechanism by which this functions is still under discussion, but linear amplification to give a circular product was previously proposed, whereas more recently, exponential amplification, to give a linear product with homologous ends and thereby enabling recombination, has been suggested (Liu and Naismith, 2008; Xia et al., 2015). Finally, iPCR with partially overlapping primers appears to be much less used, it is believed to involve production of a linear PCR product with short homologous ends that is circularised by recombination (Silva et al., 2017).

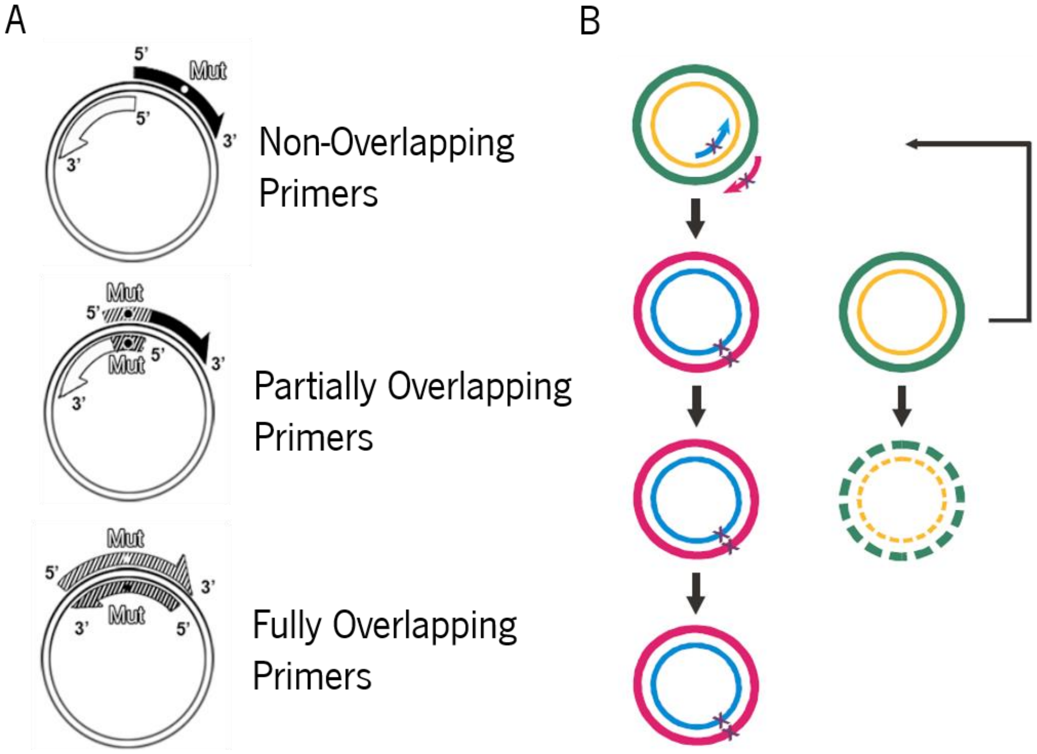


Figure 2 | Illustration of iPCR primer layouts and representation of the various steps involved in iPCR with the commercial site-directed mutagenesis kit QuickChange™ (Agilent Technologies). Use of iPCR to introduce point mutations. (A) Three different primer layouts can possibly be used: non-overlapping, partially overlapping or fully overlapping. Hatchet sections indicate overlapping regions of primers. (B) The mechanism of action suggested by the manufacturer in which linear amplification gives rise to a circular product is shown. Only one pair of primers (red and blue arrows) is required. Each primer has the desired mutation incorporated (marked by a red cross) and anneals to a different template strand (yellow and green). After amplification, the template is degraded with the restriction endonuclease *Dpn*I. Adapted from Silva et al., 2017 and taken from QuickChange™ II Site-Directed Mutagenesis Kit Instruction Manual, 200523, Agilent Technologies.

1.3 Multi-Site-Directed Mutagenesis

Introduction of mutations at single sites in proteins has already proven successful in altering protein performance, but introduction of multiple mutations, at multiple sites throughout the protein, often leads to combined and stronger effects. As an example, in the case of proteins with a reduced solubility due to a high exposed hydrophobic surface area, it is believed that mutations of complete, or almost complete hydrophobic patches, as opposed to single isolated hydrophobic residues, has a much greater efficiency in disrupting inter-protein hydrophobic interactions and thereby reducing protein precipitation (Matsui et al., 2017; Sormanni et al., 2015, 2017).

A variety of multi-site-directed mutagenesis techniques currently exist. The most basic approach for introducing multiple mutations in a protein is to carry out consecutive single site mutations one at a time, yet this approach is very laborious and time consuming. Currently, the simplest form for introducing several mutations is through the application of either of two commercially available kits: the QuikChange™ Multi Site-Directed Mutagenesis Kit (Agilent Technologies) or the GeneArt Site-Directed Mutagenesis PLUS System (Thermo Fisher Scientific). These kits represent the two mainstream PCR-mediated multi-site directed mutagenesis approaches currently used. Both kits use PCR with mutagenic primers to produce the desired mutations but differ significantly in the mechanism for preparing the final mutated circular product.

The QuikChange™ Multi Site-Directed Mutagenesis Kit is based upon the binding of two or more oligonucleotide primers containing the desired mutations to the same strand of the template construct (Figure 3). During the PCR extension phase, and depending on the number of primers used, two or more fragments are synthesised (with *Pfu* DNA polymerase) and ligated (with e.g. *Taq* DNA ligase) *in vitro*. Thereafter, the DNA template is degraded with the restriction enzyme *Dpn*I and the resulting single-stranded DNA transferred to competent cells. The cells then synthesize the complementary strand to the mutated single-strand DNA, thereby giving rise to the mutated double-stranded circular product (Bauer et al., 1995a, 1995b, 1995c; Hogrefe and Cline, 2001).

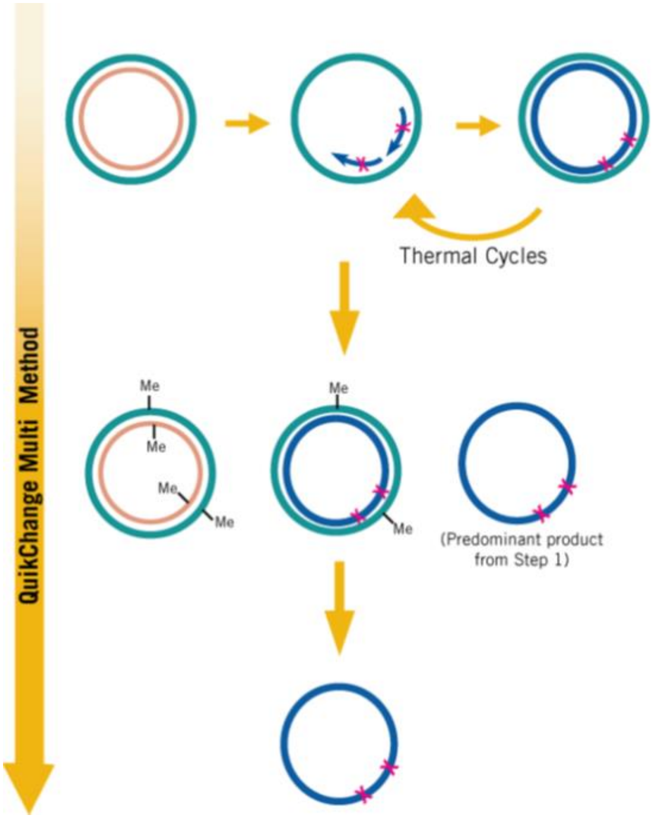


Figure 3 | Overview of the QuikChange™ Kit multi-site-directed mutagenesis kit from Agilent Technologies. Each mutation introduced (marked as a red cross) requires one mutated primer (blue arrows). A maximum of five mutations are possible per PCR without compromising efficiency. Each primer anneals to the same template strand and amplifies the segment until it encounters another primer. Template strands are then degraded by *Dpn*I treatment after amplification and fragments are ligated and introduced into the host to produce double-stranded vectors. Adapted from the QuickChange™ Multi Site-Directed Mutagenesis Kit Instruction Manual, 200514, Agilent Technologies.

The second, and more recently developed kit, by Thermo Fisher Scientific (GeneArt Site-Directed Mutagenesis PLUS System), is based on PCR and *in vitro* homologous recombination. Briefly, a multiplex, or up to three separate PCRs, are performed to amplify mutated sequences (up to three-point mutations are possible per mutagenesis round without compromising efficiency). The mutations are introduced by PCR with overlapping primers that have the centred mutated base flanked on each side by at least 10 non-altered nucleotides. This leads to products with end-terminal homologies (Figure 4) which are treated with *Dpn*I for template removal and recombined *in vitro* by use of a proprietary, manufacturer supplied, enzymatic mix. This promotes *in vitro* homologous recombination, linking fragments together and producing whole mutated circular plasmids which are then transformed to *E. coli* (Liang et al., 2012).

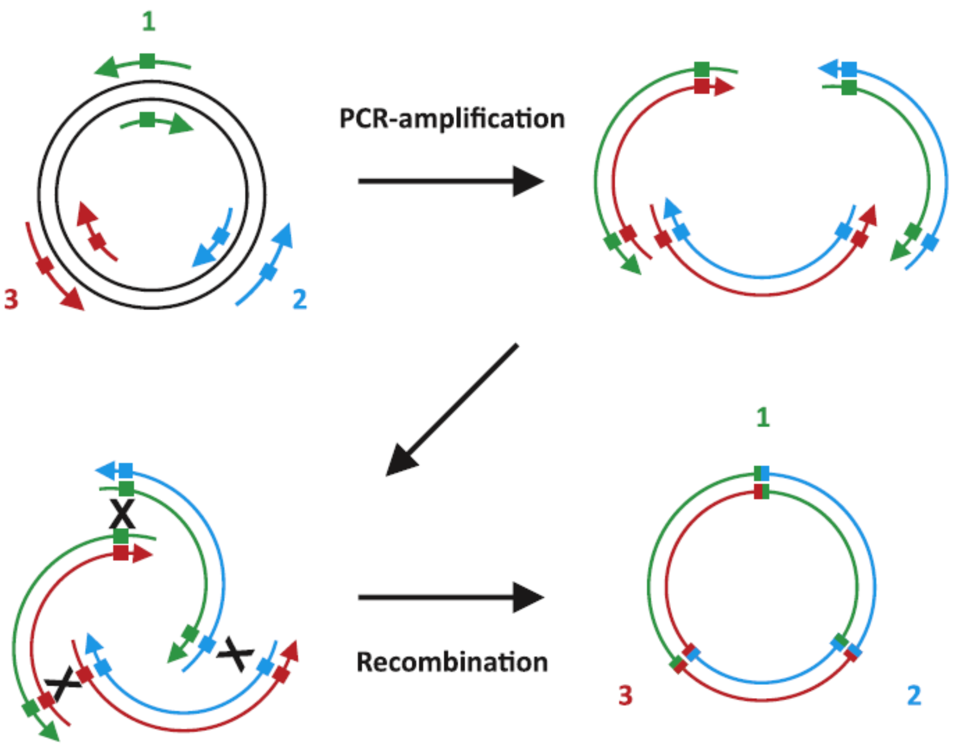


Figure 4 | Multi-site-directed mutagenesis method based on *in vitro* homologous recombination. This method can reliably introduce up to three mutations at a time, with each mutation requiring one pair of primers (coloured arrows). Each mutated base (coloured squares) is required to be flanked by at least 10 nt at each side. Reactions can be carried out in one multiplex or three separate PCRs. Amplified products have homologous end regions which are used to produce a circularised plasmid via an enzyme catalysed *in vitro* recombination. Adapted from Liang et al., 2012.

1.4 Fragment Assembly in Multi-Site-Directed Mutagenesis

From the above discussion, it can be seen that the most common site-directed mutagenesis approaches used today are multi-step PCR based methods. PCR amplification to introduce the desired mutations is followed by recircularisation or directed fragment assembly in the case of multi-site-directed mutagenesis, and then transformation to a host for propagation (Gibson et al., 2009; Hogrefe and Cline, 2001; Liang et al., 2012). Different approaches can be employed for the fragment assembly step. Historically, restriction enzymes and ligases have been used but this is sequence dependent and time consuming and hence various other novel approaches have been developed (Bhat et al., 1991; Ito et al., 1991; Scheller et al., 1977). These include enzyme based *in vitro* methods and *in vivo* or *in vitro* recombination based methods. The enzyme based *in vitro* methods rely on the use of exonuclease activities for the generation of complementary overhangs at overlapping termini of fragments and their subsequent enzyme catalysed annealing and repair e.g. Ligation Independent Cloning (LIC), Gibson Assembly and In-Fusion cloning (Takara Bio) (Gibson et al., 2009; Scholz et al., 2013). *In vitro* recombination based methods involve the direct use of proteins involved in recombination or, more probably and more economically, recombination promoting bacterial cell extracts, such as described above for the GeneArt Site-Directed Mutagenesis PLUS System (Li and Elledge, 2007; Zhang et al., 2012). All these *in vitro* methods suffer from high cost and laborious procedures, with often long overhangs being required. In contrast, *in vivo* recombination mediated fragment assembly, making use of the intrinsic recombination machinery of a biological system, should offer much potential for efficient, simpler and reduced cost protocols (Huang et al., 2017; Jacobus and Gross, 2015; Kostylev et al., 2015). Currently, this appears to be principally based on the use of yeasts, especially *Saccharomyces cerevisiae,* taking advantage of their powerful recombination abilities, yet suffers from a need for laboratory capabilities for working with eukaryotic systems and the necessity for specialised yeast plasmids (yeast compatible shuttle vectors), thereby limiting its more widespread application in protein engineering (Baudin et al., 1993; Carter and Delneri, 2010; Gibson, 2009; Güldener et al., 1996; Manivasakam et al., 1995; Oldenburg et al., 1997). Interestingly, the use of *E. coli* for *in vivo* recombination would overcome these disadvantages, yet, while the recombination capabilities of *E. coli* have been known for some time, it does not seem to have been widely adapted for fragment assembly in site-directed mutagenesis protocols. Recombination in *E. coli* depends mostly on the single-strand binding protein RecA but can also be promoted by phage systems (the RecET and Red systems) and recently a RecA-independent pathway has even been suggested (Lovett et al., 2002; Stewart et al., 1999, 1997; Zhang et al., 1998). As recombination in *E. coli* is the principal focus of the present study, the following sections will give an in-depth description of recombination and of the different types and mechanisms for achieving this in *E. coli*.

1.5 Recombination

All living cells experience minor damages throughout their lifecycle to their DNA and these can easily be repaired through a variety of processes, including nucleotide excision repair and base excision repair (Gao et al., 2017; Michaels et al., 1992). However, these processes are insufficient in the event of severe DNA damage, as in the case of double-strand breaks (broken or incomplete chromosomes), and thus requires a more sophisticated solution. To repair major DNA damages, living cells use a process called recombination – consisting of pairing homologous parts of DNA molecules to repair DNA damage.

It is suggested to be a process developed by ancestral cells since key proteins of this process exist throughout the three domains of life – Bacteria (RecA), Archaea (RadA) and Eukarya (RAD51) (Morita et al., 1993; Seitz et al., 1998; Shibata et al., 1981; Shinohara et al., 1992). In molecular biology, recombination can be useful in performing recombination-mediated genetic engineering (recombineering) for any type of mutation (insertion, deletion, substitution). It can also be used for cloning purposes, where linear DNA fragments are assembled into a plasmid, or to perform allele rescues or even gene replacements (Oldenburg et al., 1997; Rothstein, 1991). In the latter, knockout cassettes are used to replace genes with selection markers within the genome or artificial chromosome (Baudin et al., 1993; Carter and Delneri, 2010; Güldener et al., 1996).

1.5.1 Recombination Types

The main mechanism by which recombination acts is through pairing of homologous parts of DNA molecules and exchanging DNA strands (gene conversion) to re-establish integrity. Currently, three types of recombination-mediated DNA repair have been identified: broken fork repair (known as double-strand end repair or recombination-dependent replication in *E. coli*) (Figure 5), double-strand break repair (Figure 6) and gap-filling recombinational repair (known as post-replication repair in *E. coli*) (Figure 7) (Kuzminov et al., 1994; Rupp and Howard-flanders, 1968; Rupp et al., 1971; Szostak et al., 1983). The recombination process for all three types can be split into three stages. The first stage (pre-synapsis) is the preparation of single-stranded DNA for strand invasion (recombinogenic DNA, Figure 5A and B; for further clarification see Figures 8 and 9) (Heuser and Griffith, 1989; McEntee et al., 1979; Shibata et al., 1979a). In stage two (synapsis), protein-mediated strand invasion occurs and a heteroduplex is formed (Figure 5C) (Cox and Lehman, 1981a, 1981b; Kowalczykowski et al., 1987; McEntee et al., 1979; Shibata et al., 1979a). In the final stage (post-synapsis) strand exchange is disrupted (Figure 5D) or completed, producing recombinant DNA duplexes (Figure 5E and F) (Shibata et al., 1979a). Each stage requires a set of proteins, dictated by the chosen recombination pathway, which may or may not participate in more than one type of recombination-mediated DNA repair (Ivančić-Baće et al., 2005). While the mechanisms described here are in the context of *E. coli*, these also extend to many other organisms across the three domains of life, including *S. cerevisiae* (Seitz et al., 2001; Symington et al., 2014).

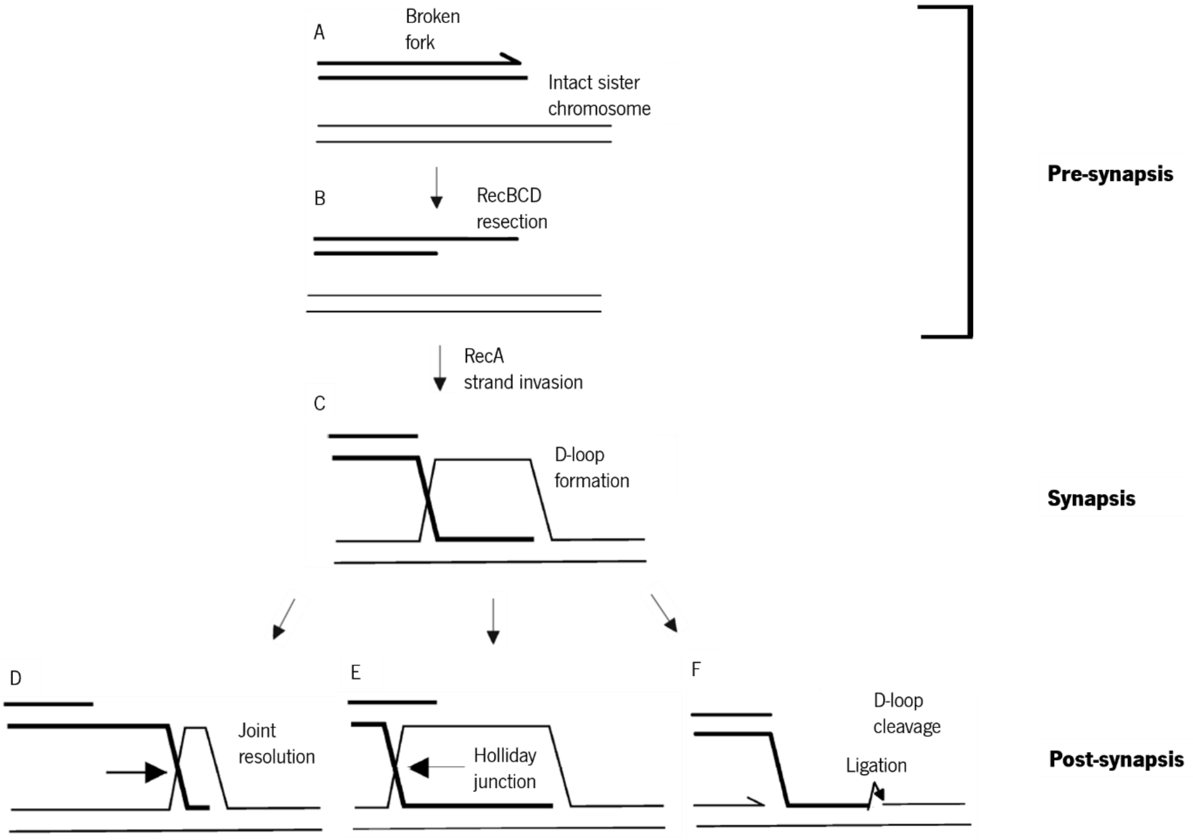

A
Broken fork
Intact sister chromosome
Pre-synapsis
B
RecBCD resection
RecA strand invasion
C
D-loop formation
Synapsis
D
Joint resolution
E
Holliday junction
F
D-loop cleavage
Ligation
Post-synapsis

Figure 5 | Broken fork repair model. Broken fork repair is initiated after the event of a collapsed replication fork. (A) To re-establish chromosome integrity a sister chromosome is required to serve as template for the missing region. (B) The protein complex RecBCD promotes a partial degradation of the dsDNA to produce 3'-overhang dsDNA (resection). (C) RecA binds to the resected DNA and searches for homologous parts within the sister chromosome (strand invasion) forming in the process a three-stranded structure called displacement-loop (D-loop). (D) Branch migration occurs in the 3'-end of the invading strand leading to joint resolution (unsuccessful recombination). (E) Branch migration occurs in the 5'-end of the invading strand and produces a four-branched structure called Holliday-junction which promotes recombination. (F) Branch migration is absent due to a D-loop cleavage. Ligation is performed and both chromosomes retain a template to re-establish chromosome integrity. Adapted from Persky and Lovett, 2008.

Broken fork repair (Figure 5) initiates when a replication fork has collapsed, for instance due to nicks in the template (Heitman et al., 1999; Kouzminova and Kuzminov, 2004; Kuzminov, 1995; Kuzminov et al., 1994). To begin broken fork repair in *E. coli*, the 5'-end of a double-strand DNA molecule is digested by the RecBCD complex to expose the complementary 3'-end DNA (resection, Figure 5A and B) (Arnold and Kowalczykowski, 2000; Kuzminov et al., 1994; Singleton et al., 2004; Wiktor et al., 2018). As resection occurs, single-strand binding proteins called RecA (note that RecA differs from single-strand binding proteins – SSBs) are loaded onto the exposed single-strand and facilitate strand invasion (Figure 5C; for further clarification see Figure 8)(Anderson and Kowalczykowski, 1997a; Arnold and Kowalczykowski, 2000; Churchill et al., 1999; Clark and Margulies, 1965; Willetts et al., 1969). Strand invasion consists of a single-strand DNA displacing one strand of the double-stranded DNA to form a heteroduplex (Cox and Lehman, 1981a, 1981b). The invasion and subsequent displacement produces a three-stranded structure called displacement-loop (D-loop, Figure 5C) that can lead to a branch migration, and can either resolve the joint, if migration is in direction of the 3'-end of the invading strand (Figure 5D), or produce a Holliday junction if migration is in direction of the 5'-end of the invading strand (Figure 5E) (Cox and Lehman, 1981a, 1981b; Kowalczykowski et al., 1987; McEntee et al., 1979; Smith, 1991). Finally, a new replication fork can be formed by either cleaving the D-Loop and subsequent ligation (Figure 5F), or by resolving the Holliday junction (Cox and Lehman, 1981a; Smith, 1991).

Double-strand break repair (Figure 6) involves a similar repair mechanism to that previously mentioned, except that instead of only recruiting one end of a chromosome, two are required to repair a section in the damaged DNA molecule (Szostak et al., 1983). Many eukaryotes, including the yeast *S. cerevisiae*, are able to repair these DNA lesions through non-homologous end joining (Kramer et al., 1994; Moore and Haber, 1996). Conventional non-homologous end joining is absent in *E. coli*, though a similar mechanism requiring small homologies named alternative end-joining could correct these double strand breaks, even though in an inefficient manner (Chayot et al., 2010; Rocha et al., 2005). While *S. cerevisiae* can make small deletions or insertions and recombine effortlessly with small homologies (<10 nt) at a potential cost of losing genetic information, *E. coli* requires recruitment of an intact homologous chromosome to repair double-strand gaps efficiently (Kramer et al., 1994; Moore and Haber, 1996). Similar to the first type of recombination-mediated DNA repair described above (broken fork repair), both ends of the damaged chromosome are digested at their 5'-ends to expose a 3'-end tail, with RecA being loaded onto it shortly afterwards (Figure 6A, B and C) (Anderson and Kowalczykowski, 1997a; Arnold and Kowalczykowski, 2000; Wiktor et al., 2018). Strand invasion initiates on one end, resulting in the displacement of one of the strands in the duplex (Figure 6C) (Resnick, 1976). If the branch migration favours the heteroduplex formation, the displaced strand is capable of aligning with the other 3'-end tail, forming in total two Holliday junctions (Figure 6D). Favourable branch migration and DNA synthesis during the process leads to the formation of two intact chromosomes once the Holliday junctions are cleaved off (Figure 6E and F) (Resnick, 1976; Smith, 1991). However, if branch migration hinders heteroduplex formation during D-Loop formation, primed DNA synthesis ceases (Figure 6C' and D') (Nassif et al., 1994; Persky and Lovett, 2008). Nevertheless, the primed DNA synthesis might still have been sufficient to anneal to the other broken segment and complete DNA repair (Figure 6E' and F') (Gumbiner-Russo and Rosenberg, 2007; Nassif et al., 1994). This alternative process is known as synthesis-dependent strand annealing (SDSA) (Nassif et al., 1994).

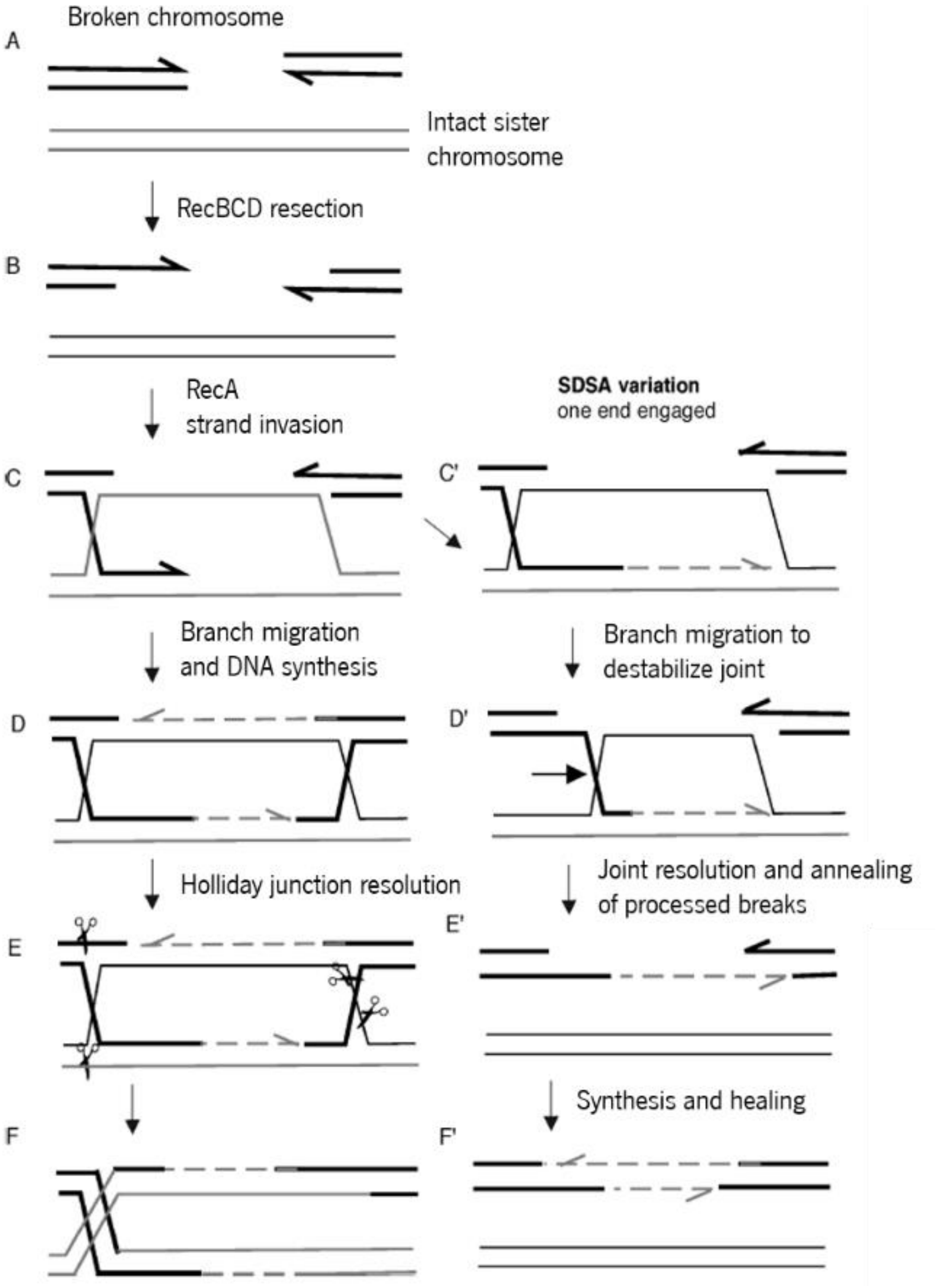

A
Broken chromosome
Intact sister chromosome
RecBCD resection
B
RecA strand invasion
SDSA variation
one end engaged
C
C'
Branch migration and DNA synthesis
Branch migration to destabilize joint
D
D'
Holliday junction resolution
Joint resolution and annealing of processed breaks
E
E'
Synthesis and healing
F
F'

Figure 6 | Double-strand break repair model. Double-strand break repair is recruited in the event of missing sections in a chromosome. (A) Both ends of the lesioned chromosome are paired with a sister chromosome. The protein complex RecBCD promotes a partial degradation of the dsDNA to produce 3'-overhang dsDNA (resection). (C) RecA binds to the resected DNA and searches for homologous parts within the sister chromosome (strand invasion), forming in the process a three-stranded structure called a displacement-loop (D-loop). (D) Branch migration is directed to the 5'-end of the invading strand, forming a double Holliday junction and enabling DNA synthesis. (E) Resolution of Holliday junctions is performed via resolvases. (F) Successful recombination can yield cross-over products. (C') Only one chromosome end engages in recombination, promoting DNA synthesis on the invading strand only; synthesis-dependent strand annealing is initiated. (D') Branch migration in the 3'-end direction destabilizes the joint. (E') Strand invasion ceases and partial or complete synthesis of the missing region of the invading strand is capable of binding to the other chromosome end. (F') DNA synthesis of missing region is performed, re-establishing chromosome integrity. Adapted from Persky and Lovett, 2008.

