ENGINEERING THE GLYCOSIDE HYDROLASE $\beta$-GLUCURONIDASE ($\beta$-GUS) TO IMPROVE A NON-NATIVE ACTIVITY FOR THE SYNTHESIS OF O-GLUCURONIDES

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY OF THE AUSTRALIAN NATIONAL UNIVERSITY

Australian National University

SHU ANN CHAN

SUBMITTED FOR EXAMINATION IN June 2016
Declaration by author

This thesis presents an original research that I have carried out under the supervision of Professor David Ollis at the Research School of Chemistry, Australian National University (ANU). Contributions to this thesis are as follows:

- Plasmid of wild-type β-GUS and glucuronylsynthase in pET28a were provided by the McLeod research group, ANU
- pJ401 backbone vector was obtained from Dr. Bradley Stevenson from the McLeod research group, ANU.
- Substrates for glucuronylsynthase reactions, α-glucuronyl fluoride and acceptor steroids were synthesized by Mr. Paul Ma from the McLeod research group, ANU
- Scaled-up reactions of the glucuronylsynthase chemoenzymatic reactions were performed by Mr. Paul Ma from the McLeod research group, ANU

The material used for the publication of this thesis has not been presented for the award of any other degree. Every effort has been made to acknowledge previously published material.

This thesis conforms to the guidelines and regulations for Higher Degree Research award set by the Australian National University (ANU).

______________________
Shu Ann Chan
June 2016
Acknowledgements

I would like to thank Professor David Ollis for the opportunity to work on this interesting and challenging project. I would also like to express my appreciation for his support and patience throughout my PhD candidature at the Research School of Chemistry. He has been a wonderful mentor who has provided critical evaluation and constructive feedback on the experimental results that are presented in this thesis.

I would also like to take this opportunity to thank Assoc. Prof. Malcolm McLeod for all his assistance and advice on the glucuronylsynthase system. His knowledge of the enzyme has been very helpful in helping us understand the workings of the enzyme. I would also like to thank Mr. Paul Ma who has kindly synthesized the substrates that were used for the engineering of the glucuronylsynthase enzyme. He has always delivered the substrates cheerfully although the amount that was required for screening was not small.

I would like to extend my gratitude to Ms. Tracy Murray for her acumen in ensuring the smooth operation of the laboratory. Her help with the molecular biology techniques had also been vital to me when I first embarked on this project. I would like to thank Dr. Bradley Stevenson for all the useful discussions and feedback on the various topics related to enzymology and screening. I would also like to thank him for his help and expertise with the LC/MS instrument. In addition, he has always shown great willingness to discuss new ideas and give constructive criticism to improve these ideas. His input has been valuable and I truly appreciate them.

I thank all past and present members of the Ollis group especially Dr. Jian-Wei Liu, now at the CSIRO, who had spent his time discussing the PCR assembly method for library construction. Special thanks is also extended to Dr Tee-Kheang Ng for all the extensive discussions on molecular biology techniques and screening designs for directed evolution experiments. I also thank the other Biological Chemistry research groups at the RSC for all the help and co-operation rendered with the various instruments and shared facilities.
To all my friends, inside and outside the RSC, and those back at home, thank you for all the support and friendship. These interactions had provided relief and kept my sanity throughout these four years.

Most importantly, to my family members who have been supportive throughout the last few years, words alone cannot express my gratitude. This journey would not have been possible without the encouragement, love and support from you all.
Abstract

The demand for O-glucuronides as potential therapeutic products and biomarkers continue to increase. However, large-scale synthesis of O-glucuronides remains a challenge for the industry, and has prompted the development of an alternative synthetic routes. This dissertation extends from a previous enzyme engineering work that had introduced a site-specific mutation E504G in β-glucuronidase (β-GUS), resulting in a functional glucuronylsynthase (Syn). However, the synthetic activity of Syn is low and leaves ample scope for improvement. The work described in this thesis aims to produce a more efficient glucuronylsynthase using different enzyme engineering approaches.

Two separate strategies were employed to achieve our objective. The first strategy engages a two-step process where the β-GUS is first engineered to have higher activity in the presence of excess substrate; 10–20 times its $K_m$. This is followed by site-specific mutation E504G to convert the β-GUS variant into a Syn. The second strategy engineers the glucuronylsynthase directly.

Chapter 3 describes the attempt to improve the activity of the native enzyme in the presence of $t$-BuOH, a solvent that was found to improve the chemistry of the glucuronylsynthase chemoenzymatic reaction. The engineering attempt produced a potential variant with a mutation at its C-terminal region, L561S, that is more active in the presence of the solvent. This mutation appears to be a determinant mutation. Biophysical characterization of the enzyme revealed that this improvement is not due to increased stability in $t$-BuOH, while our analysis of the crystal structure suggests that the mutation improved the activity by increasing loop flexibility at the C-terminal region. Subsequently, I incorporated E504G into the β-GUS variant, but this did not translate into a better glucuronylsynthase variant. Chapter 4 describes the second strategy.

Two mutations, H162Q and Y160G, at the N-terminal region were found to boost the synthetic activity but this was not accompanied by improvement in their thermostability nor solvent stability. However, combining the results from the biophysical characterization experiments and observations from the structural
examination on 3K4D, it can be inferred that the mutations promote glucuronylsynthase activity by modulating the active site of the Syn so that it would favour the glucuronyl donor substrate. Therefore, these mutations would serve as concrete starting points for further evolution program of the Syn.

Chapter 5 explores the potential reason that could account for the lack of success in transposing the potency of L561S to the glucuronylsynthase system. The work here is driven by the hypothesis that translational misincorporation introduced contaminating wild-type during enzyme expression. Essentially, this chapter highlights the potential pitfall of the glucuronylsynthase system and describes potential strategies to avoid this pitfall. Finally, Chapter 6 builds upon the results from β-GUS engineering and explores the mutational tolerance of β-GUS. Its mutational tolerance is compared with another enzyme that is structurally less complex (β-lactamase, TEM-1). In addition, its mutational tolerance with different substrate is also compared. This exercise attempts to provide insights into the elements that would drive the adaption process of the β-GUS. Consequently, we expect this study to facilitate future directed evolution studies of the β-GUS and the glucuronylsynthase.
# Table of Contents

Declaration by author ........................................................................................................ ii
Acknowledgements ........................................................................................................ iii
Abstract ............................................................................................................................ v
Table of Contents .......................................................................................................... vii
List of figures ................................................................................................................... xi
List of tables .................................................................................................................... xiv
Abbreviations and acronyms ........................................................................................... xv
Symbols and units .......................................................................................................... xvii

Nomenclature .................................................................................................................... N:1

1. Introduction ............................................................................................................. 1
   1.1 Preamble ............................................................................................................. 1
   1.2 Background ....................................................................................................... 1
      1.2.1 Glucuronide synthesis in the industry ................................................... 3
   1.3 Enzymes in the thesis ...................................................................................... 4
      1.3.1 The various forms of β-glucuronidase (β-GUS) ..................................... 4
      1.3.2 Acquisition of the glucuronylsynthase activity ..................................... 5
   1.4 Specific aims of the project ............................................................................. 7
   1.5 Basis and hypothesis for the two-step strategy ............................................... 9
   1.6 Research design and approaches .................................................................. 10
      1.6.1 A review on protein engineering ......................................................... 11
         1.6.1.1 Rational and semi-rational approaches ............................................. 11
         1.6.1.2 The evolutionary approach .............................................................. 12
      1.6.2 Combining protein engineering strategies for β-GUS evolution ........... 14
   1.7 Overview of the thesis ..................................................................................... 18
   1.8 References ...................................................................................................... 21

2. Materials and Methods ......................................................................................... 27
   2.1 Preamble .......................................................................................................... 27
   2.2 Materials .......................................................................................................... 27
   2.3 Bacterial strains .............................................................................................. 28
      2.3.1 Preparation of electrocompetent cells .................................................. 29
      2.3.2 Electrotransformation .......................................................................... 30
   2.4 Plasmids and vectors ..................................................................................... 30
      2.4.1 Description of plasmids ........................................................................ 31
      2.4.2 General restriction cloning procedures .............................................. 32
      2.4.3 Cell culture and isolation of plasmid DNA .......................................... 33
      2.4.4 DNA visualization and quantification ............................................... 33
      2.4.5 Construction of pJ401/gusA ................................................................. 34
      2.4.6 Construction of pJ401/syn ................................................................. 35
   2.5 Polymerase chain reaction (PCR) ................................................................. 36
      2.5.1 Primer design ....................................................................................... 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.2 Standard polymerase chain reaction (PCR)</td>
<td>36</td>
</tr>
<tr>
<td>2.5.3 DNA sequencing</td>
<td>37</td>
</tr>
<tr>
<td>2.6 Library construction and enzyme engineering methods</td>
<td>38</td>
</tr>
<tr>
<td>2.6.1 Site-specific mutagenesis</td>
<td>38</td>
</tr>
<tr>
<td>2.6.2 Site saturation and site directed mutagenesis</td>
<td>39</td>
</tr>
<tr>
<td>2.6.3 Error-prone PCR (ePCR)</td>
<td>40</td>
</tr>
<tr>
<td>2.6.4 Staggered extension process (StEP)</td>
<td>41</td>
</tr>
<tr>
<td>2.6.5 Staggered extension process with error-prone steps (StEP-ePCR)</td>
<td>41</td>
</tr>
<tr>
<td>2.6.6 Assembly PCR for targeted random mutagenesis</td>
<td>42</td>
</tr>
<tr>
<td>2.6.7 E-transformation for library generation</td>
<td>44</td>
</tr>
<tr>
<td>2.6.8 Analysis of mutant library</td>
<td>44</td>
</tr>
<tr>
<td>2.7 Preparation of lysates and assays for β-GUS library screening</td>
<td>45</td>
</tr>
<tr>
<td>2.7.1 Cultivation and preparation of lysates for β-GUS screen</td>
<td>45</td>
</tr>
<tr>
<td>2.7.1.1 Single colony method</td>
<td>45</td>
</tr>
<tr>
<td>2.7.1.2 Culture dilution method</td>
<td>45</td>
</tr>
<tr>
<td>2.7.2 β-GUS blue-white screen</td>
<td>47</td>
</tr>
<tr>
<td>2.7.3 β-GUS buffer screening assay</td>
<td>47</td>
</tr>
<tr>
<td>2.7.4 β-GUS tert-butanol (t-BuOH) screening assay</td>
<td>48</td>
</tr>
<tr>
<td>2.7.5 β-GUS and β-glucosidase screening assay for mutational tolerance study</td>
<td>49</td>
</tr>
<tr>
<td>2.8 Preparation of lysates and assays for glucuronylsynthase library screening</td>
<td>49</td>
</tr>
<tr>
<td>2.8.1 Cultivation and preparation of lysates for glucuronylsynthase screen</td>
<td>49</td>
</tr>
<tr>
<td>2.8.2 Glucuronylsynthase tert-butanol (t-BuOH) screening assay</td>
<td>49</td>
</tr>
<tr>
<td>2.9 Translational misincorporation study: β-GUS assay to investigate effects of expression conditions</td>
<td>50</td>
</tr>
<tr>
<td>2.10 Protein expression and purification</td>
<td>51</td>
</tr>
<tr>
<td>2.10.1 Growth and protein overexpression for β-GUS recombinants</td>
<td>51</td>
</tr>
<tr>
<td>2.10.2 Growth and protein overexpression for glucuronylsynthase recombinants</td>
<td>52</td>
</tr>
<tr>
<td>2.10.3 Protein extraction for purification</td>
<td>52</td>
</tr>
<tr>
<td>2.10.4 Metal affinity purification</td>
<td>53</td>
</tr>
<tr>
<td>2.10.5 Size exclusion purification</td>
<td>54</td>
</tr>
<tr>
<td>2.10.6 Protein visualization and quantification</td>
<td>54</td>
</tr>
<tr>
<td>2.11 Enzyme characterization using activity assays</td>
<td>55</td>
</tr>
<tr>
<td>2.11.1 Enzyme kinetics of β-GUS and variants</td>
<td>56</td>
</tr>
<tr>
<td>2.11.2 Enzyme kinetics of β-GUS in t-BuOH</td>
<td>56</td>
</tr>
<tr>
<td>2.11.3 Solvent stability study for β-GUS recombinants</td>
<td>56</td>
</tr>
<tr>
<td>2.11.4 Enzyme kinetics of glucuronylsynthase and variants</td>
<td>57</td>
</tr>
<tr>
<td>2.11.5 t-BuOH stability study for glucuronylsynthase recombinants</td>
<td>57</td>
</tr>
<tr>
<td>2.11.6 Half-lives of gluconoylsynthase variants in t-BuOH</td>
<td>58</td>
</tr>
<tr>
<td>2.12 Biophysical and biochemical properties of enzymes</td>
<td>59</td>
</tr>
<tr>
<td>2.12.1 Thermal shift assay</td>
<td>59</td>
</tr>
<tr>
<td>2.12.2 Intrinsic fluorescence assay</td>
<td>59</td>
</tr>
<tr>
<td>2.12.3 Circular dichroism (CD) spectroscopy</td>
<td>60</td>
</tr>
<tr>
<td>2.12.4 Isothermal titration calorimetry (ITC)</td>
<td>60</td>
</tr>
<tr>
<td>2.13 Computational tools</td>
<td>61</td>
</tr>
<tr>
<td>2.14 References</td>
<td>62</td>
</tr>
</tbody>
</table>

3. Directed evolution of β-glucuronidase (β-GUS) ....................................... 65
   3.1 Preamble ........................................................................................................... 65
Table of contents

3.2 Library size considerations ................................................................................ 65
3.3 Results ............................................................................................................... 67
  3.3.1 Optimization and validation of screening methodology ....................... 67
  3.3.2 Stages of β-GUS evolution ...................................................................... 69
    3.3.2.1 Stage 1 of β-GUS evolution ............................................................. 69
    3.3.2.2 Second stage evolution .................................................................... 71
  3.3.3 Library Lp6 revisited .............................................................................. 73
  3.3.4 Kinetic parameters of mutants ................................................................. 75
  3.3.5 Stability study in various denaturants ..................................................... 79
3.4 Discussion ......................................................................................................... 80
  3.4.1 Structural rationalization of L561S and other variants ......................... 80
  3.4.2 Implication of screening and evolutionary strategies that were adopted 82
3.5 Concluding remarks ........................................................................................ 84
3.6 References ......................................................................................................... 85

4. Engineering the glucuronylsynthase ............................................................... 87
  4.1 Preamble ........................................................................................................... 87
  4.2 Setting the selection pressure for screens ..................................................... 88
  4.3 Method development ....................................................................................... 89
    4.3.1 General materials and methods ........................................................ 89
    4.3.2 Tailoring the reaction time and amount of crude lysates for screening ... 90
    4.3.3 Assay design for glucuronylsynthase screens .................................... 90
      4.3.3.1 Fluoride-selective indicatory displacement assay ............................. 91
      4.3.3.2 Developing the LCMS method for library screening ....................... 92
  4.4 Results and Discussion .................................................................................... 95
    4.4.1 Glucuronylsynthase rational engineering ........................................ 95
      4.4.1.1 Site saturation mutagenesis .............................................................. 96
      4.4.1.2 Disordered loop targeted random mutagenesis ................................. 99
    4.4.2 Kinetic properties of glucuronylsynthase and its variants ................... 100
    4.4.3 Stability of glucuronylsynthase and its variants ................................... 104
      4.4.3.1 Half-lives in t-BuOH and thermostability ...................................... 104
      4.4.3.2 Inactivation of glucuronylsynthase in the presence of t-BuOH and urea 107
    4.4.4 The trend in glycosyl binding properties of Syn and variants .......... 109
  4.5 Structural deduction of mutations Y160G and H162Q ............................ 112
  4.6 Conclusion and future directions ............................................................. 114
  4.7 References .............................................................................................. 116

5. Translational misincorporation and β-GUS contamination in the glucuronylsynthase system ................................................................. 118
  5.1 Introduction ..................................................................................................... 118
  5.2 Results ............................................................................................................ 119
    5.2.1 The catalytic nucleophile of E. coli β-GUS E504 ................................ 119
    5.2.2 Eliminating β-GUS contamination .......................................................... 120
    5.2.3 Codon usage of mutation E504G affects the extent of hydrolysis by glucuronylsynthase ................................................................. 122
    5.2.4 Minimizing hydrolysis using different expression conditions ......... 125
    5.2.5 Hydrolytic activity of the glucuronylsynthase on other glycosides .... 127
  5.3 Discussion ...................................................................................................... 130
  5.4 Conclusion .................................................................................................... 133
5.5 References ....................................................................................................... 134

6. Mutational tolerance of the *Escherichia coli* β-glucuronidase ..................... 137
6.1 Introduction ..................................................................................................... 137
6.2 Results and Discussion ..................................................................................... 140
  6.2.1 Random mutagenesis on the whole gene of β-GUS ..................................... 140
  6.2.2 Tolerance of active site residues towards mutations .................................. 141
  6.2.3 Developing a statistical approach to study the mutational tolerance of single-site saturation mutagenesis library ........................................... 144
  6.2.4 Quantifying the mutational tolerance of the single-site saturation mutagenesis library ................................................................. 145
  6.2.5 Verification of activity in mutants ................................................................ 147
  6.2.6 Limitations of the statistical method used for studying mutational tolerance of single-site saturation mutagenesis libraries ...................... 149
6.3 Discussion and concluding remarks ................................................................ 150
6.4 References ....................................................................................................... 154

7. Conclusion and Future Directions ................................................................ 156

General appendices
General appendices
Appendix A: Equipments and consumable suppliers ............................................. A:1
Appendix B: Chemical suppliers ........................................................................... A:3
Appendix C: Recipes of media and buffers ............................................................ A:4
Appendix D: Oligonucleotide primers ................................................................... A:7

Chapter specific appendices
Appendix : Chapter 2 .......................................................................................... A:10
  Appendix 2-1 .............................................................................................. A:10
Appendix : Chapter 3 .......................................................................................... A:11
  Appendix 3-1 .............................................................................................. A:11
  Appendix 3-2 .............................................................................................. A:12
  Appendix 3-3 .............................................................................................. A:15
  Appendix 3-4 .............................................................................................. A:16
Appendix : Chapter 4 .......................................................................................... A:17
  Appendix 4-1 .............................................................................................. A:17
  Appendix 4-2 .............................................................................................. A:18
  Appendix 4-3 .............................................................................................. A:19
  Appendix 4-4 .............................................................................................. A:20
  Appendix 4-5 .............................................................................................. A:21
Appendix : Chapter 5 .......................................................................................... A:22
  Appendix 5-1 .............................................................................................. A:22
  Appendix 5-2 .............................................................................................. A:23
  Appendix 5-3 .............................................................................................. A:24
Appendix : Chapter 6 .......................................................................................... A:25
  Appendix 6-1 .............................................................................................. A:25
  Appendix 6-2 .............................................................................................. A:27
  Appendix 6-3 .............................................................................................. A:28
  Appendix 6-4 .............................................................................................. A:29
  Appendix 6-5 .............................................................................................. A:34
  Appendix 6-6 .............................................................................................. A:37
  Appendix 6-7 .............................................................................................. A:37
List of figures

Figure N-1: Reactions that are discussed in this thesis are the hydrolysis and synthesis of glucuronides. ................................................................. N:1
Figure N-2: Components in a glucuronide. ............................................................. N:1
Figure N-3: Numbering conventions and derivation of the β-D-glucuronic acid (GlcA). ................................................................. N:2

Figure N-4: β-D-glucuronic acid, α-D-glucuronic acid and α-D-glucuronyl fluoride N:3
Figure N-5: The aglycones discussed in this thesis. ............................................................. N:4
Figure 1-1: The fate of xenobiotic and its metabolite. .................................................. 2
Figure 1-2: Scheme depicting enzymatic mechanisms. ............................................. 6
Figure 1-3: Examples of glucuronylsynthase reactions using engineered β-GUS (E504G) enzyme. ................................................................. 8

Figure 1-4 : Two distinct approaches for the engineering of β-GUS to obtain an improved glucuronylsynthase. .................................................. 8
Figure 1-5: Scheme of the reaction in β-GUS screening assay. ..................................... 9
Figure 1-6: Typical process in directed evolution experiments. .................................. 13
Figure 1-7: Structure of E. coli β-glucuronidase. .......................................................... 15
Figure 1-8: Structures of glucuronic acid and glucaro-δ-lactam. .................................... 16
Figure 1-9: Positions of prioritized sites for saturation mutagenesis and other binding sites that are important for glycosyl binding. ........................................... 16

Figure 1-10: View of Domain III (TIM-barrel) of β-GUS from the bottom................. 17
Figure 1-11: Targeted regions for semi-rational engineering and targeted directed evolution of β-GUS. ............................................................. 19

Figure 2-1: Knocking out the gusA gene in the BW25142 cell line. ......................... 29
Figure 2-2: Plasmid maps of pDNAs used for pJ401/gusA construction. ..................... 35
Figure 2-3: Schematic illustration of PCR approaches to generate random libraries.. 42
Figure 2-4: DNA visualization of PCR amplification products for targeted random mutagenesis. ................................................................. 44