The third and final type of recombination-mediated repair (Figure 7), gap-filling recombinational repair, is thought to be employed in cases where DNA replication ceased for short durations of time due to bound proteins, DNA secondary structures or other lesions that led to the temporary arrest of the replisome (Bichara et al., 2011; Persky and Lovett, 2008; Rupp and Howard-flanders, 1968). The proposed model suggests that the intact single-strand, within the gap region, interacts with its complementary sequence in a homologous chromosome (Figure 7A and B), forming a half- or whole-Holliday junction and resulting in a sequence transfer from the homologous chromosome to the incomplete one (Figure 7C and C'). Meanwhile, primed DNA synthesis on the homologous chromosome replaces the transferred sequence (Figure 7D) (Bichara et al., 2011; Persky and Lovett, 2008). Although gaps within the DNA sequence can be completed through recombination events, an error-prone mechanism called trans-lesion DNA synthesis can be employed to fill in the gaps (Goodman and Tippin, 2000).

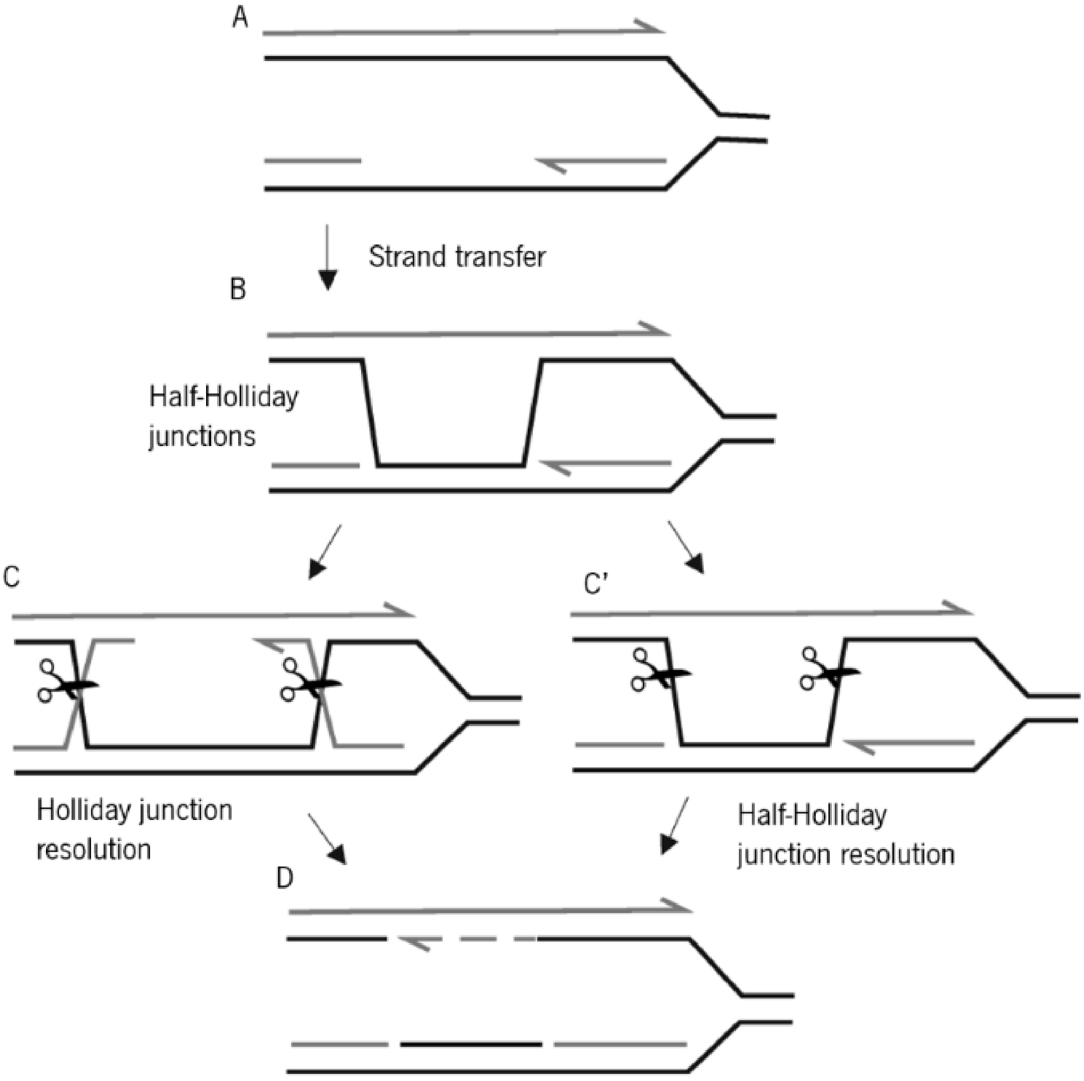

A
Strand transfer
B
Half-Holliday junctions
C
C'
Holliday junction resolution
Half-Holliday junction resolution
D

Figure 7 | Gap-filling recombinational repair model. Gap-filling recombinational repair is performed in the presence of single-strand gaps within a chromosome. (A) Intact strand of the incomplete chromosome interacts with the complementary sequence in an intact chromosome. (B) Part of the intact strand is transferred in a RecA-dependent matter to the gapped region. (C) Primed DNA synthesis on both ends of the gap interact with the transferred strand, forming a Holliday junction. (C') Non primed DNA synthesis leads to a half-holliday junction that promotes cleavage. (D) Both Holliday und Half-Holliday junctions are resolved and produce a complete intact chromosome and a DNA synthesis primed one which restores chromosome integrity. Adapted from Persky and Lovett, 2008.

While production of recombinogenic DNA in *E. coli* during the first step of recombination (pre-synapsis), for all three types of recombination, is principally promoted by the single-strand binding protein RecA (RecA-dependent recombination), RecA-independent recombination has also been identified and is only beginning to be understood. Both forms will now be discussed.

1.5.2 RecA-Dependent Recombination

E. coli performs recombination primarily through the RecA protein. RecA is a single-strand DNA binding protein, different to SSBs, which promotes recombination throughout all three stages (pre-synapsis, synapsis and post-synapsis) (Clark and Margulies, 1965; Galletto et al., 2006; McEntee et al., 1979; Shibata et al., 1979a, 1979b; Willetts et al., 1969). However, since SSBs are ubiquitous, strongly bind to ssDNA and inhibit initial RecA filament formation, a RecA loading process is required to efficiently promote recombination (Fu et al., 2013; Joo et al., 2006; Roy et al., 2009; Rupp and Howard-flanders, 1968; Thresher et al., 1988; Umezu et al., 1993). The two main pathways to load RecA onto ssDNA are the RecBCD and the RecFOR pathways. These pathways show different substrate preferences and are thought to be involved in the repair of different DNA lesions.

The RecBCD pathway is employed in the event of double strand breaks (e.g. double-strand break repair) and uses the RecBCD complex to create ssDNA and load RecA onto it (Figure 8) (Anderson and Kowalczykowski, 1997a; Arnold and Kowalczykowski, 2000; Churchill and Kowalczykowski, 2000; Singleton et al., 2004; Wiktor et al., 2018). The RecB subunit is a helicase-nuclease that is capable of digesting both the 5'- and 3'-end, while the RecD helicase attaches to the 5'-end and inhibits RecA loading by RecB (Anderson and Kowalczykowski, 1997b; Boehmer and Emmerson, 1992; Dillingham et al., 2003; Handa et al., 2005; Taylor and Smith, 1985, 2003; Yu et al., 1998). The crystal structure of RecC suggests that it is responsible for splitting duplexes through a "pin"-structure and guiding the single-strands to their respective subunits (RecB and RecD) (Singleton et al., 2004). Moreover, RecC is involved in recognising a specific octamer sequence (GCTGGTGG), known as Chi (crossover hotspot instigator, χ), within a specific sequence context (Amundsen et al., 2016; McMilin et al., 1974; Ponticelli et al., 1985; Smith et al., 1984). The Chi sequence promotes RecBCD-dependent recombination by modulating its nuclease activity. Upon passing over a Chi sequence, RecC induces an overall conformational change in the RecBCD complex that decouples the RecD subunit from the DNA while RecB changes its preferential cleave activity from the 3'-end to the 5'-end (Figure 8) (Amundsen et al., 2016; Anderson and Kowalczykowski, 1997b; Handa et al., 2005). Moreover, the overall process speed of the complex becomes reduced (Wiktor et al., 2018). As a result, a ssDNA tail emerges at the 3'-end with RecA being continuously loaded onto the tail by RecB to produce a RecA filament (Figure 8) (Anderson and Kowalczykowski, 1997a; Churchill and Kowalczykowski, 2000). Through this process, SSBs are unable to bind to the DNA and inhibit RecA filament formation (Fu et al., 2013). Interestingly, once RecA nucleation occurs, SSBs further promote RecA filament formation, presumably by removing secondary structures (Fu et al., 2013; Kowalczykowski and Krupp, 1987; Muniyappa et al., 1984). When overhangs are present and recombination-mediated repair is required, RecBCD is unable to bind due to inhibition by even small overhangs (> 25 nt) (Taylor and Smith, 1985). In this case, the alternative RecFOR pathway can be used by *E. coli* (Tseng et al., 1994; Umezu et al., 1993). Alternatively, end-blunting by single-strand exonucleases such as exonuclease I (ExoI or SbcB), RecJ or SbcCD can be performed to allow RecBCD to initiate its corresponding pathway (Connelly et al., 1997; Lehman and Nussbaum, 1964; Lovett and Kolodner, 1989; Thoms and Wackernagel, 1998).

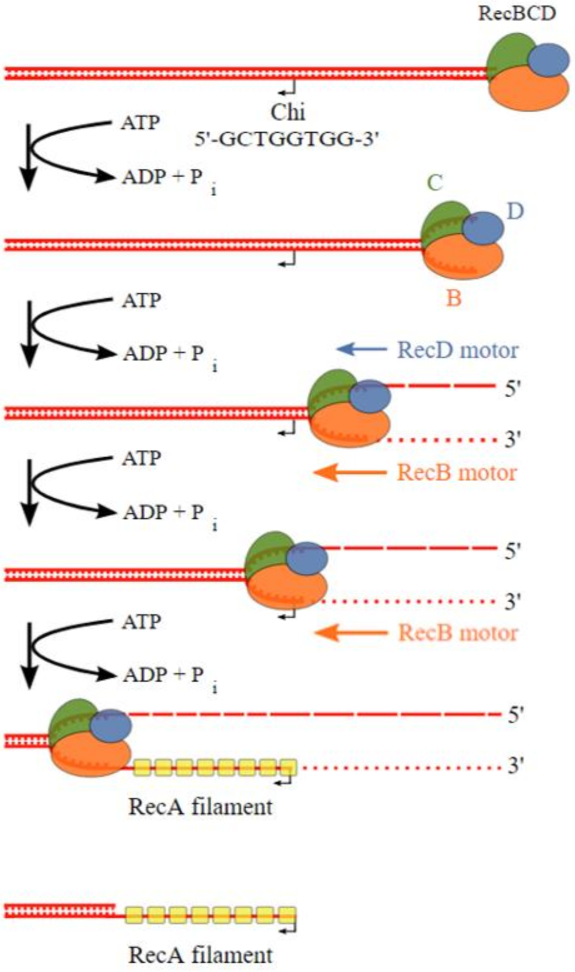


Figure 8 | The RecBCD pathway. To initiate double-strand break repair, assembly of the three subunits RecB (orange ovoid), RecC (green ovoid) and RecD (blue ovoid) is essential. The RecBCD complex combines the activity of a helicase with a nuclease and uses only blunt or near (up to 25 nt) blunt ended dsDNA as substrate. Upon binding to dsDNA and before encountering a Chi sequence (black arrow) the 3'-end is cleaved favourably in comparison to the 5'-end and the motor units of RecB and RecD are active. The recognition of a Chi sequence by RecC, disengages the RecD motor unit from the DNA. This is accompanied by a structural change in the RecBCD complex which now cleaves preferentially the 5'-end of the dsDNA. Decoupling the RecD unit from the DNA promotes RecA loading by the RecB unit, leading to the formation of a RecA filament on 3'-overhangs. Adapted from Singleton et al., 2004.

The RecFOR pathway is thought to be specialised in repairing DNA lesions unsuitable for the RecBCD pathway, namely ssDNA gaps. SSB proteins readily bind to exposed ssDNA and hinder RecA nucleation (Fu et al., 2013; Joo et al., 2006; Roy et al., 2009; Rupp and Howard-flanders, 1968; Thresher et al., 1988). The proteins of this pathway, RecF, RecO and RecR, are capable of removing SSBs bound to ssDNA and replacing it with RecA. RecF binds in an ATP-dependent and RecR modulated matter to dsDNA and is thought to direct and limit RecA nucleation to ssDNA gaps (Figure 9) (Madiraju and Clark, 1992; Webb et al., 1995, 1997). RecO has been shown to interact strongly with SSBs, and in conjunction with RecR (forming RecOR) promotes replacement of SSBs within the ssDNA gap with RecA (Figure 9). In addition, RecOR also stabilises the RecA filament to prevent RecA filament dissociation (Umezu and Kolodner, 1994). Interestingly, RecF is a dispensable protein for this process (Figure 9), pairing of RecOR is therefore sufficient to perform gap-filling recombinational repair on its own (Sakai and Cox, 2009; Umezu et al., 1993).

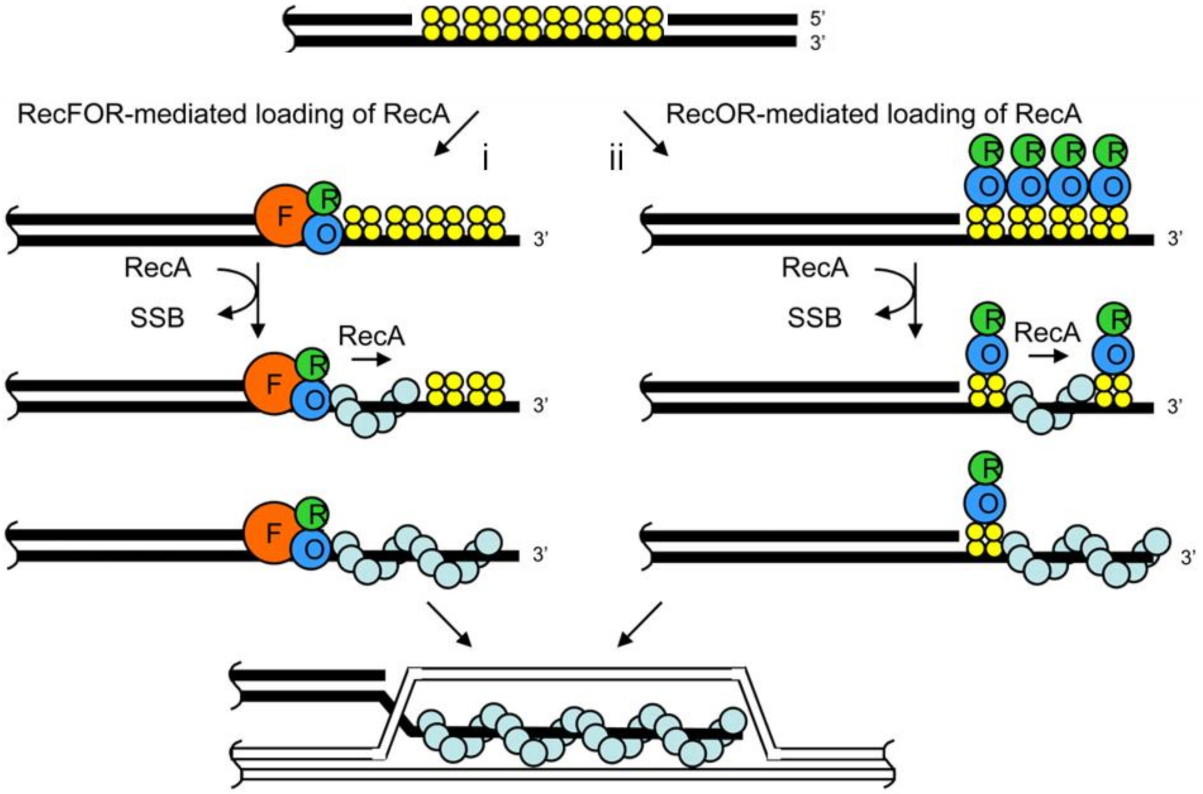

5'
3'
RecFOR-mediated loading of RecA
i
ii
RecOR-mediated loading of RecA
F
R
O
3'
RecA
SSB
RecA
→
RecA
SSB
RecA
→

Figure 9 | The RecFOR pathway model. The presence of ssDNA gaps promotes the binding of SSBs (yellow circles). Substitution of SSBs with RecA (grey circles) can be mediated by either RecF (orange circle), RecO (blue circle) and RecR (green circle) or alternatively without the presence of RecF. (i) RecF binds to the interface of ssDNA and dsDNA to contain the RecA filament within the ssDNA gap. (i and ii) RecOR promotes the release of SSBs from ssDNA, mediates RecA nucleation and further stabilises the filament formation in the 3' direction. Adapted from Handa et al., 2009.

The RecFOR pathway is not limited to ssDNA gap repairs, it is also capable of resolving double strand breaks through involvement of a synergistic mechanism performed by RecQ and RecJ (Morimatsu and Kowalczykowski, 2014). RecQ acts as a helicase to unwind the dsDNA (Umezu et al., 1990). Its substrate preference is dsDNA with 3'end overhangs (Figure 10A) or blunt-ended dsDNA (Figure 10B) (Morimatsu and Kowalczykowski, 2014; Umezu et al., 1990). On the other hand, RecJ assumes the role of a ssDNA exonuclease with preference for 5'-end overhangs (Figure 10C) (Lovett and Kolodner, 1989; Morimatsu and Kowalczykowski, 2014). Moreover, extensive digestion performed by RecJ is enhanced in the presence of SSBs and suggests a recruitment through this interaction (Han et al., 2006). Together, RecQ and RecJ process any type of DNA end (blunt, 3'-end or 5'-end overhangs) to produce 3'-ended overhangs for the RecFOR pathway to initiate (Figure 10D and E) (Bork et al., 2002; Morimatsu and Kowalczykowski, 2014; Shan et al., 1997).

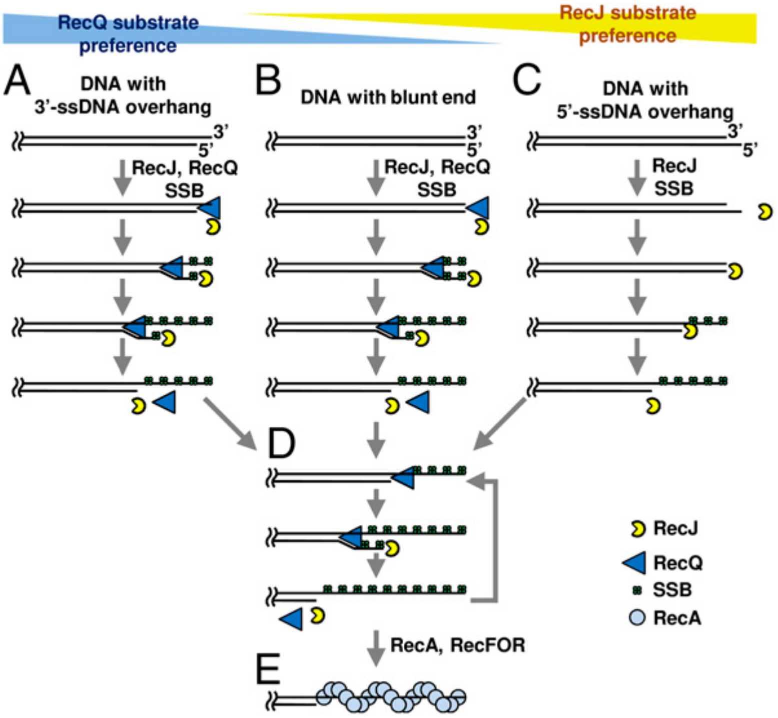

RecQ substrate preference
RecJ substrate preference
A
DNA with 3'-ssDNA overhang
B
DNA with blunt end
C
DNA with 5'-ssDNA overhang
3'
5'
RecJ, RecQ SSB
RecJ SSB
D
E
RecA, RecFOR
RecJ
RecQ
SSB
RecA

Figure 10 | Mechanism of resection of the three types of dsDNA ends by helicase RecQ and nuclease RecJ. In the absence of RecBCD, the combined action of RecQ (blue triangle) and RecJ (yellow indented circle) can process any type of dsDNA end for the RecFOR pathway. (A) dsDNA with 3'overhangs are unwound by RecQ helicase enabling RecJ exonuclease to digest the 5'-end. (B). Blunt ended dsDNA is also first processed by RecQ to enable RecJ-dependent resection. (C) 5'-overhangs do not require the presence of RecQ, instead RecJ directly processes the dsDNA 5'-end into a dsDNA with 3'overhang. (D) Repeated processing of 3'-overhangs by RecQ and RecJ increase the substrate availability for the RecFOR to operate on. (E) The RecFOR pathway mediates RecA-loading (grey circles) onto the produced dsDNAs with 3'-end tails. SSBs (green circles) are removed from the DNA substrate in the process. Adapted from Morimatsu and Kowalczykowski, 2014.

Recently a rather unconventional RecA loading mechanism has been identified. This mechanism requires the presence of the *sbcB15* mutant encoding a partial functioning exonuclease with abolished nuclease activity but retained ssDNA binding function (Buljubašić et al., 2019; Phillips et al., 1988; Thoms et al., 2008). Evidence supporting the interaction between SbcB and SSBs have led to the assumption that SbcB15's role relies not simply on its ability to protect ssDNAs, but rather in facilitating RecA loading either directly or via interaction with other mediating proteins (Buljubašić et al., 2019; Butland et al., 2005; Sandigursky et al., 1996). Besides RecQ and RecJ, none of the other proteins involved in either pathways (RecBCD and RecFOR) are required and thus the pathway was coined RecBFI (RecBCD-FOR-independent) (Buljubašić et al., 2019). Upon resection, it is postulated that SbcB15 binds to the SSB-bound ssDNA, releasing SSB in the process. From here two possibilities are envisioned: initial RecA nucleation is promoted at the site where SbcB15 binds to the ssDNA and further RecA filament extension is achieved passively (which would explain the lower recombination efficiency) (Figure 11). Alternatively, SbcB15 binds and moves through the ssDNA while also dissociating the SSB-ssDNA complex allowing for a more rapid RecA filament creation (Figure 11) (Buljubašić et al., 2019). Once a RecA filament has been formed, the RecA-DNA complex will seek homologous sequences through strand invasion (synapsis). Long homologies promote the formation of Holliday junctions and consequently strand exchange (post-synapsis). Strand exchange is then completed once resolvases such as RuvC cleave Holliday junctions and produce recombined DNA products (Iwasaki et al., 1991).

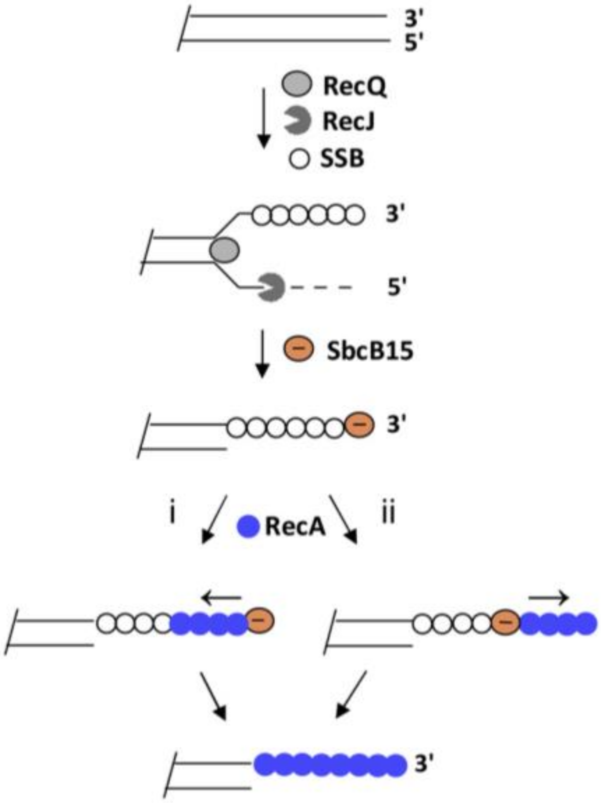


Figure 11 | Envisioned RecA-loading mechanism by SbcB15. In *recBCD* (or *recCD*) *sbcB15 recFOR* mutants dsDNA is resected by the combined activity of RecQ (grey circles) helicase and RecJ (indented grey circle) exonuclease (5'-3'). SSBs (empty circles) readily bind strongly to ssDNA. SbcB15 (orange circle) interacts with the SSB-ssDNA complex and promotes RecA (blue circles) nucleation by one of two ways. (i) SbcB15 binds stably and promotes RecA nucleation in the 3'-5' direction. (ii) SbcB15 binds and migrates along the ssDNA (3'-5'), dissociating in the process SSBs and facilitating RecA nucleation in 3'-end direction. Adapted from Buljubašić et al., 2019.

1.5.3 RecA-Independent Recombination

Evidence supporting the existence of an endogenous RecA-independent recombination system within *E. coli* started appearing during the early 1990s. Indeed, many strains used in molecular cloning e.g. DH5α and XL1-Blue, have been engineered to express a dysfunctional RecA (*recA1*) as this protein has been shown to negatively affect cell transformation efficiency and plasmid stability, yet, these cloning strains are capable of recombining without the addition of any other auxiliary system (García-Nafría et al., 2016; Huang et al., 2017; Kostylev et al., 2015; Lovett et al., 2002). It was initially thought that RecA-independent recombination mostly promoted tandem duplications and deletions, and were intramolecular in nature (Bi and Liu, 1994; Bzymek and Lovett, 2001; Lovett et al., 1994). However, recent findings support the idea of a system that promotes recombination between two different DNA molecules without the presence of RecA. In fact, this process is not exclusive to *E. coli* and is present with equal recombination efficiencies in several other proteobacteria (Swingle et al., 2010). The mechanisms and proteins by which this process occurs are however largely unknown and so far, no protein with analogous function to RecA has been identified. RecA-independent recombination relies mostly on short homologies (≈50 nt) in contrast to the large homologies (≈150 nt) required for RecA-dependent recombination (Lovett et al., 2002). In fact, the cloning strain DH5α is capable of recombining PCR fragments with homology regions as low as 17 nt (Bubeck et al., 1993; Kostylev et al., 2015). On the other hand, the frequency of recombination events is much lower in RecA-independent recombination, with events being almost four orders of magnitude lower in comparison to conventional RecA-mediated recombination (Lovett et al., 2002). Hence use of highly competent cells is critical (Bubeck et al., 1993; Kostylev et al., 2015). Deleting exonucleases within *E. coli* can increase the frequency of recombinatorial events. In particular, deleting the exonucleases RecJ and SbcB led to an up to 1000-fold increase in recombination events. Other combinations of exonuclease deletions were less effective but still led to an overall increase in recombination frequency (Dutra et al., 2007). The increases observed with RecJ and SbcB might be due to the nucleolytic activity of these, being active even on SSB coated ssDNA (Han et al., 2006; Molineux and Gefter, 1975). Indeed, the increase in RecA-independent recombination events observed upon deleting single-strand exonucleases led to the suggestion that ssDNA substrates, which lack the RecA filament and thus protective action against these exonucleases, are the main substrate for this process (Dermić et al., 2017; Dutra et al., 2007). Two models are envisioned through which recombination occurs. The first model postulates that 5'- or 3'-end overhangs are recruited to perform strand invasion on homologous sequences in the broken fork repair or double-strand break repair. The second model proposes that ssDNA anneals to single-strand regions within replication forks and becomes incorporated (Dutra et al., 2007). While RecA-dependent recombination is well characterised, RecA-independent recombination and how it operates still remains elusive.

1.6 Exogenous Recombination Pathways in *E. coli*: the Red & RecET Systems

Previous sections explained how *E. coli* uses its endogenous pathways to perform recombination. Exogenous pathways, such as those used by phages to insert their DNA into *E. coli*'s genome also exist. These exhibit much higher recombination frequencies and do not require RecA to promote recombination (Hall et al., 1993; Mosberg et al., 2010; Muyrers et al., 1999; Zhang et al., 1998). Two systems, the Red system from the lambda (λ) phage and the RecET system from the defective Rac prophage, are well studied and commonly used in performing recombineering with gram-negative bacteria.

The Red system employs a total of three recombination-mediating proteins: Gam, Exo and Beta (Kulkarni and Stahl, 1989; Murphy, 1991; Shulman et al., 1970). The Gam protein acts as an inhibitor of the RecBCD complex and the SbcCD exonuclease, so as to prevent linear dsDNA degradation of the phage DNA (Figure 12A) (Karu et al., 1975; Kulkarni and Stahl, 1989; Murphy, 1991). The Exo protein replaces the function of RecBCD and degrades linear dsDNA to provide either 3'-end ssDNA overhangs or complete ssDNA (Figure 12A) (Carter and Radding, 1971; Mosberg et al., 2010). Finally, the Beta protein binds to the DNA and is responsible for the RecA-independent pairing and recombination (Figure 12A) (Li et al., 1998; Muniyappa and Radding, 1986). This system facilitates recombination between short homology arms (36 to 50 nt) (Datsenko and Wanner, 2000; Muyrers et al., 1999; Stewart et al., 1999, 1997).

The RecET system is part of the defective Rac prophage and is carried by many *E. coli* K12-derivative strains (e.g. XL1-Blue, NZY5α and DH5α). It employs the RecE and RecT proteins only (Hall and Kolodner, 1994; Hall et al., 1993; Kaiser and Murray, 1979; Low, 1973; Noirot and Kolodner, 1998). Similar to Exo of the Red system, RecE resects the dsDNA to provide a ssDNA substrate for the recombination-mediating protein RecT (Figure 12B) (Hall and Kolodner, 1994; Hall et al., 1993; Kushner et al., 1974; Noirot and Kolodner, 1998). Unlike the Red system however, the Rec system does not encode any exonuclease inhibitors (this being the Gam protein in the case of the Red system), and its system operates efficiently in only *recBC sbcA* or *recBC sbcBC* mutants (Barbour et al., 1970; Kushner et al., 1971, 1974; Lloyd and Buckman, 1985; Yamaguchi et al., 2000). Mutation in *sbcA* leads to a more pronounced expression of the involved proteins (RecE and RecT) while mutating the exonucleases SbcB and SbcCD (*sbcBC* strain) preserves the substrates used by this system (Kushner et al., 1971, 1974; Ryder et al., 1996; Yamaguchi et al., 2000). *sbcA* mutants easily perform recombination via this system and require homology arms of at least 27 nt, but recombination was determined to be most efficient with homology arms spanning 60 nt (Fishel et al., 1981; Kaiser and Murray, 1980; Zhang et al., 1998). For non-*sbcA* mutant *E. coli* strains, plasmid constructs harbouring RecE, RecT, and additionally the Gam protein from the bacteriophage lambda, may be used to induce recombination at high efficiency despite the presence of suppressors (Murphy, 1991; Stewart et al., 1999, 1997; Zhang et al., 1998) as the Gam protein inhibits the RecET suppressors present in *E. coli* (Murphy, 1991).

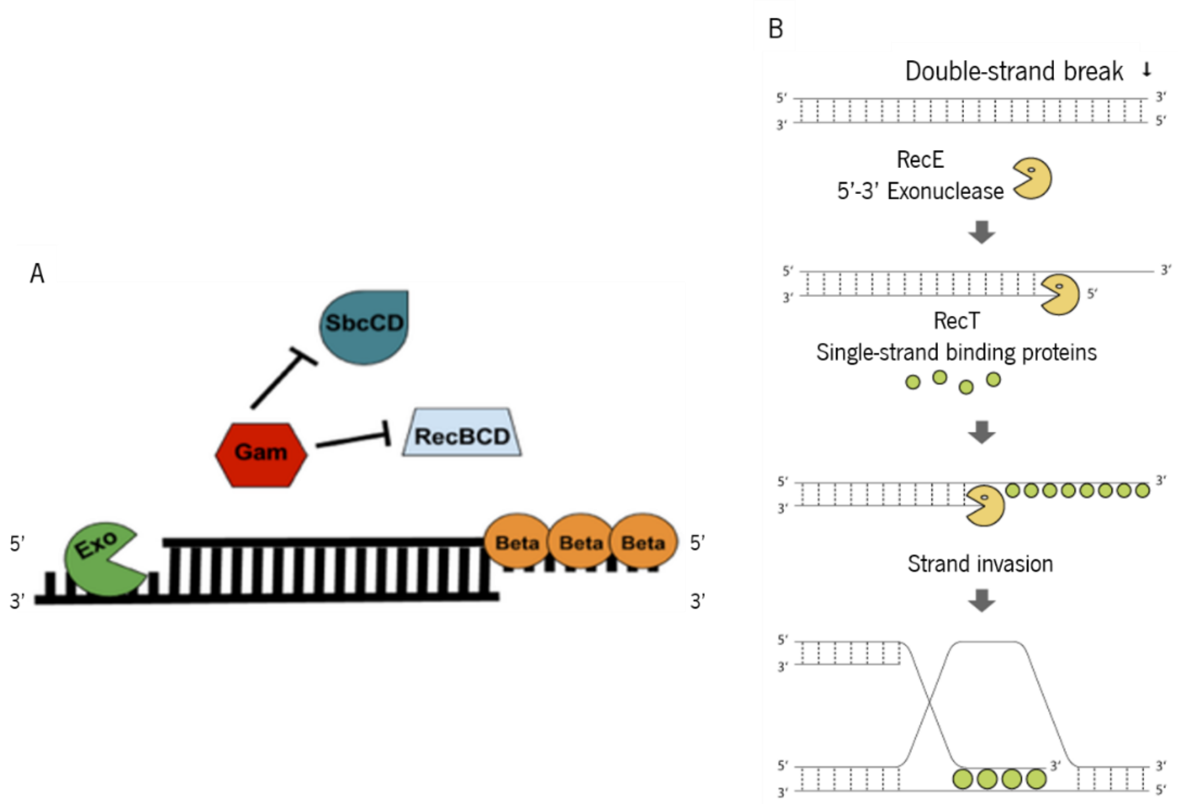


Figure 12 | Exogenous recombination systems in *E. coli*. (A) The lambda red recombination system is formed by Gam (red), Exo (green) and Beta (orange). Gam protein inhibits RecBCD (grey) and SbcCD (blue) endogenous exonucleases. Exo, a dsDNA exonuclease that degrades 5'ends, produces the ssDNA substrate that is used by the Beta protein to bind to and promote recombination. (B) The RecET pathway utilises the RecE (yellow) exonuclease to produce 3'-overhangs. RecT (green) proteins can then bind to the ssDNA and promote strand invasion. Taken from https://blog.addgene.org/lambda-red-a-homologous-recombination-based-technique-for-genetic-engineering and http://dev.genebridges.com/gb/red_et_principles.php.

pRed/ET plasmid constructs are commercially available (Genebridges, Biocat etc.) and offer either one system or the other to be expressed in many different organisms so as to produce constructs harbouring mutations or to directly mutate the hosts genome (Stewart et al., 1999, 1997). In addition, co-expression of the Red system with RecA has been shown to improve the recombination efficiency (Heermann et al., 2008; Wang et al., 2006). Both described phage systems provide easy alternatives for introduction of DNA alterations and are more efficient than *E. coli*'s endogenous recombination system while requiring only short oligonucleotides as substrates. Of all the recombination systems possible in *E. coli*, the Red system appears to be the most efficient. It is slightly more efficient than the RecET system and its recombination rate is at least 70-fold higher than that of a non-phage system (e.g. the RecBCD system) (Murphy, 1998; Muyrers et al., 1999). The biggest limitation to this system is related to its high efficiency in performing recombination; short repeats (six or more nucleotides) within fragments or linear vectors can lead to circularisation (microhomology-mediated recombination) and consequently, loss of DNA material (Zhang et al., 2000a). Interestingly, both phage systems appear to have been principally applied in the recombination of only two DNA entities, rather than in assembling multiple fragments as is required in multi-site-directed mutagenesis.

1.7 Response Surface Design (RSM)

In this project, rational experimental design with response surface methodology was used to optimise the efficacy of a multi-site-directed mutagenesis protocol and to identify interactions between the various process factors.

Designing experiments can be a challenging task, especially when many factors (variables) presumably have a significant impact on the results. Several methodologies exist to streamline the process of designing experiments and evaluating statistically the results. These should be chosen in accordance with the objective of the experiment. In the case where optimisation of a process is the objective of the operator, a simple “one factor at a time” design can be employed. This design can be used with any number of factors since the process involves testing different values for one factor while maintaining the others stable. This experimental design, however, assumes that all factors are independent of each other, meaning that no interaction between these factors occur. Yet, factors frequently do interact with each other; and therefore a rational experimental design approach such as the response surface design is appropriate (Myers and Montgomery, 1995).

Response surface methodology is generally employed when two or more interacting factors have a significant impact on a result and the objective of the experiment is to optimise a process or to measure interaction between any number of factors. Since any number of measured points can be examined, this design is often limited by the resources available and/or the degree of control over the factors. In any case, the response surface model (RSM) employs mathematical models (frequently quadratic, but also cubic models) that take into account the measured points to describe the behaviour of the process and depicts it in the form of a surface diagram or 3D-image. Note that the produced models are only an approximation to the reality and can further be improved by adding more measured points or by repeating points. For a surface response design to be practically feasible the measured points (runs) are usually selected by an experimental design such as the central composite design (CCD). This experimental design ensures that a minimum of measurements (runs) are chosen while still monitoring all the possible interactions between the factors. As a result, the response surface model in combination with the CCD are efficient in describing most systems (Myers and Montgomery, 1995).

1.8 Central Composite Design (CCD)

Central composite design is a rational experimental design first described by Box, Wilson and Hunter in the 1950s, it is used to select, out of many possible measurements (runs), a practically achievable number of runs by either employing a factorial or fractional factorial design. A factorial design uses 2^k runs, where *k* is the number of factors (variables) considered. The fractional factorial design is used when many factors are in play (e.g. $2^6 = 64$ runs) to realistically measure all combinations of interactions between all factors. In this case, only a fraction of the runs (generally ½ or ¼ of runs of the factorial design) are selected, with special care to ultimately try and infer all possible interactions within the total spectrum (NIST/SEMATECH, 2003).

The CCD is cleverly designed in a way to produce a design space with *k* dimensions (Figure 13). Inherent to this design space are -α, -1, 0 (origin), +1 and +α for each factor/dimension. These values are fixed to create the design space which should ideally have the optimal condition in the origin's (0) vicinity (optimal conditions can be inferred by preliminary experiments). Measurements that represent the origin of the design space are called centre points and are important in determining the standard deviation within the experiments. Measurements taken at -1 and +1 are factorial points while axial points are at the α extremities and their values within the space vary in function of the number of factors and desired properties in the design. Both axial and factorial points are involved in determining the interaction(s) between the factors (NIST/SEMATECH, 2003).

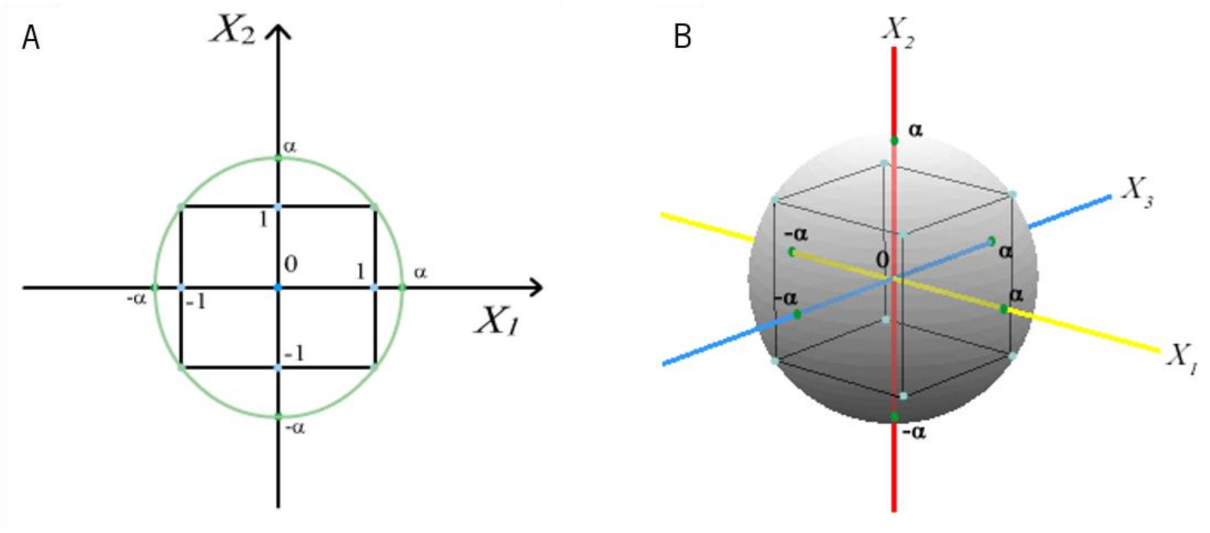


Figure 13 | Central composite design space. Centre points (dark blue dot), factorial points (light blue dots) and axial points (green dots) inhabit the design space. Design space is dependent on the value of α, which depends further on the number of factors. (A) Circular design space of a model with only two factors (X_1 and X_2). (B) Spherical design space of a model with three factors ($\mathrm{X}_1, \mathrm{X}_2$ and X_3). Taken from https://newonlinecourses.science.psu.edu/stat503/node/59/.

Thus far, one type of CCD, denoted circumscribed (CCC), has been explained above. Within the three types of CCD (CCC, CCI and CCF), CCC produces the largest design space. In contrast, the inscribed type (CCI) produces the smallest design space. Alpha values become ∓ 1 while factorial points are < 1. This is used for experiments where the indicated limits truly are limits to experimental results and cannot be extended. Finally, the face centred type (CCF) produces a squared (two factors) or cubic (three factors) design space. It is characterised by having α values equal to factorial points (∓1) (NIST/SEMATECH, 2003).

2. Rational and Objectives

The laboratory where the present project was carried out is focused on protein science, with the objective of better understanding protein structure-function relationships and in developing and engineering enzymes for application in industry. In achieving this objective, the optimisation and implementation of highly efficient, low cost in-house methods for protein engineering is essential. An efficient method for single-site directed mutagenesis has already been implemented in the laboratory, but the power of multi-site-directed mutagenesis in protein engineering has been recognised and an efficient, low cost method is required (Silva et al., 2017). Unfortunately, many of the presently available methods and commercial kits for implementing this are characterised by high costs and multi-step protocols, or require expertise in working with eukaryotic hosts. Thus, in the present study, and taking into account the ability of many commonly used laboratory strains to carry out recombination (probably via a RecA-independent pathway), the objective was to develop and optimise a PCR - *E. coli in vivo* recombination based protocol to enable the low cost, highly efficient multi-site-directed mutagenesis of proteins. *E. coli* endogenous recombination systems are believed to be poorly efficient in common laboratory cloning strains, thus, in the present study, mathematical and statistical techniques were used to optimise all process variable and identify interactions between process variables so as to enable for development of an optimised process.

In the project, the gene encoding a cold-adapted glycoside hydrolase family 8 xylanase will be used as model sequence (pXyl; European Nucleotide Archive (ENA) AJ427921; UniprotKB Q8RJN8). This protein is the subject of current studies at the laboratory aiming to enhance protein solubility by substituting numerous hydrophobic residues and hydrophobic patches to the more hydrophilic amino acid serine and/or charged amino acids glutamic acid and arginine. Response surface methodology with central composite design will be used to optimise process variables, identify interactions and identify the protocol enabling highest multi-site-directed mutagenesis efficiency.

3. Materials and Methods

3.1 Primer Design

Primers (synthesis scale: 0.01 or 0.05 µmol; purification: salt free; concentration: 100 µM) were manufactured by Eurofins Genomics (Ebersberg, Germany). Fully overlapping primer pairs were designed via *SnapGene®* (v2.3.2) for each mutation to be introduced and primer sizes ranging from 18 to 52 nt were investigated. Whenever possible, the following design rules were taken into account: a maximum of three G's and/or C's for the last five nucleotides were allowed while avoiding three G's and/or C's in a row; the primers would preferably end with a C or G, and the mutated nucleotide(s) would ideally be placed in the middle of the primers. The hydrophobic residues for mutation in the surface exposed hydrophobic patches of the cold adapted xylanase had already been identified in previous studies. *E. coli* 's codon bias and use of serine/glutamic acid/arginine codons which minimise the number of nucleotide substitutions, were taken into account during primer design. Possible secondary structures and their predicted strengths were also considered when designing the primers via the bioinformatics tool Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi).

3.2 PCR

PCR was performed with a DOPPIO thermocycler (VWR). A variety of different polymerases were investigated: Phusion® High-Fidelity polymerase (Thermo Fisher Scientific), NZYProof polymerase (NZYTech) and Xpert HighFidelity polymerase (GRiSP).

The construct pET22b-pXyl (see Annexe I), consisting of the whole 1281 bp gene sequence (including signal sequence) of the cold-adapted endo-1,4-beta-xylanase (pXyl; ENA code AJ427921) cloned in the *Nde*I and *Xho*I sites of the expression vector pET22b(+) (Novagen), was used as template in all PCRs. This template had been isolated from the dam+ *E. coli* strain XL1-Blue.

A total final volume of 20 µL was used with a final primer concentration of 0.5 µM - 1.0 µM, a total template amount varying from 0.07 ng to 8.1 ng and the DNA polymerase master mix at a final concentration of 1x (See Annex's IV and V). The remaining volume was completed with ultra-pure water. In all cases, 30 PCR cycles were employed. In Xpert HighFidelity employed PCRs, a master mix was manually prepared and contained a final concentration of PCR buffer containing dNTPs 1x (GRiSP), a total of 7 ng of template and a final polymerase concentration of 0.05 U/µL. The remaining volume (until 20 µL) was filled with ultra-pure water. Several controls were employed according to the type of experiment. Negative controls for unspecific amplification contained regular components of PCR (1 µM primers) except for DNA template. Negative controls for *Dpn*I digestion contained all components for PCR but were not run through a thermocycler (total of 7 ng template; 1 µM primers). Positive controls to confirm amplification by polymerases contained all components for PCR, including 1 µM of primers Y43E.Fwd and Y315R.Rev, 7 ng of total template, and had appropriate PCR conditions (see below), with an annealing temperature of 60.5 °C.

For the Phusion® High-Fidelity PCR Master Mix with HF Buffer (Thermo Fisher Scientific), the following PCR conditions were used: (1) initial denaturation for 3 minutes at 98 °C, (2) denaturation for 30 seconds at 98 °C, (3) annealing for 30 seconds at varying temperatures (at some conditions a 2 Step PCR was performed) (see Annex IV for details), (4) extension time was adjusted to the size of PCR fragments as recommended by the manufacturer (15 s/kb) at 72 °C, (5) final extension for 10 minutes at 72 °C.

For the NZYProof 2x Green Master Mix (NZYTech) and NZYProof 2x Colourless Master Mix (NZYTech), the following PCR conditions were applied: (1) initial denaturation for 3 minutes at 95 °C, (2) denaturation for 30 seconds at 95 °C, (3) annealing for 30 seconds at varying temperatures (see Annex V, for details), (4) extension time was adjusted to the size of PCR fragments as recommended by the manufacturer (60 s/kb) at 72 °C, (5) final extension for 10 minutes at 72 °C.

For the Xpert HighFidelity master mix, the following PCR conditions were applied: (1) initial denaturation for 1 minute at 95 °C, (2) denaturation for 15 seconds at 95 °C, (3) annealing for 15 seconds at 64.7 °C, (4) extension time for 4 minutes at 72 °C, and (5) final extension was 3 minutes at 72 °C.

PCR products were stored at -20 °C when necessary. Amplification of desired PCR products was confirmed by gel electrophoresis.

3.3 Gel Electrophoresis

Gel electrophoresis was carried on a BioRad Sub-cell GT horizontal electrophoresis system connected to a BioRad PowerPac 300 electrophoresis power supply. 1% (w/v) agarose gels containing 1.0 x TAE (40 mM Tris base, 20 mM glacial acetic acid and 1 mM EDTA) were used. Agarose gels were pre-stained with GreenSafe Premium (NZYTech) according to manufacturer specifications (1x). An initial constant 80 V current for 20 minutes was followed by 30 minutes at 120 V. Bands were visualized with a GenoSmart (VWR) transilluminator and band sizes were compared with the NZYDNA Ladder III (NZYTech).

3.4 Enzymatic Digestions

3.4.1 Plasmid Confirmation

To confirm the presence of the desired plasmid construct or as positive control for pXyl presence, restriction digestion analysis was used. pET22b-pXyl constructs were confirmed via *Nde*I and *Xho*I digestion, and pUC18 via *Bam*HI and *Nde*I digestions (Thermo Fisher Scientific, see Annex II for details). All reactions were performed with FastDigest enzymes and the FastDigest Green Buffer (1x final concentration) with ≈0.1 µg of DNA in a final volume of 5 µL, in accordance with the manufacturer´s (Thermo Fisher Scientific) specifications. Reactions were carried out in an Incucell incubator (MMM Medcenter) for 1 h at 37 °C and both digested and undigested plasmids were visualised by gel electrophoresis.

3.4.2 Template Removal

To remove methylated template plasmid (template acquired from dam+ *E. coli* strains) following PCR, *Dpn*I (NZYTech) digestion was employed. Various final concentrations (0.18 - 0.36 U/µL), incubation times (0 - 24 hours) and reaction volumes (5.5 – 11.5 µL) at 37 °C were investigated. A final total reaction volume of 5.5 µL was employed for standard digestions and in the RSM-CC experiment. The volume contained a final concentration of 0.18 U/µL *Dpn*I, was incubated for 4 h at 37 °C and then used directly to transform *E. coli*. Complete template removal by *Dpn*I digestion was confirmed by using samples containing all necessary components for PCR (total of 7 ng of template; 1 µM primers) but without passing through PCR.

Initial PCR-*Dpn*I reaction mixtures contained 1.5 µL of each PCR fragment (three in total) and 1 µL of *Dpn*I. For RSM-CCD experiments, a tube containing the correct volume ratio of big fragment to small fragments was prepared prior to the digestion, containing always 0.5 µL of the big fragment and a volume of small fragments corresponding to the desired ratio (0.5 – 5 µL). A final volume of 10.5 µL was used for all fragment ratio preparations and in cases of missing volume the remaining volume was filled with Phusion® High-Fidelity PCR Master Mix with HF Buffer (not passed through thermocycler). For *Dpn*I digestion the final volume was increased to 7.3 µL, containing 6.3 µL of PCR fragment mix and 1.4 µL of *Dpn*I (1 U/µL). Incubations were perform as described above.

3.5 *E. coli* Strains

Table 1 | *E. coli* strains used in the present work. Specification, genotype and transformation efficiencies are presented.

Strain	**Specification**	**Genotype**	**Transformation efficiency (cfu/µg) of pUC18**
XL1-Blue	Cloning strain	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q} *ZΔM15* Tn*10* (Tet^r)]	$10^6 - 10^7$
NZY5α	Cloning strain	*fhuA2Δ*(*argF-lacZ*)U169 *phoA glnV44* Φ80 *Δ*(*lacZ*)*M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	$10^7 - 10^8$

3.6 Competent Cells

3.6.1 Modified Dagert and Ehrlich Method

A single large XL1-Blue colony was pre-inoculated in 5 mL lysogeny broth (LB, 0.5% (w/v) NaCl, 0.5% (w/v) yeast extract, 1% (w/v) tryptone) at 37 °C, 200 rpm, in a Minitron incubator (Infors HT) overnight. 2 mL of pre-culture was inoculated in 200 mL LB at 37 °C, 200 rpm until reaching an optical density (at 600 nm, OD_{600}) of around 0.25 (as measured on a Genesys 20 spectrophotometer, Thermo Spectronic). The entire cell suspension was centrifuged at 4 °C for 10 minutes at 4000 *g* using the 5804 R centrifuge (Eppendorf). Obtained pellet was kept always on ice and was gently resuspended in ice cold $\mathrm{MgCl_2}$ (0.1 M) before incubating on ice for 30 minutes and centrifuging again at 4 °C for 10 minutes and at 4000 *g*. Supernatant was removed and cells were gently recollected in 10 mL of ice cold transformation buffer I (75 mM $\mathrm{CaCl_2}$, 6 mM $\mathrm{MgCl_2}$, 15% (v/v) glycerol) and always kept on ice. After centrifuging one last time at 4 °C for 10 minutes at 4000 *g*, pellet was resuspended in 3 mL of ice cold transformation buffer I and kept at 4 °C up to 24 h. Competent cells were divided into aliquots of 200 µL, preserved in liquid nitrogen and stored at -80 °C.

3.6.2 Modified Inoue et al. Method

A single large XL1-Blue colony was inoculated in 250 mL super optimal broth (SOB, 2% (w/v) tryptone, 0.5% (g/v) yeast extract, 2.5 mM KCl, 10 mM NaCl, 10 mM $\mathrm{MgSO_4}$, 10 mM $\mathrm{MgCl_2}$) for 4 – 5 days at 18 °C, 200 rpm. Cell culture was incubated on ice for 10 minutes once an $\mathrm{OD_{600}}$ of 0.6 was measured. From here, all additional steps were performed with cell culture on ice or at 4 °C. After incubation on ice, cells were centrifuged for 10 minutes at 2500 *g*, using the centrifuge 4-16KS (Sigma), and gently resuspended in 80 mL ice cold transformation buffer II (10 mM Pipes, 15 mM $\mathrm{CaCl_2}$, 250 mM KCl, 55 mM $\mathrm{MnCl_2}$, pH 6.7). Cells were incubated on ice for 10 minutes and centrifuged once again at 2500 *g* for 10 minutes. Supernatant was removed and the pellet was gently resuspended in 20 mL transformation buffer II before adding carefully DMSO to a final concentration of 7% (v/v). Cells were incubated one last time on ice for 10 minutes before being aliquoted in 200 µL and frozen in liquid nitrogen. Competent cells were stored at -80 °C.

3.6.3 In-house Magnesium Chloride and Calcium Chloride Method

A streak of XL1-Blue cells was pre-inoculated in 10 mL LB at 37 °C, 200 rpm overnight. 1 mL of overnight culture was then inoculated in 100 mL LB at 37 °C, 200 rpm until an $\mathrm{OD_{600}}$ between 0.4 – 0.6 was obtained. Upon reaching desired OD, 80 mL of culture was incubated on ice for 10 minutes and for the remaining steps, culture was always kept on ice or at 4 °C. Following incubation, cells were centrifuged for 10 minutes at 4000 *g*, resuspended very gently in 40 mL of ice cold $\mathrm{MgCl_2}$ (0.1 M), before being again centrifuged for 10 minutes at 4000 *g*. Pellet was resuspended very gently in 4 mL of ice cold $\mathrm{CaCl_2}$ (0.1 M) and left in ice for 1 h and 30 minutes. DMSO was added slowly to cells until obtaining a final concentration of 7% (v/v), aliquoted in 200 µL and immediately cryopreserved in liquid nitrogen, before storing at -80 °C.