Figure 2-5: Poisson distribution of cells in liquid culture when µ = 4. .......................... 46
Figure 3-1: Peptide fragments and protein feature view of the targeted fragments chosen for random mutagenesis. .................................................. 66
Figure 3-2: Investigating the interplate variation. ...................................................... 68
Figure 3-3: Discrimination of wells containing 2-fold improved hits over the expected average activity in a plate when pooling 3 or 4 cfu/well. ................... 69
Figure 3-4: Confirmation of potential hits from Stage I evolution................................ 71
Figure 3-5: The second stage of β-GUS evolution. ....................................................... 72
Figure 3-6: Variants from Round 5 evolution............................................................. 73
Figure 3-7: The performance of the variants. .............................................................. 76
Figure 3-8: The trend in evolutionary progress. .......................................................... 77
Figure 3-9: Thermal stability values of β-GUS variants. .......................................... 78
Figure 3-10: Heat map depicting the degree of enhancement conferred by mutations relative to wild-type...................................................... 80
Figure 3-11: Positions of the mutations in the crystal structure 3K4D............................ 81
Figure 3-12 Desired kinetic profiles from our hypothesis. ........................................... 83
Figure 4-1: Two engineering approaches used in attempts to improve the SynWT system. ................................................................. 87
Figure 4-2: The reaction involved in an indicator displacement assay.......................... 91
Figure 4-3: Quantifying the limit of detection in the fluorescence indicator displacement assay................................................................. 92
Figure 4-4: The pH values of indicator displacement assays before and after the addition of lysis buffer .......................................................... 93
Figure 4-5: Calibration curve of CMO-DHEA glucuronide for the quantification and validation of the analytical method ........................................... 95
Figure 4-6: Sites (turquoise and blue) targeted for single-site saturation mutagenesis are labeled ................................................................. 97
Figure 4-7: Screening of targeted random mutagenesis library .......................................................... 99
Figure 4-8: Plots of initial rates versus increasing CMO-DHEA concentration for Syn (blue) and SynH162Q (black) and the curve-fit parameters from applying the Michaelis-Menten model ........................................... 101
Figure 4-9: Kinetic characterization of Syn and Syn variants .......................................................... 102
Figure 4-10: Investigating the relative performance of Syn variants with different steroid acceptor substrates ................................................ 103
Figure 4-11: Alteration in the secondary structures of Syn and Syn H162Q monitored by circular dichroism (CD) ......................................................... 105
Figure 4-12: Effects of 1FGlcA on Tm .......................................................... 106
Figure 4-13: Inactivation of the enzyme solutions in t-BuOH and urea ................................................ 107
Figure 4-14: The dissociation constants (Kd) for Syn, Syn Y160G and Syn H162Q ................................................ 110
Figure 4-15: Trends in Kd and kcat/Km of the Syn variants .................................................. 112
Figure 4-16: Interactions between Y160 and S557 ........................................................................... 112
Figure 4-17: Position and spatial orientation of residue 162 and 163 in the crystal structure .......................... 113
Figure 5-1: Confirming E504G as the nucleophilic residue .......................................................... 121
Figure 5-2: Eliminating external sources as potential contaminant .............................................. 122
Figure 5-3: Kinetics of the hydrolytic activity exhibited by glucuronylsynthase encoded with rare and common codons ................................................. 124
Figure 5-4: Competing hydrolysis by glucuronylsynthase encoded rare codon (E504G) or common codon (E504G) .................................................. 125
Figure 5-5: Comparing the effects of different induction duration and temperatures on the β-GUS activity by glucuronylsynthase ....................... 126
Figure 5-6: The glycosyl rings of the three glycoside substrates studied ............................................... 128
Figure 5-7: Hydrolytic kinetic parameters of the glucuronylsynthase and the wild type with different glycosides ................................................. 129
Figure 6-1: Rate of fitness decline in response to increasing mutation rate μ for β-GUS follows an exponential function defined by Wn = K exp(-Dμ) ........................................ 141
Figure 6-2: Interactions between substrate-binding residues and pNPGlcA ........................................ 143
Figure 6-3: Interactions of several binding site residues in the β-GUS crystal structure ....................... 144
Figure 6-4: Simulation of activity distribution profile following data sorting of activity levels relative to wild-type in a library (hypothetical) ................. 146
Figure 6-5: Mutational tolerance of sites chosen for site-saturation mutagenesis study ......................... 147
Figure 6-6: Relative improvement of purified enzymes at selected substrate concentrations .......................................................... 148
Figure 6-7: Tm of T509A ........................................................................ 149
Figure A2-1: The standard curves used to calculate the extinction coefficient values for pNP in 50 mM NaPi, pH 7.4 .................................................. A:10
Figure A2-2: The standard curves used to calculate the extinction coefficient values for pNP in 50 mM NaPi, pH 7.4 containing 20% t-BuOH ......................... A:10
Figure A4-1: Optimization of the LC/MS ionization parameters .................................................. A:16
Figure A4-2: Resolution of the free CMO-DHEA and CMO-DHEA glucuronide peaks using only the guard column at two difference lysate loads. A:18
Figure A4-3: ITC binding isotherms of glucuronylsynthase and its variants (a) SynWT (b) SynH162Q (c) Syn Y160G. A:20
Figure A4-4: Reaction scheme for the glucuronidation of CMO-DHEA (k₁) and the competing hydrolytic reaction (k⁻₁). A:21
Figure A5-1: SDS-Page analysis of the expression levels of Syn expressed at different temperature and length of time on two different days (D1 and D2). A:22
Figure A6-1: Mutation rate increased in a linear manner in response to the amount of MnCl₂. A:26
Figure A6-2: Location of residues targeted for site-saturation mutagenesis in the E. coli β-GUS crystal structure (PDB ID: 3K46). A:27
Figure A6-3: Comparison of profile slopes (m4) with different number of data points. A:28
Figure A6-4: Activity distribution curves of site-saturation libraries from screening data with pNPglcA substrate (left) and pNPglc substrate (right). A:29
Figure A6-5: Michalis-Menten kinetic curves of purified enzymes (WT, M447G/S, R562S/T and T509A) from with pNPglcA substrate (left) and pNPglc substrate (right). A:34
Figure A6-6: Comparison of activity distribution data between raw screening data and a subset of screening data. A:37
Figure A6-7: Curve-fitting data to obtain input values for unpaired t-test analysis. A:38
Figure A6-8: The fitness decline of β-GUS fitted with an epistasis parameter (β). A:38
Figure A6-9: Interactions between domains and between subunits in β-GUS. A:39
List of tables

Table 1-1: Priority mutations to be considered for site-saturation mutagenesis that were derived from structural analyses .............................................................. 16
Table 2-1: Thermo cycling conditions for standard and analyses ........................................ 37
Table 2-2: Thermo cycling conditions for Sanger sequencing ........................................ 37
Table 2-3: Thermo cycling conditions for site-specific mutagenesis using Stratagene's QuikChange® .......................................................... 38
Table 2-4: Thermo cycling conditions for site-saturation and site-directed mutagenesis. .......... 40
Table 2-5: Thermo cycling conditions for StEP ................................................................. 41
Table 2-6: Thermo cycling conditions for StEP-ePCR ....................................................... 42
Table 2-7: Components of SDS-PAGE gel ................................................................. 54
Table 3-1: Libraries that were generated for the directed evolution of β-GUS ............ 70
Table 3-2: Some diversity of Lp6 potential hits ............................................................. 74
Table 3-3: Kinetic parameters of variants selected for kinetics characterization and the mutations that they harboured ............................................. 75
Table 3-4: Kinetic parameters in the presence of 20% t-BuOH and in 50 mM NaPi, pH7.4 buffer ................................................................................ 75
Table 4-1: The different detector parameters optimized ............................................... 94
Table 4-2: Optimization of the mobile phase to resolve the peaks of glucuronide and free steroid (CMO-DHEA) ...................................................... 94
Table 4-3: Validation for LCMS analytical method using Syn and hypothetical hit (2x Syn) ................................................................................ 95
Table 4-4: Summary of Syn SSM library screening ................................................. 98
Table 4-5: Relative improvement of various variants in secondary screens .................... 98
Table 4-6: Variants that were kinetically characterized and their mutations .............. 100
Table 4-7: Half-lives of Syn, SynY160G and SynH162Q in 20% t-BuOH from three replicates (n=3) ............................................................................ 104
Table 4-8: The concentration of t-BuOH at which 50% of the activity is lost for SynWT and the variants and the concentration of urea at which 50% of the enzymes unfold in urea ................................................................. 107
Table 4-9: The values of the c-parameter in Syn and the variants .................................. 111
Table 5-1: The turnover rate kcat of purified glucuronidase and glucuronylsynthase activities by wild-type, Syn (rare) and Syn (common) ..................... 124
Table 6-1: Solvent accessibility areas ....................................................................... 142
Table 6-2: The sites chosen for randomization and their roles .................................... 143
Table 6-3: The kinetic parameters of WT and T509A for glucosidase and glucuronidase activities ................................................................................................. 149
Table A-1: Table listing the oligonucleotide primers used ........................................... A:7
Table A3-1: Library sizes and the oversampling size required to achieve 95% library coverage in directed evolution experiments ................................................. A:11
Table A3-2: Mutations accumulated in β-GUS evolution ........................................ A:12
Table A3-3: Mutations accumulated in revisited Lp6 libraries after the addition of t-BuOH (Rd 1) .............................................................................. A:15
Table A3-4: Mutations accumulated in revisited Lp6 libraries after the addition of t-BuOH (Rd 2) .............................................................................. A:16
Table A4-1: Library sizes for various site-saturation libraries .................................... A:19
Table A6-1: Deviation of empirical (experimentally determined) and theoretical SD, and the quartiles of the mutation rates in each library ........................... A:25
Table A6-2: Summary of sequenced mutations from site saturation libraries ......... A:32
Table A6-3: Unpaired two-tailed t-test analysis of TEM-1 and β-GUS α-values .... A:37
Abbreviations and acronyms

-F : Fluorine leaving group  
1FGlcA : α-D-glucuronyl fluoride  
β-GAL : β-galactosidase  
β-gluc : β-glucuronide  
β-GUS : β-glucuronidase, wild-type enzyme  
2xYT : 2x-yeast tryptone media  
AA : Amino acid  
Ac : Acetyl  
AEBSF : 4-(2-aminoethyl) benzene sulfonyl fluoride hydrochloride  
ataS : Pseudomonas Aeruginosa arylsulfatase  
CD : Circular dichroism  
CMO-DHEA : Dehydroepiandosterone O-(carboxymethyl)oxime  
CMO-DHEA gluc : Dehydroepiandosterone O-(carboxymethyl)oxime-β-D-glucuronide  
DHEA : Dehydroepiandosterone  
DMSO : Dimethylsulfoxide  
DNA : Deoxyribonucleic acid  
E. coli : Escherichia coli  
E504A : β-GUS with position E504 mutated to alanine  
E504G : β-GUS with position E504 mutated to glycine  
E504S : β-GUS with position E504 mutated to serine  
ePCR : Error-prone PCR  
ePCR-StEP : Error-prone PCR – staggered extension PCR  
eq. : Equivalence  
ESI : Electron spray ionization  
E total amount of enzyme  
Et : Ethyl  
F : Fluorescence  
F⁻ : Fluoride ion  
GH : Glycoside hydrolase  
H₃PO₄ : Phosphoric acid, 85%  
HPLC : High performance liquid chromatography  
IPA : Isopropyl alcohol  
IPTG : Isopropyl β-D-thiogalactopyranoside  
ITC : Isothermal titration calorimetry  
K₂HPO₄ : Dipotassium hydrogen phosphate; potassium phosphate (dibasic)  
KH₂PO₄ : Potassium dihydrogen phosphate; potassium phosphate (monobasic)  
LB : Luria-Bertani media
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBA&lt;sup&gt;100&lt;/sup&gt;</td>
<td>LB media or LB agar plate with 100 μg/mL of ampicillin</td>
</tr>
<tr>
<td>LBK&lt;sup&gt;50&lt;/sup&gt;</td>
<td>LB media or LB agar plate with 50 μg/mL of kanamycin</td>
</tr>
<tr>
<td>LCMS or LC/MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>lit.</td>
<td>Literature values</td>
</tr>
<tr>
<td>M/Z</td>
<td>Mass to charge ratio</td>
</tr>
<tr>
<td>M9</td>
<td>Minimal media M9 salts</td>
</tr>
<tr>
<td>Me</td>
<td>Methyl</td>
</tr>
<tr>
<td>MM</td>
<td>Michaelis-menten</td>
</tr>
<tr>
<td>mQH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>Milli-Q water</td>
</tr>
<tr>
<td>Na&lt;sub&gt;2&lt;/sub&gt;HPO&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Disodium hydrogen phosphate; sodium phosphate (dibasic)</td>
</tr>
<tr>
<td>NaH&lt;sub&gt;2&lt;/sub&gt;PO&lt;sub&gt;4&lt;/sub&gt;</td>
<td>Sodium dihydrogen phosphate; sodium phosphate (monobasic)</td>
</tr>
<tr>
<td>NaOH</td>
<td>Sodium hydroxide</td>
</tr>
<tr>
<td>NaPi</td>
<td>Sodium phosphate buffer</td>
</tr>
<tr>
<td>Ph</td>
<td>Phenyl</td>
</tr>
<tr>
<td>pNPGlcA/pNPGA</td>
<td>Para-nitrophenyl-β-D-glucuronide</td>
</tr>
<tr>
<td>pNPgal</td>
<td>Para-nitrophenyl-β-D-galactoside</td>
</tr>
<tr>
<td>pNPGlC/pNP</td>
<td>Para-nitrophenol-β-D-glucoside</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>rxn</td>
<td>Reaction</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate - polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>StEP</td>
<td>Staggered extension PCR</td>
</tr>
<tr>
<td>Syn</td>
<td>Glucuronylsynthase</td>
</tr>
<tr>
<td>t-BuOH</td>
<td>t-butanol</td>
</tr>
<tr>
<td>t&lt;sub&gt;1/2&lt;/sub&gt;</td>
<td>Half-life</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>t&lt;sub&gt;ret&lt;/sub&gt;</td>
<td>Retention time</td>
</tr>
<tr>
<td>UDP</td>
<td>Uridine 5'-diphosphate</td>
</tr>
<tr>
<td>UDPGT/UGT</td>
<td>Uridine 5'-diphospho-glucuronosyltransferase</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet spectrum</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>X-gluc</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid</td>
</tr>
<tr>
<td>YENB</td>
<td>Yeast extract-nutrient broth media</td>
</tr>
</tbody>
</table>
Symbols and units

[ ] : concentration
°C : degree Celsius
ΔH : binding enthalpy
λ_{em} : emission maximum wavelength
λ_{ex} : excitation maximum wavelength
Å : Angstrom
au : arbitrary units

c : shape of sigmoidal curve of an isothermal titration calorimetry (ITC) binding isotherm (= e^c_1 x K_a)
d : days
Da : Dalton
g : gram
g : acceleration due to gravity
h : hour
Hz : Hertz
K : Kelvin
K_a : association constant
kb : kilobase pairs
k_{cat} : catalytic turnover constant
k_{cat/K_m} : substrate specificity constant
K_d : dissociation constant
kDa : kiloDalton
K_m : Michaelis constant, substrate concentration when at ½
\( V_{max} \)
L : litre
M : molar (moles per litre)
mg : milligram
mM : millimolar
mm : millimetre
mmol : millimole
mol : mol
mA : milliAmpere
mV : millivolt
nM : nanomolar
nm : nanometre
nmol : nanomole
ppm : parts per million
RFU : relative fluorescence units
rpm : revolutions per minute
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Unit/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>correlation coefficient in non-linear regression</td>
</tr>
<tr>
<td>$R^2$</td>
<td>coefficient of determination in linear regression</td>
</tr>
<tr>
<td>s</td>
<td>second</td>
</tr>
<tr>
<td>$\mu$L</td>
<td>microlitre</td>
</tr>
<tr>
<td>$\mu$M</td>
<td>micromolar</td>
</tr>
<tr>
<td>$\mu$m</td>
<td>micrometer</td>
</tr>
<tr>
<td>$\mu$mol</td>
<td>micromole</td>
</tr>
<tr>
<td>$\mu$s</td>
<td>microsecond</td>
</tr>
<tr>
<td>V</td>
<td>volts</td>
</tr>
<tr>
<td>vol.</td>
<td>volume</td>
</tr>
<tr>
<td>$v$</td>
<td>velocity</td>
</tr>
<tr>
<td>$v_0$</td>
<td>initial velocity</td>
</tr>
<tr>
<td>$v/v$</td>
<td>volume per volume</td>
</tr>
<tr>
<td>$V_{max}$</td>
<td>maximum velocity of the enzymes</td>
</tr>
<tr>
<td>w/w</td>
<td>weight per weight</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
</tbody>
</table>
Nomenclature

A. Reactions

The substrates used in this thesis are structurally complex. They exhibit a very rich chemistry due to the large number of stereoisomers that they possess. This thesis generally follows the nomenclature outlined by the IUPAC guidelines. Nonetheless, we will provide a summary of the nomenclature used for the chemical compounds and the main reactions that will be encountered throughout the thesis. The reactions that concern this thesis are shown in Figure N-1.

Figure N-1: Reactions that are discussed in this thesis are the hydrolysis and synthesis of glucuronides.

“Glucuronides” is a trivial name for a class of glycosides. Glucuronide substrates have two moieties, generally called the glycone and aglycone components. The two components are linked by a glycosidic bond, which takes the form of -O bridge. Specifically, glucuronides consists of a glucuronoyl and a general aglycone component, which are linked by an –O glycosidic bond. The hydrolysis of glucuronides generates a glucuronic acid and a general aglycone. The reverse process synthesizes the –O glycosidic bond. This will be referred to as conjugation or glucuronylation. The nomenclature for the two components, as depicted in Figure N-2, is detailed in the next two sections.
B. Glycosyl groups

The monosaccharide that concerns this thesis is β-D-glucopyranosiduronic acid or β-D-glucuronic acid (GlcA), which differs from β-D-glucose (Glc) at C6 position (Figure N-3), where the hydroxyl group has been oxidized to a carboxylic acid. This is one of the hydrolysis products that will feature frequently in this thesis. Figure N-3 shows the open chain (Fisher projection) and the 6-membered cyclic configuration of the two compounds.

Following IUPAC guidelines, the numbering for the monosaccharides begins at the aldehyde chain in the Fisher projection. The stereochemistry at the highest-number chiral centre (C5) provides the reference point for the definition of the stereochemistry at the other stereocentres. The symbols “D” and “L” refer to the configuration at this carbon itself. A monosaccharide is classed as “D” sugar if the hydroxyl group on this carbon is positioned on the right. “D” and “L” are related as enantiomers and the “D”
form is the more common of the two in nature. It is the form that is applicable to this
thesis. Therefore, only the “D” sugars were represented in Figure N-1.

Simple sugars often exist in their 6-membered cyclic configuration (pyran). This
is generated when the C5 hydroxyl group forms an intramolecular bond with the
aldehyde (C1) resulting in ring closure. This forms a new chiral centre at C1, which is
also known as the anomeric carbon. The stereochemistry at C1 of the pyran form is
defined by prefixes “α” and “β”. Its absolute configuration is also defined by its
orientation relative to C5. If the hydroxyl group on this anomeric carbon is cis to the
substituent C5, the sugar will be known as the β-anomer. The corresponding glycosides
derived from these monosaccharides will be known as glucosides and glucuronides
respectively (Fig N-2).

The hydroxyl group at the anomeric carbon can be substituted for another
functional group. Nevertheless, the naming convention remains the same. For example,
the α-D-glucuronyl fluoride (1FGlcA) is similar to the α-D-glucuronic acid except that
its anomeric substituent is a -F rather than a -OH group (Figure N-4). The 1FGlcA is the
synthetic substrate used for this work and will be also be called the glucuronyl donor
in glucuronylation reactions. Where reference is made to general glycone group donors,
they are simply termed glycosyl donor.

\[
\begin{align*}
\beta-\text{D-Glucuronic acid} & \quad \alpha-\text{D-Glucuronic acid} & \quad \alpha-\text{D-Glucuronyl fluoride} \\
\end{align*}
\]

Figure N-4: β-D-glucuronic acid, α-D-glucuronic acid and α-D-glucuronyl fluoride

C. Aglycone groups

The other component in the β-D-glucuronide is the aglycone. This is either the
leaving group in the hydrolysis reaction, or an acceptor substrate for the conjugation
reaction. They are termed acceptor substrates because they accept the sugar ring in
glycosylation reactions. For glucuronylation, it is generally known as glucuronyl
acceptors. There are two types of acceptors—alcohol acceptors and steroid acceptors. In both cases, the center for the conjugation reaction happens at a\(-OH\) functional group. **Figure N-5** depicts some acceptors that will be referred to in the thesis. The two main steroids used for this work are dehydroepiandrosterone \(O\)-(carboxymethyl)oxime (CMO-DHEA) and testosterone. Since the systematic names are too long, the common names and abbreviations will be used.