3.6.4 Rubidium Chloride Method

XL1-Blue cells were pre-pre-inoculated in 10 mL of LB at 37 °C, 200 rpm overnight. A pre-culture was inoculated with 400 µL of pre-pre-culture and 10 mL of LB at 37 °C, 180 rpm until reaching $\mathrm{OD_{600}}$ between 0.3 and 0.4. Finally, 4 mL of pre-culture were incubated in 100 LB at 37 °C, 180 rpm until an $\mathrm{OD_{600}}$ between 0.3 and 0.4 was obtained. The entire culture content was incubated on ice for 5 minutes and in all following steps, cells were kept either on ice or at 4 °C. Cells were harvested by centrifuging at

2500 *g* for 5 minutes, very gently resuspended in ice cold transformation buffer III (30 mM CH_3COOK, 50 mM $MnCl_2$, 100 mM RbCl, 10 mM $CaCl_2$, 15% (v/v) glycerol, pH 5.8) and then centrifuged again for 5 minutes at 2500 *g*. The transformation buffer IV (10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl, 15% (v/v) glycerol, pH 7) was used to resuspend cells before being aliquoted in 100 or 200 µL. Liquid nitrogen was used to instantaneously cryopreserve, and cells were stored at -80 °C.

3.7 Transformation

In-house prepared *E. coli* XL1-Blue (100 and 200 µL aliquots) and commercial *E. coli* NZY5α (NZYTech) competent cells (Table 1) were used for transformation via heat shock. *E. coli* NZY5α (NZYTech) cells were stored at -80 °C and carefully aliquoted in 100 µL or 95 µL volumes before transforming. Transformation procedures recommended by the manufacturer for *E. coli* NZY5α (NZYTech) were followed with both strains. Various volumes (1 - 15 µL) of material to be transformed (pUC18) were investigated and optimised. For the RSM-CCD experiment, a total volume of 5.5 µL of the PCR-*Dpn*I mix was added to 100 µL or 95 µL cells. A control with 1 µL pUC18 (0.01 ng/µL) was used to determine the transformation efficiency during all runs. Following incubation on ice for 30 minutes, heat shock at 42 °C for 40 seconds and immediate incubation on ice for 2 minutes, cells were incubated in 0.9 mL of room temperature super optimal broth with catabolite repression (SOC, 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 2.5 mM KCl, 10 mM NaCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$, 20 mM D-Glucose) for 1 h at 37 °C, 225 rpm. Following incubation, cells were centrifuged for 1 minute at 5000 rpm with a MiniSpin Plus (Eppendorf) and recovered in 150 µL of SOC. Finally, the entire volume was plated on lysogeny broth (0.5% (w/v) NaCl, 0.5% (w/v) yeast extract, 1% (w/v) tryptone, 2% (w/v) agar) with ampicillin (amp, 200 µg/mL) and incubated at 37 °C overnight. For pUC18 transformations, 100 µL were previous to centrifugation removed and plated, while the remaining volume was treated as previously described.

3.8 Bacterial Cell Storage: Glycerol Cell Stocks

E. coli cells harboring the pET22b-pXyl mutant constructs were stored at -80 °C in 8% (v/v) glycerol. Cells were pre-cultured in 5 mL LB (0.5% (w/v) NaCl, 0.5% (w/v) yeast extract, 1% (w/v) tryptone) containing 200 µg/mL amp overnight at 37 °C, 200 rpm. This pre-culture was used to inoculate 20 - 25 mL LB, containing 200 µg/mL amp, at an initial OD_{600} of 0.1. Cultures were then incubated at 37 °C, 200 rpm until a final OD_{600} of 0.6 - 0.8 and 900 µL aliquots were added to 100 µL autoclave sterilised 80% glycerol, gently mixed and immediately stored at -80 °C.

3.9 Plasmid Extraction

A single transformed *E. coli* colony was pre-inoculated in 5 mL LB containing 200 µg/mL amp and incubated overnight at 37 °C, 200 rpm. *E. coli* plasmids were extracted using the NZYMiniprep kit (NZYTech), following the manufacturer specifications. Briefly, 3 mL cell suspensions were pelleted using 14100 *g* for 30 seconds and resuspended in 250 µL of cold (4 °C) buffer A1 (resuspension solution) by vigorously vortexing. Next, 250 µL of buffer A2 (lysis solution) was added followed by gently inverting tubes 6 times and quick addition of 300 µL buffer A3 (neutralisation solution) before finishing again by inverting tubes gently 6 times. Lysates were centrifuged for 10 minutes at 14100 *g* before loading supernatant carefully onto a column to be centrifuged again at 11000 *g* for 1 minute. The produced flow-through was discarded and 500 µL of buffer AY was loaded onto the column followed by a centrifugation for 1 minute at 14100 *g*. Flow-through was discarded before adding 600 µL of buffer A4, centrifuging for 1 minute at 141000 *g* and 2 minutes at the same speed. Finally, the column was inserted into a new tube and 35 µL of ultra-pure water were added to the centre of the column. Incubation for 1 minute was followed by a final centrifugation step for 1 minute at 14100 *g*. Nucleic acid concentration was subsequently quantified.

3.10 DNA Quantification

Plasmid DNA was quantified by absorption readings of 2 µL samples at 260 nm with a NanoDrop ND-1000 (Alfagene) spectrophotometer. Absorption reading at 230 nm and 280 nm were also recorded to determine sample purity.

3.11 Sequencing

Sequencing of the pXyl gene was used to determine mutagenesis efficacy. This was performed by Eurofins Genomics (Cologne, Germany) with either the T7 promoter or T7 terminator primers (Annexe III), with samples containing 80 – 100 ng/µL of extracted plasmid and 5 µM of primer in a final volume of 10 µL.

For sequencing of the RSM-CCD transformants, custom DNA sequencing with the SupremeRun 96 service at Eurofins Genomics was used. 96-well plates containing LB and amp (150 µg/mL) provided by Eurofins Genomics were inoculated with transformants, incubated overnight at 37 °C and sequenced using the standard primer pET-RP (listed in this document as T7 terminator short) provided by Eurofins Genomics.

3.12 Central Composite Design

Statistical analysis and estimation of optimal condition for recombination of three fragments was performed by using a surface response model within a central composite design (CCC) aided by the software program *Design Expert* (v7.0.0). Three numeric factors: total template (ng), homology (nt) and large PCR fragment to small PCR fragments ratio (1:n:n, where n ∈ small fragment ratios of {1, 2, 5, 8, 10}) and 0 categoric factors were selected for this analysis. Experiment was divided into three blocks within the CCD and only one response (recombination efficiency (%)) was analysed for data evaluation. Order and Model (within the Evaluation category) was set to Quadratic and Polynomial, respectively. ANOVA was applied to assess significance of the model and variables. The optimal condition was determined with settings of a total template amount range of 0.1 ng – 8.1 ng, an 18 to 52 nt homology range, and fragment ratio range 1:1:1 and 1:10:10 and efficiency to maximize.

4. Results

The necessity for developing a time- and cost-efficient multi-site-directed mutagenesis protocol emerged within our group during structure-function studies of a psychrophilic xylanase isolated from *Pseudoalteromonas haloplanktis* TAH 3a wherein the preparation of a relatively large selection of multi-mutated sequences was required. The previous reports of an endogenous recombination systems in many commonly used *E. coli* cloning strains, a well-established bacterial model in our group, seemed fitting for this objective to avoid accessory biological models or even the use of recombination enhancing constructs or/and enzyme cocktails that would increase costs. As discussed in the introduction of the present work (see section 1.4), previous studies had indicated this endogenous system to be extremely inefficient, yet we believed that use of highly competent cells in combination with mathematical and statistical techniques to optimise all process variables would enhance efficiency and enable for effective mutagenesis.

The strategy adapted (see Figure 14) was based on the protocol used in the GeneArt Site-Directed Mutagenesis PLUS System described by Liang et al., 2012 (See Section 1.3 and Figure 4 for a description). In the present study however, the *in vitro* enzyme catalysed recombination reaction and transformation steps were replaced by direct transformation with *in vivo* recombination in a common *E. coli* cloning strain. To demonstrate the potential of in-house *E. coli* strains for use in the protocol, an initial attempt was made at introducing three mutations in the pXyl gene sequence with use of *E. coli* NZY5α and XL1-Blue strains. The three mutations to be introduced were determined in a previous bioinformatics study (not published) attempting to identify mutations enabling enhanced pXyl solubility. The amino acids tyrosine 43, tyrosine 315 and threonine 403 were selected for mutation to glutamic acid (Y43E) or arginine (Y315R, T403R). Overlapping primers, forward (Fwd) and reverse (Rev), of sizes 24 bp (Y315R), 26 bp (Y43E) and 31 bp (T403R), with the mutation located in the middle, had already been designed (See Annex III for primer details). As depicted in Figure 14, three separate PCRs were carried out to produce 3 fragments (F) with overlapping termini: fragments F1 (primers Y43E.Fwd & Y315R.Rev), F2 (primers Y315R.Fwd & T403R.Rev) and F3 (primers T403R.Fwd & Y43E.Rev). Two of the fragments (F1 - 840 bp and F2 - 295 bp) harboured sequences from the inserted pXyl gene sequence, whereas the third fragment (F3 - 5669 bp) amplified the remaining pXyl gene sequence and the whole plasmid backbone to give a much larger fragment containing the ampicillin selection marker. Indeed, it is important to note that any multi-site-directed mutagenesis study of any inserted gene sequence with this protocol would always give rise to one large fragment, containing the vector and hence selection marker sequence, and a number of much smaller fragments, whose number and size depends on the number of mutations and the insert size. For PCR, the high-fidelity DNA polymerase Phusion was used with 0.7 ng of the pET22b-pXyl template as described in the Materials and Methods and agarose gel electrophoresis indicated amplification of single bands at the expected sizes (Figure 15A). 1.5 µL of each of the three PCR products were then added together to 1 µL *Dpn*I (1 U/µL) and incubated at 37 °C for 4 h before transformation and plating on selective media. By skipping the *in vitro* step, time and more importantly costs are reduced as the PCR-*Dpn*I mix is directly transferred to the cloning host and fragments assembled dependent on the capability of the host. 101 cfu were obtained with *E. coli* NZY5α, which is reported by the manufacturer to have a transformation efficiency of ≥ 10^9 cfu/µg with the plasmid pNZY28, but was observed to have an efficiency of 10^7 - 10^8 cfu/µg of pUC18 in our control study. For the preparation of competent *E. coli* XL1-Blue cells, various methods were examined and indicated the rubidium chloride protocol to be the most successful, producing competent cells with an efficiency of $\approx 10^7$ cfu/µg of pUC18 and giving rise to 75 transformants with the multi-site-directed mutagenesis protocol. Sequencing of 10 *E. coli* NZY5α transformants indicated 8 sequences with the desired triple mutations, one truncated sequence and one wild-type sequence. Sequencing of 3 *E. coli* XL1-Blue transformants also indicated success, with two correct triple mutants and one single mutant. Hence, this preliminary study indicated the high potential of the protocol used, with mutagenesis efficiencies of 66 - 80% even without optimisation.

4.1 Method Development

Following the preliminary experiments, subsequent studies were focused on developing the protocol and determining the limits and most appropriate methods to be used in the RSM-CCD process optimisation study. Various parameters were examined, namely, PCR conditions, DNA polymerase, template concentration, *Dpn*I digestion protocol, fragment ratios for mixing and transformation volume.

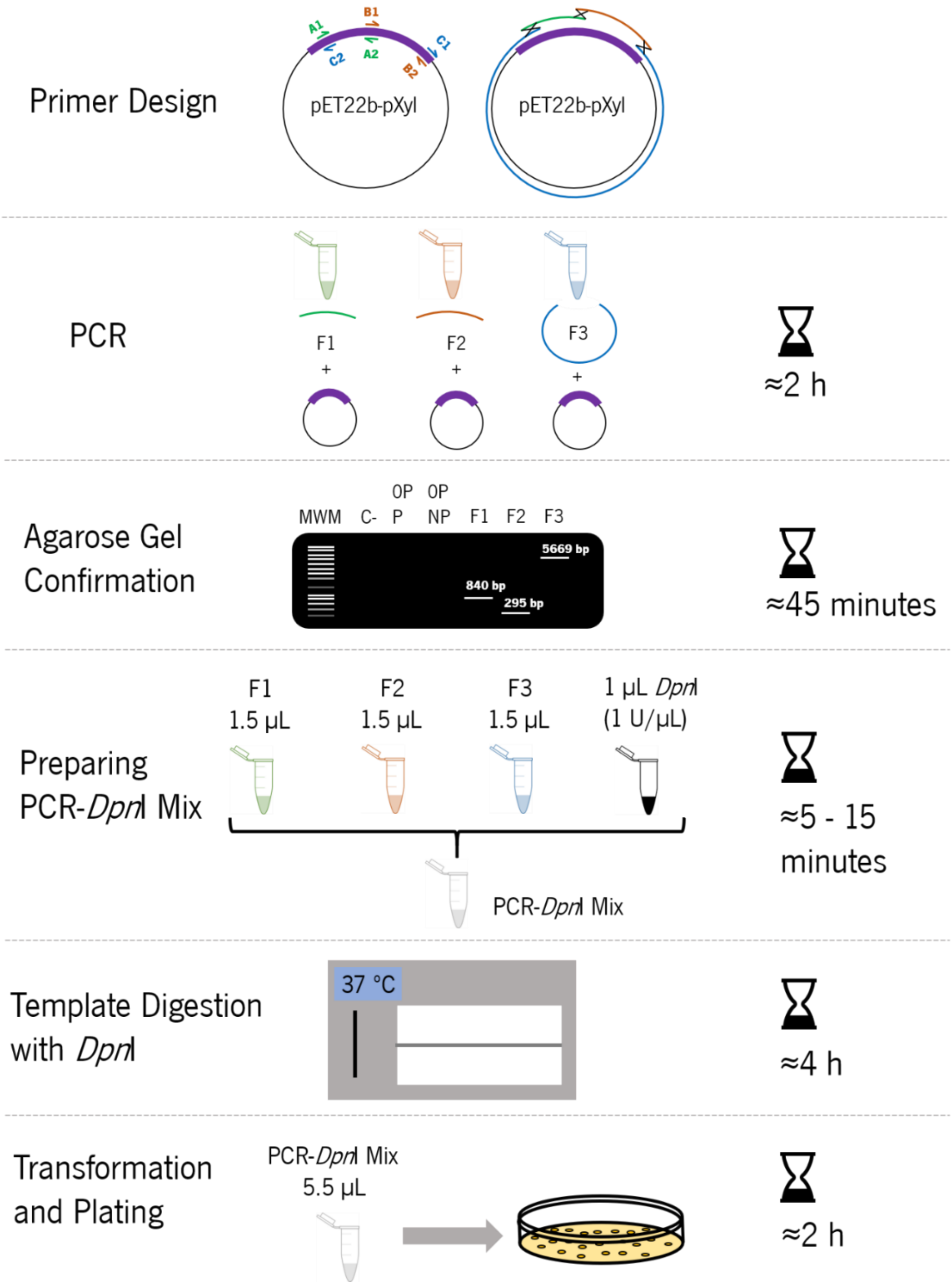

Primer Design
A1
B1
C1
C2
A2
B2
pET22b-pXyl
pET22b-pXyl
PCR
F1
+
F2
+
F3
+
≈2 h
Agarose Gel Confirmation
MWM
C-
OP P
OP NP
F1
F2
F3
5669 bp
840 bp
295 bp
≈45 minutes
Preparing PCR-*DpnI* Mix
F1 1.5 µL
F2 1.5 µL
F3 1.5 µL
1 µL *DpnI* (1 U/µL)
PCR-*DpnI* Mix
≈5 - 15 minutes
Template Digestion with *DpnI*
37 °C
≈4 h
Transformation and Plating
PCR-*DpnI* Mix 5.5 µL
≈2 h

Figure 14 | Flowchart of the developed multi-site-directed mutagenesis protocol by *in vivo* recombineering in *E. coli*. A summary of the developed protocol along with approximate times to perform each step are shown. Primers should be designed as indicated, with each mutation (not shown) requiring one pair of primers. Primers A1 & A2 (green), B1 & B2 (brown) and C1 & C2 (blue) produce corresponding PCR products which are assembled via recombination (indicated by crosses) through homologies at their termini. Agarose gel confirmation of fragments along with proper controls is followed by the preparation of the PCR-*DpnI* mix, incubation for 4 h at 37 °C to remove the template (pET22b-pXyl, vector backbone indicated by the black portion and insert shown as purple section) and transformation of appropriate *E. coli* cloning strains with 5.5 µL of PCR-*DpnI* mix. Finally, cells are plated on appropriate selective media and incubated overnight.

4.1.1 PCR Conditions

To better identify the effects of template concentration, the same protocol was repeated with three different template amounts for the PCR reactions: 0.07 ng, 0.7 ng and 7 ng. Figure 15 shows the PCR results where it can be seen that single bands of the expected size were observed for all reactions (Figure 15A) and that band intensities increased for all fragments according to the template concentration used (Figure 15B). Furthermore, it was also observed that fragment F2 (295 bp) showed generally a less intense band compared to the other two bands at all concentrations examined and this is believed to be associated with the non-optimised PCR conditions used. Therefore, for all subsequent PCR reactions, identical conditions were to be employed for all fragments, but with use of extension times appropriate to fragment length, and optimisation of primer concentration (0.5 – 1 µM) and annealing temperature only as required to give single bands of the correct size with similar intensities for all fragments as monitored by agarose gel electrophoresis.

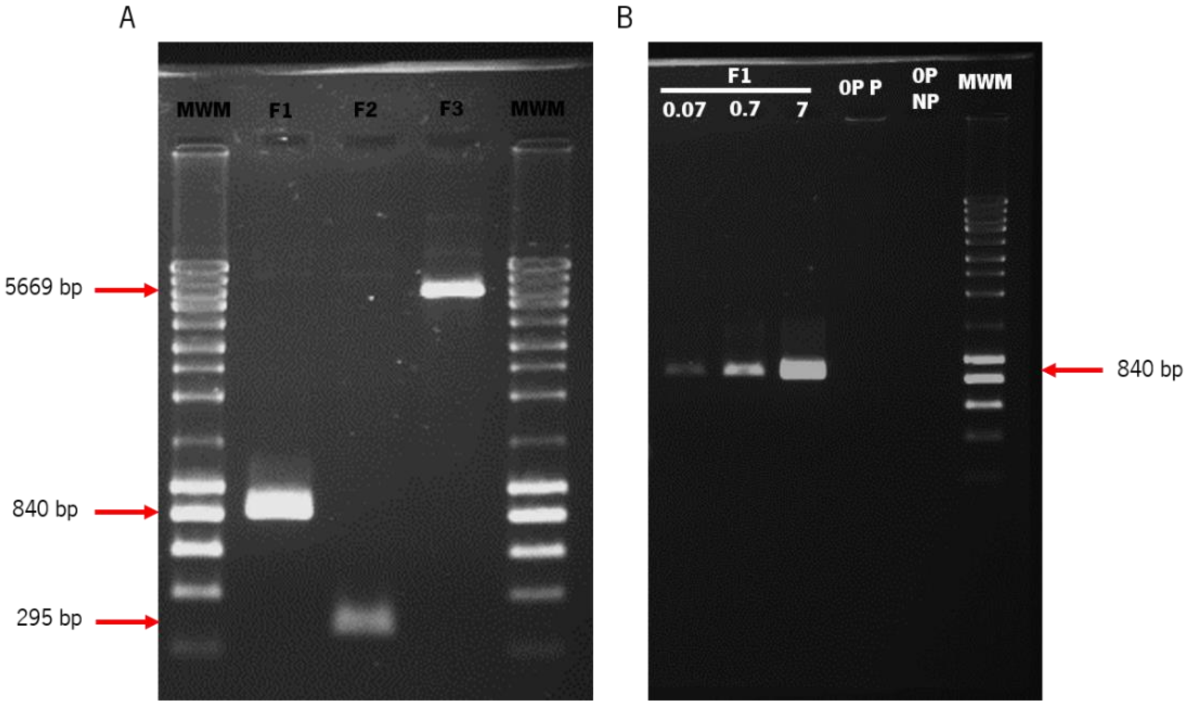


Figure 15 | Amplification of F1, F2 and F3. (A) PCR products of fragments F1 (840 bp), F2 (295 bp) and F3 (5669 bp, selection marker present), harbouring point mutations Y43E & Y315R, Y315R & T403R and T403R & Y43E, respectively. (B) Band intensity of F1 increases as template amount for PCR reaction increases. No bands are visible for both negative controls. 0P P: PCR reaction without primers run through PCR thermal cycles; 0P NP: PCR reaction without primers not run through PCR. MWM: Molecular weight marker NZYLadder III (NZYTech).

4.1.2 Template Concentration

Figure 16 shows the number of transformants and proportion of correct triple mutants obtained (the recombination efficiency), as confirmed by sequencing of the pXyl insert gene (Table 2), for the template concentration study. The number of transformants increased with increasing template concentration (40, 101 and 134 cfu at 0.07, 0.7 and 7 ng total template, repectively) but recombination efficiencies varied from a maximum of 80% at 0.7 ng to 40% at 0.07 ng and 30% at 7 ng total template. The subsequent RSM-CCD experiment was to be designed to encompass this same range of template concentrations.

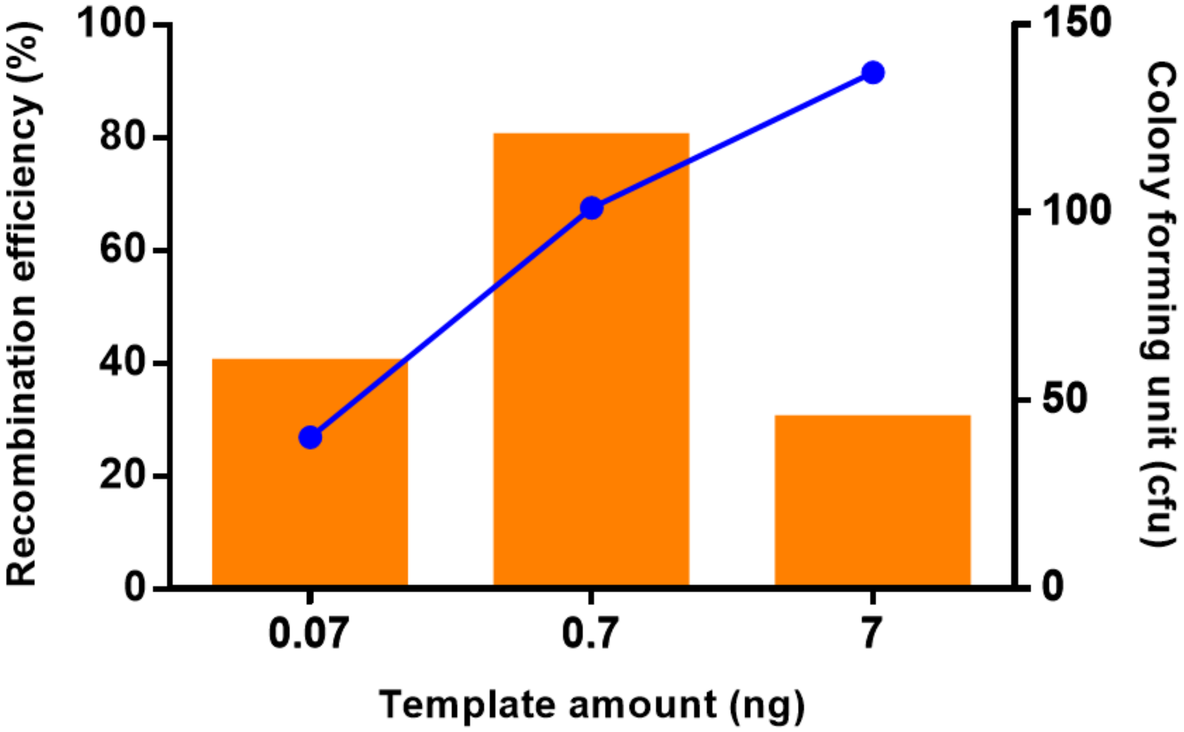


Figure 16 | Effect of PCR template concentration on recombination efficiency and number of transformants. Recombination efficiency (orange columns), as determined by sequencing of 10 transformants at each template concentration, and colony forming units (blue dots) as a function of the final amount of template used in the PCR reactions.

Agarose gel analysis of digested and non-digested plasmids isolated from 10 transformants (Figure 17) at each template concentration indicated a higher occurrence of truncation events with very low template concentrations. 5 truncated fragments were observed when using the lowest template concentration, one truncated fragment with each of the two other template concentrations investigated and one plasmid of larger size with 0.7 ng total template for PCR.

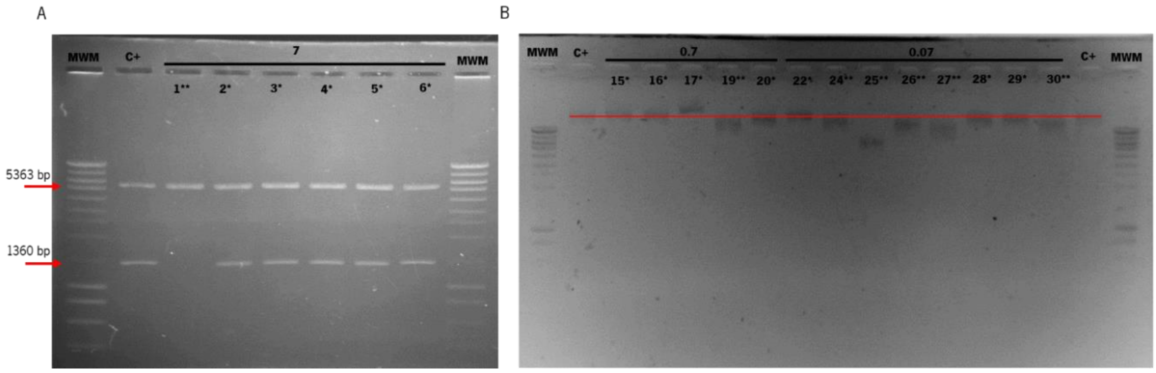


Figure 17 | Molecular weight analysis of pET22b-pXyl mutants. (A) Extracted pET22b-pXyl constructs from triple mutant population at 7 ng total template. Constructs were digested with restriction enzymes *Nde*I and *Xho*I, producing the vector fragment (5363 bp) and the pXyl insert (1360 bp). (B) Extracted pET22b-pXyl constructs from triple mutant population at 0.7 ng and 0.07 ng total template. Constructs were not digested. Red line (B) indicates the vertical position of C+. C+: pET22b-pXyl digested (A) or undigested (B) that serves as positive control; MWM: Molecular weight marker NZYDNA Ladder III; ***** - correct assembly of pXyl gene confirmed by sequencing; ****** - truncation or complete absence of pXyl gene confirmed by sequencing. The remaining plasmid isolates of the original 10 selected transformants were confirmed to be non-truncated and are not shown here.

Table 2 shows the complete results for the analysis of the sequences determined with the T7 promoter and terminator primers for each of the 10 selected transformants. This confirmed the restriction digestion-agarose gel identification of truncated fragments and indicated these truncations to be principally found at the 6xHis-tag related to repetitive sequences promoting deletions (Figure 18). Truncation removed much more frequently the T7 promoter than the T7 terminator and thereby lower the sequencing success by the T7 promoter primer.

Table 2 | Analysis of insert sequences. Sequencing of 10 transformants for each template amount used. Number of triple mutants, wild-types, truncations, increased plasmid size and double mutants are shown. Sample with increased size was also a triple mutant.

Total template (ng)	Triple mutants	Wild-type	Truncated	Increased size	Double mutants
0.07	4	1	5	0	0
0.7	8	1	1	1	0
7	3	5	1	0	1

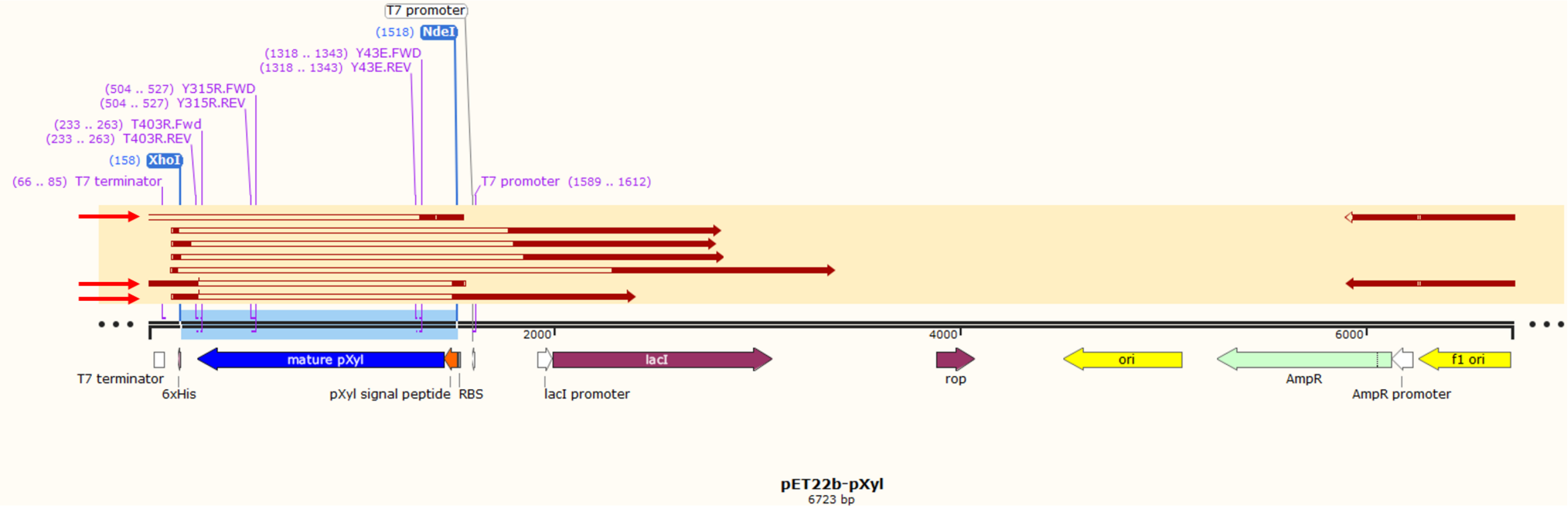

T7 promoter
(1518) NdeI
(1318 .. 1343) Y43E.FWD
(1318 .. 1343) Y43E.REV
(504 .. 527) Y315R.FWD
(504 .. 527) Y315R.REV
(233 .. 263) T403R.Fwd
(233 .. 263) T403R.REV
(158) XhoI
(66 .. 85) T7 terminator
T7 promoter (1589 .. 1612)
2000
4000
6000
mature pXyl
lacI
rop
ori
f1 ori
T7 terminator
6xHis
pXyl signal peptide
RBS
lacI promoter
AmpR
AmpR promoter
pET22b-pXyl
6723 bp

Figure 18 | pET22b-pXyl truncations. pET22b-pXyl is represented with its main features on the bottom. The pXyl gene is inserted between the *Xho*I and *Nde*I restriction sites. Primers used to either amplify fragments (only those for F1, F2 and F3 are shown) or to sequence (T7 terminator, T7 promoter) are displayed in pink. Sequenced constructs of Y43E, Y315R, T403R pXyl mutants containing pXyl truncations are aligned with the construct (long red arrows). Missing sequence within the aligned sequences are displayed in white. Short red arrows (left side) indicate aligned sequences where truncation occurred at a mutagenic primer annealing location. Samples are ordered from top to bottom: 1, 19, 24, 25, 27, 30 (sequenced via T7 terminator) and 30 (sequenced via T7 promoter). Sequencing of sample 26 gave no data.

4.1.3 False Positives: *Dpn*I Digestion

It can be seen from Table 2 that higher PCR template concentrations led to a higher number of non-mutated wild-type sequences; 7 ng template resulted in 5 (50%) wild-type sequences whereas only one was observed for each of the other concentrations examined. Such observations are somewhat unexpected as the *Dpn*I digestion step was incorporated to remove all wild-type template. Indeed, the negative control (titled 0P NP in Table 3), wherein a PCR reaction mix without primers is directly treated with *Dpn*I and transformed as for the mutagenesis protocol, resulted in no transformants (see Table 3). Multiple repetitions of this negative control with 7 ng template always gave zero transformants. Shorter incubation times (down to 1 hour) or increased template concentrations (up to 10-times the amount used) gave non to very little (maximum of 6 cfu for 10-times the normal template amount) transformants, thereby confirming the efficacity of the *Dpn*I digestion step in removing template under the conditions used: 0.18 U/µL *Dpn*I final concentration in final volume of 5.5 µL, incubated for 4 hours at 37 °C. A second negative control was therefore designed wherein the PCR reaction mix without primers was subjected to PCR thermal cycling. As expected, no bands were visible on agarose gel following PCR (see Figure 15B, 0P P), but, interestingly, transformants were obtained and the number of transformants increased with increasing template concentration (Table 3). To investigate this further, we repeated the PCR thermal cycling using a reaction mix without any polymerase and here again no transformants were produced, indicating that it is the activity of the polymerase during the PCR thermal cycling, and not the high temperatures or any other component, that apparently interferes with the subsequent *Dpn*I digestion step.

4.1.4 DNA Polymerase

To investigate whether the perseverance of false positive wild-type sequences was polymerase specific we repeated the control experiment with thermal cycling (0P P) with a number of different DNA polymerases: Phusion, Xpert HighFidelity and NZYProof. Duplicate experiments resulted in transformants with both Phusion (6 and 7 cfu) and Xpert HighFidelity (25 and 54 cfu), both of which are differentiated by the presence of an additional DNA binding clamp domain, whereas no colonies were observed with the remaining polymerase tested. For this reason, NZYProof DNA polymerase was to be used in subsequent mutagenesis experiments with the objective of eliminating background wild-type sequences in transformants.

Table 3 | Transformation associated recombination results of Y43E, Y315R, T403R mutants. Number of cfu are displayed in accordance to the amount of template used in PCR reaction and number of fragments used for the recombination events in competent NZY5α *E. coli* cells. In all cases, three fragments must recombine to produce a circular DNA product with pXyl Y43E, Y315R, T403R point mutations. 0 fragment samples were transformed with PCR reaction mixtures lacking primers, with or without passing through a thermocycler.

		Transformation results (cfu)		
Total template (ng)		0.07	0.7	7
3 Fragments	F1 F2 F3	40	101	134
2 Fragments	F2 F3	13	17	234
1 Fragment	F2	2	5	7
	F3	18	30	57
0 Fragments	PCR (0P P)	0	12	57
	No PCR (0P NP)	0	0	0

In addition to the controls described above, controls in which only 1 or 2 fragments were transformed were also carried out. Interestingly, relatively high numbers of transformants were observed in all cases, with the number being influenced by the template concentration but also by the presence of the large fragment containing the ampicillin selection marker (fragment F3) (Table 3). Such results indicate the potential for partial plasmid assembly, influenced by the presence of wild-type sequence, in leading to false postive transformants.

These method development experiments described above allowed for the obtention of a number of single (Y43E; Y315R; T403R) double (Y43E-T403R) and triple (Y43E-Y315R-T403R) pXyl mutants which will be of use for the structure-function studies of this protein.

4.1.5 Transformation Volume

The effect of transformation volume on the transformation efficiency of the NZY5α cells was also investigated and indicated that lower volumes enabled highest efficiency (Figure 19). A significantly higher transformation efficiency (≈5-fold) was observed with 1 µL volume but we chose 5.5 µL for our subsequent RSM-CCD experiments so as to overcome pipetting errors associated with very low volumes.

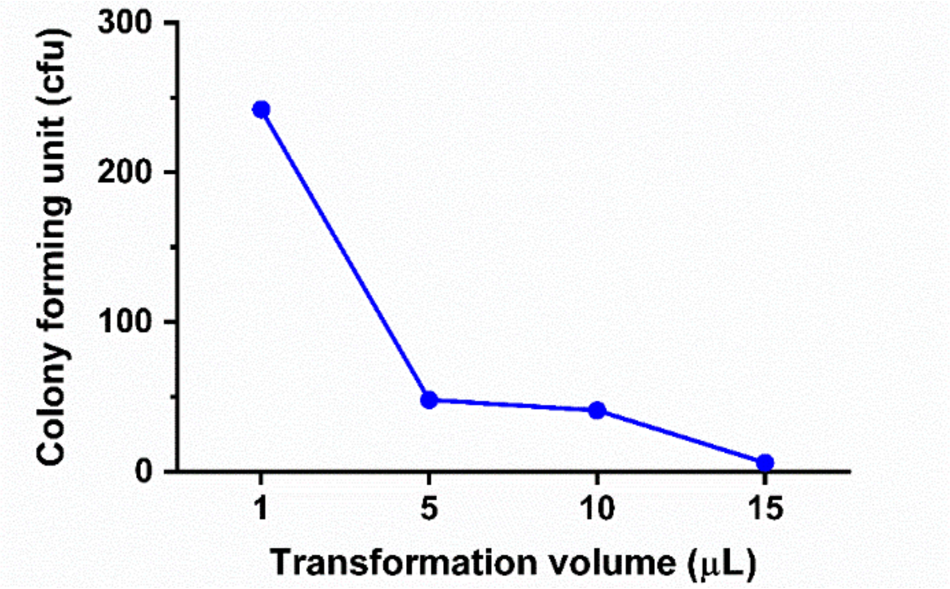


Figure 19 | Transformation efficiency of varying volumes. Transformation of competent NZY5α cells with 1 µL, 5 µL, 10 µL and 15 µL, containing always 0.01 ng of pUC18.

4.1.6 Fragment Ratio

In all types of cloning experiments, it is known that the relative fragment concentrations have a major influence on cloning efficiency and that most typically higher amounts of the smaller insert fragment are required as compared to the larger vector fragment (Oldenburg et al., 1997). Typically, ratios of large fragment to small fragment range from 1:1 and 1:10 and therefore this range was to be investigated in our RSM-CCD optimisation study (Jacobus and Gross, 2015; Kostylev et al., 2015).

4.1.7 Primer Size

A meta-study of already existing recombination protocols identified homology length, as defined in our protocol by the primer length, as an important factor in recombination efficiency. RecA-independent cloning has been reported to be effective with homologies at ≈50 bp to as low as 17 bp (Bubeck et al., 1993; Kostylev et al., 2015). Thus, in the present study a primer range from 18 to 52 bp was chosen, the latter being near the size limit for simplified, reduced cost primer design.

4.2 NZYProof DNA Polymerase

As described above, the DNA polymerase NZYProof was chosen for subsequent studies and was initially used to investigate a second set of primers so as to demonstrate that mutation site specificity was absent with the mutagenesis protocol. Previous bioinformatics studies had indicated the following triple mutations as potentially enabling an enhanced pXyl solubility: Y194S, A272S, Y378S. PCR was carried out with NZYProof and the designed primers (see Annex III and V), with a total template amount of 0.7 and 7 ng and enabled successful amplification of the three fragment at the correct sizes: F4 (Y194S.Fwd & A272S.Rev; 258 bp), F5 (A272S.Fwd & Y378S.Rev; 350 bp) and F6 (Y378S.Fwd & Y194S.Rev; 6195 bp) (Figure 20A). *Dpn*I treatment and transformation of competent NZY5α cells yielded 279 transformants with 7 ng template and 9 with 0.7 ng template. Only three plasmids were extracted at each template concentration investigated and agarose gel analysis indicated one of these to be truncated (0.7 ng template) (Figure 20B).

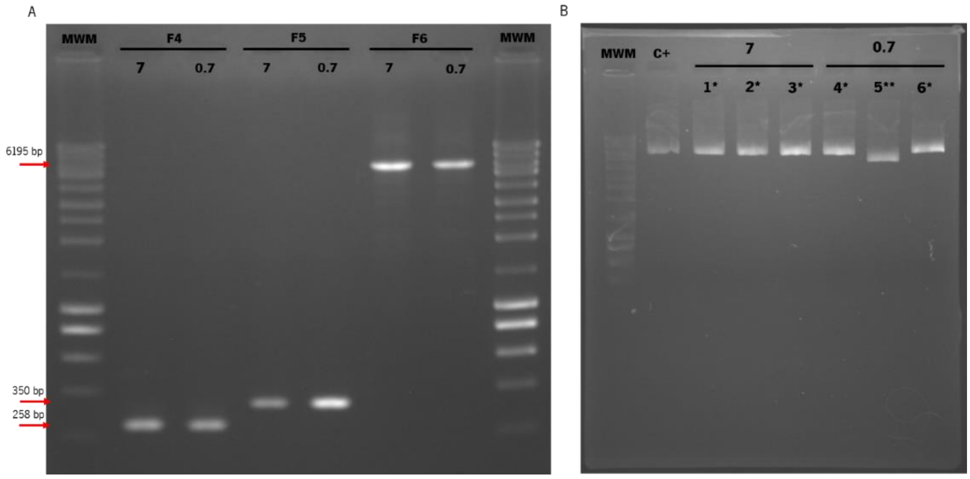


Figure 20 | Amplification of fragments F4, F5 and F6. (A) PCR products of fragments F4 (258 bp), F5 (350 bp) and F6 (6195 bp, selection marker present), harbouring point mutations Y194S & A272S, A272S & Y378S and Y378S & Y194S, respectively. (B) Undigested pET22b-pXyl constructs at 7 ng and 0.7 ng total template. C+: undigested pET22b-pXyl; MWM: Molecular weight marker NZYDNA Ladder III; ***** - correct assembly of pXyl gene confirmed by sequencing; ******- truncation or complete absence of pXyl gene confirmed by sequencing.

Sequencing of inserts implied the total absence of wild-type sequence and thereby further indicated the suitability of NZYProof for eliminating wild-type background false positives. Nevertheless, with the higher template concentration, no triple mutant (among the 3 sequenced) was identified and one single mutant (Y194S) and two double mutants (Y194S-Y378S) were observed. With 0.7 ng template, two triple mutants were indeed identified (67% recombination efficiency) but the remaining plasmid was truncated in the pXyl gene. Thereafter, the NZYProof polymerase was investigated for use in the RSM-CCD experiment and while it exceled in eliminating wild-type templates, limitations were observed with the amplification of large fragments, fragment F6 in this case. Variable amounts of a smeared high molecular weight DNA material often appeared and negatively influenced the band intensity of fragment F6 (Figure 21). Various experiments under various conditions were carried out to try and overcome this. Initially, it was believed that this smearing was the result of contamination problems but after numerous experiments systematically replacing all reaction components and also comparing to the use of Phusion polymerase, this possibility was excluded. Thereafter, optimisation studies, including PCR thermal cycling optimisation, template concentration optimisation, were investigated but variable smearing with effects on F6 intensity continued to be observed. Hence it was concluded that this polymerase was not optimal to amplify fragment F6 under the conditions desired for the RSM-CCD experiment. Phusion DNA polymerase was therefore chosen for all future studies. While being aware of the problems with false positive wild-type sequences with this polymerase, it was believed that process optimisation by RSM-CCD would allow for identification of those conditions in which such false positives were absent.

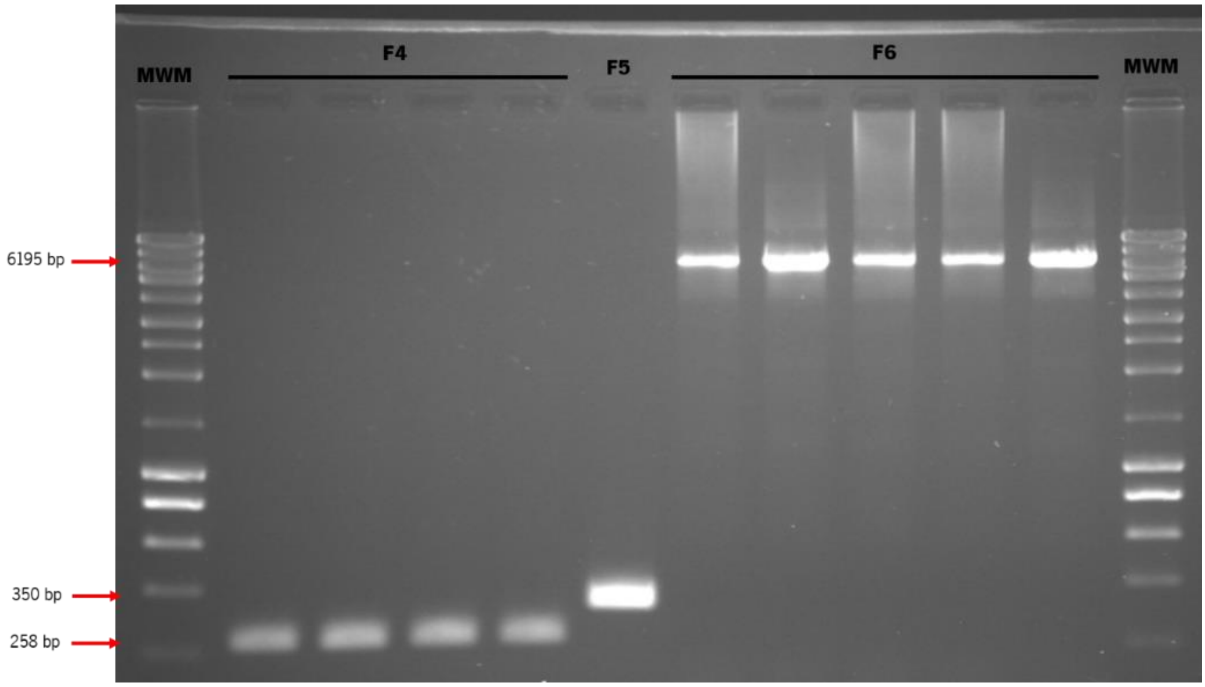


Figure 21 | High molecular DNA entities produced by NZYProof. PCR products of fragments F4 (258 bp), F5 (350 bp) and F6 (6195 bp) produced with NZYProof. Fragments F4 and F5 were amplified using an extension time of 45 seconds and showed no signs of high molecular entities. Fragment F6 was amplified using an extension time of 7 minutes and shows varying band intensities in function of the presence of high molecular entities. MWM: Molecular weight marker NZYDNA Ladder III.

4.3 RSM-CCD for Protocol Optimisation

RSM with a CCD was used to optimise the multi-site-directed mutagenesis protocol so as to identify those conditions maximising the proportion of correctly mutated sequences in the transformants obtained. RSM-CCD analysis consists of 5 principal steps: 1) definition of output response and input factors (process variables); 2) experiment design and evaluation of design; 3) run experiment (measure output response for each run); 4) fit, diagnose and interpret model; and 5) confirm model.

4.3.1 Output Response and Input Factors

As described above, initial experiments identified the total amount of template as a key factor governing recombination efficiency of the multi-site-directed protocol while a meta-study of already existing recombination protocols identified homology length and vector to insert ratio as further important factors. Therefore, these three process factors (template amount, homology length and vector to insert ratio) were selected as input factors, and recombination efficiency was selected as the output response.

4.3.2 Experiment Design and Analysis

The software *Design Expert* (v7.0.0) was used in conjunction with the results of the method development studies described above for experimental design with the CCD matrix shown in Table 4.

Table 4 | Central composite design matrix. Variables, units and values of - α, -1, 0, +1 and α used to design the experiment. n ∈ {1, 2, 5, 8, 10}.

	Variable	Units	-α	-1	0	+1	+α
A	Template	ng	0.1	0.6	3.4	6.2	8.1
B	Homology	nt	18	25	35	45	52
C	Ratio	1:n:n	1:1:1	1:2:2	1:5:5	1:8:8	1:10:10

The final designed experiment consisted of 26 experimental runs, including 6 replicates of one condition (the center points) and 2 replicates for 6 conditions (axial points) to determine experimental error (Table 5 and Table 6). Evaluation of the designed experiment indicated an appropriate degree of freedom for lack of fit (11), while pure error was slightly below optimal (3 instead of 4). Standard errors were equal or below 0.35 and were similar (within 0.06 of range) within types of coefficient (see Table 5). Variance inflation factor remained near the ideal value (1.00), ensuring that lack of orthogonality does not have a significant impact on the variance of the model. Ri-squared means were also close to the ideal value (0.0) indicating lack of correlation between terms and therefore higher chances for producing an adequate model.

Table 5 | Evaluation of the designed experiment. Standard errors (StdErr), variance inflation factor (VIF) and Ri-squared are shown for each variable/term.

Term	StdErr	VIF	Ri-squared
A	0.26	1.13	0.1116
B	0.23	1.00	0.0000
C	0.25	1.06	0.0572
AB	0.35	1.00	0.0000
AC	0.35	1.00	0.0000
BC	0.35	1.00	0.0000
A^2	0.27	1.24	0.1939
B^2	0.21	1.13	0.1164
C^2	0.25	1.18	0.1514

4.3.3 Experiment Run

All 26 experimental runs were then carried out according to the experimental conditions dictated by each run (as defined in Table 6 for primer length, total template amount, and large fragment to small fragment ratio). For PCR, Phusion polymerase was used with optimisation of both the annealing temperature via a 2 °C temperature gradient and the primer concentration for each PCR for each of the three fragments (see Figure 22 for examples). To achieve the fragment volume ratios defined by each run, fragment mixtures containing 0.5 µL of the large fragment (F6) and the respective volume of the small fragments (F4 and F5) was used in a final volume of 10.5 µL, e.g. for a ratio of 1:5:5, 0.5 µL F6 + 2.5 µL F4 + 2.5 µL F5 was used with 5 µL PCR reaction mix without primer (0P NP). For *Dpn*I digestion, 6.3 µL of the fragment mix and 1.4 µL of *Dpn*I (1 U/µL) were incubated for 4 hours at 37 °C. 5.5 µL of this mix was then used to transform 95 µL of competent *E. coli* NZY5α before plating to ampicillin containing medium for selection and propagation. 10 transformants from each of the 26 runs were then randomly selected and the pXyl sequence inserts of the isolated plasmids sequenced.

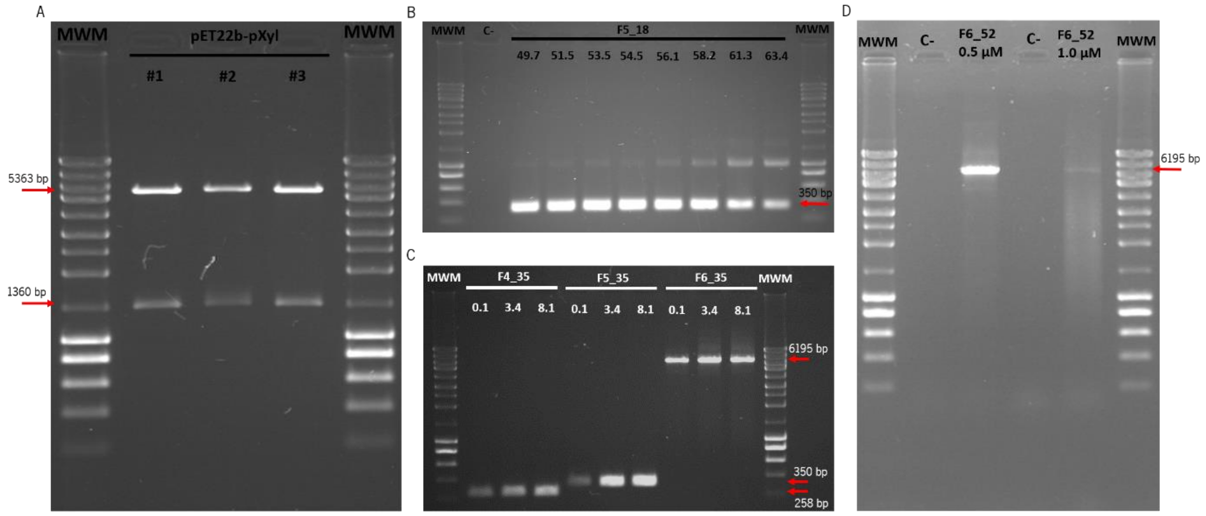

A
MWM
pET22b-pXyl
MWM
#1
#2
#3
5363 bp
1360 bp
B
MWM
C-
F5_18
MWM
49.7
51.5
53.5
54.5
56.1
58.2
61.3
63.4
350 bp
C
MWM
F4_35
F5_35
F6_35
MWM
0.1
3.4
8.1
0.1
3.4
8.1
0.1
3.4
8.1
6195 bp
350 bp
258 bp
D
MWM
C-
F6_52
0.5 µM
C-
F6_52
1.0 µM
MWM
6195 bp

Figure 22 | Template preparation, F4, F5 and F6 fragment amplification and optimization with Phusion. (A) Digested pET22b-pXyl templates to confirm correct construct (vector: 5363 bp; insert: 1360 bp). (B) Optimization of annealing temperature for fragment F5_18 (≈258 bp, fragment 5, 18 nt primers). (C) Amplification of F4_35 (≈258 bp), F5_35 (≈350 bp) and F6_35 (≈6195 bp) with different template quantities. (D) Optimization of primer concentration for fragment F6_52 (≈6195 bp) amplification. MWM: molecular weight marker NZYDNA Ladder III; C-: negative control, no template.

Table 6 | Results corresponding to the RSM-CCD experiment. Colony and sequencing results of transformations related to the RSM-CCD. Number of identified positives (Y194S-A272S-Y378S), wild-types (WT), pET22b-pXyl truncations (truncated), single (Y194S; A272S; Y378S) and double (Y194S-A272S; Y194S-Y378S; A272S-Y378S) mutants are displayed. Conditions are ordered with increasing homology, template and ratio, respectively. Competence of NZY5α cells was 10^7 unless otherwise stated.

Conditions homology: total template: vector to insert ratio	**cfu**	**Positives**	**Negatives**							
			WT	**truncated**	**Single**			**Double**		
					Y194S	A272S	Y378S	Y194S-A272S	Y194S-Y378S	A272S-Y378S
18; 3.4; 1:5	94	0	0	0	2	0	1	0	7	0
18; 3.4; 1:5 *	476	1	1	0	2	1	1	0	4	0
25; 0.6; 1:2	12	4	1	3	0	0	1	0	1	0
25; 0.6; 1:8	28	5	0	1	0	0	2	0	2	0
25; 6.2; 1:2	124	1	0	0	2	1	2	0	4	0
25; 6.2; 1:8	102	1	2	0	1	0	2	0	4	0
35; 0.1; 1:5	43	9	0	1	0	0	0	0	0	0
35; 0.1; 1:5 *	250	7	0	3	0	0	0	0	0	0
35; 3.4; 1:1	66	1	6	0	0	0	3	0	0	0
35; 3.4; 1:1 *	118	6	0	1	0	0	0	0	3	0
35; 3.4; 1:5	23	10	0	1	1	0	1	0	0	0
35; 3.4; 1:5	12	4	0	1	3	0	1	0	1	0
35; 3.4; 1:5	9	4	0	2	2	0	0	0	1	0
35; 3.4; 1:5	44	10	0	2	1	0	0	0	0	0
35; 3.4; 1:5	50	6	1	2	0	0	0	0	1	0
35; 3.4; 1:5	20	6	1	3	0	0	0	0	0	0
35; 3.4; 1:10	26	9	0	1	0	0	0	0	0	0
35; 3.4; 1:10 *	311	7	1	1	0	0	1	0	0	0
35; 8.1; 1:5	30	5	2	3	0	0	0	0	0	0
35; 8.1; 1:5 *	120	5	1	0	1	0	1	0	2	0
45; 0.6; 1:2	61	5	0	4	0	0	0	0	1	0
45; 0.6; 1:8	80	7	0	3	0	0	0	0	0	0
45; 6.2; 1:2	36	5	0	1	0	0	0	2	2	0
45; 6.2; 1:8	33	5	0	1	2	0	1	0	1	0
52; 0.1; 1:10 *, **	187	10	0	0	0	0	0	0	0	0
52; 0.1; 1:10 *, **	210	10	0	0	0	0	0	0	0	0
52; 0.1; 1:10 *, **	157	9	0	1	0	0	0	0	0	0
52; 3.4; 1:5	18	9	0	1	0	0	0	0	0	0
52; 3.4; 1:5 *	308	5	2	1	1	0	0	0	1	0
Total	-	166	18	37	18	2	17	2	35	0
					37			37		

***** - Competence of NZYα cells 10^8

****** - Optimal conditions suggested by the quadratic model

4.3.4 Model Fit, Diagnose and Interpretation

The 260 sequencing results (see Table 6) were integrated into the surface response model to evaluate different models. A confidence level of 95% was used for all statistical analysis and the linear model was determined as the most significant with a p value of 0.0013. The data was also compatible with a quadratic model (Figure 23) which showed a p value of 0.0148. For both models, lack of fit was not significant. Although the linear model has a lower p value, the model which best explains the variation encountered in the results is the quadratic model, which has an adjusted R^2 of 0.5109 versus the 0.4676 of the linear model (Table 7).