(a) Alcohol acceptor

![2-phenylethanol](image)

2-phenylethanol

![para-nitrophenol](image)

*para*-nitrophenol

(b) Steroid acceptor

![Dehydroepiandrosterone](image)

Dehydroepiandrosterone (DHEA)

![Dehydroepiandrosterone](image)

Dehydroepiandrosterone \(O\)-(carboxymethyl)oxime (CMO-DHEA)

![Testosterone](image)

Testosterone

Figure N-5: The aglycones discussed in this thesis. Acceptors of glucuronyl donors in glucuronylation reactions or leaving group in the hydrolysis reaction. *para*-nitrophenol (\(p\)NP) is the leaving group in the substrate used to monitor hydrolysis in this thesis.
1. Introduction

1.1. Preamble

This chapter details the aims of this thesis and the strategies that will be employed to achieve these aims. We will also give a background on the enzymes that concern this thesis. In the interest of readers who may not be familiar with the naming conventions of the glycone and aglycone groups in glucuronides and glycosides, a description of these and their chemical transformations has been provided in the preceding section—Nomenclature. The last part of this chapter aims to provide an overview of the thesis.

1.2. Background

Glucuronide synthesis plays an important role during detoxification in the human body\(^1-^3\). This process is known as glucuronidation or glucuronylation. It is a major biochemical reaction in Phase II xenobiotic metabolism where exogenous substances are converted to glucuronides. Xenobiotics may take the form of drugs or pharmaceutics, dietary intake, or various kinds of chemicals. Figure 1-1 illustrates the fate of xenobiotics and the metabolites in the xenobiotic metabolism pathway.

The biochemical reaction for glucuronylation in eukaryotes is catalyzed by the uridine 5'-diphosphoglucurunosyl transferases (also known as uridine glucuronosyl transferase), UDPGTs/UGTs (E.C. 2.4.1.17)\(^3,^4\). UDPGT is a glycosyltransferase that catalyzes the conjugation of the metabolites from Phase I to glucuronides\(^5,^6\). This modification generates a polar conjugate with increased hydrophilicity and serves to facilitate elimination via biliary excretion\(^7\). Alternatively, biliary excretion (excretion by bile) may lead to the recycling of xenobiotic metabolites in an enterohepatic cycle\(^2,^8-^10\). This occurs in the intestines where glucuronides from Phase II are degraded by the enzymes of the intestinal microflora\(^11,^12\). This regenerates the glucuronic acid that will be used as a nutrient source and the native xenobiotic. Native xenobiotic will be reabsorbed into the bloodstream to re-enter the liver and the metabolic pathway. This recycling prolongs a drug therapeutic life and is an important feature in drug
development for the assessment of novel drug potency (pharmacokinetics/pharmacodynamics studies).\textsuperscript{8,13}

Since glucuronides are major products of xenobiotic metabolism, its presence can serve as an indicator of drugs or xenobiotics consumption. Thus, they are extensively applied as standards and biomarkers. Detection and quantification of glucuronide conjugates have become established protocols for monitoring of drug intake and pharmacological evaluation in various fields including sports drug testing\textsuperscript{14}, detection of agricultural residues\textsuperscript{15,16} and in drug development\textsuperscript{8}. Consequently, this creates a significant demand for various glucuronide conjugates. There has also been growing interest in the development of glucuronide pro-drugs for therapeutic purposes\textsuperscript{17–19} in recent years. Glucuronide metabolites or $\beta$-glucuronides can sometimes appear more potent than the drugs that are currently available. Evidence of this can be found in the glucuronide of morphine conjugate, morphine-6-glucuronide\textsuperscript{20,21}, which is currently in phase III clinical trials for application in post-operative analgesic therapy. Glucuronides therefore offer renewed options in medicinal chemistry. As such, the demand for glucuronide synthesis continues to exist, creating a need for cleaner and more efficient synthesis of glucuronides.

---

Figure 1-1: The fate of xenobiotic and its metabolite. After Phase II metabolism, glucuronylated products can be eliminated or used as a feedstock by intestinal bacteria. Inset shows biotransformation in Phase II.
1.2.1. Glucuronide synthesis

The production of glucuronides using traditional methods of chemical synthesis is challenging\textsuperscript{22,23}. Chemical syntheses of glucuronides are generally based on derivations of the Koenigs-Knorr reaction\textsuperscript{24} but these often suffer from poor yields (\textasciitilde 30\%)\textsuperscript{23,25–28}. These methods are complicated by undesired side reactions due to the low reactivity of glucuronic acid as a glycosyl donor. In addition, this approach requires multiple protection/de-protection steps of the glucuronic acid to provide control of regioselectivity in the glucuronylation.

Understanding the glucuronylation process in human body has opened up new options for the preparation of glucuronides using enzymatic methods. The use of glycosyltranferases\textsuperscript{29} for the synthesis of glycoproducts was the first chemoenzymatic route to be considered. UDPGTs are glycosyl specific and they are able to catalyze the glycosidic bond formation with a high degree of control over the regio- and stereospecificity. Although this approach provides a method for mild and stereospecific single-step synthesis, its application has been limited by two major reasons. Firstly, UDPGTs occur naturally as membrane bound enzymes that are anchored to the endoplasmic reticulum membrane\textsuperscript{22,30}. Hence, they are very sensitive and unstable under laboratory conditions. Secondly, they are substrate-specific to the acceptors, which limits the range of substrates that can be used as starting materials\textsuperscript{31}. The nature of its machinery coupled with its low stability imposes a limit to its scalability for process chemistry and restricts the procedure to small-scale syntheses (\textasciitilde1 mg). Furthermore, the UDPGTs are Leloir enzymes that depend on sugar-nucleotide donors (uridine 5’-diphosphoglucurunosyl, UDP) to drive the glycosidic bond formation\textsuperscript{32,33}. Enzymatic methods using UDPGTs may require the synthesis of natural or unnatural diphosphate sugar donor analogue\textsuperscript{34,35}, which are difficult to prepare. The yields of these reactions are mostly around 20–40\%\textsuperscript{35,36} and their preparation is encumbered by the instability of the diphosphate sugar donors. This further restricts the applicability of UDPGTs for industrial application.

The drawbacks of using UDPGTs can be circumvented using an alternative enzymatic approach – the glycosynthase. Glycosynthases\textsuperscript{37} have been developed by the Withers’ group to synthesize β-glycoside linkages\textsuperscript{38–40}. These enzymes are derived from
the engineering of β-glycosidases so that their hydrolytic pathway is rendered inactive and that the hydrolytic function can subsequently be reversed. This concept had been adopted by the McLeod group to design a feasible enzyme from β-glucuronidase (β-GUS) that could be used in chemoenzymatic glucuronylation of alcohol acceptor substrates\textsuperscript{41,42}. This thesis will expound on these early investigations.

1.3. Enzymes: β-GUS and glucuronylsynthase

This thesis will first engineer a native enzyme, β-glucuronidase (β-GUS) (EC 3.2.1.31) to have different kinetic properties. A β-GUS variant will then be converted to a synthetic enzyme (glucuronylsynthase) \textit{via} a site-specific mutation E504G. This section describes the two enzymes and how their functions and mechanisms relate to each other.

1.3.1. The various forms of β-glucuronidase (β-GUS)

β-GUS is an exoglycosidase and a member of the glycosidase enzyme family that catalyzes the hydrolysis of glycosidic bonds. It is a carbohydrate active enzyme and is classified in the Carbohydrate-Active Enzymes (CAZy) database as a member of Glycoside Hydrolase Family 2 (GH 2)\textsuperscript{43,44}. The enzyme cleaves a terminal β-glucuronic acid residue. As in most glycosidases, β-GUS is glycosyl specific \textit{i.e.} specific for the glucuronic acid, but is promiscuous with the aglycone moiety. Hence, key binding residues are those that interact with the glycosyl ring. As such, the binding behavior, kinetics and catalytic rate can be altered by modifying these residues or nearby residues.

There are two commonly studied forms of β-GUS: the human form\textsuperscript{45} and the bacterial \textit{E. coli} form\textsuperscript{46}. The two isoforms have similar overall fold and share 50% sequence similarity\textsuperscript{45,47}. However, there are several key differences. The human β-GUS exists as a lysosomal enzyme in mammalian cells. It is synthesized in the endoplasmic reticulum and then transported into the lysosomes\textsuperscript{48} with the aid of phosphotransferases\textsuperscript{49}. It breaks down complex glucuronides known specifically as glycosaminoglycans (GAGs). It also has a narrow pH range of pH 3–5. We were more
interested in an enzyme that works with simpler glucuronides and under milder condition. Hence, the human β-GUS was not suitable for our experiments but it was used as a reference for sequence and structure comparison with the bacteria ortholog.

The bacterial form of β-GUS was more suitable for our work. It is commonly expressed by gut bacteria such as \textit{E. coli} that uses it to harvest the glucuronidated products that enter the enterohepatic cycle after Phase II xenobiotic metabolism. This was demonstrated by Novel et al. in experiments that showed that β-GUS synthesis in \textit{E. coli} was induced in the presence of β-glucuronide and glucuronates\textsuperscript{50,51}. Hence, the \textit{E. coli} β-GUS caters to simpler substrates compared to the human ortholog. Its substrate range includes simple glucuronyl conjugates of chemicals such as alcohols, drugs and steroids. These are similar to the glucuronides that we would like to investigate. Therefore, the \textit{E. coli} version of the β-GUS was used for the work that will be discussed in this thesis.

1.3.2. Glucuronylsynthase activity of engineered β-GUS

The \textit{E. coli} β-glucuronidase (β-GUS) contains two catalytic residues, E413 as a general acid/base catalyst, and E504 as a nucleophile in its substrate binding pocket (Figure 1-2a)\textsuperscript{47,52}. The catalytic machinery is similar to the other members of retaining glycoside hydrolases\textsuperscript{53,54}, and occurs through a hybrid S\textsubscript{n}1/S\textsubscript{n}2 mechanism\textsuperscript{55–57}. The carboxylate anion E504 initiates the catalysis via S\textsubscript{n}2 attack on the anomeric carbon of glucuronide. This displaces the aglycone residue and results in the formation of a glucuronyl-enzyme covalent intermediate of a different anomeric configuration. E413 acts as a general acid/base catalyst that first acts as the proton donor to the leaving group. In the second step, it deprotonates a water molecule or a hydroxylated compound for displacement of the glucuronyl-enzyme bond. Thus, the β-GUS enzyme is regenerated with a net retention in configuration of the β-D-glucuronic acid (GlcA) product.

The glucuronylsynthase is generated by site-specific mutagenesis at E504 to a non-nucleophilic residue that renders the residue incapable of nucleophilic attack at the C1 position. Glucuronylation is achieved through substrate-assisted catalysis in the presence of a glycosyl donor of an opposite anomeric configuration relative to the
native substrate and that has a good leaving group at C1 such as a fluoride substituent. The E413 residue activates alcohol substrates (acceptor) that would displace the anomeric leaving group. Since the E504 is incapable of nucleophilic attack, the hydrolytic pathway is also disabled. This allows the glucuronide product to accumulate (Figure 1-2b). The McLeod group investigated the mutations E→A/G/S to derive the glucuronylsynthase system using α-D-glucuronyl fluoride (1FGlcA) as the glycosyl donor42. In their study, they have found that the mutation E→G at position 504 was the most efficient. E504G variant was finally used as the glucuronylsynthase (Syn).

Figure 1-2: Schemes depicting enzymatic mechanisms. (a) Mechanism of β-GUS. (b) Proposed mechanism of the glucuronylsynthase.

In cells, synthetic reactions are often coupled with reactions that would release large amounts of energy, usually in the form of ATP hydrolysis to drive the reactions. For example, the synthesis of glycosides occurs through catalysis by transferases that require activated sugar nucleotide diphosphates as co-factors. In the case of the
glucuronylsynthase, the reaction is driven by the F- substituent as a good leaving group, where the 1FGlcA mimics the covalent intermediate in the wild-type reaction. Hence, the availability of the 1FGlcA substrate to drive the synthetic reaction is important. It is also used in excess, primarily to overcome product inhibition\textsuperscript{41}. The efficiency of the glycosynthase reaction is partly dependent on the absence of competing hydrolytic reaction. In particular, minute presence (1\%) of the parent hydrolase may reduce the efficiency of the synthetic reaction as it hydrolyses the glucuronide products. Therefore, it is important to eliminate the sources of hydrolysis.

A major source of the parent hydrolase in the glucuronylsynthase system is translational misincorporation, which had been quoted for several glycosynthase systems\textsuperscript{38,58}. However, its effects have not been known to adversely impact the synthetic reactions. At the start of this project, this had not been known to have an impact on the glucuronylsynthase system. However, during the course of our work, we have found that the effects of translational misincorporation can be a relevant issue for the glucuronylsynthase system and we have found ways to reduce its extent.

1.4. Specific aims of the project

The general aim of the project is to improve the glucuronylsynthase enzyme. Using the recombinant enzyme E504G, Wilkinson et al. was able to achieve glucuronylation of simple acceptor substrates such as of 2-phenylethanol (Figure 1-3) with yields of 96\% (~ 700 mg)\textsuperscript{41}. However, the turnover rate ($k_{cat}$) and efficiency ($k_{cat}/K_m$) are low; the reaction takes three days. In addition, the current glucuronylsynthase system does not demonstrate the same efficiency for different acceptor alcohols. Although its $k_{cat}$ with 2-phenylethanol is 1.4 min\textsuperscript{-1}, its turnover rate for the more structurally complex steroid (CMO-DHEA) is ten-times slower. This leaves considerable scope for improving the glucuronylsynthase system. Hence, this thesis will be aimed towards the following:

(i) To develop a better enzyme for the synthesis of β-glucuronides by evolving \textit{E. coli} β-glucuronidase using two different approaches (Scheme I and II) of Figure 1-4. We will prioritize the first approach (Scheme I), which will involve
two steps. At the same time, we will work on the development of a suitable screening assay for the second approach.

(ii) To characterize the enzymatic properties of the mutant enzymes and gain better understanding of how they function.

(iii) To gain a better understanding of the glucuronylsynthase system and generalize our understanding to facilitate future studies on glycosynthases.

Figure 1-3: Examples of glucuronylsynthase reactions using engineered β-GUS (E504G) enzyme.

Figure 1-4: Two distinct approaches for the engineering of β-GUS to obtain an improved glucuronylsynthase.
1.5. Basis and hypothesis for the two-step strategy

As part of the 2-step approach that was outlined in Section 1.4, the E. coli β-GUS enzyme will first be evolved. The wild-type β-GUS has a low $K_m (200 \mu M)^{59-61}$, which allows it to bind and process substrates rapidly at low concentrations. Evolving the wild-type β-GUS to have higher activity at high substrate concentrations typical of synthesis experiments (10–100 mM) will possibly generate a variant that is more suitable for industrial application. The major advantage of this approach is that the assay is simple, quick and sensitive. Screening can be performed using β-D-para-nitrophenyl-glucuronide (pNPGlcA), which is commercially available. Upon hydrolysis, this liberates chromogenic para-nitrophenolate (pNP) that can be monitored on a spectrophotometer at absorbance wavelength 405 nm (yellow) (Figure 1-5). The disadvantage, however, is the cost of the substrate. This prohibits us from screening at 100 mM substrate concentration, but still allows us to screen at concentrations that is 10 to 20-times the $K_m$ of the wild-type β-GUS (2–4 mM). In the second step of the two-step approach, an improved β-GUS enzyme can be converted to a glucuronylsynthase by incorporating a mutation at position 504. This can be done by site-specific mutagenesis.

![Figure 1-5: Scheme of the reaction in β-GUS screening assay. Screening assay for β-GUS can be performed using para-nitrophenol glucuronide (pNPGlcA) by monitoring the liberation of p-nitrophenolate at 405 nm](image)

Our basis for the two-step approach is grounded on observations that the enzyme has specificity constant $k_{cat}/K_m$ approaching diffusion limit. However, not all enzymes with $k_{cat}/K_m$ approaching diffusion limit have high $K_m$, suggesting that there may be room for improvement in the $k_{cat}$ and $K_m$ of the enzyme. If we can increase the $K_m$ of the wild-type by 4 to 5-fold without changing the enzyme specificity constant $k_{cat}/K_m$, we may be able to simultaneously increase the $k_{cat}$ of the enzyme by the same degree. An enhancement in the $k_{cat}$ might indicate that there is an increase in the hydrolysis rate in
the parent enzyme, which may indicate more efficient base catalysis by E413. Higher base catalytic rate by E413 may translate to higher synthetic rate when we incorporate the E504G mutation for the acquisition of synthetic activity.

The two-step strategy outlined in Scheme I has not been attempted before. Attempts to improve glycosynthases had often been done by directly evolving the recombinant i.e. the glycosidase with inactivated nucleophilic residue. Considerable effort had been put into the development of a suitable high-throughput screening assay for the glycosynthase libraries. However, many of the methods that have been developed are also specifically tailored for the detection of the glycosyl substrates or type of glycosides involved in a particular glycosynthase system. To date, there is no viable high-throughput screening assay system that can be adopted for the glucuronylsynthase system. Development of high-throughput screening assays for glucuronyl systems (results unpublished) had been complicated by the low activity of glucuronylsynthase. This means that an assay has to be sensitive enough to overcome interfering effects, or a more active glucuronylsynthase has to be obtained for the development of a high-throughput screening assay.

The two-step strategy that is proposed is a unique approach and has never been reported before. More importantly, it offers an alternative that circumvents the need for the development of a high-throughput assay for glucuronylsynthase. The most reliable screening method for the glucuronylsynthase method would be the high-performance liquid chromatography (HPLC) or liquid chromatography mass spectrophotometry (LCMS) method equipped with 96-well plate automation. The latter became available in the middle of the project, and enabled the initiation of the second approach (Scheme II, Figure 1-4).

1.6. Research Design and Approaches

We will apply protein engineering techniques to achieve the aims outlined in Section 1.4. Many of these techniques have been described in the literature. This section gives a review of the technologies that are available and details how we will adopt them to design suitable engineering approaches.
1.6.1. A review on protein engineering

Enzymes in nature can offer unparalleled sophistication as biocatalysts because they are highly specific and efficient. However, like the glucuronylsynthase, many natural enzymes require modifications before they can be used for practical applications. The application of natural enzymes may often be constrained by stability issues or incompatible function. Consequently, they may require adaptation for application in chemoenzymatic reactions.

Attempts at adapting or improving enzymes for practical applications can be chronicled in three stages. The first stage involved the isolation and application of natural enzymes in various industries, e.g. therapeutic drugs and fine chemicals production. The second was set off by the advancement in genetic manipulation methods that allows site-specific mutations to be encoded. As a result, natural enzymes could be tailored for generating synthetic intermediates in multi-step reaction schemes and the use of biocatalysts became more prevalent. The latest development in biocatalysis has been dominated by efforts to develop molecular biology techniques for the manipulation of genes to enhance enzyme properties. This has led to increased capability in protein engineering. There are three approaches to engineer an enzyme. These are the rational, semi-rational or bioinformatics approach and evolutionary approach.

1.6.1.1. Rational and semi-rational approaches

The rational approach is “knowledge-based” and incorporates mutations specifically. This is usually referred to as site-specific mutagenesis. Usually, rational approach is applied using a 3D enzyme structure or a reliable model. It requires detailed structural information and a good understanding of the enzyme function. Most of the time, the approach relies heavily on the computational predictions of an enzyme’s molecular dynamics and behavior. This was the approach used to create the glucuronylsynthase, where E→A/G/S were created specifically at E504. We will use this method to introduce interesting mutations that were found in the literature and in the second step of the two-step approach when we convert an improved β-GUS into a glucuronylsynthase.
Likewise, semi-rational approaches also utilize computational prediction powers and a structure to guide a researcher\textsuperscript{76,77}. However, it is less specific. Some knowledge of the enzyme structure and function is required so that appropriate sites can be selected for mutagenesis. Examples of semi-rational approach include site-saturation mutagenesis or site-directed mutagenesis. It is considered to be semi-rational approach in that it requires knowledge of the enzyme structure and mechanism, but it “blindly” randomizes the positions of interest. Prediction and modeling softwares such as CUPSAT\textsuperscript{78}, FTMap\textsuperscript{79} and Swiss-Model\textsuperscript{80,81} are used to facilitate decision making on target choice.

There are several strategies that can be used to randomize positions of interest in site-saturation mutagenesis (SSM). These exploit the codon degeneracy of amino acids. The most common strategy introduces mutations using the NNK codons (N = a/c/g/t, K = c/g) and encodes the canonical 20 amino acids\textsuperscript{82}. The codons can be changed according to preference to incorporate fewer codon combinations and amino acids. For example, the NDT codon (D = a/g/t/) encodes only 12 amino acids, but represents a balanced mix of the diverse chemical structure of the amino acids. There are also options to randomize several sites at the same time. The strategy depends on the purpose of the library and the screening effort that the researcher can manage. An NNK codon library will involve more screening effort than the NDT library\textsuperscript{83}. Libraries that randomize several sites at the same time will also require more screening effort. A program such as CAStef 2.0 assists such endeavors by optimizing codon usage and calculating the optimal library screening size\textsuperscript{84}.

1.6.1.2. The evolutionary approach

The third approach to protein engineering is an evolutionary approach. It is also known as directed evolution\textsuperscript{85–87}. This approach imitates natural evolution and is a powerful engineering technique. It is the least reliant on computational predictions and modeling. It does not require prior knowledge of an enzyme structure\textsuperscript{88}. Instead, it may facilitate understanding of an enzyme structure and function.

Directed evolution involves two steps that are iterated (Figure 1-6). The first step generates the diversity of a library using random mutagenesis methods. These are typically error-prone PCR (ePCR)\textsuperscript{89,90} and DNA recombination of libraries\textsuperscript{91–95}. Error-
prone PCR is a highly efficient method for introducing variations into a genetic pool with a good control over the mutation rates. The libraries are typically created using a mutagen (such as high Mn$^{2+}$ concentrations) in the PCR reactions. The second step applies an assay to screen or select for improved clones (potential hits). Improved clone(s) are used to create the diversity for subsequent generations. There are three options for this:

(i) Improved clones can either be recombined using a DNA recombination technique such as DNA shuffling$^{95,96}$ or Staggered Extension PCR (StEP)$^{93,97}$.

(ii) A combination of StEP and ePCR (ePCR-StEP) can be employed to simultaneously recombine mutations from previous cycles and introduce new random mutations.

(iii) The most improved clone is used as the template to generate another ePCR library.

![Figure 1-6: Typical process in directed evolution experiments. Experiments involve repetitive cycles of library generation and screening/selection](image)

Typically, one complete cycle is one round of selection or screening of a generation where a variant with a desired property is picked for subsequent iterations.
The process is iterated until a desired property is obtained or when convergence (fixation of mutations and marginal improvements/diminished returns) is reached.

The advantages of directed evolution lie in the fact that it does not require any prior knowledge of enzyme structure and mechanism. It sometimes reveals interesting mutations that may not have been picked up by rational or semi-rational approaches. Recombination during iterative cycles is an effective way to accumulate beneficial mutations. However, the success of in vitro evolution is dependent on various factors such as recombination efficiency, mutation rates, library size and screening capacity. This is linked to the availability of a rapid but reliable screening procedure and the enzyme being investigated. A larger multi-domain enzyme is likely to involve more effort. Attempting to enhance the $k_{\text{cat}}$ and $K_m$ of an enzyme that is already high will also be challenging. In such cases, an important aspect of the engineering endeavor will involve strategies to adapt the three different approaches to optimize the screening effort.

1.6.2. Combining protein engineering strategies for β-GUS evolution

The β-GUS enzyme is a large tetrameric protein (68 kDa, 603 amino acids per subunit). Its evolution would involve larger screening effort compared to other enzymes in our laboratory. To achieve our aims in Section 1.3, we will adopt all three approaches to design libraries that can optimize engineering effort. Rational engineering will be used to incorporate an E504G to produce the glucuronylsynthase. Rational engineering will also be used to introduce mutations from other directed evolution experiments that were performed on the β-GUS. We will be using a combination of semi-rational engineering and a hybrid of semi-rational design/directed evolution (or targeted evolution) approach to evolve the enzyme. Targeted evolution applies random mutagenesis on specific regions of the enzyme. Mutations identified in the different regions can then be strategically recombined. Hence, targeted evolution would allow us to sample a sequence space more thoroughly.

For both approaches, we will use the structure of E. coli β-GUS, obtained by Wallace et al. to guide us (Figure 1-7a). The bacterial enzyme essentially comprises three domains: (i) the N-terminal resembling a sugar-binding domain that is 180 residues long, (ii) an immunoglobulin-like β-sandwich domain that is 96 residues
long, and (iii) the C-terminal domain made up of residues 274 through to 603, which forms the αβ-barrel (TIM-barrel) (Figure 1-7b). The first two domains are not known to have catalytic function but they are important structural elements of the enzyme. However, a small section of the loop from Domain I interacts with the active site (Figure 1-7a, inset). The C-terminal domain (TIM-barrel) is the catalytic domain and contains the active-site residues E413 and E504.

β-GUS is glycosyl specific. Hence, mutating sites that interact with the glycosyl moiety will likely change its activity. These sites were identified from the interaction of the *E. coli* β-GUS with an inhibitor glucaro-δ-lactam (PDB ID: 3K4D). The structure of this inhibitor is very similar to the glucuronic acid (Figure 1-8), and serves as a suitable
enzyme-glycosyl complex model to deduce important glycosyl binding sites. Table 1-1 summarizes the initial target residues. The position of the target residues relative to the inhibitor is shown in Figure 1-9.

Table 1-1: Priority mutations to be considered for site-saturation mutagenesis that were derived from structural analyses

<table>
<thead>
<tr>
<th>Residue</th>
<th>Codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>N412</td>
<td>NNK</td>
</tr>
<tr>
<td>D163</td>
<td>NNK</td>
</tr>
<tr>
<td>L361</td>
<td>NNK</td>
</tr>
<tr>
<td>R562</td>
<td>NNK</td>
</tr>
<tr>
<td>M447</td>
<td>NNK</td>
</tr>
<tr>
<td>T509</td>
<td>NNK</td>
</tr>
</tbody>
</table>

Figure 1-8: Structures of glucuronic acid and glucaro-δ-lactam.

Figure 1-9: Positions of prioritized sites for saturation mutagenesis and other binding sites that are important for glycosyl binding. Inset: Position of N412 relative to E413 and glucaro-δ-lactam (GDL).

Residue D163 is from an extruding Domain I loop that is implicated in glycosyl binding. It forms salt-bridge interaction with R562, which is also implicated in glycosyl binding. L361 marks the start of the disordered region in the crystal structure that forms an overhang over the active site. Residue N412 is a polar residue next to the general acid/base catalyst E413 and may yield interesting observation. Targeting these residues should change the kinetic parameters of the enzyme at high concentrations of substrate. V446, M447 and F448 residues are not involved in glycosyl binding but they form hydrophobic interaction with the leaving group (aglycone). One of these, M447 is positioned at the surface (mouth of the barrel) and may facilitate departure of the leaving group product. Hence, this was chosen as a target as well. T509 was picked up from the literature\textsuperscript{59,99,101,102}. This residue appears to change the specificity of the
enzyme for other glycoside substrates, e.g. galactoside, xyloside, and glucoside\textsuperscript{59,99,102}, and to increase thermostability\textsuperscript{101}.

Since we aim to change the catalytic activity, random mutagenesis will be performed on regions that are in the TIM-barrel (Figure 1-10). This is with the exception of Domain 1 loop, which points into the active site and that is directly involved in glycosyl binding. The αβ barrel offers considerable potential for altering catalytic activity because of the numerous flexible loops that connects the α-helices and β-sheets, which form the frame of the TIM-barrel. The dynamics of flexible loop has often been associated with increased enzyme plasticity and altered affinity\textsuperscript{104–107}. Targeting such region is very likely to alter kinetic parameters. We will exploit the flexibility offered by β-GUS loops in the active site to tailor the enzyme for higher activity. The regions (Figure 1-11) that will be targeted for random mutagenesis are the (i) disordered loop, (ii) the loop from Domain I that interacts with the active site and (iii) regions in the TIM barrel domain.

![Figure 1-10: View of Domain III (TIM-barrel) of β-GUS from the bottom. α-helices and core β-sheets form the αβ barrel and is held by a hairpin loop at the base. Insets show the approximate position of the active site in the barrel and the side view of the TIM-barrel reveals the interconnecting flexible loops.](image)

The disordered loop is made up of residues 360-370. It is interesting not only because it is highly flexible, but it is also implicated in interactions with symmetry mates.\textsuperscript{47} It interacts with the active site of an adjacent subunit. In addition, this loop is also unique to the bacterial ortholog. Jain et al. have found that in the sequence
alignments of the *E. coli* β-GUS and the human β-GUS, the residues 360-377 in the bacterial ortholog is absent from the human sequence\(^{45}\). These two observations suggest that the disordered loop is important for catalytic activity. Targeting this loop either by targeting the loop in itself or by engineering the loop as part of the TIM-barrel region may yield interesting observations (mutations that improve activity). The TIM barrel can be further divided into two parts: from residue K286 up to the nucleophilic residue E504 and the C-terminal region from E504 onwards (Figure 1-11). Dissecting the region to a smaller fragment may allow for more thorough search of the sequence space at this region.

The β hairpin structure at the bottom of the barrel may also be investigated by itself or as part of a larger sequence space (Figure 1-11). The C-terminal contains an extended mobile loop (V548-S579) that is ~30 amino acid long. It is flexible and this region seems to be important for modulating activity. In several other directed evolution experiments on β-GUS, mutations were found at N566-K568, Q585 S550, G559, F582\(^{60,99}\) and various other sites in this region. Mutations from different regions can be recombined in the penultimate round. Finally, we can consider random mutagenesis on the whole gene and screen a small fraction of the library to see if there are any interesting mutations from other parts of the enzyme that is not targeted.

### 1.7. Overview of the thesis

The aim of the thesis is to evolve the β-GUS so that it can be converted to a more competent glucuronylsynthase. There are two approaches for this: a two-step indirect approach and a one-step direct approach. The approaches and the basis were described in Sections 1.3 and 1.4. The two-step indirect approach is conceptually different from other glycosynthase engineering approaches. It was appealing because it could potentially overcome screening limitations in the one-step direct approach. The shortcoming of this approach is that its outcome depended on whether the enhancement observed for the β-GUS was translated across to the glucuronylsynthase enzyme. We would only know if the approach was successful at the end, when we incorporate the E504G mutation.
The work for the β-GUS evolution is described in Chapter 3, where we evolved the glucuronidase activity in two different stages. The first stage evolved β-GUS under mild conditions (30–37 °C, pH 7.0–7.5), which had been used for the development of the glucuronylsynthase system. The second stage was conducted in harsher conditions...
when it was discovered that the addition of t-BuOH as a solvent improved the process chemistry of the glucuronylsynthase enzyme. Although we did not find any improved variants from the first stage, the second yielded several promising mutations that could be converted to a glucuronylsynthase (Chapter 4).

When we converted one of the improved variants from Chapter 3 into a glucuronylsynthase, we did not see improved synthase activity. This led to two separate experiments. The first continued the engineering work using the second approach (one-step direct approach). This is described in Chapter 4. The second study sought to understand the possible reason behind this failure. Our work here was first led by observations that different batches of purified glucuronylsynthase enzymes had different synthetic activity. In some cases, the synthetic activity could not be observed. This led to a review on the growth and expression systems/conditions that had been used. As a result, we found that translational misincorporation may limit the efficiency of the glucuronylsynthase enzyme. The systematic study that led to this conclusion is detailed in Chapter 5.

The penultimate chapter (Chapter 6) consolidates the observations from site-saturation mutagenesis (SSM) of β-GUS that was performed prior to the commencement of the β-GUS evolution. Our effort with SSM had not yielded any enhanced β-GUS variant but we could draw several conclusions about their roles of the sites. We substantiated this deduction by expanding the study to investigate the effects of targeting the same sites to enhance its promiscuous activity. This was done by conducting parallel screens on β-D-glucuronide and β-D-glucoside that led to a clearer understanding on the active site of β-GUS, particularly its tolerance towards mutations (mutational tolerance). We extended the cause of this work to study the mutational tolerance of the entire enzyme. This lends several insights on how the tetrameric, multi-domain enzyme tolerates amino acid changes. Finally, Chapter 7 provides an outlook for the glucuronylsynthase system based on the work described in Chapters 3 through 6.
1.8. References


Lutz, S. Beyond directed evolution-semi-rational protein engineering and design.


2. Materials and methods

2.1. Preamble

This chapter describes the procedures that were used to perform the experiments for this dissertation. The equipment, chemicals and reagents are also described in this chapter. A complete list of suppliers and manufacturers of equipment, kits, chemicals and reagents is attached in Appendix A and B. Recipes of culture media, buffers and other reagents used in this study are given in Appendix C. The preparation of most of these reagents was adopted from the laboratory manual by Sambrook and Russell\(^1\). These will be referred to in relevant sections.

2.2. Materials

Generally, all chemicals and reagents used for preparative work were of analytical grade. Milli-Q water (mQH\(_2\)O) was obtained from Reagent Water System (Millipore). Chromatography gradient-grade methanol was used for Liquid Chromatographic-Mass Spectrophotometry (LCMS) analyses. Solvents and buffers are passed through 0.4 \(\mu\)m filter using vacuum pump. All reagents, plastic wares and glasswares used for molecular biology work were sterilized by autoclaving using ASB270BT autoclave (Astell Scientific). Chemicals that are heat sensitive or flammable were filter sterilized using 0.2 \(\mu\)m membranes (Sartorius Minisart).

Oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, USA). Appendix D lists the primers and the corresponding chapters where they have been used. Kits that were used for preparative molecular biology work were obtained from Qiagen, Bioline and Promega. Restriction enzymes used for cloning work were purchased from New England Biolabs (NEB) Inc.

The substrates used for the directed evolution of \(\beta\)-glucuronidase (\(\beta\)-GUS) in Chapter 3 are \(para\)-\(\beta\)-D-nitrophenyl glucuronide (\(p\)NPGlcA) and 5-bromo-4-chloro-3-indolyl-\(\beta\)-D-glucuronic acid, cyclohexylammonium salt (X-gluc). \(p\)NPGlcA was purchased from Acros Organics and X-gluc was purchased from Thermo Fischer.
Scientific. The glycosyl substrate, α-D-glucuronyl fluoride (1FGlcA), and the steroid acceptor substrates – testosterone and O-carboxymethyloxime dehydroepiandrosterone (CMO-DHEA) – were synthesized by the McLeod research group (Australian National University, ANU).

2.3. Bacterial strains

Cell lines with gusA gene knocked-out were used for growth and expression. These were BW25142 and GMS407(DE3). BW25142 was the primary cell line used. It has a recA+ genotype, which eliminates the possibility of recombination between chromosomal and plasmid genes. This strain also has the non-specific endonuclease endA gene knocked-out (endA- genotype), which eliminates problems associated with plasmid preparation.

The cell line was purchased from E. coli Genetic Stock Centre (Yale) and stored as 50% glycerol stock at -80 °C. BW25142 was constructed by allele replacement technique4, which took advantage of conditional requirement of a π protein for replication. This resulted in the loss of a major portion of the gusA gene resulting from the insertion of the pir, and the disorientation of the transcription direction, which generated a gusA- strain (β-GUS incompetent).

The π factor was supplied by a vector containing the conditional replicon oriR6Kγ or by strains carrying pir gene, which transcribes the π protein4,5. A mutant pir gene (pir-116) was inserted into an M13 cloning vector that contains gusA and kanamycin resistant genes (M13KmR), but lacked oriR6Kγ replicon6. The mutation at 116 confers a higher copy number, but is otherwise similar to pir. The insertion of pir gene disrupted the gusA gene and removed three quarter of the gene (Figure 2-1). It gave rise to M13KmRΔgusA phage upon transduction with a reversed transcription direction. Transfecting the M13 into an E. coli rep mutant yields recombinant ΔgusA, KmS as a result of primary and secondary cross-over events during integration of the M13 into the E. coli chromosome4,7.
GMS407(DE3) was used initially to prepare the glucuronylsynthase enzyme for process chemistry development in the organic laboratory. The strain is derived from BL21(DE3), which has a recA+ genotype. Except for harbouring gusA+, its genotype resembles that of BL21(DE3). It is commonly used for high-level protein expression. Since it harbours a lacUV5 promoter and the gene for T7 RNA polymerase, it is compatible with T7-based expression vectors (pET vectors) (Novagen, Darmstadt, Germany). BL21 is a lysogen of bacteriophage λDE3, which contains the lacUV5 promoter and the gene for T7 RNA polymerase.

![Figure 2-1: Knocking out the gusA gene in the BW25142 cell line. The BW25142 cell line was created by inserting a pir gene between MluI at 596 and MluI at 1294 in the reverse direction.](image)

2.3.1. Preparation of electrocompetent cells.

This procedure was adopted from the laboratory manual by Sambrook and Russell. Frozen glycerol stock of bacterial cells (BW25142 or GMS407(DE3)) from -80 °C was streaked onto Luria-Bertani (LB) agar plate with no antibiotics (Appendix C). The plate was incubated at 37 °C for 16 h to allow the cells to grow. One single colony was picked and inoculated in 10 mL of LB media without antibiotics. Inoculum was grown for 16 h in an orbital shaker set at a shaking speed of 200 rpm and 37 °C. 4 mL of overnight culture was then subcultured in a 2 L baffled flask containing 800 mL of fresh of YENB (Appendix C) medium. The culture was grown in a 37 °C orbital shaker with an agitation speed of 200 rpm until the optical density at 600 nm (OD$_{600}$) reached 0.4-0.6 OD$_{600}$ reading was taken on CO8000 cell density meter (WPA Biowave). The culture was rapidly cooled to 4 °C on ice and left to stand for 30 min. Meanwhile, the centrifuge, centrifuge bottles and all solutions that would be used in
proceeding steps were pre-chilled to 4 °C in a refrigerator. The culture was then centrifuged at 4,000 rpm in pre-chilled VWR R9A rotor for 20 min at 4 °C. The clear supernatant was discarded. The cell pellet was washed by resuspending it first in 250–350 mL of chilled sterile mQH₅O in gentle swirling motion on ice. The volume was then topped-up to 800 mL with chilled sterile mQH₂O and cells were pelleted again using the same centrifugation setting. The washing step was repeated using cold sterile 10% glycerol. The cell pellet after this step was gently resuspended in 500 μL of cold sterile 10% glycerol. Homogenous cell resuspension was then distributed in 50 μL aliquots and snap frozen in ethanol dry ice bath. The tubes containing the electrocompetent cells were stored for no longer than nine months in -80 °C, and retrieved when they were needed. The competency of the cell was checked by transforming 1–10 ng of pJWL1030/gusA plasmid using electroporation method¹²,¹³. All batches of e-competent cells prepared were also checked for contamination by diluting one tube of e-competent cells in 500 μL LB and plating 50 μL on LBA¹⁰⁰ and LBK⁵⁰ agar plates.

2.3.2. Electrotransformation

Plasmid DNA made up to 10 ng/μL or double stranded DNA from ligation reactions were e-transformed into 50 μL of e-competent cells with v/v ratios of 1:50 and 1:10 respectively in 1 cm electroporation cuvette. Electroporation was performed using BioRad Micropulser™ preset at V = 2.5 kV and electroporation time τ = 5 ms. E-transformed mixtures were immediately recovered in 1 mL of YENB shaking at 200 rpm, 37 °C for 1 h. 50 μL of recovered e-transformation mixtures were then plated on LB or M9 (Appendix C) agar plates supplemented with 50 μg/mL of kanamycin (LBK⁵⁰ or M9K⁵⁰). Agar plates were then incubated overnight at 37 °C for isolation of single colonies.

2.4. Plasmids and vectors

This section describes the general procedures for plasmid manipulation, visualization and quantification. Then, a description of the vectors used in our experiments will be given. This will be followed by a description of the construction of
pJ401/gusA plasmid, which is the parent gene for the evolution and engineering program of this dissertation. The general procedures for plasmid manipulation, visualization and quantification are also covered. Lastly, we will describe the construction of pJ401/gusA.

2.4.1. Description of expression vectors

Expression vector pET28a

The expression vector pET28a(+) belongs to the series of pET28 vectors from Novagen® (Merck KGaA, Darmstadt, Germany), which are T7-based. It carries a kanamycin resistance gene, T7lac promoter and an N-terminal 6×His-Tag extension. It is a low-copy, non-leaky expression vector. The lac operon (lacUV5) and the T7 bacteriophage transcription system control its transcription. The lac operon is the weaker transcription factor that can be influenced by lactose and the T7 promoter is a strong promoter, which can be influenced irreversibly by isopropyl-β-D-1-thiogalactopyranoside (IPTG). Therefore, it is suitable for high-yield protein expression using IPTG.

Expression vector pJWL1030

In-house constitutive vector pJWL1030 modified from pCY76 and pJJKmf is a high-copy number vector. The constitutive expression cassette was isolated from pCY76, which was constructed to overexpress non-toxic genes constitutively in E. coli. The isolated expression cassette was then ligated to the backbone of pJJKmf, which conferred it with kanamycin resistance. It has a leaky lac promoter, which allows constitutive expression of a target gene that had been cloned into it. It also has a strong ribosome binding site (RBS) and translation initiation region that can improve protein expression. It is compatible with the BW25142 cell line but it does not have a His-Tag to facilitate purification using affinity chromatography method that will be described in Section 2.7.1.2. It was initially used to prepare plasmids but was later phased out when we stopped using the pET28a/GMS407(DE3) system for expression.

Expression vector pJ401

The main expression vector used was pJ401. The vector is compatible with both BW25142 and GMS407(DE3) host cells. It uses kanamycin as a selectable marker and
it is a high copy plasmid because its replication origin is pUC. Expression is repressed by the product of a lacI gene and is inducible by IPTG under the influence of a T5 promoter. Like the pET28a and pJWL1030, it has a strong ribosome binding site (RBS) to improve protein expression. The vector backbone pJ401 was obtained from pJ401/ataS (arylsulfatase from Pseudomonas Aeruginosa) plasmid that was purchased by the McLeod group from DNA2.0 Inc.