Table 7 | Model summary statistics for recombination efficiency. Standard deviation, R^2, adjusted R^2, predicted R^2 and p values are listed for their respective models.

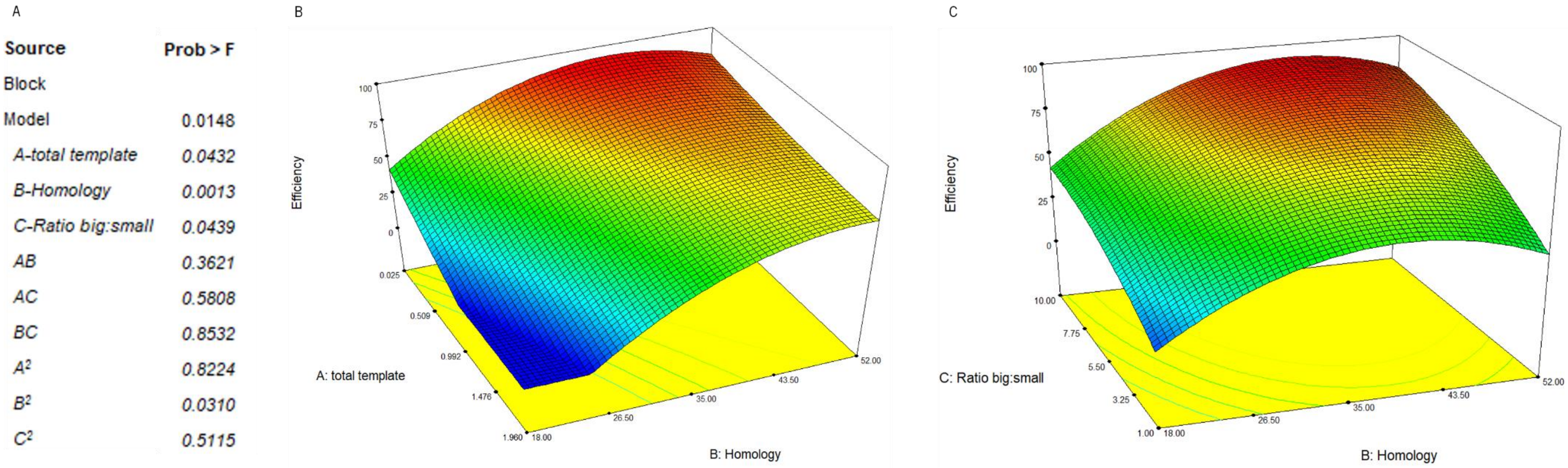
Model	Std. Dev.	R-Squared	Adjusted R-Squared	Predicted R-Squared	p value	Lack of Fit
Linear	19.58	0.5371	0.4676	0.2303	0.0013	0.6713
2FI	20.62	0.5635	0.4094	-0.0624	0.0162	0.6145
Quadratic	18.76	0.7023	0.5109	-0.0319	0.0148	0.7080
Cubic	21.42	0.7783	0.3626	-	0.1866	0.5320

The input variables (total template, homology, vector to insert ratio) were statistically analysed with the parametric ANOVA test and all variables are predicted to have a significant impact on recombination efficiency (A, B, C, Figure 23A) while interactions between factors are not (AB, AC, BC). Furthermore, B^2 (Figure 23A) is significant, which indicates that optimisation through this process variable (homology size/primer size) has been exhausted.

Sequence analysis showed in accordance to previous results a slightly higher tendency to produce truncated pXyl variants at lower template concentrations (Table 6). Interestingly, the single mutant A272S was only encountered twice in a total of 37 sequenced single mutants. Moreover, Y194S and Y378S single mutants were equally represented, with 18 and 17, respectively. For double mutants, similar results were found. Y194S-Y378S mutations represent roughly 95% of all sequenced double mutants. Mutants containing Y194S and A272S were only identified twice, while A272S-Y378S mutants were never found in a total of 37 sequenced double mutants. It is worth noting that mutants classified as single or double mutants do not contain truncations but rather wild-type sequence between the mutations.

4.3.5 Model Confirmation

Optimal conditions determined by setting ranges for terms A, B and C between 0.1 – 8.1 ng, 18 – 52 nt and 1:1:1 – 1:10:10, respectively, enabling maximum recombination efficiency were calculated with this quadratic model as using a primer size of 52 bp, 0.1 ng template and a PCR volume ratio of vector to inserts 1:10:10 (Figure 23B and C). To confirm this, the multi-site-directed mutagenesis experiment was repeated in triplicate with these conditions; 29 plasmids of the 30 sequenced contained all three desired mutations, indicating a recombination efficiency of 97% (Table 6). Estimated efficiency by the model was 82%, showing a 15% difference, probably due to its relatively poor predicted R-squared value (Table 7). When data from optimal conditions were introduced to complete the model, adjusted R-squared and predicted R-squared rose to 0.6470 and 0.3478, respectively. Recombination efficiency for optimal conditions estimated by the complete model (95%) was in good agreement with observed results.


A

Source	Prob > F
Block	
Model	0.0148
A-total template	*0.0432*
B-Homology	*0.0013*
C-Ratio big:small	*0.0439*
AB	*0.3621*
AC	*0.5808*
BC	*0.8532*
A^2	*0.8224*
B^2	*0.0310*
C^2	*0.5115*

Figure 23 | Surface response model for recombination efficiency. (A) Statistical analysis of variables by ANOVA using a quadratic model. (B) Quadratic surface response model shown at a large fragment to small fragments ratio of 1:10:10. (C) Quadratic surface response model shown at a total template amount of 0.1 ng. Blue indicates low recombination efficiency. Red indicates high recombination efficiency.

The final quadratic equation applied was:

$$y = 59.01 - 10.92A + 16.92B + 10.33C + 6.25AB - 3.75AC + 1.25BC + 1.14A^2 - 9.27B^2 - 3.17C^2$$

5. Discussion

Site-directed mutagenesis is a powerful tool in protein engineering used to make specific alterations to a specific gene and gene product. A number of protocols exist for both single-site-directed mutagenesis and multi-site-directed mutagenesis but are characterised by multiple steps and high costs. *E. coli* is a well-established biological model used worldwide with well characterised processes that partake in recombination. Amplified, suitably designed, mutagenic fragments can be assembled into a functional plasmid by these processes and hence this host offers a potential for simplifying and reducing the costs of mutagenesis protocols. The present study was focused on the optimisation of such a protocol.

5.1 Possible Epigenetic Interference by DNA Polymerases

An initial interesting observation of the present study was the high number of wild-type unmutated sequences that were identified in the final transformants, even after *Dpn*I digestion, when Phusion DNA polymerase was used. Phusion is a recombinant protein that combines the family B DNA polymerase from *Pyrococcus furiosus* with the proliferating cell nuclear antigen (PCNA) homolog from *Sulfolobus solfataricus* (Wang, 2000). PCNA homologs are non-specific dsDNA binding proteins that act as sliding clamps and mediate metabolic processes involved in DNA replication and/or DNA repair by interacting with polymerases and/or nucleases (Burkhart et al., 2017; Cann et al., 1999; Kong et al., 1992). The sliding clamp is crucial for high processivity during DNA replication (Stukenberg et al., 1991; Wang et al., 2004). Indeed, the Sso-7d PCNA homolog confers Phusion with its high processivity while the *Pfu* DNA polymerase mediates thermostable DNA replication in a high-fidelity matter (Choli et al., 1988; Cline, 1996; Greagg et al., 1999; Wang et al., 2004). Phusion's high processivity and high-fidelity makes it attractive for site-directed mutagenesis and PCRs where the amplification of large fragments is necessary. In the present study, Phusion was employed with mutagenic primers to produce mutated amplicons of up to 6.2 kb long which were then treated with *Dpn*I to digest the methylated GATC sequences of template DNA isolated from dam+ *E. coli* strains and thereby remove the template. Our results indicate that Phusion activity interfered with this *Dpn*I activity, enabling the persistence of template DNA and suggesting that Phusion could manipulate epigenetic marks, specifically adenosine methylations within GATC specific sequences. Investigation of several high-fidelity DNA polymerases (Phusion, Xpert HighFidelity, NZYProof) showed that Xpert HighFidelity also enabled template persistence. Interestingly, this DNA polymerase also contains a clamping domain which is fused to an Archaean DNA polymerase to combine high-fidelity and increased processivity. On the other hand, the NZYProof DNA does not contain such a clamping domain.

PCNA are known to mediate epigenetic inheritance by interacting with methyltransferases and other involved proteins (Chuang et al., 1997; Dionne and Bell, 2005; Moggs et al., 2000; Zhang et al., 2000b). Interaction between epigenetic marks and PCNA are not however described in the literature. An alternative explanation, though less likely, would be that both Phusion and Xpert HighFidelity remain tightly bound to DNA and impede *Dpn*I activity by physically obstructing the GATC sites. Sliding clamps show high affinity towards dsDNA and could therefore anchor the polymerase to the specific sites (Baumann et al., 1994). Further studies investigating possible epigenetic interference of specific polymerases or attached domains are called for.

5.2 Microhomology-Mediated Recombination and Fragment Bias

A second observation from our initial method development studies was the presence of constructs containing pET22b-pXyl truncations in which pXyl was frequently completely removed from constructs (Figure 18). Interestingly, these truncations were most frequent when low template concentrations were used (0.07 ng template) and indeed in our later RSM-CCD study truncation frequency was found to be slightly more frequent with decreasing template concentrations. Sequence analysis of adjacent truncation sites revealed the presence of microhomologies of between 5 and 8 nucleotides. Frequently, truncations occurred at the 6xHis-tag site (Figure 18), suggesting that repetitive sequences could interfere with the efficiency of this protocol. Also, investigation of a sample with an increase in molecular weight (Figure 17, Sample 17) revealed an insertion of 291 nt between Y266 and I267. The inserted sequence was identified as part of the fragment 5 PCR product. Similar to truncation events, this insertion was also most likely caused by a microhomology of ≈10 nt. The idea that both truncations and insertions are performed by microhomologies is further supported by the recent proposed mechanism of alternative end-joining (Chayot et al., 2010). Evidence in the past had hinted to a DSB repair mechanism that promotes end-joining in the absence of large homologous regions (Albertini et al., 1982; Meddows et al., 2005). On the other hand, alternative end-joining could promote large deletions through resection by RecBCD and possibly other nucleases which would continue until microhomologies were encountered by ligase A and thereby facilitate microhomology-mediated recombination as seen in the present study (Chayot et al., 2010).

During the studies, a number of single and double mutants were observed with a bias towards mutations involving the large, selection marker containing fragment, i.e. F3 and F6. Single mutants are most likely the result of linear amplification wherein *E. coli* cells are transformed with the linear amplification product of any given primer. Such linear amplifications produce relatively low levels of amplicons containing the whole vector information with only one mutation (Figure 24A). Following *Dpn*I treatment the linear amplified product containing the single mutation is then transformed into *E. coli* cells along with the predominant PCR product (double mutant fragment). Since most template is removed by the digestion step, the linear amplification product is in a ssDNA form. *E. coli* synthesises the complementary strand to this and repairs nicks to produce circular dsDNA with a single mutation as observed in our study. In the case of the production of double mutants, it is believed that this can arise through recombination of predominant PCR products (Figure 24B) with either wild-type plasmids or single mutation plasmids.

Data analysis of mutants from the RSM-CCD study showed that A272S single mutants were strongly underrepresented when compared to Y194S and Y378S single mutants (Table 6). An equal distribution of single mutants would be expected since linear amplification products are produced in similar quantities during PCR. Interestingly, this pattern continued in double mutants, with Y194S-A272S double mutants only identified two times and A272S-Y378S double mutants not being identified once (Table 6). In contrast, the majority of double mutants were Y194S-Y378S, with 35 hits. The cause for this bias towards specific mutants is unclear but is probably related to the large size and possibly also presence of the selection marker gene of the fragment involved in the production of these mutants.

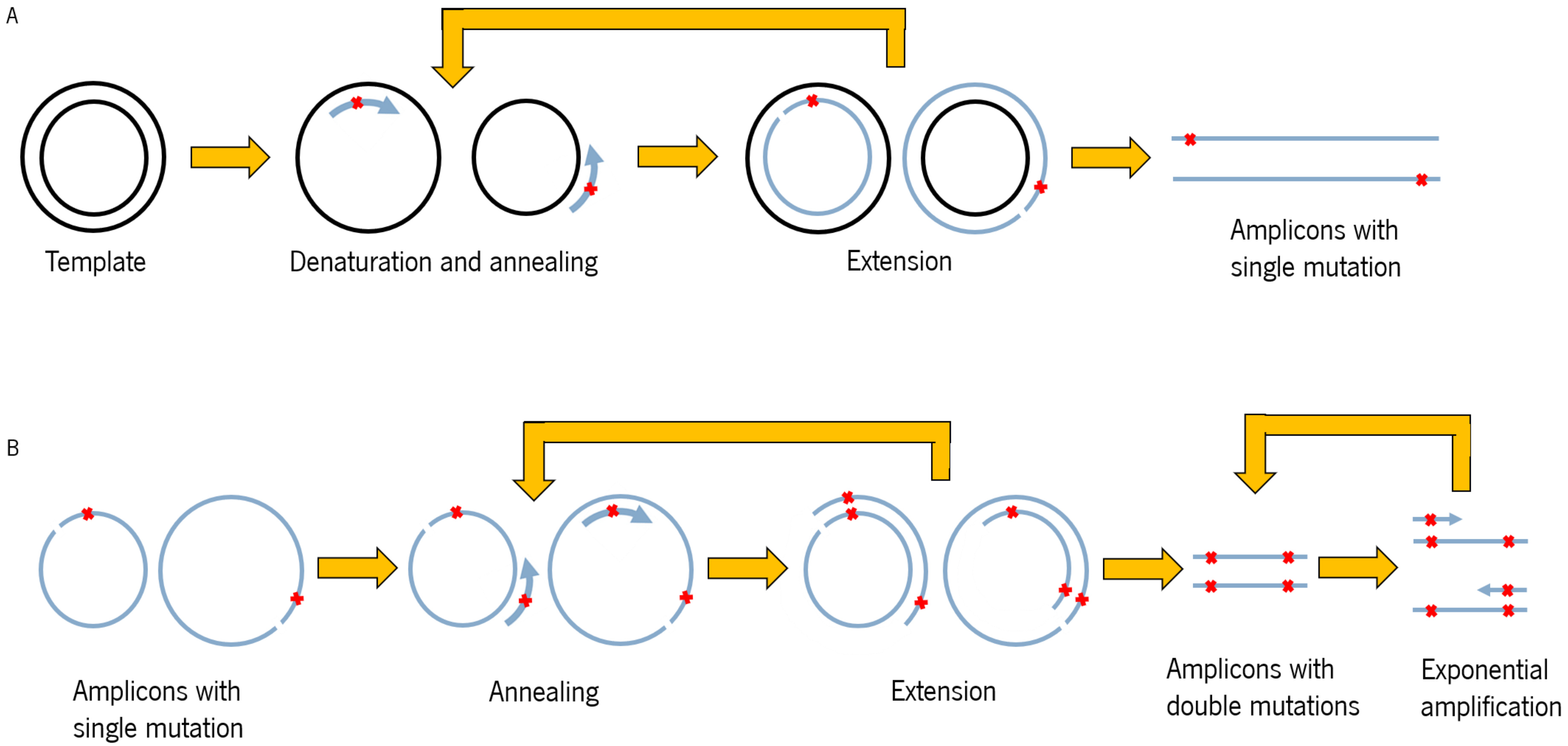

A
Template
Denaturation and annealing
Extension
Amplicons with single mutation
B
Amplicons with single mutation
Annealing
Extension
Amplicons with double mutations
Exponential amplification

Figure 24 | Mechanism of production of single and double mutants. Amplification of fragments using circular DNA template and primer setup previously stated. (A) First PCR cycle produces circular nicked amplicons with only one mutation. Amplification of this PCR product is linear since this product is only produced when primers (blue curved arrows) anneal to the template. (B) PCR product of the first PCR cycle serves as template for future cycles to produce the desired fragment, containing two mutations. PCR fragments are also used in subsequent PCR cycles as templates to produce desired fragments, leading to exponential amplification. Bold lines indicate wild-type constructs; Blue lines indicate PCR products; Red crosses indicate mutations.

5.3 Identification of Crucial Factors for Mutagenesis Efficiency

Initial experiments identified the amount of total template as a key factor governing mutagenesis efficiency. To identify additional crucial factors a meta-analysis of conditions used in both PCR and recombination was conducted. Identified factors were: (1) number of PCR cycles; (2) total template amount used in PCR; (3) strategies to remove wild-type template; (4) minimal fragment size required for recombination; (4) length of homologous regions; (5) maximum number of fragments assembled; and (6) vector to insert ratio. Based on these parameters, three factors, believed to be crucial for efficient recombination, were selected for use in our RSM-CCD study: template concentration, size of homologous region (i.e. primer size), and fragment ratio. The number of PCR cycles was discarded as an input factor since different template amounts would roughly regulate the amount of PCR product produced. *Dpn*I digestion is the most popular approach used for removing template, its digestion site ($\mathrm{G^{m}ATC}$) occurs at a relatively high frequency (≈1 per 256 bp in DNA of random sequence), and template methylation is simply achieved by propagation in dam+ *E. coli* strains (García-Nafría et al., 2016; Jacobus and Gross, 2015; Kostylev et al., 2015; Liu and Naismith, 2008). Hence *Dpn*I digestion was employed in our study. Minimum fragment size and in particular maximum number of fragments do appear to be important in recombination but study of these was believed to be outside the practical limits of the present study (Mitchell et al., 2013; Taniguchi et al., 2013). Length of the homologous regions and vector to insert ratio had already been identified as crucial factors and optimal conditions for both these factors in vector recombination have been described through the "one factor at a time" methodology (Jacobus and Gross, 2015; Kostylev et al., 2015; Lovett et al., 2002; Watt et al., 1985). It was shown that recombination efficiency tends to improve with increasing homology regions, with 30 to 50 nt generally being used. Meanwhile, vector to insert ratio improves greatly until a ratio of 1:10, while up to 1:20 improvements are subtle (Jacobus and Gross, 2015; Kostylev et al., 2015). These observations however did not consider any interaction between the factors, which may or may not lead to different effects.

We sought to investigate whether interactions between template amount, size of the homologous regions and vector to insert ratios occurs while simultaneously producing a model that would aid in maximising mutagenesis efficiency with the multi-site-directed mutagenesis protocol conditions used. This was carried out using surface response methodology which was employed in combination with a CCD to design experimental conditions. Three significant models were obtained with this framework, the most significant being a linear model (Table 7). Nevertheless, despite the linear model being more significant, the quadratic model was preferred for data analysis and investigation of optimal conditions due to a higher R^2 and adjusted R^2. R^2 indicates the accuracy of the model based on the determined points, whereas the adjusted R^2 indicates how much variation within the experiment is explained by the model. Predicted R^2 was lower than in the linear model however, meaning that prediction via this model is slightly less accurate than by the linear model. The two-factor interaction (2FI) model scored worse in almost every aspect comparative to linear and quadratic models and was therefore not adopted. The use of the cubic model was strongly discouraged since aliasing was detected. The significance of numeric factors was examined by ANOVA with a confidence level of 95% and determined in all cases as significant (Figure 23A). Interactions between factors (AB, AC, BC) were not detected. Use of the model identified optimal mutagenesis conditions with the protocol as being 0.1 ng template, a fragment ratio of 1:10:10 and between 40 and 50 base homologies. Triplicate investigations with these conditions demonstrated a 97% efficiency in introducing all three desired mutations. This being similar to the predicted efficiency of 95% by the complete model and thereby inferring the suitability of the model for estimating mutagenesis efficiency for any combination of conditions within the design space.

5.4 *recA1 E. coli* Strains as Hosts for Recombination *In Vivo*

Recombination in *E. coli* is primarily performed by the single-strand binding protein RecA (Shibata et al., 1979b). It promotes the formation of recombinogenic DNA, induces strand invasion and completes recombination by exchanging strands (Dunderdale et al., 1991; Kowalczykowski et al., 1987; Shibata et al., 1979b). The mechanisms by which RecA is loaded onto DNA is well characterised. Here, two main pathways exist which are thought to have different roles. The RecBCD pathway employs the RecBCD helicase-nuclease complex to convert dsDNA into a recombination suitable substrate (ssDNA-RecA) (Amundsen et al., 1986; Churchill and Kowalczykowski, 2000; Singleton et al., 2004; Smith, 1991). The RecFOR pathway, composed of RecF, RecO and RecR, is believed to have evolved to repair DNA lesions unsuitable for RecBCD, such as ssDNA gaps (Churchill and Kowalczykowski, 2000; Morimatsu and Kowalczykowski, 2014; Smith, 1991; Tseng et al., 1994). *E. coli* cloning hosts (e.g. XL1-Blue, DH5α, NZY5α) generally contain the *recA1* genotype, which renders the RecA protein inactive and improves plasmid stability. Loss of function resides in the substitution of glycine 160 by an aspartic acid residue where RecA1 is no longer able to interact competitively with ssDNA against SSBs and as a result, *recA1* mutants show significant recombination deficiency (Bryant, 1988; Kowalczykowski and Krupp, 1987).

Despite the occurrence of a *recA1* genotype, *E. coli* cloning strains have been found to still be capable of performing significant recombination. In fact, our results indicate that given proper optimised conditions *E. coli* is remarkably efficient in assembling three fragments. Indeed, recombination in *recA1*

E. coli strains has been reported for some time but appears to have been little heeded by the scientific community (Bubeck et al., 1993; García-Nafría et al., 2016; Jacobus and Gross, 2015; Kostylev et al., 2015). The type of recombination involved is believed to be different from the phage-induced RecA-independent recombination (Dutra et al., 2007; Swingle et al., 2010). The RecET phage system is only functional in *recBC sbcA* or *recBC sbcBC* mutants, and in *recA1* or Δ*recA* this phage system is supressed (Fishel et al., 1981; Kushner et al., 1974). The Red system is not naturally present in cloning strains and no other phage system that performs recombination is known to inhabit *E. coli*. Therefore, a different endogenous mechanism is thought to exist which is capable of performing recombination at moderate frequencies (Dutra et al., 2007; Lovett et al., 2002; Ozgenc et al., 2005). *E. coli*'s RecA-dependent recombination requires large homologies (≈150 bp) while RecA-independent recombination, requires smaller homologies, reaching near maximum efficiency at ≈50 bp as shown in our study. Recombinant frequencies for two fragment assemblies with RecA-independent recombination varies between 10^{-8} and 10^{-7} per viable cell, far lower than RecA-dependent recombination rates of $10^{-6}$and 10^{-4} per viable cell (Lovett et al., 2002). Interestingly, deleting the *ruvC* gene, a gene encoding a recombinase that resolves Holliday junctions (resolvase), has been shown to stimulate RecA-independent recombination (Dunderdale et al., 1991; Lovett et al., 2002). The reason behind it is unknown but it is speculated that RuvC cleaves critical intermediates required for more efficient RecA-independent recombination (Lovett et al., 2002). In another report, the efficiency of RecA-independent recombination was found to be competitive with that of RecA-dependent recombination by deleting single-strand exonucleases, such as RecJ and SbcB (Dutra et al., 2007). Indeed, mutating several single-strand exonucleases and *recA* increases the recombination rate 10-fold and cross-over recombination can by the same means be increased up to 65-fold. The most dramatic increases however were found in gene conversion recombination, where deleting *recJ* and *sbcB* exonucleases led to a 1000-fold increase. Interestingly, mutating only one given exonuclease produces comparable rates of gene conversion events compared to wild-type, hence suggesting exonuclease redundancy in terms of removing substrates for RecA-independent recombination. The low recombination frequency encountered in many cases seems to be linked to low single-strand substrates, rather than to an inefficient RecA-independent recombination system (Dutra et al., 2007). Indeed, the recent RecBFI pathway suggests that protection of ssDNA is important for DNA repair by recombinatorial means (Buljubašić et al., 2019). In yeast the absence of the RecA homolog RAD51, a crucial player in eukaryotic homologous recombination, can still promote recombination through the ssDNA protective action of RAD52 (Malacaria et al., 2019; Malkova et al., 1996; Tsukamoto et al., 2003). Taken together, studies imply that RecA-independent recombination occurs through its own pathway, rather than in a defective RecA-dependent manner or even by another unidentified phage system (Dutra et al., 2007; Lovett et al., 2002; Ozgenc et al., 2005).

The presence of RuvC and several exonucleases in common cloning strains (e.g. XL1-Blue, DH5α, NZY5α) could explain the requirement of *E. coli* cells with high competence (minimum of 10^7 cfu/μg) since these proteins might be involved in removing substrates required for RecA-independent recombination. The current hypothesis implies that RecA-independent recombination is possibly performed via strand invasion of 3'- or 5'-end tailed DNA molecules that displace and anneal to homologous sequences or alternatively by ssDNA molecules binding to single-strand regions within replication forks (Bzymek and Lovett, 2001; Dutra et al., 2007). The later hypothesis appears to be more appropriate in our case. The blunt-ended double-stranded DNA molecules prepared during the PCR reactions could be processed by the functional RecBCD complex in the NZY5α and XL1-Blue cells. Since the pET22b-pXyl construct features no Chi sequences, the double-stranded PCR fragments would be processed into ssDNA. Alternatively, the combined action of RecQ and RecJ could also allow for PCR products to be processed into ssDNA. In *recA* wild-type strains, the formation of the RecA filament protects the 3'-ends of ssDNA against exonucleases (Dermić et al., 2017). In the case of NZY5α and XL1-Blue, the RecA1 mutant is unable to bind to ssDNA and thus, cannot prevent its degradation from exonucleases such as RecJ and SbcB. As a result, substrates required for RecA-independent recombination might be present in only low amounts and hence reduces the recombination efficiency, requiring highly competent cells for obtention of only low numbers of colonies, albeit in sufficient number for mutagenesis purposes (Jacobus and Gross, 2015; Kostylev et al., 2015).

5.5 Protocol Mutagenesis Efficiency

Efficient site-directed mutagenesis protocols are crucial to the scientific community so that more resources can be allocated to the research of protein function and structure and to protein engineering for enhancing protein properties. Several commercial kits exist to simplify the process but require considerable investment for relatively few mutations, especially if large mutant libraries are to be made (Bauer et al., 1995a, 1995b, 1995c; Hogrefe and Cline, 2001; Liang et al., 2012). These kits come with their own limitations and some protocols have improved upon them (Liu and Naismith, 2008; Zheng et al., 2004). Many multi-site-directed mutagenesis protocols seem overcomplicated: requiring multiple PCR rounds, employing special enzyme cocktails or multiple steps to assemble mutagenised constructs. The method by which these mutagenic PCR products are assembled into one large DNA entity has since the

1990s shifted from restriction and ligation cloning to recombination (Bubeck et al., 1993; Jacobus and Gross, 2015; Jones and Winistorfer, 1991; Kostylev et al., 2015; Zhang et al., 2012).

Using the optimised conditions, a mutagenesis efficiency of 97% was achieved for introduction of three mutations with our protocol. Commercial kits from Agilent Technologies and Thermo Fisher Scientific showed lower (>55%) and comparable (≈95%) efficiencies, respectively but require additional steps and costs. *S. cerevisae* has also been shown to be an efficient host for recombination with efficiencies of ≈100% with three fragments, but this host is characterised by the necessity for expertise in working with eukaryotic systems, a reduced growth rate and a capacity for performing non-homologous end-joining which can lead to false positives and thereby requires a more elaborate amplification strategy to increase recombination efficiency (Kuijpers et al., 2013; Shao et al., 2009). Considering that such high efficiency multi-site-directed mutagenesis was performed in our study using a standard RecA negative *E. coli* cloning host without recombination enhancing systems is remarkable and demonstrates the power of mathematical and statistical techniques for experimental design and data set analysis for process optimisation.

6. Conclusions and Future Perspectives

A multi-site-directed mutagenesis protocol was optimised that is both highly reliable and practical. The protocol makes use of PCR based introduction of mutations, *Dpn*I digestion and *E. coli*'s endogenous recombination machinery to assemble three mutagenic fragments into one functional construct *in vivo*. The protocol was optimised by employing a response surface methodology along with a central composite design. Over 300 sequencings were performed to seek optimal conditions given the variables: total template amount (ng), homology length (nt) and large fragment to small fragments ratio (1:n:n). A recombination efficiency of 97% was obtained when using 0.1 ng template, 52 base primers and a fragment ratio of 1:10:10 (0.5 µL of large fragment PCR product with 5 µL of each smaller fragment). Both *E. coli* NZY5α and XL1-Blue strains were shown to function with the protocol developed and it is expected that this can be extended to other common cloning strains e.g. DH5α, DH10B, XL10-Gold and DB3.1. An essential requirement for success of the protocol is the use of highly competent cells with a competency of at least 10^7 or 10^8 cfu/µg.