2.4.2. General restriction cloning procedures

Restriction endonuclease enzymes and calf intestinal alkaline phosphatase (CIP) were bought from New England Biolabs Inc. (NEB) and supplied by e-f freezer. T4 DNA ligase was bought from NEB, Fermentas or Invitrogen Technologies (Invitrogen).

Approximately 3–4 μg of supercoiled plasmid DNA or polymerase chain reaction (PCR) products were double digested with 10–20 U of appropriate restriction enzymes at 37 °C for 3 h. Before commencing digestion of PCR products, they were first treated with 20 U of DpnI at 37 °C for an hour to remove DNA templates. As the reaction buffer for DpnI digestion was compatible with the restriction enzymes used in this thesis, the DpnI treated samples could be used directly for double digestions. In addition, for the digestion of PCR products, an additional 2.5–5 U of restriction enzyme was supplemented after three hours of initial digestion, and the reaction continued for an additional 1 h. This step was added to ensure optimal digestion of the PCR products because the sizes of cut and uncut products of PCR products are hard to discriminate. Digestion product that would be used to provide the vector backbone was subjected to CIP treatment. Digestion reaction of the vector backbone was first stopped by heat inactivating the restriction enzymes at 65 °C for 30 min. 10 U of CIP was then added to the heat inactivated product and incubated at 37 °C for 1 h. As in the case for DpnI treatment, the reaction buffer for CIP treatment was compatible with the digestion reaction. Hence, purification was not required between the two steps.

Digested products and CIP-treated digested products were extracted from DNA agarose gels that had been subjected to DNA gel electrophoresis. Bromophenol blue (BPB) loading dye for DNA agarose gel electrophoresis (Appendix C) was added to the
DNA products in volume ratio of 1:5 (BPB:DNA). Samples were then loaded into 0.8% 1× SB-agarose gel of dimensions 125 mm x 80 mm x 30–50 mm (width x length x depth) (Appendix A) supplemented with 5 µg/L of ethidium bromide or 1x RedSafe™. Electrophoresis was performed for 40 min at 120 V and 300 mA in 1× SB buffer in a horizontal mini tank. The gels were then viewed on blue LED array box for fluorescence. The bands corresponding to the size of the desired fragments were excised and gel purified. Gel purification was performed using Promega Wizard® SV Gel and PCR Clean-Up kit. Products were eluted in 30 µL of nuclease-free water supplied in the kit.

The purified vectors and inserts were ligated using T4 DNA ligase from Thermo Scientific and NEB. A molar ratio of 3:1 (insert:vector) was used for the ligation reaction. 100 ng of the vector was used. The insert amount was calculated using the following equation (Equation 2-1):

\[
\text{100 ng of vector} \times \frac{\text{size of insert in kb}}{\text{size of vector in kb}} \times \frac{3}{1}
\]  

(2-1)

20 µL of ligation reaction mixture containing 1.5 U of T4 DNA ligase, 1× Ligase reaction buffer supplied along with the ligase, 100 ng of vector and the calculated amount of insert, and sterile water was set up. Ligation was performed at 16 °C for 18 h. Ligation product was then PCR purified using the Wizard® SV Gel and PCR Clean-Up System kit from Promega. Purified ligation products were eluted in 30 µL of water. Ligation products were e-transformed into the appropriate host cell using the protocol described in Section 2.3.2.

### 2.4.3. Cell culture and isolation of plasmid DNA

Single colonies or a stab of frozen glycerol stock was inoculated in 4 mL of LB growth media supplemented with 50 µg/mL of Kanamycin. Cells were grown overnight (18–20 h) at 37 °C in an orbital shaker with an agitation speed of 200 rpm. For cultivation of bacterial cells harboring low copy plasmid vector backbone (pET28a), 10 mL of cultures were grown in sterile 50 mL conical tubes under the same condition as the 4 mL cultures. 50% glycerol stocks of cell cultures were made for long-term storage in -80 °C by adding 625 µL of sterile 80% glycerol to 375 µL of cultures.
Supercoiled DNA plasmids were extracted from 1.5 mL (medium and high copy plasmid, pJ401 and pJWL1030) or 8 mL (low copy plasmid, pET28a) overnight cultures using QIAprep® Spin Miniprep kit (QIAGEN) or Bioline Isolate II DNA extraction kit. DNA plasmids were eluted in sterile mQH₂O and stored long-term in -20 °C freezer.

2.4.4. DNA visualization and quantification

DNA visualizations were performed using 0.8–1% agarose gel electrophoresis as described in Section 2.4.2. For visualization, gels were viewed under strong UV (254 nm) and the images were taken using UVI Pro electronic system.

Concentrations of the DNA were quantified using NanoDrop® ND-1000 spectrophotometer by observing the absorbance at 260 nm. A plasmid DNA volume of 2 μL was used for quantification. A260/A280 and A260/230 ratios were noted to ensure minimal protein and carbohydrate contamination. Ratios ~ 1.8 were taken as indication of acceptable quality.

2.4.5. Construction of plasmid pJ401/gusA

The plasmid pJ401/gusA was constructed from pET28a/gusA and pJ401/ataS with C-terminal HisTag (sulfatase gene from Pseudomonas Aeruginosa). pET28a/gusA was first constructed from pET28a/gusA(L2V) that was obtained from the McLeod group⁹. The plasmid pET28a/gusA(L2V) was first PCR amplified and reversed engineered (V→L) by standard polymerase chain reaction (PCR) as described in Section 2.5.1 using high-fidelity Phusion-HF polymerase and oligodeoxy-nucleotide primers carrying NdeI and EcoRI restriction endonuclease sites – pgusf (5’–cgcggccagcataatgttacgtcctgtagaaacc–3’) and pgusr (5’–ggtggtggtggaattctcattgtttgcc–3’) (restriction sites underlined) – to obtain pET28a/gusA. The PCR product corresponding to the band fragment at ~ 1.8kb was purified by 1% DNA agarose gel electrophoresis. Purified PCR products and pET28a/gusA(L2V) were double digested with NdeI and EcoRI. Double digestion product of pET28a/gusA(L2V) was then CIP-treated. Ligation was performed as described in Section 2.4.2. Ligation products were e-transformed into BW25142. Eight single colonies were picked and their plasmids were isolated. Clones
harbouring pET28a/gusA were used for the construction of pJ401/gusA with N-terminal HisTag. Plasmids pJ401/ataS and pET28a/gusA were double digested in XbaI and EcoRI and cloned according to the cloning procedures described in Section 2.4.2. Figure 2-2 maps the plasmids of pET28a/gusA, pJ401/ataS and the resulting pJ401/gusA, and the positions of NdeI, EcoRI, XbaI and the 6×HisTag in the plasmid.

Figure 2-2: Plasmid maps of pDNAs used for pJ401/gusA construction. Plasmid maps of pET28a/gusA, pJ401/ataS and resulting pJ401/gusA. The fragment between XbaI and EcoRI in pET28a/gusA contains the RBS, 6×-HisTag and the gene. This fragment was inserted into pJ401 between XbaI and EcoRI to replace the ataS gene.

2.4.6. Construction of plasmid pJ401/syn

Plasmid pJ401/syn was constructed using the dual-tube megaprimer method that will be described in Section 2.6.2. The primers used to generate two types of glucuronylsynthase (rare and common) in Chapter 5 are GUS-E504G(common) R’ (5’–ccgcgtatccggtgatgataatcggc–3’) and GUS-E504G(rare) R’ (5’–acgcgtatccggtgatgataatcggc –3’); underlined are the codons encoding glycine.
2.5. Polymerase chain reaction (PCR)

2.5.1. Primer design

Primers of ~ 30 base pairs (bp) in length were designed with the help of OligoCalc Properties Calculator, a web-based computational software. This software computes the melting temperature $T_m$ and primer GC content. The $T_m$ of primers were around 50–60 °C and the GC contents were around 50%. Where primer pairs were used, they were partially overlapped but fully complementary. The forward and reverse primers for assembly PCR that were used for library construction of targeted random mutagenesis were designed so that the adjoining fragments had significant overlap. The primers used are listed in Appendix D.

2.5.2. Standard Polymerase Chain Reaction (PCR)

Standard PCR was performed to amplify a gene for the purpose of checking new constructs, to create flanking non-mutagenic PCR fragments for library construction and to check for the presence of inserts in DNA plasmids. Colony PCR was performed to check for the presence of inserts after constructs were made. For amplification to build new constructs or libraries, high-fidelity polymerase Phusion-HF (Finnzyme, NEB) was used. Taq polymerase (Roche) or BioTaq™ (Bioline) were used to check for the presence of inserts in plasmids or in colony PCR.

Standard PCR was carried out in reaction mix of 50 μL total volume containing 25–50 ng plasmid DNA, 1.25 μM of forward and reverse primer, 200 μM dNTP mix, 1× Taq reaction buffer containing 1.5 mM MgCl₂, and 2.5 U of DNA polymerase. PCR cycling was done on any of the thermocyclers listed in Appendix A. Reaction mixture was heated at 95 °C for 2 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and extension at 68 °C or 72 °C for 1–2 min (30 s/kb for Phusion-HF and BioTaq, and 1 min/kb for Taq). A final extension at 72 °C was performed for 10 min and the PCR products were stored at 4 °C. For colony PCR, 1 μL of resuspended cell mixture prepared by resuspending one single colony in 10 μL of sterile water was used as the template in place of plasmid DNA. The protocol used was
similar to the basic PCR except that a longer initial denaturation step of 5 minutes was used to ensure lysis of the cell. Table 2-1 lists the cycling conditions.

### Table 2-1: Thermo cycling conditions for standard and colony PCR.

<table>
<thead>
<tr>
<th>Step</th>
<th>Conditions</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial denaturation</strong></td>
<td>95 °C or (98 °C for Phusion)</td>
<td>2 – 5 mins</td>
</tr>
<tr>
<td><strong>Denaturation</strong></td>
<td>95 °C or (98 °C for Phusion)</td>
<td>30 s</td>
</tr>
<tr>
<td><strong>Annealing</strong></td>
<td>Average Tₘ of primers ~ 55°C</td>
<td>30 s</td>
</tr>
<tr>
<td><strong>Extension</strong></td>
<td>72 °C or (68 °C for Phusion)</td>
<td>1 min/kb for Taq amplification, 30 s/kb for Phusion and BioTaq™ amplification</td>
</tr>
<tr>
<td><strong>Final extension</strong></td>
<td>72 °C or (68 °C for Phusion)</td>
<td>10 min</td>
</tr>
<tr>
<td><strong>Hold</strong></td>
<td>72 °C or (68 °C for Phusion)</td>
<td>∞</td>
</tr>
</tbody>
</table>

#### 2.5.3. DNA Sequencing

DNA sequencing was done using Sanger sequencing method. Each PCR reaction consisted in 20 μL mixtures of sterile mQH₂O, 150–350 ng of plasmid DNA, 3.2 pmol primer, 1 μL BigDye® Terminator and 1× sequencing buffer supplied by the ACRF Biomolecular Resource Facility (BRF), The John Curtin School of Medical Research (JCSMR) in the Australian National University (ANU). The cycling conditions are listed in Table 2-2.

### Table 2-2: Thermo cycling conditions for Sanger sequencing.

<table>
<thead>
<tr>
<th>Step</th>
<th>Conditions</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial denaturation</strong></td>
<td>95 °C</td>
<td>5 min</td>
</tr>
<tr>
<td><strong>Denaturation</strong></td>
<td>96 °C</td>
<td>10 s</td>
</tr>
<tr>
<td><strong>Annealing</strong></td>
<td>50 °C</td>
<td>5 s</td>
</tr>
<tr>
<td><strong>Extension</strong></td>
<td>60 °C</td>
<td>4 min</td>
</tr>
<tr>
<td><strong>Hold</strong></td>
<td>4 °C</td>
<td>∞</td>
</tr>
</tbody>
</table>

PCR products were cleaned up using ethanol precipitation method detailed in the guidelines given by BRF. Sequencing reaction clean-up reagent was prepared fresh by mixing 62.5 μL 95% ethanol, 3 μL 3 M sodium acetate pH 4.6 and 14.5 μL of sterile mQ H₂O. 80 μL of the sequencing clean-up reagent was mixed thoroughly with 20 μL of sequencing PCR products. The mixture was then left to stand at room temperature (RT), in the dark, for 15 min. DNA is pelleted by maximum speed centrifugation for 30 min at RT. The supernatant was then discarded. Pellets were washed with 200 μL of freshly prepared 70% ethanol and centrifuged at maximum speed for 20 min. The supernatant was removed and air-dried in the dark for 1–2 hours. Cleaned-up
sequencing samples were then sent to BRF for sequence processing. The DNA sequencing chromatograms were then viewed and analyzed using Sequencher 4.10.1 (Gene Codes Corp).

2.6. Library construction and enzyme engineering methods

This section describes the procedures to engineer sites specifically or to generate libraries randomized at single sites (site-saturation mutagenesis) or random libraries of a region in the gene. This is followed by a description of library construction methods for the introduction of random point mutations using various random mutagenesis methods. This includes the fragment assembly procedure that was used to assemble fragments for targeted random mutagenesis. The post processes (e-transformation and analysis of mutant libraries) is described at the end of this section.

2.6.1. Site-specific mutagenesis

Stratagene’s QuikChange® method (QCM) was used to introduce site-specific mutations for rational design of variant T509A (Chapter 3) and the codon bias usage experiment of E504G (Chapter 4). High fidelity polymerase Phusion-HF was used for amplification. Reaction mixture of 50 μL containing 25 ng of plasmid template, 1 μM of forward and reverse primer, 200 nM of dNTP, 1× Phusion-HF buffer containing 1.5 mM MgCl₂ and 2 U of Phusion-HF was first set up on ice. The reaction was then carried out in a thermocycler using the cycling conditions listed in Table 2-3.

<table>
<thead>
<tr>
<th>Initial denaturation</th>
<th>98 °C</th>
<th>2 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 cycles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>98 °C</td>
<td>30 s</td>
</tr>
<tr>
<td>Annealing</td>
<td>58 °C (± 3 °C calculated T_m)</td>
<td>60 s</td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>5 s</td>
</tr>
<tr>
<td></td>
<td>68 °C</td>
<td>20 s/kb</td>
</tr>
<tr>
<td>Final extension</td>
<td>68 °C</td>
<td>10 min</td>
</tr>
<tr>
<td></td>
<td>72 °C</td>
<td>10 min</td>
</tr>
<tr>
<td>Hold</td>
<td>4 °C</td>
<td>∞</td>
</tr>
</tbody>
</table>

PCR product was purified using Wizard® SV Gel and PCR Clean-Up System kit from Promega and eluted using nuclease-free water. Wild-type plasmid template was
then removed by treating the purified sample with 10 U of \textit{DpnI} at 37 °C for 3 h. 2 U was added and the treatment was continued for a further 3 h at 37 °C. \textit{DpnI} treated product was PCR-purified and eluted in 30 μL of nuclease-free water. 5 μL of the purified PCR product was e-transformed into BW25142 e-competent cells.

\section*{2.6.2. Site-saturation and site-directed mutagenesis}

Site saturation mutagenesis was performed to create single-site saturation libraries used for the study of mutational tolerance in \textit{Chapter 6}. The mutagenic primers were encoded with NNK codon degeneracy (N = all four deoxyribonucleotides adenine (a)/cytosine (c)/guanine (g)/thymine (t), and K = g/t). Two megaprimer PCR methods developed separately by Fang’s (dual-tube) and Reetz’s (single-tube) groups were attempted.

The first method attempted was the single-tube method reported by Sanchis \textit{et al.}\textsuperscript{20} from Reetz’s group. For this method, the PCR reactions were performed in single tubes using a 2-stage amplification protocol. Reactions were set up in final volumes of 50 μl containing 50–100 ng of template and 5 pmol of forward and reverse primers, 200 nM of each dNTP and 1 U of \textit{Phusion-HF} in 1× \textit{Phusion-HF} buffer containing 1.5 mM MgCl\textsubscript{2}. The amplification protocol was as such: initial denaturation at 98 °C for 3 min, 5 cycles of 30 s denaturation at 98 °C and 30 s annealing at 55 °C (± 3 °C of the calculated T\textsubscript{m} of the oligonucleotides designed), extension 30 s/kb according to the megaprimer size at 72 °C. This was followed by 20 cycles of initial denaturation and extension using cycling conditions as described for the first stage. A final extension of 2 min/kb at 68 °C was performed. The mutation frequency obtained was around 50–60% and the library sizes generated from this method were between 100–300 transformants. This was not very efficient and attempts to optimize various parameters including using \textit{Pfu} polymerase and increasing the extension duration of the thermo cycling conditions did not improve the efficiency of the method significantly. Hence, we attempted the dual-tube method developed by Tseng \textit{et al.}\textsuperscript{21}

The dual-tube method was adapted to generate libraries with mutagenesis efficiencies between 75% and 90%. The library sizes generated using this method were between 2000 and 4000 transformants. Reactions were set up in final volumes of 50 μl
containing 50–100 ng of wild-type plasmid DNA template and 1 pmol of each mutagenic primer and flanking primer, 200 nM of each dNTP and 1 U of Phusion-HF in 1× Phusion-HF buffer containing 1.5 mM MgCl₂. Thermocycling was performed in two stages. After the first stage, 2.5 μL of the reaction mixture was added to another tube 47.5 μL of the following mixture: 25 ng of wild-type plasmid DNA template, 200 nM of each dNTP and 1 U of Phusion-HF in 1× Phusion-HF buffer containing 1.5 mM MgCl₂. The cycling conditions are summarized in Table 2-4.

Table 2-4: Thermo cycling conditions for site-saturation and site-directed mutagenesis.

<table>
<thead>
<tr>
<th>Initial denaturation</th>
<th>98 °C</th>
<th>2 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 cycles (Stage I)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>98 °C</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>55–58 °C (± 3 °C calculated Tₘ)</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>30 s</td>
</tr>
<tr>
<td>Hold until Stage II</td>
<td>4 °C</td>
<td>60 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30 s/kb</td>
</tr>
<tr>
<td></td>
<td></td>
<td>megaprimer</td>
</tr>
<tr>
<td>25 cycles (Stage II)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>98 °C</td>
<td>30 s</td>
</tr>
<tr>
<td>Annealing and extension of megaprimer</td>
<td>68 °C</td>
<td>20 min</td>
</tr>
<tr>
<td>Final extension</td>
<td>68 °C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72 °C</td>
<td></td>
</tr>
<tr>
<td>Hold</td>
<td>4 °C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>∞</td>
</tr>
</tbody>
</table>

PCR product was purified using Wizard® SV Gel and PCR Clean-Up System kit from Promega and eluted using nuclease-free water. Wild-type plasmid template was then removed by treating the purified sample with 10 U of DpnI at 37°C for 3 h. 2 U was added and the treatment was continued for a further 3 h at 37 °C. DpnI treated product was PCR-purified and eluted in 30 μL of nuclease-free water. 5 μL of purified PCR product was e-transformed into BW25142 e-competent cells.

2.6.3. Error-prone PCR (ePCR)

Random mutations were introduced in the gene using error-prone PCR (ePCR) with unbalanced concentrations of MgCl₂ and MnCl₂ and high amounts of low-fidelity Taq polymerase. The ePCR protocol was adapted from Cadwell et al. and Arnold et al. 5 mM of MgCl₂ was used compared to the concentration of 1.5 mM in standard PCR reactions to stabilize non-complementary pairs that resulted from random introduction of mutations. A higher amount of Taq polymerase was also used to encourage nucleotide extension beyond the positions of the base mismatch. MnCl₂
was used to reduce the specificity and fidelity of the DNA polymerase, thereby increasing the error rates.

An ePCR reaction was set up in the PCR reaction buffer without MgCl\textsubscript{2} supplied by Roche at a final concentration of 1×. Each reaction consisted of 50 ng of plasmid DNA, 1.25 μM of forward and reverse primers each, 5 mM of MgCl\textsubscript{2}, 0.2 mM of each dNTP mix and 5 U of Taq polymerase (Roche). 0.05–0.5 mM of MnCl\textsubscript{2} was added depending on the desired mutation rate. The cycling conditions were identical to the standard PCR cycling parameters described in Section 2.5.2.

2.6.4. Staggered extension process (StEP)

StEP is an in vitro recombination process that was first described by Zhao et al.\textsuperscript{24,25} A short extension time was used to generate short fragments from a pool of mutant DNA due to incomplete extension, which were then used as annealing templates in the subsequent cycle. This allowed for template swapping. Mutations from different generations or fragments were recombined using this procedure. Proofreading polymerases Phusion-HF and Pfu were used to maintain the fidelity during shuffling. A StEP reaction consisted of a pool of plasmid DNA in equal proportion (total amount 50 ng), 1.25 μM of forward and reverse primers each, 5 mM of MgCl\textsubscript{2}, 0.2 mM of each dNTP, 1.5 mM MgCl\textsubscript{2} (Phusion) or 2.5 mM MgSO\textsubscript{4} (Pfu) and 1 U of Phusion-HF or 2.5 U of Pfu polymerase (Roche). The cycling conditions are listed in Table 2-5.

Table 2-5: Thermo cycling conditions for StEP.

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94 °C (or 98 °C for Phusion)</td>
<td>5 min</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94 °C (or 98 °C for Phusion)</td>
<td>10 s</td>
</tr>
<tr>
<td>Annealing</td>
<td>55 °C</td>
<td>5 s</td>
</tr>
<tr>
<td>Extension</td>
<td>72 °C</td>
<td>3-5 s</td>
</tr>
<tr>
<td>Final annealing</td>
<td>55 °C</td>
<td>2 min</td>
</tr>
<tr>
<td>Final extension</td>
<td>72 °C</td>
<td>10 min</td>
</tr>
</tbody>
</table>

2.6.5. Staggered extension process with error-prone steps (StEP-ePCR)

StEP-ePCR combines shuffling and the introduction of random mutations in one reaction. The PCR reactions were set up as described for StEP procedure. The cycling conditions were identical to ePCR with the exception that an extension time of 5–8 s
rather than 30 s was used. The extension time depended on the fragment shuffled. Longer extension time was used to shuffle longer fragments.

2.6.6. **Assembly PCR for targeted random mutagenesis**

Assembly PCR was performed to assemble the mutagenic and non-mutagenic PCR products for the creation of targeted random mutagenesis libraries. Non-mutagenic PCR products were created using the standard PCR method described in Section 2.5.2. These would flank the mutagenic fragment on either or on both sides. The 5’–flank was created using pgus401f (Appendix D) as the forward primer and a reverse primer that partially overlapped the 5’–end of the mutagenic fragment. The 3’–flank was created using pgusr as the reverse primer and a forward primer that partially overlapped the 3’–end of the mutagenic primer. Mutagenic fragments were created using ePCR, StEP or StEP-ePCR described above. When the mutagenic fragment was found at either end of the gene, the forward or reverse primer used in the mutagenic PCR steps was pgus401f or pgusr while the other primer would correspond to the start or the end to the targeted region. **Figure 2-3** illustrates the assembly process for two cases: when the mutagenic fragments are between non-mutagenic fragments, and when the mutagenic fragment was found at one end.

The fragments were separated using DNA gel electrophoresis, excised and gel purified using Wizard® SV Gel and PCR Clean-Up System kit from Promega. Fragments that were 300–1000 base pairs in length were separated on 1% SB-agarose. Fragments that were 100–300 base pairs in length were separated on 1.6% SB-agarose.

Assembly reactions were set up using 20 ng of each gel purified fragments from the first PCR reactions as the templates. The templates were annealed by the overlap between the fragments. Amplification of the annealed products was initiated by forward primer for the 5’ fragment pgus401f and pgusr primer (reverse primer for the 3’ fragment). The extension step in the assembly PCR reaction was carried out by 2 U of *Pfu* polymerase and 1× *Pfu* buffer containing 2.5 mM MgSO₄. The cycling conditions are listed in **Table 2-6**.
Table 2-6: Thermo cycling conditions for StEP-ePCR.

<table>
<thead>
<tr>
<th>Step</th>
<th>Conditions</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial denaturation</strong></td>
<td>95 °C</td>
<td>3 mins</td>
</tr>
<tr>
<td><strong>30–35 cycles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>94 °C</td>
<td>15 s</td>
</tr>
<tr>
<td>Annealing</td>
<td>Average T&lt;sub&gt;m&lt;/sub&gt; of primers ~ 55 °C</td>
<td>40 s</td>
</tr>
<tr>
<td>Extension</td>
<td>68 °C, 72 °C</td>
<td>2 min (~1 min/kb)</td>
</tr>
<tr>
<td>Final extension</td>
<td>68 °C, 72 °C</td>
<td>10 min</td>
</tr>
<tr>
<td>Hold</td>
<td>4 °C</td>
<td>∞</td>
</tr>
</tbody>
</table>

Figure 2-3: Schematic illustration of PCR approaches to generate random libraries. epPCR: error-prone PCR, nmPCR: non-mutagenic PCR. Target regions for random mutagenesis are illustrated in light blue and the error-prone PCR amplification products of these regions are illustrated in the same light blue, with dotted foreground. Flanking regions for nmPCR are illustrated in dark blue. Overlapping regions are shaded in diagonal stripes. Arrows depict the forward (→) and reverse (←) primers used to generate the fragments for assembly PCR. Vertical errors represent the restriction sites. (a) The megaprimer approach is performed in two PCR reactions, and (b-c) the triple step procedure involving separate generation of error-prone and non-mutagenic products and subsequent assembly of the fragments. Two different assembly methods are shown—one involving the assembly of two fragments (shown in (b)), which is useful for targeted mutagenesis at a terminal region. The other, shown in (c) assembles three fragments and applies to libraries directing evolution within the gene. In green: vector backbone.
Figure 2-4 shows the failed attempt to generate sufficient products from whole-gene megaprimer method (Figure 2-4a), a failed assembly PCR attempts (Figure 2-4b) and successful attempts (optimized conditions) in generating assembly products for library generation (Figure 2-4(c-d)) using mutagenic fragments from various concentrations of MnCl$_2$ (Figure 2-4d). Assembled products corresponding to the band at ~ 1.8 kb were then gel purified (Figure 2-4(c-d)). They were then cut with NdeI and EcoRI restriction endonucleases and cloned into pJ401(+6×HisTag) vector backbone.

![Figure 2-4: DNA visualization of PCR amplification products for targeted random mutagenesis. (a) Inefficient mutagenic PCR in generating whole-gene from megaprimers with random mutations. (b) Failed attempt at assembly PCR reaction produced fragments of various sizes. (c) Successful attempt at assembly PCR reaction with a desired band at 1.8 kb. (d) Assembled PCR product and mutagenic fragments (~ 700 bp) generated using different concentrations of mutagen MnCl$_2$.](image)

2.6.7. E-transformation for library generation

Ligation or PCR products were e-transformed into BW25142 e-competent cells according to the procedures in Section 2.3.2. Recovered e-transformation mixtures were pelleted and resuspended in 800 µL sterile 10% glycerol. Library glycerol mixtures were then snap frozen in 100 µL aliquots and kept in -80 °C. These were thawed when needed for library growth and expression.

2.6.8. Analysis of mutant library

Colony PCR or basic PCR was performed to ensure the presence of plasmids harbouring the mutant genes. The average mutation frequencies of the libraries were estimated by sequencing eight to twelve randomly selected colonies. The mutation frequency was expressed as the means and standard deviation of the number of mutations found.
2.7. Preparation of lysates and assays for β-GUS library screening

This section describes the library screening procedures for β-GUS engineering and directed evolution. Two different methods for the cultivation and preparation of cell lysates are first given. This is followed by a description of the X-gluc colony agar plate screening assay, the pNPGlcA screening assay in the presence and absence of t-BuOH, and para-nitrophenyl-β-D-glucoside (pNPGlc) screening assay.

2.7.1. Cultivation and preparation of lysates for β-GUS screen

Libraries for directed evolution were grown in sterile 96-well microtitre plates by inoculating single colonies per well or using the culture dilution method developed by Dr. Bradley Stevenson during his time in this laboratory\(^\text{26}\). Generally, overnight cultures of libraries were grown in sterile 96-well microtitre plates at 37 °C for 16–18 hours in an orbital shaker shaking at 200 rpm (pre-culture plates). 10 µL of overnight cultures were then inoculated in 150 µL of LBK\(^\text{50}\) and grown for 3 h at 37 °C, shaking at 200 rpm, to a confluence of ~ 0.6 OD\(_{600}\). 50 µL of 2 mM IPTG was added into the sub-saturated culture. IPTG induction was performed at 30 °C overnight in an orbital shaker set to a speed of 180 rpm. Subculture plates were then frozen at -80 °C.

2.7.1.1. Single colony method

In the single colony method, single colonies were picked and inoculated into 96-well round bottom microtitre plates containing 200 µL of LBK\(^\text{50}\). Libraries were grown at 37 °C in orbital shakers shaking at 200 rpm for 16–18 h. 10 µL of overnight cultures were then sub-cultured and induced with 0.5 mM IPTG at 30 °C for 12–14 h. Pre-culture plates were then stored at 4 °C so that potential hits can be isolated.

2.7.1.2. Culture dilution method

The culture dilution method was employed to screen ePCR and ePCR-StEP libraries where the library sizes were large. For this method, the cell density of the recovered e-transformation mixture was first determined. Serial dilution mixtures 1:1, 1:10, 1:20 and 1:50 of a 10% library transformant glycerol aliquot were made and 50 µL of each diluted mixtures was propagated on M9 agar plates supplemented with
50 μg/mL of kanamycin (M9K\(^{50}\)). The number of colony forming units (cfu) per μL of transformation mixtures were enumerated from the dilutions that gave 20–300 cfu and averaged. The remaining library of transformants was then diluted to 1 or 2 cfu per 100 μL of LBK\(^{50}\). Diluted cell culture was dispensed into sterile 96-well microplates so that each well contained 200 μL of dilution mixtures (2–4 cfu per well). These plates were the primary screening plates. The number of sterile wells (wells that did not contain any colonies) after overnight growth was enumerated. The number of colonies per well was confirmed by substituting this number into the Poisson distribution function given by Equation 2-2.

\[
P_r = \frac{e^{-\mu} \mu^r}{r!}
\]  
(2-2)

where \(P_r\) is the probability of a well having \(r\) number of cells when the mean number of cells per well is \(\mu\). The actual mean number of cells per well was calculated based on the number of sterile wells, \textit{i.e.} when \(r = 0\). The probability of encountering a well with \(r = 0–10\) transformants for a library grown with mean number of cells per well, \(\mu = 4\), is plotted in Figure 2-5a. The cumulative distribution function is plotted in Figure 2-5b.

![Figure 2-5: Poisson distribution of cells in liquid culture when \(\mu = 4\). (a) The distribution of cells containing \(r\) number of cells in a well, and (b) the cumulative distribution of the distribution.](image)

Secondary screen was performed on the wells from primary screening that harboured potential hits. Culture from selected wells were streaked and grown on M9K\(^{50}\) plates. Single colonies were then picked and grown in 200 μL LBK\(^{50}\) as in the single-colony method. The number of colonies to be inoculated from each well depended on the mean number of cells per well in the primary screen. As illustrated in Figure 2-5b, for a mean of four cells per well, ~ 99% of wells would contain nine or fewer clones. To maximize the chance of picking all clones present in a selected
primary well, ten colonies from each primary wells were isolated and inoculated for secondary screen.

2.7.2. **β-GUS blue-white screen**

X-gluc colony agar screen was used to check if clones were active or inactive. An aliquot of culture containing 300 transformants was evenly distributed on M9K agar plates supplemented with 10 μM of IPTG to induce expression and 4–8 mg/100 mL of X-Gluc. 4–8 mg/mL of X-gluc solutions were prepared in neat DMSO of molecular biology grade and then mixed in molten M9 agar solution containing 50 μg/mL of kanamycin and 10 μM of IPTG. Colonies were grown at 37 °C overnight. Colonies that turned blue were considered as clones that have β-GUS activity.

2.7.3. **β-GUS buffer screening assay**

Thawed culture from subculture plates was diluted 20-times (10 μL into 190 μL) in 50 mM NaPi, pH 7.4 buffer (Appendix C). Diluted subculture was then further diluted eight times in 70 μL of 50 mM NaPi, pH 7.4 buffer. Cells was lysed with 80 μL of 0.025 kU/μL of rLysozyme® in 50 mM NaPi, pH 7.4. Lysis was performed at room temperature on plate rocking shakers set at a speed of 1,000 rpm/s for 15 min. Whole cell lysates were kept temporarily on ice to ensure that they were not inactivated. If plates were used to compare with tert-butanol (t-BuOH) screen for residual activity, cell debris were pelleted down by centrifugation at 3,700 rpm for 5 min at room temperature. Whole cell lysates or cleared cell lysates were kept temporarily on ice to ensure that they were not inactivated.

Whole cell lysate or clarified cell lysate was added to the assay buffer (50 mM NaPi, pH7.4) at a 1:10 v/v ratio. To start the reaction, 100 μL of 2–4 mM pNPGlcA prepared in the assay buffer was dispensed into each well using an 8-channel micropipette. The progress of reaction was followed using wavelength 405 nm on SpectraMax® M2e microplate reader.
2.7.4. β-GUS tert-butanol (t-BuOH) screening assay

Thawed culture from subculture plates was diluted 20-times (10 µL into 190 µL) in 50 mM NaPi, pH 7.4 buffer. The diluted culture was then further diluted 2–3 times with 40 µL of 50 mM NaPi, pH 7.4 buffer. Cells was lysed win 80 µL of 0.025 kU/µL of rLysozyme® in 50 mM NaPi, pH7.4. Lysis was performed at room temperature on plate rocking shakers set at a speed of 1, 000 rpm/s for 15 min. Cell debris were pelleted down by centrifugation at 3, 700 rpm for 5 min at room temperature. Cleared cell lysates were kept temporarily on ice to ensure that they were not inactivated.

Cleared cell lysate was added to 80 µL of assay buffer containing 25 or 31.25% t-BuOH in 50 mM NaPi, pH7.4 at a v/v ratio of 1:4. Samples were incubated at 28 °C (room temperature) for 20 min. To start the reaction, 100 µL of 2–4 mM pNPGlcA prepared in assay buffers containing 20% or 25% t-BuOH was dispensed into each well using 8–channel micropipette. The progress of reaction was followed for 10–20 minutes using wavelength 405 nm on SpectraMax® M2e microplate reader. For residual activity screening, 20 µL of t-BuOH incubated sample was diluted in 80 µL of 50 mM NaPi, pH7.4 buffer. The reaction was initiated by adding 100 µL of 2–4 mM pNPGlcA prepared in buffer without t-BuOH and the activities with and without t-BuOH were compared.

2.7.5. β-GUS and β-glucosidase screening assay for mutational tolerance study

Thawed culture from subculture plates was diluted 20-times (10 µL into 190 µL) in 50 mM NaPi, pH 7.4 buffer. Cells was lysed in 100 µL of 0.025 kU/µL of rLysozyme® in 50 mM NaPi, pH7.4. Lysis was performed at room temperature on plate rocking shakers set at a speed of 1, 000 rpm/sec for 15 min and then kept on ice temporarily. For β-GUS assay, whole cell lysates were further diluted five-times and then added to 80 µL of buffer at a v/v ratio of 1:4. The reaction was initiated by adding 100 µL of 1.6 mM pNPGlcA. The β-glucosidase assay was carried out using 100 µL whole cell lysate and 100 µL of 9 mM pNPGlc. Progress of the reactions were followed at 405 nm on SpectraMax® M2e microplate reader.
2.8. Preparation of lysates and assays for glucuronylsynthase library screening

This section describes the library screening procedures for glucuronylsynthase engineering and directed evolution. The method for library cultivation and preparation of cell lysates are first given. This is followed by a description of the screening assay in 20% \( t \)-BuOH to isolate potential hits that may be more kinetically competent.

2.8.1. Cultivation and preparation of lysates for glucuronylsynthase screen

Single colonies were picked from M9K\(^{50}\) agar plates supplemented with IPTG and X-gluc as described in Section 2.7.2. Single colonies were inoculated into 96-well round bottom microtitre plates containing 200 \( \mu \)L of LBK\(^{50}\). Libraries were grown at 37 °C in orbital shakers shaking at 200 rpm for 16–18 h. Preculture plates were then stored at 4 °C so that potential hits can be isolated. Subculture plates were made by transferring 40 \( \mu \)L preculture into 360 \( \mu \)L 2xYT\(^{50}\) (Appendix C) and then induced with 0.5 mM IPTG at 30 °C for 12–14 h in sterile deep well-plates. Subculture was first grown for 3 h at 37 °C, shaking at 200 rpm to a confluence of \( \sim 0.6 \) OD\(_{600}\) and then adding 100 \( \mu \)L of 2 mM IPTG. IPTG induction was performed in an orbital shaker set to a speed of 180 rpm. Induced cell culture was then harvested at maximal speed for 20 min and frozen at -80 °C.

2.8.2. Glucuronylsynthase \( t \)-butanol (\( t \)-BuOH) screening assay

Into each well, 100 \( \mu \)L of 50 mM 3-(N-morpholino)propanesulfonic buffer (MOPS-NH\(_4^+\)), pH 7.4 buffer (Appendix C) was added to resuspend the cell pellets. Cells was lysed by adding 100 \( \mu \)L of 0.05 kU/\( \mu \)L of rLysozyme\(^\oplus\), prepared in the assay buffer, into the resuspension mixtures. Lysis was performed at room temperature on plate rocking shakers set at a speed of 1, 000 rpm/sec for 20 min. Cell debris was pelleted down by centrifugation at 3, 700 rpm for 10 min at room temperature. Cleared cell lysates were kept temporarily on ice to ensure that they were not inactivated.

Cleared cell lysates were transferred to reaction assay plates (96-deep well plates). 400 \( \mu \)L of reaction mixture containing 2 mM acceptor substrate CMO-DHEA (\( \sim 0.6 \)
mg/mL) and 6 mM α-D-glucuronyl fluoride as the glucuronyl donor substrate was dispensed into each well containing the cell lysates. The substrates were prepared in the assay buffer containing 21% t-BuOH so that the final concentration of t-BuOH in the reaction was 20%. The reaction was performed for 14–18 h in a plate incubator shaker set at 30 °C with agitation speed of 300 rpm. 100 µL of the reactions were quenched with 50% MeOH. The products in the quenched reaction were monitored using liquid chromatography-mass spectrophotometry (LCMS). The peak areas of product glucuronides were determined using M/Z value of 546.30 (CMO-DHEA glucuronide) or 463.20 (testosterone glucuronide). The peak area of each well relative to the average peak areas of the positive controls (glucuronylsynthase parent) was then computed to isolate potential hits.

2.9. Translational misincorporation study: β-GUS assay to investigate effects of expression condition

The plasmid pJ401/syn was e-transformed into BW25142 e-competent cells and propagated on M9K\textsuperscript{50} agar plates. Colonies were grown at 37 °C for 16 h. Single colonies were picked and inoculated into 10 mL of LBK\textsuperscript{50}. Starter culture was grown at 37 °C for 16 h, shaking at 200 rpm. 100 µL of starter cultures were subcultured in 50 mL of 2xYTK\textsuperscript{50} in sterile 150 mL Erlenmeyer flasks for 3 h at 37 °C until the OD\textsubscript{600} reached ~ 0.6–0.8. 0.5 mM IPTG was then added to induce the cells at 18 °C, 30 °C and 37 °C for a period of 4 h, 8 h and 16 h with agitation at 200 rpm on orbital shakers. Cells were then harvested and the cell pellets were frozen at -80°C.

Cell pellets were resuspended in 1 mL of 50 mM NaPi, pH 7.4. BugBuster® 10× Protein Extraction Reagent (Merck, EMD) was diluted with 62.5 mM NaPi, pH 7.4 to obtain 2× BugBuster\textsuperscript{TM}. 2× BugBuster\textsuperscript{TM} was added to the cell resuspension in a volume ratio of 1:1. Lysis was performed at room temperature on orbital shakers by gentle shaking at speed ~ 50 rpm.

β-GUS activity of the crude lysate samples was measured using 150 µL of crude lysate from the 4 h expression trial and 150 µL of 1/10 diluted crude lysate samples from the 8 h and overnight expression trials. Reactions were initiated with 100 µL of
2 mM pNPGlcA. The progress of the reaction was followed at wavelength 405 nm on SpectraMax® M2e microplate reader over a period of three days at 25 °C.

Bovine Serum Albumin (BSA) standard curve was made using 0.0625–2.5 g/L of BSA in buffer. 5 µL of prepared BSA solution was added into 100 µL of mQH₂O and 100 µL of 20% Bradford reagent (Appendix C). The components were mixed thoroughly and left to stand at room temperature for ~ 5 min. Endpoint absorbance was taken at 595 nm on the SpectraMax® M2e microplate reader. Crude enzyme lysate was diluted in mQH₂O and the amount of crude enzymes was determined in the same way as the BSA standard curve determination. Diluted crude enzyme lysates (10–20 µL) were also analyzed on SDS-PAGE according to the procedures in Section 2.10.6.

2.10. Protein expression and purification

This section describes the protein expression and purification method used to purify enzymes for kinetics and biophysical characterization. The growth and expression method for the β-GUS recombinants and glucuronylsynthase recombinants are first described in separate sections. This is followed by a description of the lysis procedure. Two purification methods are described- Nickel affinity purification using 5 mL HisTrap FF columns and size exclusion chromatography (Superdex S-200).