Current limitations associated with this protocol are the long *Dpn*I digestion time (4 hours), false positives related to template wild-type and use of Phusion DNA polymerase, and the requirement for high cell competence to obtain sufficient colonies. Under the optimal process conditions with low template concentrations (0.1 ng total template), *Dpn*I digestion could possibly be reduced and should be investigated. Our investigation of different DNA polymerase hinted at the possibility that the clamping domains of both Phusion and Xpert HighFidelity could possibly be involved in the persistence of wild-type template when these polymerases are used. Future experiments should therefore seek to investigate this further and to better understand the role, if any, of sliding clamps in interfering with DNA methylations and/or in reducing subsequent *Dpn*I activity. The requirement of highly competent cells is most certainly related to the *recA1* genotype that cloning hosts generally have, but also to the existence of many single-strand exonucleases simultaneously present, that degrade recombination substrates. To reduce the requirement for high cell competence, expression or silencing of *rec* and/or *sbc* genes could possibly be envisioned. Nevertheless, the high mutagenesis efficiencies attained in the present study and the relative ease with which cells with competencies of $\approx 10^7$ or 10^8 cfu/µg are prepared, probably negates the need for this. Additional further studies could involve testing different cloning strains and even expression strains (e.g. BL21) so as to expand the use of this protocol. Finally, investigating the minimal size of fragments required for recombination and, more importantly, the maximum number of mutagenic fragments that can be assembled in an efficient manner should be investigated.

7. References

Albertini, A.M., Hofer, M., Calos, M.P., and Miller, J.H. (1982). On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. Cell *29*, 319–328.

Amundsen, S.K., Taylor, A.F., Chaudhury, A.M., and Smith, G.R. (1986). *recD*: The gene for an essential third subunit of exonuclease V. Proc. Natl. Acad. Sci. U. S. A. *83*, 5558–5562.

Amundsen, S.K., Sharp, J.W., and Smith, G.R. (2016). RecBCD enzyme “Chi Recognition” mutants recognize chi recombination hotspots in the right DNA context. Genetics *204*, 139–152.

Anderson, D.G., and Kowalczykowski, S.C. (1997a). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. Cell *90*, 77–86.

Anderson, D.G., and Kowalczykowski, S.C. (1997b). The recombination hot spot χ is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. Genes Dev. *11*, 571–581.

Antikainen, N.M., and Martin, S.F. (2005). Altering protein specificity: Techniques and applications. Bioorganic Med. Chem. *13*, 2701–2716.

Arnold, F.H. (1993). Engineering proteins for nonnatural environments. FASEB J. *7*, 744–749.

Arnold, D.A., and Kowalczykowski, S.C. (2000). Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme. J. Biol. Chem. *275*, 12261–12265.

Barbour, S.D., Nagaishi, H., Templin, A., and Clark, A.J. (1970). Biochemical and genetic studies of recombination proficiency in *Escherichia coli*. II. Rec+ revertants caused by indirect suppression of rec- mutations. Proc. Natl. Acad. Sci. U. S. A. *67*, 128–135.

Baudin, A., Ozier-kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. Nucleic Acids Res. *21*, 3329–3330.

Bauer, J.C., Wright, D.A., Braman, J.C., and Geha, R.S. (1995a). Circular site-directed mutagenesis. US Patent No. US5789166A. Retrieved from Google Patents.

Bauer, J.C., Wright, D.A., Braman, J.C., and Geha, R.S. (1995b). Circular site-directed mutagenesis. US Patent No. US5932419A. Retrieved from Google Patents.

Bauer, J.C., Wright, D.A., Braman, J.C., and Geha, R.S. (1995c). Circular site-directed mutagenesis. US Patent No. US6391548B1. Retrieved from Google Patents.

Baumann, H., Knapp, S., Lundback, T., Ladenstein, R., and Hard, T. (1994). Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*. Nat. Struct. Biol. *1*, 808–819.

Bhat, G.J., Lodes, M.J., Myler, P.J., and Stuart, K.D. (1991). A simple method for cloning blunt ended DNA fragments. Nucleic Acids Res. *19*, 398.

Bi, X., and Liu, L.F. (1994). *recA*-independent and *recA*-dependent intramolecular plasmid recombination. Differential homology requirement and distance effect. J. Mol. Biol. *235*, 414–423.

Bichara, M., Meier, M., Wagner, J., Cordonnier, A., and Lambert, I.B. (2011). Postreplication repair mechanisms in the presence of DNA adducts in *Escherichia coli*. Mutat. Res. - Rev. Mutat. Res. *727*, 104–122.

Boehmer, P.E., and Emmerson, P.T. (1992). The RecB subunit of the *Escherichia coli* RecBCD enzyme couples ATP hydrolysis to DNA unwinding. J. Biol. Chem. *267*, 4981–4987.

Bork, J.M., Cox, M.M., and Inman, R.B. (2002). The RecOR proteins modulate RecA protein function at 5′ ends of single-stranded DNA. EMBO J. *20*, 7313–7322.

Bryant, F.R. (1988). Construction of a recombinase-deficient mutant *recA* protein that retains single-stranded DNA-dependent ATPase activity. J. Biol. Chem. *263*, 8716–8723.

Bubeck, P., Winkler, M., and Bautsch, W. (1993). Rapid cloning by homologous recombination *in vivo*. Nucleic Acids Res. *21*, 3601–3602.

Buljubašić, M., Hlevnjak, A., Repar, J., Đermić, D., Filić, V., Weber, I., Zahradka, K., and Zahradka, D. (2019). RecBCD- RecFOR-independent pathway of homologous recombination in *Escherichia coli*. DNA Repair (Amst). *83*.

Burkhart, B.W., Cubonova, L., Heider, M.R., Kelman, Z., Reeve, J.N., and Santangelo, T.J. (2017). The GAN exonuclease or the flap endonuclease Fen1 and RNase HII are necessary for viability of *Thermococcus kodakarensis*. J. Bacteriol. *199*.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature *433*, 531–537.

Bzymek, M., and Lovett, S.T. (2001). Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U. S. A. *98*, 8319–8325.

Cann, I.K.O., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K., and Ishino, Y. (1999). Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. J. Bacteriol. *181*, 6591–6599.

Carter, D.M., and Radding, C.M. (1971). The role of exonuclease and beta protein of phage lambda in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. J. Biol. Chem. *246*, 2502–2512.

Carter, Z., and Delneri, D. (2010). New generation of *loxP*-mutated deletion cassettes for the genetic manipulation of yeast natural isolates. Yeast *27*, 765–775.

Chayot, R., Montagne, B., Mazel, D., and Ricchetti, M. (2010). An end-joining repair mechanism in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *107*, 2141–2146.

Choli, T., Henning, P., Wittmann-Liebold, B., and Reinhardt, R. (1988). Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the Archaebacterium, *Sulfolobus solfataricus*. BBA - Gene Struct. Expr. *950*, 193–203.

Chuang, L.S.H., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F.L. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21(WAF1). Science (80-.). *277*, 1996–2000.

Churchill, J.J., and Kowalczykowski, S.C. (2000). Identification of the RecA protein-loading domain of RecBCD enzyme. J. Mol. Biol. *297*, 537–542.

Churchill, J.J., Anderson, D.G., and Kowalczykowski, S.C. (1999). The RecBC enzyme loads *recA* protein onto ssDNA asymmetrically and independently of χ, resulting in constitutive recombination activation. Genes Dev. *13*, 901–911.

Clark, A.J., and Margulies, A.D. (1965). Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. Proc. Natl. Acad. Sci. U. S. A. *53*, 451–459.

Cline, J. (1996). PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. *24*, 3546–3551.

Clore, A., Reinertson, B., and Rose, S. (2011). Experimental overview, protocol, troubleshooting. In Mutagenesis Application Guide, J. Sabel, ed. pp. 5–22.

Connelly, J.C., De Leau, E.S., Okely, E.A., and Leach, D.R.F. (1997). Overexpression, purification, and characterization of the SbcCD protein from *Escherichia coli*. J. Biol. Chem. *272*, 19819–19826.

Cox, M.M., and Lehman, I.R. (1981a). Directionality and polarity in *recA* protein-promoted branch migration. Proc. Natl. Acad. Sci. U. S. A. *78*, 6018–6022.

Cox, M.M., and Lehman, I.R. (1981b). *recA* protein of *Escherichia coli* promotes branch migration, a kinetically distinct phase of DNA strand exchange. Proc. Natl. Acad. Sci. U. S. A. *78*, 3433–3437.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. *97*, 6640–6645.

Dermić, E., Zahradka, D., Vujaklija, D., Ivanković, S., and Dermić, D. (2017). 3'-Terminated overhangs regulate DNA double-strand break processing in *Escherichia coli*. G3 Genes, Genomes, Genet. *7*, 3091–3102.

Dillingham, M.S., Spies, M., and Kowalczykowski, S.C. (2003). RecBCD enzyme is a bipolar DNA helicase. Nature *423*, 893–897.

Dionne, I., and Bell, S.D. (2005). Characterization of an archaeal family 4 uracil DNA glycosylase and its interaction with PCNA and chromatin proteins. Biochem. J. *387*, 859–863.

Dunderdale, H.J., Benson, F.E., Parsons, C.A., Sharpies, G.J., Lloyd, R.G., and West, S.C. (1991). Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC proteins. Nature *354*, 506–510.

Dutra, B.E., Sutera, V.A., and Lovett, S.T. (2007). RecA-independent recombination is efficient but limited by exonucleases. Proc. Natl. Acad. Sci. U. S. A. *104*, 216–221.

Fishel, R.A., James, A.A., and Kolodner, R. (1981). RecA-independent general genetic recombination of plasmids. Nature *294*, 184–186.

Forloni, M., Liu, A.Y., and Wajapeyee, N. (2018). Random mutagenesis using error-prone DNA polymerases. Cold Spring Harb. Protoc. *2018*, 220–230.

Fu, H., Le, S., Chen, H., Muniyappa, K., and Yan, J. (2013). Force and ATP hydrolysis dependent regulation of RecA nucleoprotein filament by single-stranded DNA binding protein. Nucleic Acids Res. *41*, 924–932.

Galletto, R., Amitani, I., Baskin, R.J., and Kowalczykowski, S.C. (2006). Direct observation of individual RecA filaments assembling on single DNA molecules. Nature *443*, 875–878.

Gao, Y., Mutter-Rottmayer, E., Zlatanou, A., Vaziri, C., and Yang, Y. (2017). Mechanisms of post-replication DNA repair. Genes (Basel). *8*.

García-Nafría, J., Watson, J.F., and Greger, I.H. (2016). IVA cloning: A single-tube universal cloning system exploiting bacterial *In Vivo* Assembly. Sci. Rep. *6*.

Gibson, D.G. (2009). Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res. *37*, 6984–6990.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Goodman, M.F., and Tippin, B. (2000). Sloppier copier DNA polymerases involved in genome repair. Curr. Opin. Genet. Dev. *10*, 162–168.

Greagg, M.A., Fogg, M.J., Panayotou, G., Evans, S.J., Connolly, B.A., and Pearl, L.H. (1999). A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. Proc. Natl. Acad. Sci. U. S. A. *96*, 9045–9050.

Güldener, U., Heck, S., Fiedler, T., Beinhauer, J., and Hegemann, J.H. (1996). A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. *24*, 2519–2524.

Gumbiner-Russo, L.M., and Rosenberg, S.M. (2007). Physical analyses of *E. coli* heteroduplex recombination products *in vivo*: On the prevalence of 5' and 3' patches. PLoS One *2*.

Hall, S.D., and Kolodner, R.D. (1994). Homologous pairing and strand exchange promoted by the *Escherichia coli* RecT protein. Proc. Natl. Acad. Sci. U. S. A. *91*, 3205–3209.

Hall, S.D., Kane, M.F., and Kolodner, R.D. (1993). Identification and characterization of the *Escherichia coli* RecT protein, a protein encoded by the *recE* region that promotes renaturation of homologous single-stranded DNA. J. Bacteriol. *175*, 277–287.

Han, E.S., Cooper, D.L., Persky, N.S., Sutera, V.A., Whitaker, R.D., Montello, M.L., and Lovett, S.T. (2006). RecJ exonuclease: Substrates, products and interaction with SSB. Nucleic Acids Res. *34*, 1084–1091.

Handa, N., Bianco, P.R., Baskin, R.J., and Kowalczykowski, S.C. (2005). Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after χ recognition. Mol. Cell *17*, 745–750.

Heermann, R., Zeppenfeld, T., and Jung, K. (2008). Simple generation of site-directed point mutations in the *Escherichia coli* chromosome using Red®/ET® recombination. Microb. Cell Fact. *7*.

Heitman, J., Ivanenko, T., and Kiss, A. (1999). DNA nicks inflicted by restriction endonucleases are repaired by a RecA- and RecB-dependent pathway in *Escherichia coli*. Mol. Microbiol. *33*, 1141–1151.

Heuser, J., and Griffith, J. (1989). Visualization of RecA protein and its complexes with DNA by quick-freeze/deep-etch electron microscopy. J. Mol. Biol. *210*, 473–484.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51–59.

Hogrefe, H., and Cline, J. (2001). Multi-site mutagenesis. US Patent No. US20060051748A1. Retrieved from Google Patents.

Huang, F., Spangler, J.R., and Huang, A.Y. (2017). *In vivo* cloning of up to 16 kb plasmids in *E. coli* is as simple as PCR. PLoS One *12*.

Ito, W., Ishiguro, H., and Kurosawa, Y. (1991). A general method for introducing a series of mutations into cloned DNA using the polymerase chain reaction. Gene *102*, 67–70.

Ivančić-Baće, I., Salaj-Šmic, E., and Brčić-Kostić, K. (2005). Effects of *recJ*, *recQ*, and *recFOR* mutations on recombination in nuclease-deficient *recB recD* double mutants of *Escherichia coli*. J. Bacteriol. *187*, 1350–1356.

Iwasaki, H., Takahagi, M., Shiba, T., Nakata, A., and Shinagawa, H. (1991). *Escherichia coli* RuvC protein is an endonuclease that resolves the Holliday structure. EMBO J. *10*, 4381–4389.

Jacobus, A.P., and Gross, J. (2015). Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli*. PLoS One *10*.

Jones, D.H., and Winistorfer, S.C. (1991). Site-specific mutagenesis and DNA recombination by using PCR to generate recombinant circles *in vitro* or by recombination of linear PCR products *in vivo*. Methods *2*, 2–10.

Joo, C., McKinney, S.A., Nakamura, M., Rasnik, I., Myong, S., and Ha, T. (2006). Real-time observation of RecA filament dynamics with single monomer resolution. Cell *126*, 515–527.

Kaiser, K., and Murray, N.E. (1979). Physical characterisation of the “Rac prophage” in *E. coli* K12. Mol. Gen. Genet. *175*, 159–174.

Kaiser, K., and Murray, N.E. (1980). On the nature of *sbcA* mutations in *E. coli* K12. Mol. Gen. Genet. *179*, 555–563.

Karu, A.E., Sakaki, Y., Echols, H., and Linn, S. (1975). The γ protein specified by bacteriophage λ. Structure and inhibitory activity for the *recBC* enzyme of *Escherichia coli*. J. Biol. Chem. *250*, 7377–7387.

Kong, X.P., Onrust, R., O’Donnell, M., and Kuriyan, J. (1992). Three-dimensional structure of the β subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. Cell *69*, 425–437.

Kostylev, M., Otwell, A.E., Richardson, R.E., and Suzuki, Y. (2015). Cloning should be simple: *Escherichia coli* DH5α-mediated assembly of multiple DNA fragments with short end homologies. PLoS One *10*.

Kouzminova, E.A., and Kuzminov, A. (2004). Chromosomal fragmentation in dUTPase-deficient mutants of *Escherichia coli* and its recombinational repair. Mol. Microbiol. *51*, 1279–1295.

Kowalczykowski, S.C., and Krupp, R.A. (1987). Effects of *Escherichia coli* SSB protein on the single-stranded DNA-dependent ATPase activity of *Escherichia coli* RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-. J. Mol. Biol. *193*, 97–113.

Kowalczykowski, S.C., Clow, J., and Krupp, R.A. (1987). Properties of the duplex DNA-dependent ATPase activity of *Escherichia coli recA* protein and its role in branch migration. Proc. Natl. Acad. Sci. U. S. A. *84*, 3127–3131.

Kramer, K.M., Brock, J.A., Bloom, K., Moore, J.K., and Haber, J.E. (1994). Two different types of double-strand breaks in *Saccharomyces cerevisiae* are repaired by similar RAD52-independent, nonhomologous recombination events. Mol. Cell. Biol. *14*, 1293–1301.

Kuijpers, N.G.A., Solis-Escalante, D., Bosman, L., van den Broek, M., Pronk, J.T., Daran, J.M., and Daran-Lapujade, P. (2013). A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. Microb. Cell Fact. *12*.

Kulkarni, S.K., and Stahl, F.W. (1989). Interaction between the *sbcC* gene of *Escherichia coli* and the *gam* gene of phage lambda. Genetics *123*, 249–253.

Kushner, S.R., Nagaishi, H., Templin, A., and Clark, A.J. (1971). Genetic recombination in *Escherichia coli*: the role of exonuclease I. Proc. Natl. Acad. Sci. U. S. A. *68*, 824–827.

Kushner, S.R., Nagaishi, H., and Clark, A.J. (1974). Isolation of exonuclease VIII: The enzyme associated with the *sbcA* indirect suppressor. Proc. Natl. Acad. Sci. U. S. A. *71*, 3593–3597.

Kuzminov, A. (1995). Collapse and repair of replication forks in *Escherichia coli*. Mol. Microbiol. *16*, 373–384.

Kuzminov, A., Schabtach, E., and Stahl, F.W. (1994). Chi sites in combination with RecA protein increase the survival of linear DNA in *Escherichia coli* by inactivating *exoV* activity of RecBCD nuclease. EMBO J. *13*, 2764–2776.

Lehman, I.R., and Nussbaum, A.L. (1964). The deoxyribonucleases of *Escherichia Coli*. V. On the specificity of exonuclease I (phosphodiesterase). J. Biol. Chem. *239*, 2628–2636.

Li, M.Z., and Elledge, S.J. (2007). Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. Nat. Methods *4*, 251–256.

Li, Z., Karakousis, G., Chiu, S.K., Reddy, G., and Radding, C.M. (1998). The beta protein of phage λ promotes strand exchange. J. Mol. Biol. *276*, 733–744.

Liang, X., Peng, L., Li, K., Peterson, T., and Katzen, F. (2012). A method for multi-site-directed mutagenesis based on homologous recombination. Anal. Biochem. *427*, 99–101.

Liu, H., and Naismith, J.H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. *8*, 91.

Lloyd, R.G., and Buckman, C. (1985). Identification and genetic analysis of *sbcC* mutations in commonly used *recBC sbcB* strains of *Escherichia coli* K-12. J. Bacteriol. *164*, 836–844.

Lovett, S.T., and Kolodner, R.D. (1989). Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the *recJ* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *86*, 2627–2631.

Lovett, S.T., Gluckman, T.J., Simon, P.J., Sutera, V.A., and Drapkin, P.T. (1994). Recombination between repeats in *Escherichia coli* by a *recA*-independent, proximity-sensitive mechanism. Mol. Gen. Genet. *245*, 294–300.

Lovett, S.T., Hurley, R.L., Sutera, V.A., Aubuchon, R.H., and Lebedeva, M.A. (2002). Crossing over between regions of limited homology in *Escherichia coli*: RecA-dependent and RecA-independent pathways. Genetics *160*, 851–859.

Low, B. (1973). Restoration by the rac locus of recombinant forming ability in recB- and recC- merozygotes of *Escherichia coli* K-12. Mol. Gen. Genet. *122*, 119–130.

Madiraju, M.V.V.S., and Clark, A.J. (1992). Evidence for ATP binding and double-stranded DNA binding by *Escherichia coli* RecF protein. J. Bacteriol. *174*, 7705–7710.

Malacaria, E., Pugliese, G.M., Honda, M., Marabitti, V., Aiello, F.A., Spies, M., Franchitto, A., and Pichierri, P. (2019). RAD52 prevents excessive replication fork reversal and protects from nascent strand degradation. Nat. Commun. *10*.

Malkova, A., Ivanov, E.L., and Haber, J.E. (1996). Double-strand break repair in the absence of RAD51 in yeast: A possible role for break-induced DNA replication. Proc. Natl. Acad. Sci. U. S. A. *93*, 7131–7136.

Manivasakam, P., Weber, S.C., Mcelver, J., and Schiestl, R.H. (1995). Micro-homology mediated PCR targeting in *Saccharomyces cerevisiae*. Nucleic Acids Res. *23*, 2799–2800.

Matsui, D., Nakano, S., Dadashipour, M., and Asano, Y. (2017). Rational identification of aggregation hotspots based on secondary structure and amino acid hydrophobicity. Sci. Rep. *7*.

McEntee, K., Weinstock, G.M., and Lehman, I.R. (1979). Initiation of general recombination catalyzed *in vitro* by the *recA* protein of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *76*, 2615–2619.

McMilin, K.D., Stahl, M.M., and Stahl, F.W. (1974). Rec mediated recombinational hot spot activity in bacteriophage lambda. I. Hot spot activity associated with Spi-deletions and bio substitutions. Genetics *77*, 409–423.

Meddows, T.R., Savory, A.P., Grove, J.I., Moore, T., and Lloyd, R.G. (2005). RecN protein and transcription factor DksA combine to promote faithful recombinational repair of DNA double-strand breaks. Mol. Microbiol. *57*, 97–110.

Michaels, M.L., Miller, J.H., Tchou, J., and Grollman, A.P. (1992). A repair system for 8-oxo-7, 8-dihydrodeoxyguanine. Biochemistry *31*, 10964–10968.

Mitchell, L.A., Cai, Y., Taylor, M., Noronha, A.M., Chuang, J., Dai, L., and Boeke, J.D. (2013). Multichange isothermal mutagenesis: A new strategy for multiple site-directed mutations in plasmid DNA. ACS Synth. Biol. *2*, 473–477.

Moggs, J.G., Grandi, P., Quivy, J.-P., Jonsson, Z.O., Hubscher, U., Becker, P.B., and Almouzni, G. (2000). A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. Mol. Cell. Biol. *20*, 1206–1218.

Molineux, I.J., and Gefter, M.L. (1975). Properties of the *Escherichia coli* DNA-binding (unwinding) protein interaction with nucleolytic enzymes and DNA. J. Mol. Biol. *98*, 811–825.

Moore, J.K., and Haber, J.E. (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *16*, 2164–2173.

Morimatsu, K., and Kowalczykowski, S.C. (2014). RecQ helicase and RecJ nuclease provide complementary functions to resect DNA for homologous recombination. Proc. Natl. Acad. Sci. U. S. A. *111*, E5133–E5142.

Morita, T., Yoshimura, Y., Yamamoto, A., Murata, K., Mori, M., Yamamoto, H., and Matsushiro, A. (1993). A mouse homolog of the *Escherichia coli recA* and *Saccharomyces cerevisiae* RAD51 genes. Proc. Natl. Acad. Sci. U. S. A. *90*, 6577–6580.

Mosberg, J.A., Lajoie, M.J., and Church, G.M. (2010). Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. Genetics *186*, 791–799.

Muniyappa, K., and Radding, C.M. (1986). The homologous recombination system of phage lambda. Pairing activities of beta protein. J. Biol. Chem. *261*, 7472–7478.

Muniyappa, K., Shaner, S.L., Tsang, S.S., and Radding, C.M. (1984). Mechanism of the concerted action of *recA* protein and helix-destablizing proteins in homologous recombination. Proc. Natl. Acad. Sci. U. S. A. *81*, 2757–2761.

Murphy, K.C. (1991). Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of *Escherichia coli* RecBCD enzyme. J. Bacteriol. *173*, 5808–5821.

Murphy, K.C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. J. Bacteriol. *180*, 2063–2071.

Muyrers, J.P.P., Zhang, Y., Testa, G., and Stewart, A.F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. *27*, 1555–1557.

Myers, R.H., and Montgomery, D.C. (1995). Response surface methodology: Process and product optimization using designed experiments (New York: John Wiley & Sons).

Nassif, N., Penney, J., Pal, S., Engels, W.R., and Gloor, G.B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell. Biol. *14*, 1613–1625.

NIST/SEMATECH (2003). e-Handbook of statistical methods.

Noirot, P., and Kolodner, R.D. (1998). DNA strand invasion promoted by *Escherichia coli* RecT protein. J. Biol. Chem. *273*, 12274–12280.

Oldenburg, K.R., Vo, K.T., Michaelis, S., and Paddon, C. (1997). Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. Nucleic Acids Res. *25*, 451–452.

Ozgenc, A.I., Szekeres, E.S., and Lawrence, C.W. (2005). *In vivo* evidence for a *recA*-independent recombination process in *Escherichia coli* that permits completion of replication of DNA containing UV damage in both strands. J. Bacteriol. *187*, 1974–1984.

Persky, N.S., and Lovett, S.T. (2008). Mechanisms of recombination: Lessons from *E. coli*. Crit. Rev. Biochem. Mol. Biol. *43*, 347–370.

Phillips, G.J., Prasher, D.C., and Kushner, S.R. (1988). Physical and biochemical characterization of cloned *sbcB* and *xonA* mutations from *Escherichia coli* K-12. J. Bacteriol. *170*, 2089–2094.

Ponticelli, A.S., Schultz, D.W., Taylor, A.F., and Smith, G.R. (1985). Chi-dependent DNA strand cleavage by RecBC enzyme. Cell *41*, 145–151.

Resnick, M.A. (1976). The repair of double-strand breaks in DNA: A model involving recombination. J. Theor. Biol. *59*, 97–106.

Robinson, S.J., Tang, L.H., Mooney, B.A., McKay, S.J., Clarke, W.E., Links, M.G., Karcz, S., Regan, S., Wu, Y.Y., Gruber, M.Y., et al. (2009). An archived activation tagged population of *Arabidopsis thaliana* to facilitate forward genetics approaches. BMC Plant Biol. *9*.

Rocha, E.P.C., Cornet, E., and Michel, B. (2005). Comparative and evolutionary analysis of the bacterial homologous recombination systems. PLoS Genet. *1*, 0247–0259.

Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. *10*, 866–876.

Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: Integrative DNA transformation in yeast. Methods Enzymol. *194*, 281–301.

Roy, R., Kozlov, A.G., Lohman, T.M., and Ha, T. (2009). SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. Nature *461*, 1092–1097.

Rupp, W.D., and Howard-flanders, P. (1968). Discontinuities in the DNA synthesized in an excision-defective strain of *Escherichia coli* following ultraviolet irradiation. J. Mol. Biol. *31*, 291–304.

Rupp, W.D., Wilde, C.E., Reno, D.L., and Howard-Flanders, P. (1971). Exchanges between DNA strands in ultraviolet-irradiated *Escherichia coli*. J. Mol. Biol. *61*, 25–44.

Ryder, L., Sharpies, G.J., and Lloyd, R.G. (1996). Recombination-dependent growth in exonuclease-depleted *recBC sbcBC* strains of *Escherichia coli* K-12. Genetics *143*, 1101–1114.

Sakai, A., and Cox, M.M. (2009). RecFOR and RecOR as distinct RecA loading pathways. J. Biol. Chem. *284*, 3264–3272.

Sandigursky, M., Mendez, F., Bases, R.E., Matsumoto, T., and Franklin, W.A. (1996). Protein-protein interactions between the *Escherichia coli* single-stranded DNA-binding protein and exonuclease I. Radiat. Res. *145*, 619.

Scheller, R.H., Dickerson, R.E., Boyer, H.W., Riggs, A.D., and Itakura, K. (1977). Chemical synthesis of restriction enzyme recognition sites useful for cloning. Science (80-.). *196*, 177–180.

Scholz, J., Besir, H., Strasser, C., and Suppmann, S. (2013). A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. BMC Biotechnol.

Seitz, E.M., Brockman, J.P., Sandler, S.J., Clark, A.J., and Kowalczykowski, S.C. (1998). RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. Genes Dev. *12*, 1248–1253.

Seitz, E.M., Haseltine, C.A., and Kowalczykowski, S.C. (2001). DNA recombination and repair in the Archaea. Adv. Appl. Microbiol. *50*, 101–169.