2.10.1. Growth and protein overexpression for β-GUS recombinants

Plasmid DNAs of β-GUS parent or β-GUS variants were e-transformed into BW25142 e-competent cells and propagated on M9K⁵₀ agar plates. Colonies were grown at 37 °C for 16 h. Single colony was picked and inoculated into 10 mL of LBK⁵₀. Starter culture was then grown at 37°C for 16 h, shaking at 200 rpm. 1.6 mL of starter culture was inoculated into 500 mL of 2xYTK⁵₀ in 2 L sterile baffled flasks. Subculture was grown for 3 h until it reached OD₆₀₀ of ~ 0.6–0.8. Subculture was then induced with 0.5 mM IPTG and left to continue shaking at 30 °C, 200 rpm for 16 h in water bath orbital shakers.
2.10.2. Growth and protein overexpression for glucuronylsynthase recombinants

Plasmid DNAs of glucuronylsynthase (Syn) parent or glucuronylsynthase variants were e-transformed into BW25142 e-competent cells and propagated on M9K\textsuperscript{50} agar plates supplemented with 8 mg/100 mL of X-gluc. Colonies were grown at 37 °C for 16 h. Singles colonies were picked and inoculated into 10 mL of LBK\textsuperscript{50}. Starter culture was grown at 37 °C for 16 h, shaking at 200 rpm. Subculture was prepared by inoculating 3.2 mL starter culture into 1 L of 2xYTK\textsuperscript{50} in 2 L sterile baffled flasks and allowing the subculture to grow for 3 h until it reached a confluence of ~ 0.6-0.8 OD\textsubscript{600}. Subculture was then induced with 0.5 mM IPTG and left to continue shaking at 28 °C, 200 rpm for 16 h in water bath orbital shakers.

2.10.3. Protein extraction for protein purification

Induced cultures were harvested at 4 °C by centrifugation at 5, 000 rpm for 20 min in a VWR R9A rotor. Supernatants were discarded and the cells pellets were washed with cold 200 mL of 50 mM NaPi, pH 7.4 buffer. Resuspended cells was pelleted at 4 °C by centrifugation at 4, 000 rpm for 30 min and transferred into 50 mL conical tubes for storage at -80 °C.

Cell pellets were resuspended in 30 mL of HisTrap column binding buffer supplemented by 1 mM of AEBSF to minimize proteolysis. The column binding buffer was 50 mM NaPi, pH 7.4 containing 500 mM sodium chloride (NaCl). Cell suspensions were pre-treated with 60 μL of 40 mg/mL hen egg-white lysozyme prepared in the same binding buffer per gram of cell pellet. Pre-treated samples were incubated with gentle agitation on orbital shakers at 4 °C for 15 min. Samples were then lysed using Omni Sonic Ruptor\textsuperscript{®} equipped with OT-T-375 probe from Omni Sonic. This was performed at RT with the samples kept cold in ice bath. Cells were disrupted at 50% power and 50% pulse length for 2 min. This was repeated four times. In between each sonication repeat, the sonication probe was rinsed with cold water to keep it from warming up to minimize protein denaturation. Cell lysates were clarified by centrifugation at 14, 000 rpm in a VWR R20A2 rotor kept at 4 °C for 40 min. Clarified cell lysates were then applied to purification columns using a peristaltic pump or a syringe.
2.10.4. Metal affinity purification

5 mL FF HisTrap® columns charged with Ni$^{2+}$ were used for 6×HisTag enzyme samples. Purification was performed using AKTA™ FPLC System at 4 °C or by manual bench-top purification at room temperature. Columns were equilibrated with 50 mM NaPi, pH 7.4 containing 500 mM NaCl (binding buffer, buffer A).

For purification using FPLC, unbound fractions were washed out of the column with 5 column volumes (CV) of buffer A. Weakly bound fractions were then washed out of the column by 3 CV of 10% elution buffer. The elution buffer was 50 mM NaPi, pH 7.4 containing 500 mM NaCl and 500 mM imidazole (buffer B). Samples were eluted using a linear gradient of 10% to 100% B over 10–20 CV. Fractions corresponding to the highest peak on the FPLC chromatogram were analyzed on SDS-polyacrylamide gel electrophoresis in 12% polyacrylamide gels to check their purity. Fractions that were > 95% pure were pooled and dialyzed overnight against 50 mM NaPi, pH 7.4 buffer. Alternatively, the fractions were concentrated to 1 mL using Amicon® ultra centrifugal filters with 30 kDa molecular weight cutoff cellulose membrane by centrifugation at 4, 000 rpm at 4–8 °C. This was then applied to a 5 mL HiTrap® desalting column pre-equilibrated with 50 mM NaPi buffer, pH 7.4. Samples were then eluted using the same buffer. Eluted samples were collected in 0.5 mL fractions. Fractions that had the enzyme were pooled and concentrated again to 500 μL. 10 mL of phosphate buffer was added gently to the concentrated sample to further dilute any trace amount of residual salt that might have been retained after the desalting column. Samples were concentrated again to 250 μL and stored in -20 °C as 25% glycerol stocks.

For purifications using manual bench-top method, washing buffers A1 and A2 were prepared by mixing buffer A and buffer B in volume ratio of 1:9 (A1) and 4:6 (A2). Columns were equilibrated with 5 CV of A1 containing 5 mM of imidazole prior to cell lysates loading. Unbound and weakly bound fractions were washed out of the column with 4 CV of A2. Samples were then eluted with 20 mL of buffer B. Eluted samples were collected in fractions of 2.5 mL, 2.5 mL and 5 mL thereafter. Buffers were kept as cold as possible and were passed through the column either using peristaltic pump or syringe. This method was sometimes used for purification of β-GUS.
variants and all the time used for purification of glucuronylsynthase variants. Eluates were analyzed, concentrated and stored as described for the FPLC purification.

2.10.5. Size-exclusion purification

1 mL of concentrated sample from HisTrap purification was loaded onto Superdex S-200 column (bed dimension 16 (d) x 600 mm (l), bed volume of 120 mL) and purified at a flow rate of 0.5 mL/min using the FPLC system. The column was first equilibrated with 120 mL (1 CV) of 50 mM NaPi, pH 7.4 before loading the samples. Sample was eluted over 1 CV in 50 mM NaPi, pH 7.4. Fractions were collected in 0.5 mL fractions and analyzed on SDS-PAGE. Size-exclusion purification was performed on β-GUS for kinetics characterization in Chapter 3 and for an isothermal titration calorimetry (ITC) experiment that was set-up to investigate the binding between β-GUS and β-D-glucuronic acid (1GlcA) (data not presented in this thesis).

2.10.6. Protein visualization and quantification

Enzyme samples were visualized using Laemmli SDS-PAGE system. A two-component SDS-PAGE gel consisting a 5% (w/v) stacking gel at pH 6.8 and a 15% (w/v) resolving gel at pH 8.8 was used. 10 lane gels were prepared using combs with well widths of 5 mm. The two components of the gel were prepared as in Table 2-7.

<table>
<thead>
<tr>
<th>Component</th>
<th>15% (w/v) Resolving gel</th>
<th>5% (w/v) Stacking gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acryl/Bis 37.5: 1, 40% (w/v)</td>
<td>3000 µL</td>
<td>375 µL</td>
</tr>
<tr>
<td>0.5 M Tris, pH 6.8</td>
<td>-</td>
<td>833 µL</td>
</tr>
<tr>
<td>1.5 M Tris, pH 8.8</td>
<td>2000 µL</td>
<td>-</td>
</tr>
<tr>
<td>10% (w/v) SDS</td>
<td>80 µL</td>
<td>33.3 µL</td>
</tr>
<tr>
<td>dH2O</td>
<td>2850 µL</td>
<td>2069 µL</td>
</tr>
<tr>
<td>TEMED</td>
<td>10 µL</td>
<td>3 µL</td>
</tr>
<tr>
<td>10% (w/v) APS</td>
<td>60 µL</td>
<td>20 µL</td>
</tr>
<tr>
<td>Total volume</td>
<td>8000 µL</td>
<td>3333.3 µL</td>
</tr>
</tbody>
</table>

Low range SDS-PAGE molecular weight standard was used as a standard. The marker contains six proteins with molecular weights ranging between 14.4 kDa and 97.4 kDa. The marker was purchased as a 20× stock from BioRad, and is diluted 20-fold with β-mercaptoethanol and stock sample buffer as recommended by the supplier’s protocol. The protein standard was heated for five minutes at 95 °C and stored at -20 °C.
Enzyme samples were mixed with SDS-reducing sample buffer (Appendix A) at a ratio of 1:2 and heat-treated at 95 °C for 5 min. Incubated samples were left to cool to room temperature for 5 min prior to loading and 10–20 μL of the enzyme samples were loaded per well. Electrophoresis was performed using the SDS running buffer system described in Appendix A. A constant current of 30 mA per gel and variable voltage was used. Running water was passed through the system to keep it cool. Separation was run for 35–45 min until the dye front migrated to the end of the gel.

Protein gels were stained using BioRad Coomasie G-250 dye. Prior to staining, SDS was removed by boiling the gel three times in water, changing the water in between each treatment, and then washing it with cold water. Gels were stained for 40 min if a fresh staining solution was used, or at least 6 hours if a recycled staining solution was used. The gels were then de-stained in water overnight. Images of the gel were scanned. The amount of protein was estimated using ImageJ (NIH).

Enzyme concentration was determined by NanoDrop® by measuring the absorbance at 280 nm. 2 μL of enzyme sample was used for each measurement. Three measurements were taken for each sample and averaged. Measurements were repeated if the standard deviation of the replicate measurements was more than 5%. The protein concentration was calculated form the $A_{280}$ values using the Beer-Lambert equation $A = εcl$; where $A$ is the absorbance, $l$ is the cell path length (cm) and $ε$ is the molar absorption coefficient ($M^{-1}cm^{-1}$) estimated by ProtParam tool (http://web.expasy.org/protparam/).

2.11. Enzyme characterization using activity assays

The purified enzymes of β-GUS and glucuronylsynthase variants were characterized using activity assay. This section describes the procedures used to determine their kinetic parameters.
2.11.1. Enzyme kinetics of β-GUS and variants

HisTrap purified enzymes ranging between 20–50 nM were reacted with 50 μM to 16 mM pNPGlcA or pNPGlc in 50 mM NaPi, pH 7.4 buffer. The liberation of pNP product from pNPGlcA in a reaction volume of 240 μL was followed for 20 min on SpectraMax® M2e microplate reader at a wavelength of 405 nm at 25 °C. The liberation of pNP product from pNPGlc in a reaction volume of 240 μL was followed for 60 min on SpectraMax® M2e microplate reader at a wavelength of 405 nm. The extinction coefficient of pNP was determined to be 7625 M⁻¹cm⁻¹ in 50 mM NaPi, pH 7.4 (Figure A2-1, Appendix 2-1). Kinetic parameters of the wild-type and mutant enzymes were obtained by fitting the turnover rates to the Michaelis-Menten equation using Kaleidagraph 4.5 (Synergy Software, Reading, PA).

2.11.2. Enzyme kinetics of β-GUS in t-BuOH

HisTrap purified enzyme samples (20–50 nM) were pre-incubated in 90 μL 50 mM NaPi, pH 7.4 assay buffer containing 20% t-BuOH for 20 min at 30 °C, in a total volume of 120 μL. Pre-incubated enzyme mixtures were then reacted with pNPGlcA at concentrations ranging from 10–50 μM in 50 mM NaPi, pH 7.4 buffer containing 20% t-BuOH. Hydrolysis of pNPGlcA in a reaction volume of 240 μL was followed for 20 min on SpectraMax® M2e microplate reader at a wavelength of 405 nm at 25 °C. The extinction coefficient of pNP was determined to be 8259 M⁻¹cm⁻¹ in the assay buffer with 20% t-BuOH β-GUS kinetics (Figure A2-2, Appendix 2-1). Kinetic parameters of the wild-type and mutant enzymes were obtained by fitting the initial velocities to the Michaelis-Menten equation.

2.11.3. Solvent stability study for β-GUS recombinants

HisTrap purified enzymes (50–100 nM) were pre-incubated in 80 μL of 50 mM NaPi, pH 7.4 containing solvents at a v/v ratio of 1:4 and total volume of 100 μL for 20 min at 30 °C. A control reaction mixture was also performed by incubating 20 μL of 50–100 nM of purified enzyme with 80 μL of 50 mM NaPi, pH7.4 20 min at 25 °C. The pre-incubated mixture was then added into 80 μL of 50 mM NaPi pH 7.4 (assay buffer) at a v/v ratio of 1:2. The reaction was initiated by adding 120 μL of 2 mM pNPGlcA
prepared in the assay buffer. Hydrolysis of pNPGlcA in a reaction volume of 240 µL was followed for 20 min on SpectraMax® M2e microplate reader at a wavelength of 405 nm at 28 °C. Residual activity was obtained by dividing the activity of solvent pre-incubated sample with the activity of the control sample. The residual activity was then plotted as a function of the solvent concentration to generate a sigmoidal curve. The curve was fitted to a Boltzmann function (Equation 2-3) using sigmoidal curve-fitting function on Kaleidagraph. The solvent concentration at which the enzyme loses half its activity, [solvent]_{50}, was obtained from the inflection point per Equation 2-3.

\[ y = y_{\text{min}} + \frac{y_{\text{max}} - y_{\text{min}}}{1 + e^{\frac{(S - S_{50})}{\lambda}}} \]  \hspace{1cm} (2-3)

where \( y_{\text{min}} \) and \( y_{\text{max}} \) are the values of the minimum and maximum activity, \( S \) is the concentration of solvents, \( S_{50} \) is the [solvent]_{50}, and \( \lambda \) denotes the slope of the transition between active and inactive enzyme.

2.11.4. Enzyme kinetics of glucuronylsynthase variants

HisTrap purified enzymes of concentrations ranging between 0.1–0.4 µM were reacted with 1.75 mM α-D-glucuronyl fluoride and CMO-DHEA at concentrations ranging from 0 µM to 12 mM in 50 mM NaPi, pH 7.4 reaction buffer containing 20% t-BuOH at 30 °C. Reactions of total volume 1.2 mL were performed in 1.5 mL HPLC vials with resealable lids. Reaction mixtures were left in the LCMS sample compartment maintained at 30 °C. Reaction progress was monitored by quantifying CMO-DHEA glucuronide in real time via automated sampling of 10 µL reaction every 4 h. A C-18 Poroshell 120 column with dimensions 2.1 x 30 mm or 50 mm and 2.7 µm pore size was used to resolve the glucuronide produce and acceptor substrate peaks. The method development for the LCMS analytical method will be described in Chapter 4.

2.11.5. t-BuOH stability study for glucuronylsynthase recombinants

HisTrap purified enzymes of a concentration 4 µM were pre-treated with 100 µL of 50 mM NaPi pH 7.4 containing 0–65% t-BuOH for 20 min at 30 °C in a final volume of 125 µL. A control reaction mixture was also performed by incubating the same amount of enzyme with 160 µL 50 mM NaPi, pH 7.4 for 20 min at 30 °C. Pre-incubated
samples were diluted 10 times in reaction mixtures containing 1.75 mM of α-D-glucuronyl fluoride and 2 mM CMO-DHEA 50 mM NaPi pH 7.4. Their activities were determined using the same method for glucuronylsynthase enzyme kinetics in Section 2.11.3. Residual activities and $[\text{t-BuOH}]^{50}$, the concentration of t-BuOH at which 50% of an enzyme sample was inactivated, were determined as in Section 2.11.4.

2.11.6. Half-lives of glucuronylsynthase variants in t-BuOH

HisTrap purified enzymes of a concentration 1 mM were pre-treated at 30 °C with 160 μL 50 mM NaPi, pH 7.4 containing 10% t-BuOH either in the absence of presence of 1 mM α-D-glucuronyl fluoride, in a final volume of 200 μL. At various time points between 30 min and 4 h, 5 μL samples were diluted 10 times in 50 mM NaPi buffer, pH 7.4. Diluted samples were kept at 4 °C temporarily prior to being used for the reaction assay. A control reaction was also performed by incubating the same amount of enzyme sample with 160 μL of 50 mM NaPi, pH 7.4 for 20 min at 28 °C. Incubated samples were diluted 10 times with 50 mM NaPi pH 7.4. The residual activity was determined as in Section 2.11.5. The $[\text{t-BuOH}]^{50}$ was determined as in Section 2.11.4. The residual activity was then plotted as a function of the incubation time and fitted to a Boltzmann function (Equation 2-4) using sigmoidal curve fitting on Kaleidagraph. The time point at which the enzyme lose half its activity, half-life $t_{1/2}$, was obtained from the inflection point.

$$y = y_{\text{min}} + \frac{y_{\text{max}} - y_{\text{min}}}{1 + e^{\frac{t - t_{1/2}}{\lambda}}}$$  \hspace{1cm} (2-4)

where $y_{\text{min}}$ and $y_{\text{max}}$ are the values of the minimum and maximum activity, $t$ is the incubation time, $t_{1/2}$ is the half-life, and $\lambda$ denotes the slope of the transition between active and inactive enzyme.

2.12. Biophysical and biochemical properties of enzymes

Thermal shift assay (differential scanning fluorimetry), intrinsic fluorescence assay, and circular dichroism were used to probe the biophysical properties of the enzyme. The ligand binding affinity of the glucuronylsynthase and its variants were determined using isothermal titration calorimetry. This section describes the procedures used in these experiments.
2.12.1. Thermal shift assay

Thermal shift assay was used to determine the temperature denaturation profiles of the enzyme. The method utilizes a fluorescent dye, Sypro Orange®, which binds non-specifically to hydrophobic surfaces and is quenched by water. As the protein denatures, the hydrophobic core becomes solvent exposed. The fluorescence of the dye increases as they bind to the exposed hydrophobic surface.

Reaction mixtures of 20 µL containing 5 µM of purified enzyme samples (final concentration) and 5× Sypro Orange® diluted in 25 mM NaCl (final concentrations) were set-up. The thermal effect on enzyme unfolding was observed for every 0.1 °C between 20 °C and 95 °C at a rate of 1 °C min⁻¹ in a 7900HT Fast Real-Time PCR machine at the BRF (JCSMR, ANU) equipped with 488 nm argon-ion laser excitation source. Emission fluorescence was monitored using continuous wavelength (λ) detection from 500-660 nm. The fluorescence at 605 nm was recorded and plotted as a function of temperature. The protein unfolding transition between the lower limit and the upper limit of the denaturation profile in response to increasing temperature was fitted to the Boltzmann function using the sigmoidal curve-fitting function of Kaleidagraph as in Equation 2-4. The term t₁/₂ was replaced by Tₘ, the temperature midpoint between protein transition.

2.12.2. Intrinsic fluorescence assay

The β-GUS enzyme contains 13 tryptophan and tyrosines, which enabled the monitoring of its intrinsic fluorescence. Intrinsic fluorescence was used to follow the unfolding of enzyme in the presence of chemical denaturant²⁹,³⁰ (urea and t-BuOH). Purified enzyme solutions of 10 mg/mL in 50 mM NaPi, pH7.4 buffer were incubated in different concentration of solvents (0–8 M of urea) at a volume ratio of 1:5 (enzyme:solvent). Incubations were done at room temperature for 30 min. Incubated samples were then diluted 7–8 times to 4 µM with 50 mM NaPi, pH 7.4. Fluorescence of 4 µM enzyme solutions was monitored using λexcitation = 274 nm and 280 nm, and continuous λemission between 310 nm and 500 nm were recorded. Protein unfolding transition was observed from the decrease in fluorescence intensity and from a red shift from 325 nm to 345 nm. The natural logarithmic function of the fluorescence intensity ratios at 325 nm and 345 nm were plotted against solvent concentration. A sigmoidal curve with two-state transition was generated. The [solvent]⁵⁰ was obtained by curve-fitting the transition profile to the Boltzmann function as in Section 2.11.4.


2.12.3. Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectroscopy was used to monitor the secondary structure perturbation of enzyme solutions incubated with 20% \( t \)-BuOH in the presence or absence of a ligand (\( \alpha \)-D-glucuronoyl fluoride). Enzyme solutions with concentrations ranging between 0.25–0.5 mg/mL were incubated at room temperature in 10 mM NaPi buffer, pH7.4 containing 0% or 20% \( t \)-BuOH. To monitor unfolding in the presence of ligand, 1 mM of \( \alpha \)-D-glucuronoyl fluoride was added to the incubation mixture. At different time-points between 0 h and 8 h, the incubated samples were transferred into a UV-transparent quartz cuvette with a pathlength of 0.1 cm. CD spectra were recorded at room temperature (25 °C) using Chirascan set with a time constant of 100 ms. Data (in millidegrees) was collected for wavelengths between 190 and 260 nm, at intervals of 0.5 nm, 0.5 s per point. A blank spectra of the buffer in 20% \( t \)-BuOH was collected for blank subtraction. At least two scans were performed for each spectrum and averaged. Collected measurements were converted to the mean residue ellipticity \( [\theta]_{MRE} \) (deg cm\(^2\) dmol\(^{-1}\)) by normalizing the raw elliptical data to the pathlength (in cm) and amount of protein (in dmol) used.