Shan, Q., Bork, J.M., Webb, B.L., Inman, R.B., and Cox, M.M. (1997). RecA protein filaments: End-dependent dissociation from ssDNA and stabilization by RecO and RecR proteins. J. Mol. Biol. *265*, 519–540.

Shao, Z., Zhao, H., and Zhao, H. (2009). DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res. *37*.

Shibata, T., Cunningham, R.P., DasGupta, C., and Radding, C.M. (1979a). Homologous pairing in genetic recombination: Complexes of *recA* protein and DNA. Proc. Natl. Acad. Sci. U. S. A. *76*, 5100–5104.

Shibata, T., DasGupta, C., Cunningham, R.P., and Radding, C.M. (1979b). Purified *Escherichia coli recA* protein catalyzes homologous pairing of superhelical DNA and single-stranded fragments. Proc. Natl. Acad. Sci. U. S. A. *76*, 1638–1642.

Shibata, T., DasGupta, C., Cunningham, R.P., Williams, J.G., Osber, L., and Radding, C.M. (1981). Homologous pairing in genetic recombination. The pairing reaction catalyzed by *Escherichia coli recA* protein. J. Biol. Chem. *256*, 7565–7572.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992). RAD51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. Cell *69*, 457–470.

Shulman, M.J., Hallick, L.M., Echols, H., and Signer, E.R. (1970). Properties of recombination-deficient mutants of bacteriophage lambda. J. Mol. Biol. *52*, 501–520.

Silva, D., Santos, G., Barroca, M., and Collins, T. (2017). Inverse PCR for point mutation introduction. In Methods in Molecular Biology, pp. 87–100.

Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C., and Wigley, D.B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. Nature *432*, 187–193.

Smith, G.R. (1991). Conjugational recombination in *E. coli*: Myths and mechanisms. Cell *64*, 19–27.

Smith, G.R., Amundsen, S.K., and Chaudhury, A.M. (1984). Roles of RecBC enzyme and chi sites in homologous recombination. Cold Spring Harb. Symp. Quant. Biol. *49*, 485–495.

Sormanni, P., Aprile, F.A., and Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. J. Mol. Biol. *427*, 478–490.

Sormanni, P., Amery, L., Ekizoglou, S., Vendruscolo, M., and Popovic, B. (2017). Rapid and accurate *in silico* solubility screening of a monoclonal antibody library. Sci. Rep. *7*.

Stewart, A.F., Zhang, Y., and Muyrers, J.P.P. (1999). Methods and compositions for directed cloning and subcloning using homologous recombination. US Patent No. US6355412B1. Retrieved from Google Patents.

Stewart, F., Zhang, Y., and Buchholz, F. (1997). DNA cloning method relying on the *E. coli* RecE/RecT recombination system. US Patent No. US6509156B1. Retrieved from Google Patents.

Stukenberg, P.T., Studwell-Vaughan, P.S., and O'Donnell, M. (1991). Mechanism of the sliding β-clamp of DNA polymerase III holoenzyme. J. Biol. Chem. *266*, 11328–11334.

Swingle, B., Markel, E., Costantino, N., Bubunenko, M.G., Cartinhour, S., and Court, D.L. (2010). Oligonucleotide recombination in Gram-negative bacteria. Mol. Microbiol. *75*, 138–148.

Symington, L.S., Rothstein, R., and Lisby, M. (2014). Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. Genetics *198*, 795–835.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25–35.

Taniguchi, N., Nakayama, S., Kawakami, T., and Murakami, H. (2013). Patch cloning method for multiple site-directed and saturation mutagenesis. BMC Biotechnol. *13*.

Taylor, A.F., and Smith, G.R. (1985). Substrate specificity of the DNA unwinding activity of the RecBC enzyme of *Escherichia coli*. J. Mol. Biol. *185*, 431–443.

Taylor, A.F., and Smith, G.R. (2003). RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. Nature *423*, 889–893.

Thoms, B., and Wackernagel, W. (1998). Interaction of RecBCD enzyme with DNA at double-strand breaks produced in UV-irradiated *Escherichia coli*: Requirement for DNA end processing. J. Bacteriol. *180*, 5639–5645.

Thoms, B., Borchers, I., and Wackernagel, W. (2008). Effects of single-strand DNases ExoI, RecJ, ExoVII, and SbcCD on homologous recombination of recBCD+ strains of *Escherichia coli* and roles of SbcB15 and XonA2 ExoI mutant enzymes. J. Bacteriol. *190*, 179–192.

Thresher, R.J., Christiansen, G., and Griffith, J.D. (1988). Assembly of presynaptic filaments. Factors affecting the assembly of RecA protein onto single-stranded DNA. J. Mol. Biol. *201*, 101–113.

Tseng, Y.C., Hung, J.L., and Wang, T.C.V. (1994). Involvement of RecF pathway recombination genes in postreplication repair in UV-irradiated *Escherichia coli* cells. Mutat. Res. Repair *315*, 1–9.

Tsukamoto, M., Yamashita, K., Miyazaki, T., Shinohara, M., and Shinohara, A. (2003). The N-terminal DNA-binding domain of RAD52 promotes RAD51-independent recombination in *Saccharomyces cerevisiae*. Genetics *165*, 1703–1715.

Turanli-Yildiz, B., Alkim, C., and Petek, Z. (2012). Protein engineering methods and applications. In Protein Engineering, p.

Umezu, K., and Kolodner, R.D. (1994). Protein interactions in genetic recombination in *Escherichia coli*. Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. J. Biol. Chem. *269*, 30005–30013.

Umezu, K., Nakayama, K., and Nakayama, H. (1990). *Escherichia coli* RecQ protein is a DNA helicase. Proc. Natl. Acad. Sci. U. S. A. *87*, 5363–5367.

Umezu, K., Chi, N.W., and Kolodner, R.D. (1993). Biochemical interaction of the *Escherichia coli* RecF, RecO, and RecR proteins with RecA protein and single-stranded DNA binding protein. Proc. Natl. Acad. Sci. U. S. A. *90*, 3875–3879.

Wang, Y. (2000). Nucleic acid modifying enzymes. US Patent No. US6627424B1. Retrieved from Google Patents.

Wang, J., Sarov, M., Rientjes, J., Fu, J., Hollak, H., Kranz, H., Xie, W., Stewart, A.F., and Zhang, Y. (2006). An improved recombineering approach by adding RecA to λ red recombination. Mol. Biotechnol. *32*, 43–53.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. Nucleic Acids Res. *32*, 1197–1207.

Watt, V.M., Ingles, C.J., Urdea, M.S., and Rutter, W.J. (1985). Homology requirements for recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *82*, 4768–4722.

Webb, B.L., Cox, M.M., and Inman, R.B. (1995). An interaction between the *Escherichia coli* RecF and RecR proteins dependent on ATP and double-stranded DNA. J. Biol. Chem. *270*, 31397–31404.

Webb, B.L., Cox, M.M., and Inman, R.B. (1997). Recombinational DNA repair: The RecF and RecR proteins limit the extension of RecA filaments beyond single-strand DNA gaps. Cell *91*, 347–356.

Wiktor, J., Van Der Does, M., Büller, L., Sherratt, D.J., and Dekker, C. (2018). Direct observation of end resection by RecBCD during double-stranded DNA break repair *in vivo*. Nucleic Acids Res. *46*, 1821–1833.

Willetts, N.S., Clark, A.J., and Low, B. (1969). Genetic location of certain mutations conferring recombination deficiency in *Escherichia coli*. J. Bacteriol. *97*, 244–249.

Wong, T., Zhurina, D., and Schwaneberg, U. (2006). The diversity challenge in directed protein evolution. Comb. Chem. High Throughput Screen. *9*, 271–288.

Xia, Y., Chu, W., Qi, Q., and Xun, L. (2015). New insights into the QuikChangeTM process guide the use of Phusion DNA polymerase for site-directed mutagenesis. Nucleic Acids Res. *43*, e12.

Yamaguchi, H., Hanada, K., Asami, Y., Kato, J.I., and Ikeda, H. (2000). Control of genetic stability in *Escherichia coli*: The *sbcB* 3'-5' exonuclease suppresses illegitimate recombination promoted by the RecE 5'-3' exonuclease. Genes to Cells *5*, 101–109.

Yu, M., Souaya, J., and Julin, D.A. (1998). The 30-kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *95*, 981–986.

Zhang, Y., Buchholz, F., Muyrers, J.P.P., and Francis Stewart, A. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. Nat. Genet. *20*, 123–128.

Zhang, Y., Muyrers, J.P.P., Testa, G., and Stewart, A.F. (2000a). DNA cloning by homologous recombination in *Escherichia coli*. Nat. Biotechnol. *18*, 1314–1317.

Zhang, Y., Werling, U., and Edelmann, W. (2012). SLiCE: A novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. *40*.

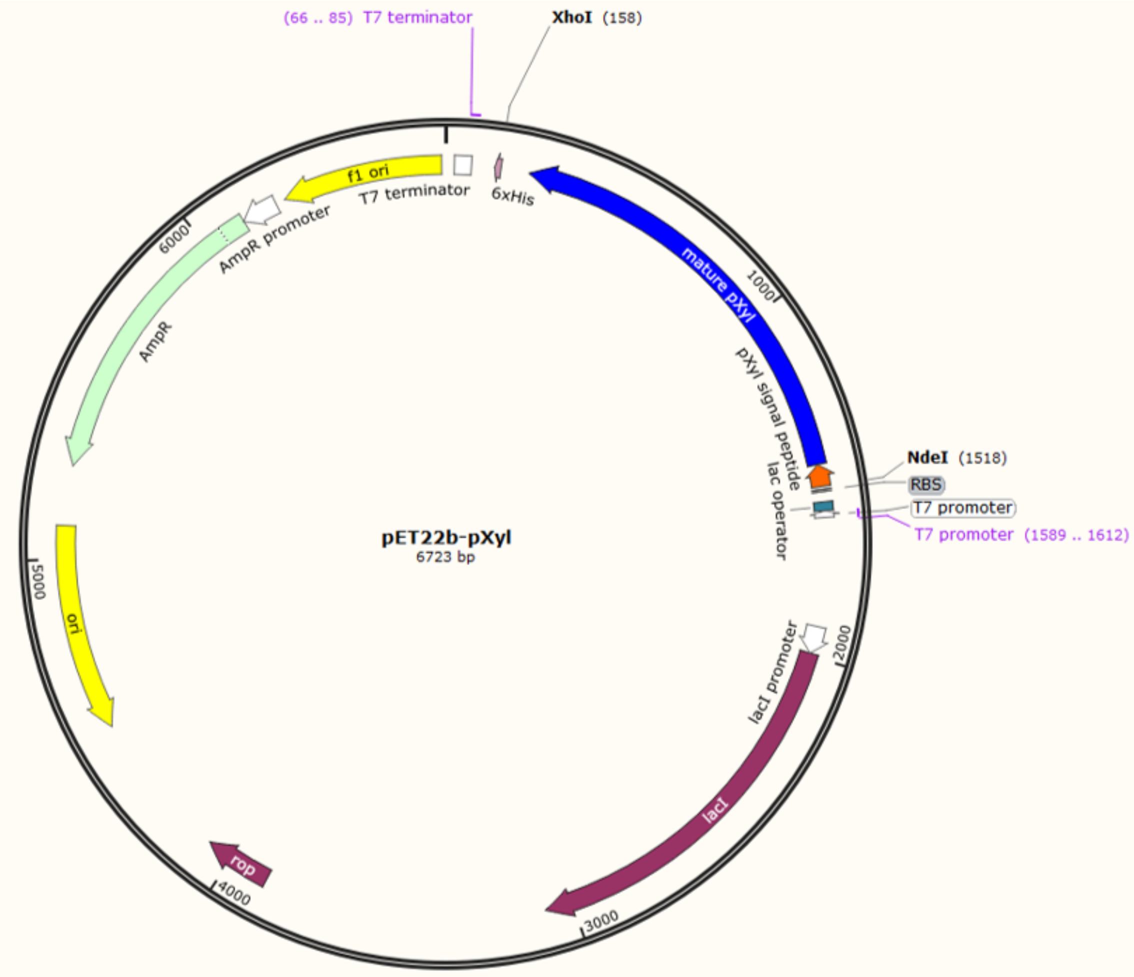
Zhang, Z., Shibahara, K.I., and Stillman, B. (2000b). PCNA connects DNA replication to epigenetic inheritance in yeast. Nature *408*, 221–225.

Zheng, L., Baumann, U., and Reymond, J.L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res. *32*.

Zoller, M.J. (1991). New molecular biology methods for protein engineering. Curr. Opin. Struct. Biol. *1*, 605–610.

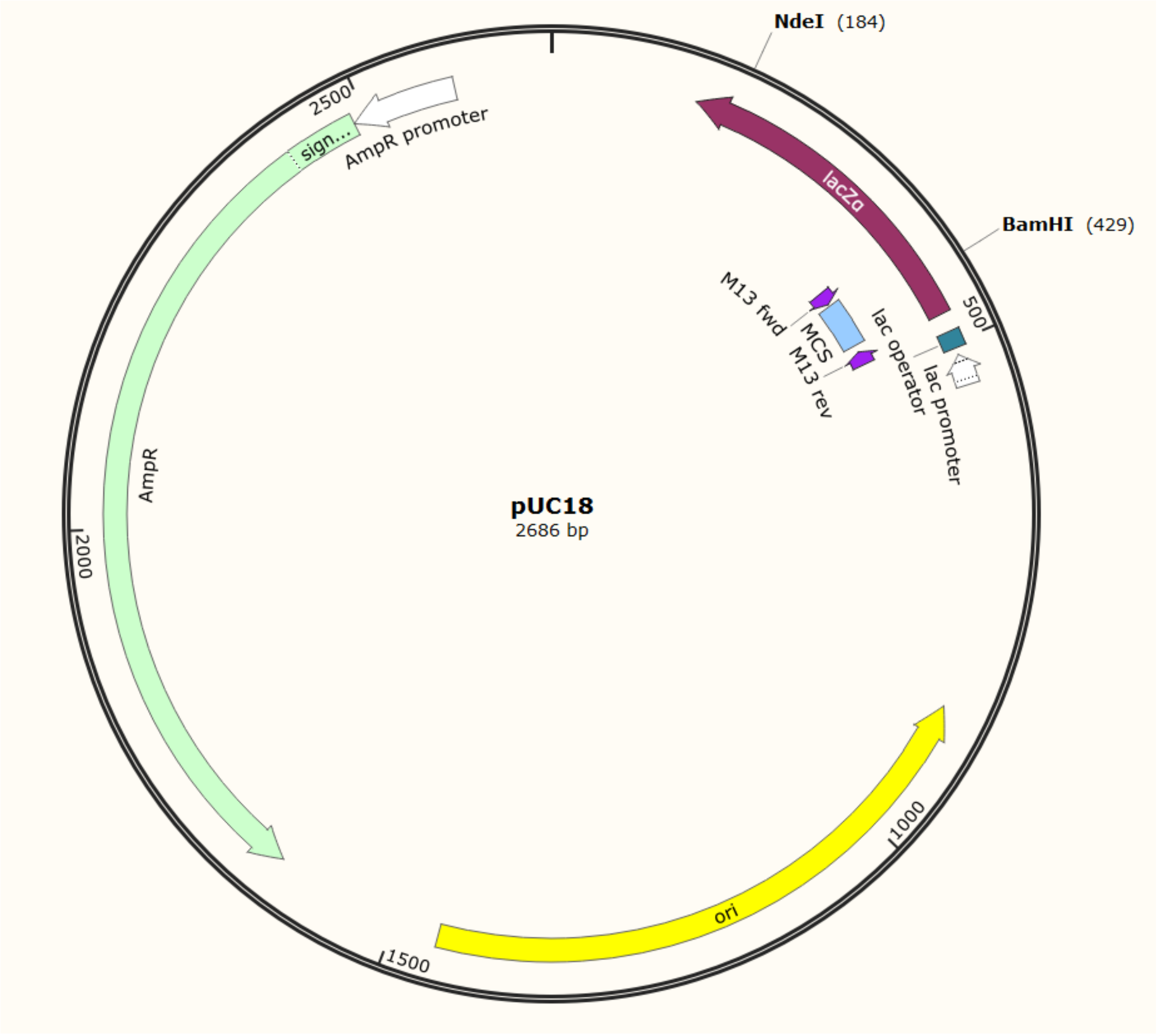
Annexe I – Map of Construct pET22b-pXyl

Gene sequence of *Pseudoalteromonas haloplanktis* xylanase (pXyl, European Nucleotide Archive code AJ427921), including signal peptide and mature sequence, is inserted in the *Nde*I and *Xho*I sites of the multiple cloning site (MCS) of the 5363 bp expression vector pET22b(+) (Novagen). The natural stop codon of pXyl was included and no tags were added. Construct contains the resistance gene for ampicillin (amp^{R}), an origin of replication (Ori-orp), lacI repressor gene (*lacI*), multiple cloning site (MCS) and lac promoter and operator.


(66 .. 85) T7 terminator
XhoI (158)
f1 ori
T7 terminator
6xHis
AmpR promoter
6000
AmpR
mature pXyl
1000
pXyl signal peptide
lac operator
NdeI (1518)
RBS
T7 promoter
T7 promoter (1589 .. 1612)
pET22b-pXyl
6723 bp
5000
ori
lacI promoter
2000
lacI
rop
4000
3000

Annexe II – Map of Cloning Vector pUC18

Map of vector pUC18 showing *Nde*I and *Bam*HI restriction digestion sites used in plasmid digestion. Ampicillin resistance gene (amp^R), origin of replication (Ori), multiple cloning site (MCS) and lac promoter and operator are also shown.


NdeI (184)
BamHI (429)
2500
sign...
AmpR promoter
lacZα
M13 fwd
MCS
M13 rev
lac operator
lac promoter
500
AmpR
pUC18
2686 bp
2000
1000
ori
1500

Annexe III – Primers

List of all primers used in the study, with specification of primer length, melting temperature (*Tm*, calculated via the Thermo Fisher Scientific website https://www.thermofisher.com/pt/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html), GC content and sequence. All primers were obtained from Eurofins Genomics and target the pET22b-pXyl construct. Mutated nucleotides are shown in bold.

Name	Length (nt)	*Tm* (°C)	GC (%)	Sequence
Y43E.Fwd	26	59.6	38	CAA TAT GTT TGG C**GA A**AA CAA CAC AC
Y43E.Rev	26	59.6	38	GTG TGT TGT T**TT C**GC CAA ACA TAT TG
Y315R.Fwd	24	57.8	42	GGG CTT TTT AAG T**CG** TGC AAA AAC
Y315R.Rev	24	57.8	42	GTT TTT GCA **CG**A CTT AAA AAG CCC
T403R.Fwd	31	51.6	23	GTT TTA TAA CAA C**CG C**TT TAA TTA AAA TAA C
T403R.Rev	31	51.6	23	GTT ATT TTA ATT AAA **GCG** GTT GTT ATA AAA C
Y194S.Fwd	24	58.5	42	GCT TTT CAC CTT **C**CA TTG ATA ACC
Y194S.Rev	24	58.5	42	GGT TAT CAA TG**G** AAG GTG AAA AGC
A272S.Fwd	29	65.5	45	CAT TTT TAA CGG CCA A**TC T**AA CCC AGG TC
A272S.Rev	29	65.5	45	GAC CTG GGT T**AG A**TT GGC CGT TAA AAA TG
Y378S.Fwd	27	63.2	48	CAA CGG GTG ACT **C**CC GTT ACT ATA ATG
Y378S.Rev	27	63.2	48	CAT TAT AGT AAC GG**G** AGT CAC CCG TTG
Y194S.Fwd_18	18	49.3	44	CAC CTT **C**CA TTG ATA ACC
Y194S.Rev_18	18	49.3	44	GGT TAT CAA TG**G** AAG GTG
A272S.Fwd_18	18	56.5	50	AAC GGC CAA **TCT** AAC CCA
A272S.Rev_18	18	56.5	50	TGG GTT **AGA** TTG GCC GTT
Y378S.Fwd_18	18	54.8	56	GGG TGA CT**C** CCG TTA CTA
Y378S.Rev_18	18	54.8	56	TAG TAA CGG **G**AG TCA CCC
Y194S.Fwd_25	25	59.9	40	GCT TTT CAC CTT **C**CA TTG ATA ACC T
Y194S.Rev_25	25	59.9	40	AGG TTA TCA ATG **G**AA GGT GAA AAG C
A272S.Fwd_25	25	64.1	48	TTT AAC GGC CAA **TCT** AAC CCA GGT C
A272S.Rev_25	25	64.1	48	GAC CTG GGT T**AG A**TT GGC CGT TAA A
Y378S.Fwd_25	25	62.1	48	ACG GGT GAC T**C**C CGT TAC TAT AAT G
Y378S.Rev_25	25	62.1	48	CAT TAT AGT AAC GG**G** AGT CAC CCG T
Y194S.Fwd_35	35	67.2	40	AAT CCG CTT TTC ACC TT**C** CAT TGA TAA CCT AAC AG
Y194S.Rev_35	35	67.2	40	CTG TTA GGT TAT CAA TG**G** AAG GTG AAA AGC GGA TT
A272S.Fwd_35	35	70.0	43	CAT TTT TAA CGG CCA A**TC T**AA CCC AGG TCA ATG GT
A272S.Rev_35	35	70.0	43	ACC ATT GAC CTG GGT T**AG A**TT GGC CGT TAA AAA TG
Y378S.Fwd_35	35	69.6	49	CAA CCA ACG GGT GAC T**C**C CGT TAC TAT AAT GGT TC
Y378S.Rev_35	35	69.6	49	GAA CCA TTA TAG TAA CGG **G**AG TCA CCC GTT GGT TG
Y194S.Fwd_45	45	71.3	42	CCA AAT AAT CCG CTT TT**C** ACC TTC CAT TGA TAA CCT AAC AGA CCC
Y194S.Rev_45	45	71.5	42	GGG TCT GTT AGG TTA TCA ATG **G**AA GGT GAA AAG CGG ATT ATT TGG
A272S.Fwd_45	45	73.3	42	GGC TAC ATT TTT AAC GGC CAA **TCT** AAC CCA GGT CAA TGG TAT GAA
A272S.Rev_45	45	73.3	42	TTC ATA CCA TTG ACC TGG GTT **AGA** TTG GCC GTT AAA AAT GTA GCC
Y378S.Fwd_45	45	73.1	47	GTC TTT ATC GCA ACC AAC GGG TGA CT**C** CCG TTA CTA TAA TGG TTC
Y378S.Rev_45	45	73.1	47	GAA CCA TTA TAG TAA CGG **G**AG TCA CCC GTT GGT TGC GAT AAA GAC
Y194S.Fwd_52	52	73.3	40	GGA AAA CCA AAT AAT CCG CTT TT**C** ACC TTC CAT TGA TAA CCT AAC AGA CCC T
Y194S.Rev_52	52	73.5	40	AGG GTC TGT TAG GTT ATC AAT G**G**A AGG TGA AAA GCG GAT TAT TTG GTT TTC C
A272S.Fwd_52	52	73.7	40	GGC TAC ATT TTT AAC GGC CAA **TCT** AAC CCA GGT CAA TGG TAT GAA TTT GAT G
A272S.Rev_52	52	73.7	40	CAT CAA ATT CAT ACC ATT GAC CTG GGT T**AG A**TT GGC CGT TAA AAA TGT AGC C
Y378S.Fwd_52	52	75.4	44	TTG GTC TTT ATC GCA ACC AAC GGG TGA CT**C** CCG TTA CTA TAA TGG TTC GTT A
Y378S.Rev_52	52	75.4	44	TAA CGA ACC ATT ATA GTA ACG G**G**A GTC ACC CGT TGG TTG CGA TAA AGA CCA A
T7 promoter	24	54.5	33	GAA ATT AAT ACG ACT CAC TAT AGG
T7 terminator	20	55.4	40	TAT GCT AGT TAT TGC TCA GC
T7 terminator (short)	18	55.2	50	CTA GTT ATT GCT CAG CGG

Annexe IV – PCR Conditions used with Phusion® High-Fidelity Polymerase

Description of PCR conditions used to amplify each specific fragment for recombination experiments. Name of the fragment, primers used, final primer concentration, total template amount and annealing temperature are specified.

Fragment	Primers	Final primer concentration (µM)	Total template (ng)	Annealing T° (°C)	Fragment	Primers	Final primer concentration (µM)	Total template (ng)	Annealing T° (°C)
F1	Y43E.Fwd and Y315R.Rev	1	7	64.7	F5_35	A272S.Fwd_35 and Y378S.Rev_35	1	0.1	*, **
F1	Y43E.Fwd and Y315R.Rev	1	0.7	64.7	F6_35	Y378S.Fwd_35 and Y194S.Rev_35	1	0.1	*, **
F1	Y43E.Fwd and Y315R.Rev	1	0.07	64.7	F4_35	Y194S.Fwd_35 and A272S.Rev_35	1	3.4	*, **
F2	Y315R.Fwd and T403R.Rev	1	7	61.5	F5_35	A272S.Fwd_35 and Y378S.Rev_35	1	3.4	*, **
F2	Y315R.Fwd and T403R.Rev	1	0.7	61.5	F6_35	Y378S.Fwd_35 and Y194S.Rev_35	1	3.4	*, **
F2	Y315R.Fwd and T403R.Rev	1	0.07	61.5	F4_35	Y194S.Fwd_35 and A272S.Rev_35	1	8.1	*, **
F3	T403R.Fwd and Y43E.Rev	1	7	61.5	F5_35	A272S.Fwd_35 and Y378S.Rev_35	1	8.1	*, **
F3	T403R.Fwd and Y43E.Rev	1	0.7	59.7	F4_45	Y194S.Fwd_45 and A272S.Rev_45	1	0.6	*, **
F3	T403R.Fwd and Y43E.Rev	1	0.07	59.7	F5_45	A272S.Fwd_45 and Y378S.Rev_45	1	0.6	*, **
F4_18	Y194S.Fwd_18 and A272S.Rev_18	1	3.4	56.4 **	F6_45	Y378S.Fwd_45 and Y194S.Rev_45	1	0.6	*, **
F5_18	A272S.Fwd_18 and Y378S.Rev_18	1	3.4	39.7 **	F4_45	Y194S.Fwd_45 and A272S.Rev_45	1	6.2	*, **
F6_18	Y378S.Fwd_18 and Y194S.Rev_18	1	3.4	53.5 **	F5_45	A272S.Fwd_45 and Y378S.Rev_45	1	6.2	*, **
F4_25	Y194S.Fwd_25 and A272S.Rev_25	0.5	0.6	63.4 **	F6_45	Y378S.Fwd_45 and Y194S.Rev_45	1	6.2	*, **
F5_25	A272S.Fwd_25 and Y378S.Rev_25	0.5	0.6	64.6 **	F4_52	Y194S.Fwd_52 and A272S.Rev_52	1	3.4	*, **
F6_25	Y378S.Fwd_25 and Y194S.Rev_25	0.5	0.6	66.1 **	F5_52	A272S.Fwd_52 and Y378S.Rev_52	1	3.4	*, **
F4_25	Y194S.Fwd_25 and A272S.Rev_25	0.5	6.2	63.4 **	F6_52	Y378S.Fwd_52 and Y194S.Rev_52	0.5	3.4	*, **
F5_25	A272S.Fwd_25 and Y378S.Rev_25	0.5	6.2	64.6 **	F4_opt	Y194S.Fwd_52 and A272S.Rev_52	1	0.1	*, **
F6_25	Y378S.Fwd_25 and Y194S.Rev_25	0.5	6.2	66.1 **	F5_opt	A272S.Fwd_52 and Y378S.Rev_52	1	0.1	*, **
F4_35	Y194S.Fwd_35 and A272S.Rev_35	1	0.1	*, **	F6_opt	Y378S.Fwd_52 and Y194S.Rev_52	0.5	0.1	*, **

***** - no annealing step was used, instead a 2 step PCR was performed

****** - conditions used for the RSM-CCD experiment

Annexe V – PCR Conditions used with NZYProof Polymerase

Description of PCR conditions used to amplify each specific fragment for recombination experiments. Name of the fragment, primers used, final primer concentration, total template amount and annealing temperature are specified.

Fragment	Primers	Final primer concentration (µM)	Total template (ng)	Annealing T° (°C)
F4	Y194S.Fwd and A272S.Rev	1	7	56.1
F5	A272S.Fwd and Y378S.Rev	1	7	61.3
F6	Y378S.Fwd and Y194S.Rev	1	7	56.1
F4	Y194S.Fwd and A272S.Rev	1	0.7	56.1
F5	A272S.Fwd and Y378S.Rev	1	0.7	61.3
F6	Y378S.Fwd and Y194S.Rev	1	0.7	56.1