2.12.4. Isothermal calorimetry titration (ITC)

Isothermal titration calorimetry (ITC) experiments were performed to study the ligand affinity of the glucuronosynthase enzyme with glucuronoyl donor, \( \alpha \)-D-glucuronoyl fluoride. Experiments were performed on ITC2000 (TA Instruments, Texas, USA). Purified enzyme sample of 100 \( \mu \)M concentration and 4–5 mM of the ligand were used to conduct the experiment. Purified enzyme and ligand solutions were degassed prior to the experiment for 10 min to avoid bubble formation in the calorimeter cell. A stable baseline was achieved prior to ligand injection. Experiments were performed at 25 °C with a stirring rate of 300 rpm to equilibrate the titration mixtures. Equilibration time between each injection was set as 150 s. The data obtained was corrected for dilution heat by subtracting excess heat at high molar ratio of ligand to protein. The thermodynamics parameters – enthalpy (\( \Delta H^\circ \)), binding affinity (\( K \)) and entropy (\( \Delta S^\circ \)) – were interpreted using the NanoAnalyze® software provided by TA instruments. The association (\( K_a \)) and dissociation (\( K_d \)) constants were determined using an independent (1:1) model. The binding standard molar Gibbs energy change (\( \Delta G^\circ \)) and standard molar entropy contribution (\( \Delta S^\circ \)) were calculated according to the
thermodynamics relationships $\Delta G^\circ = -RT \ln K_a$ and $\Delta G^\circ = H^\circ - T\Delta S^\circ$, where $T$ is 298 K and $R$ is the gas constant in kJ mol$^{-1}$K$^{-1}$.

### 2.13. Computational tools

DNA sequences were analysed using Sequencher 4.10.1. Multiple sequence alignments of DNA and protein sequences were performed using Clustal-Omega\(^\text{35}\) (http://www.ebi.ac.uk/Tools/msa/clustalo/). Pymol 1.3 (DeLano Scientific) was used to visualize protein structures and to generate the figures of the structures. SWISS-MODEL\(^\text{36–38}\) (http://swissmodel.expasy.org/) was used to generate models of $\beta$-GUS and glucuronylsynthase, and their variants using $\beta$-GUS WT as template. Evolutionary conservation of amino acids and phylogenetic relationship were calculated using the ConSurf-DB server\(^\text{37,38}\) (http://bental.tau.ac.il/new_ConSurfDB/). Enzyme-solvent interactions were predicted using FTMap\(^\text{41,42}\) (http://ftmap.bu.edu/). CUPSAT\(^\text{43,44}\) was used to predict destabilization $\Delta\Delta G$ of single-site mutations (http://cupsat.tu-bs.de/). Molecular docking of $p$NPGlcA and 1GlcA in Chapter 6 was performed using the pair-fit command in Pymol 1.3. Heat map was generated using open source server Plotly (https://plot.ly/). All curve fitting was performed using Kaleidagraph 4.5 (Synergy Software, Reading, PA).
2.14. References


3. Directed evolution of β-glucuronidase (β-GUS)

3.1. Preamble

The main aim of this thesis is to improve the glucuronylsynthase system developed by the McLeod group. We had intended to do this in two steps as outlined in Section 1.4. This chapter describes the evolution of β-GUS. Our hypothesis was that a β-GUS variant with higher activity in the presence of high substrate concentration (10–20 times $K_m$) could be converted into a glucuronylsynthase (Syn) that is more suitable for synthetic application.

During the course of evolution, it was noted that the McLeod group observed improvement in the process chemistry for the glucuronylsynthase reactions when t-BuOH was added. However, the amount of t-BuOH that could be added to the system was only 10% due to enzyme stability issues. We envisioned that a β-GUS variant with enhanced performance in the presence of t-BuOH would also be converted to an improved glucuronylsynthase variant. Hence, we added t-BuOH to the screening assay in the latter evolutionary rounds, while maintaining the substrate concentration at 20-times the β-GUS $K_m$.

To differentiate the evolutionary rounds conducted under different conditions, we classified those performed without t-BuOH as Stage I β-GUS evolution. Stage II β-GUS evolution continues from Stage I, with the presence of t-BuOH. We will describe the screening strategies that were employed for the two-stages of evolution and the outcome of the evolution attempt.

3.2. Library size considerations

One of our earlier concerns in evolving the β-GUS was the sheer size of the enzyme. It is a multi-domain homo-tetramer with more than 600 amino acids per chain. Compared to the other enzymes that are investigated in this laboratory, evolving this enzyme will involve larger screening effort (Appendix 3-1, Table A3-1). To reduce the
library size, we considered an evolution strategy where segments of the gene were evolved separately. We used the crystal structure of the *E. coli* β-GUS obtained by Wallace *et al.*\(^2\) in 2010 to choose our target regions (Figure 3-1).

The identified fragments were located in the TIM-barrel domain that contained the active site. The largest of these fragments consists of residues K286–T503 (Lp2 libraries). A shorter peptide fragment within Lp2 that focused on residues 429 to 503 was also engineered (Lp4). These residues are proximal to the catalytic residues, E413 and E504 (Figure 3-1, inset). Hence, targeting this region may yield a variant that has altered binding and kinetic properties. Q547–S580 is a long loop consisting 40 amino acid residues that overhangs the active site. This loop (Lp6) may facilitate substrate entrapment and product release. Hence, this loop was also targeted.

![Figure 3-1: Peptide fragments and protein feature view of the targeted fragments chosen for random mutagenesis. (a) Peptide fragments chosen for targeted random mutagenesis. Inset on the right shows Lp4 in green. (b) Protein feature view of β-GUS shows the relative positions of the three domains and targeted loops.](image)
3.3. Results

3.3.1. Optimization and validation of screening methodology

The procedures for library growth and expression for β-GUS evolution was described in Chapter 2.7. We had applied the pooling strategy described in Section 2.7.1.2 to increase screening efficiency. However, to apply this strategy, two important aspects had to be considered. The first was the number of colonies to be pooled in each well. The second was the setting of appropriate criteria to select wells for secondary screens. We chose to dilute four colonies per well (4 cfu/well) based on deductions that were derived from estimates of screening errors (inter-plate variation) and the expected activity distribution of a library.

Experiments were set-up to establish the basis for assuming the choice of diluting 4 cfu/well. We first determined from six plates containing only wild-type, that the inter-plate variation (% CV) that would arise from screening would be around 20% (Figure 3-2a), with the lowest variation at 12% and the highest at 23%. We also determined that a library with a mutation rate of one amino acid per gene would have median activity of 0.8-fold (Figure 3-2b). Approximately 25% of the library appeared to be inactive. Based on these numbers, we assumed the following. In most of the wells, one of the four clones pooled would likely be inactive while the other three clones would have ~ 0.8x control activity. These wells would display around 0.6-fold the control activity. By contrast, a well with one 2-fold improved variant and one inactive variant would display 0.9-fold control activity (Figure 3-3). This would be significantly higher than the rest of the plate. Thus, we would be able to single out wells that contain potential hits i.e. at least 2-fold improved for secondary screening. We compared this to pooling 3 cfu/well, which is used by the others in the laboratory. There was no significant difference when the number of colonies in a well was reduced to three. Hence, we decided that pooling 4 cfu/well would allow maximum screening efficiency. Figure 3-3 describes four types of wells (A–D) that will most likely be encountered in the screening plates when we pool 3 cfu/well or 4 cfu/well.
Figure 3-2: Investigating the inter-plate variation. (a) The expected inter-plate variation (% CV) caused by systematic and random errors during screening. Red wells and blue wells are positions that are susceptible to being picked as false negative (blue) or false positive (red). (b) The activity distribution in a library with mutation rate of 1 amino acid per gene.
3.3.2. Stages of E-GUS evolution

Two stages of evolution were performed on E-GUS. The first stage was performed without solvent, and the second stage was performed with t-BuOH. Libraries for the two stages of evolution were created using error-prone PCR (ePCR) and StEP-ePCR as described in Chapter 2. Table 3-1 summarizes the libraries that were created and the screening method that was employed to screen them. The mutation rates and coverage of the library size are also listed in the table. Appendix 3-2 (Table A3-2) summarizes the mutations that were found in the two stages of evolution.

3.3.2.1. Stage 1 of β-GUS evolution

Libraries for three different fragments (Lp2, Lp4 and Lp6) were screened in the first stage evolution. The locations of these fragments were pointed out in Figure 3-1. No improved clones were detected in Lp6 library and this library was left for revisititation in a latter section. Lp4, which was meant to search part of Lp2 region more thoroughly, also did not yield any improved clones. Hence, we did not pursue this
region further. However, we continued the evolution done on Lp2 where we made several observations.

In the first round, two different types of Lp2 libraries were screened. One of these had an average mutation rate of 1.2 amino acid change per gene. The other library had an average mutation rate of 2.2 amino acid change per gene. Although we did not find any clones that were significantly improved, we managed to isolate several neutral mutations. These were subsequently recombined using StEP-ePCR.

Table 3-1: Libraries that were generated for the directed evolution of β-GUS. The libraries that were screened in the presence of t-BuOH (Stage-2) are highlighted in blue. The level of recombination was estimated by counting the number of recombination events observed in randomly selected sequences.

<table>
<thead>
<tr>
<th>Lib</th>
<th>Name/Region</th>
<th>Region Length (AA)</th>
<th>Library creation method</th>
<th>Mutation rate or recombination</th>
<th>Screening mtd</th>
<th>No. of clones screened</th>
<th>Theoretical Library Size</th>
<th>% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>1LowLp2</td>
<td>220</td>
<td>ePCR</td>
<td>1.2 (±0.8)</td>
<td>Pool 4 cfu/well</td>
<td>2400</td>
<td>7920</td>
<td>30</td>
</tr>
<tr>
<td>L2</td>
<td>1HighLp2</td>
<td>220</td>
<td>ePCR</td>
<td>2.2 (±1.1)</td>
<td>Pool 4 cfu/well</td>
<td>3200</td>
<td>867240</td>
<td>0.4</td>
</tr>
<tr>
<td>L3</td>
<td>2aLp2</td>
<td>220</td>
<td>ep-SIEP (recombination)</td>
<td>0.85 (±1.2);</td>
<td>Pool 4 cfu/well</td>
<td>1600</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Recombination*: &gt; 2/30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L4</td>
<td>2bLp2</td>
<td>220</td>
<td>ep-SIEP (recombination)</td>
<td>0.5 (±1.1);</td>
<td>Pool 4 cfu/well</td>
<td>3200</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Recombination*: &gt; 10/32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L5</td>
<td>1Lp4</td>
<td>90</td>
<td>ePCR</td>
<td>1.2 (±0.79)</td>
<td>Pool 4 cfu/well</td>
<td>1200</td>
<td>3240</td>
<td>37</td>
</tr>
<tr>
<td>L6</td>
<td>1Lp6</td>
<td>100</td>
<td>ePCR</td>
<td>1.5 (±0.8)</td>
<td>Pool 4 cfu/well</td>
<td>2400</td>
<td>3600</td>
<td>67</td>
</tr>
<tr>
<td>L7</td>
<td>3Lp2+6</td>
<td>320</td>
<td>ep-SIEP (recombination)</td>
<td>1.12 (±0.64);</td>
<td>Pool 2 cfu/well</td>
<td>1600</td>
<td>11520</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Recombination*: &gt; 2/10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L8</td>
<td>4Lp2+6</td>
<td>320</td>
<td>SIEP</td>
<td>Recombination*: &gt; 3/8</td>
<td>1 cfu/well</td>
<td>2400</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>L9</td>
<td>5-GUS</td>
<td>603</td>
<td>ePCR</td>
<td>2.25 (±1.16)</td>
<td>Pool 4 cfu/well</td>
<td>12000</td>
<td>6534108</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Two different recombinant libraries (2aLp2 and 2bLp2) were screened because our first attempt generated a library that had very low recombination frequency. The second library, 2bLp2, showed more recombined sequences compared to our first attempt. Nevertheless, we screened both libraries and after screening nearly 5000 clones from both libraries, we did not find any recombined mutations that were better than the wild-type. Some of the mutations from the first round disappeared. However, seven of the fourteen mutations that were found in the first round libraries (1LowLp2 and 1HighLp2) were retained. Interestingly, with the exception of mutation K495R, none of the other mutations from 1HighLp2 library were retained. The other two mutations that had appeared together with K495R also disappeared, but K495R reappeared in the final Lp2 library as a single mutant variant.
Mutants that appeared in the final round were subjected to tertiary screen to reconfirm their activity (Figure 3-4a). This was done by re-transforming the variants and expressing them in 1 mL scale. The activities were determined using the same assay used in primary and secondary screen and then normalized based on the cell density. The best clone (A426S) and a mediocre clone (S376G) were purified. Their activities were compared with the wild-type (Figure 3-4b). A426S showed very marginal improvement over wild-type, but this was within the error range. Nonetheless, this mutation was added to the pool of templates for the subsequent library, where t-BuOH was introduced.

Figure 3-4: Confirmation of potential hits from Stage I evolution. (a) Relative activity of clones from Stage I library compared to wild-type in tertiary screen. (b) The activities of purified enzymes - A426S and S376G – compared to the wild-type.

3.3.2.2. Second-stage of β-GUS evolution

Round 3 marked the start of the second-stage evolution of β-GUS. Solvent t-BuOH was added to the screening assays so that we could isolate a variant that performed better in t-BuOH. Before we commenced screening, a quick investigation was conducted to determine the concentration of t-BuOH that should be used for the screens. Purified β-GUS enzyme was incubated in 5–60% t-BuOH for 20 min. The incubated samples were then assayed using 1 mM of pNPGlcA substrate. The enzyme lost most of its activity but it was not completely inactivated in the presence of 20% t-BuOH (Figure 3-5a). We decided to use this amount of t-BuOH in the screening assay.
Three additional rounds of evolution with incremental \( t\)-BuOH (20–25%) were conducted using this assay. Variants from both 2aLp2 and 2bLp2 were used as templates for ePCR-StEP to generate the library for Round 3. At this stage, we also expanded the mutagenesis region to encompass the entire TIM-barrel domain (Lp2+6). Mutagenesis was subsequently extended to include the entire protein in the final round. Figure 3-5b summarizes the evolutionary route that was taken.

Following the expansion of the mutagenesis region to include the entire TIM-barrel, several new mutations were found in the C-terminal region (Lp6 region). One of these was L561S, which was conserved through to the final round. Most of the mutations from Stage I disappeared but, mutations K277Q, H313L and K495R were retained. Mutation A426S from Stage I was not retained. Of these three conserved mutations, H313L and K495R were retained in the final variant, along with L561S.

The final round was created using triple mutant H313L/K495R/L561S (4GUS1) as the template for random mutagenesis. This variant had more than two-times the activity of wild-type in the presence of \( t\)-BuOH (Appendix 3-2, Table A3-3) and was chosen rather than 4GUS12 as the single mutation A431V appeared to be non-beneficial. Screening the final round library led to the final variant (5GUS12) consisting of eight mutations (Figure 3-6 (a)), and that had nearly two-times the activity of the parent triple mutant 4GUS1 (Figure 3-6 (b)). Interestingly, the four random mutations that were introduced in this round were found in the first domain and no new mutations were introduced in the catalytic domain.
3.3.3. Library Lp6 revisited

The Lp6 segment consists of a mobile region made up of 20 amino acids (AA550–578) and remained interesting because the flexibility of this region may be useful for tailoring enzymatic activity. Ensuing from the positive selection of the mutation L561S after introducing 20% \(t\)-BuOH, we decided to revisit this region using the library that was created for Stage I even though we could not identify any hits during the first stage evolution. Two screening assays were performed – with and without \(t\)-BuOH – so that we could compare the activities under different solvent conditions and compute the \(t\)-BuOH stability. Table 3-2 lists some of the mutations that were found in this library. The full diversity can be found in Appendix 3-3 (Table A3-4).

As was the case for Stage I-II evolution, we did not find any mutations that were more active than the wild-type in the absence of \(t\)-BuOH. In the presence of \(t\)-BuOH, mutation L561S appeared again among Lp6 diversity and showed higher activity and residual activity. There were several outliers that showed high relative residual activity, but their activities were relatively poor. As a result, these appeared to be more stable. However, deeper analysis of the data revealed that this was an artifact caused by low initial activities in buffer. As we preferred to maintain as much native activity (activity in buffer) as possible, we removed them from further consideration.
Table 3-2: Diversity of Lp6 potential hits from round 1 \(t\)-BuOH screen. Marked in red are artifacts, which showed high relative residual activities because their initial activities in buffer were low. Rel Act (bfe): activity relative to wild-type in 50 mM NaPi, pH 7.4 buffer; Rel Act (t-BuOH): activity relative to wild-type in 20% \(t\)-BuOH, Rel residual: Residual activity in the presence of 20% \(t\)-BuOH relative to wild-type. Full list from Lp6 evolution can be found in Appendix 3-3 and 3-4. (False-R: considered as false and removed, NR: not removed). Mutations were recombined to generate second generation library (2Lp6).

| Mutants   | Rel Act (bfe) | Rel Act (t-BuOH) | Rel residual | Comments | A511 | Y633 | V543 | Q647 | S557 | I560 | L561 | G654 | N696 | K576 | F592 | G694 | K596 |
|-----------|---------------|------------------|--------------|----------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1aLp6.22  | 1.25          | 1.18             | 0.94         | WT/neutral | I    | T    | S    |       |      |      |      |      |      |      |      |      |
| 1aLp6.11  | 0.42          | 3.41             | 8.04         | Hit       |      |      |      |       |      |      |      |      |      |      |      |      |
| 1aLp6.12  | 0.12          | 0.66             | 5.65         | False-R   |      |      |      |       |      |      |      |      |      |      |      |      |
| 1aLp6.14  | 0.49          | 3.93             | 8.08         | Hit       |      |      |      |       |      |      |      |      |      |      |      |      |
| 1aLp6.17  | 0.13          | 0.56             | 4.21         | False-R   |      |      |      |       |      |      |      |      |      |      |      |      |
| 1aLp6.23  | 0.04          | 0.18             | 3.93         | False-R   |      |      |      |       |      |      |      |      |      |      |      |      |
| 1aLp6.26  | 0.18          | 0.41             | 2.25         | False-R   |      |      |      |       |      |      |      |      |      |      |      |      |
| 1aLp6.30  | 0.60          | 2.60             | 4.31         | Hit       |      | L    | S    |      |      |      |      |      |      |      |      |      |
| 1aLp6.31  | 1.08          | 4.06             | 3.75         | Hit       |      |      |      |       |      |      |      |      |      |      |      |      |
| 1aLp6.27  | 0.24          | 0.73             | 3.08         | Potentially false-NR |      | S    | C    |      |      |      |      |      |      |      |      |      |
| 1aLp6.10  | 0.61          | 1.62             | 2.66         | Hit       |      |      |      |       |      |      |      |      |      |      |      |      |

Variants that retained more than 20% wild-type (WT) activity in native condition, and that were 2-times more stable than wild-type in the presence of \(t\)-BuOH (variants that have > 2-times relative residual activity with \(t\)-BuOH) were recombined using StEP. A library of \(10^4\) clones was generated. We checked the library quality by determining the number of new unique sequences (sequences that showed shuffling or recombination events) out of 12 randomly picked clones. From 12 sequences, four of these were new and unique. We screened approximately 10% (800 clones) of the library in the presence of 25% \(t\)-BuOH. Clones that had more than 3-times WT activity were sequenced. The diversity we found in this round (round 2Lp6) was significantly reduced. Mutation L561S was found in all the recombinants. Appendix 3-4 lists the full diversity from the second generation library of Lp6 \(t\)-BuOH evolution (2Lp6). Recombinants that contained neighbouring mutations (I560T/L561S) and mutations that frequently appeared (F593S and K596R) were selected for characterization and reassigned as in Table A3-5 (Appendix 3-4). We also selected double mutant S557L/L561S because substitution at S557 had been reported in the literature\(^5\) to accommodate other glycosyl rings. In addition, we were surprised over the absence of mutation T509A in our evolution program. This mutation had been prevalent in literatures and had been reported to be beneficial for activity on other \(\beta\)-glycosides\(^5–9\) and for thermostability\(^6,10\). It was not clear to us why this mutation did not turn up in
both the earlier evolution (Stage I-II) and in this targeted Lp6 evolution. Hence, we made the variant T509A site specifically for characterization.

### 3.3.4. Kinetic parameters of mutants

Kinetic parameters of the variants from Round 3–5 in Stage II and the selected mutants from Lp6 were obtained. Table 3-3 lists the mutations of the variants that were characterized. It also summarizes some of the observations for the more interesting variants. Table 3-4 summarizes the kinetic parameters of the variants in buffer and in the presence of 20% t-BuOH. The t-BuOH tolerance was determined by determining the % t-BuOH where the enzymes lost half their activity \([t\text{-BuOH}]^{50}\). The turnover rate with and without t-BuOH, and the stability in \([t\text{-BuOH}]^{50}\) are compared in Figure 3-7.

#### Table 3-3: Variants that were selected for kinetics characterization and the mutations that they harbour. A summary of their activities relative to wild-type in the screening experiments are given where possible; ND: not determined, NA: not available.

<table>
<thead>
<tr>
<th>Source</th>
<th>Variants</th>
<th>Mutations</th>
<th>Native activity</th>
<th>t-BuOH activity</th>
<th>Residual activity in the presence of 20 - 25% t-BuOH</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-GUS</td>
<td>Native</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Stage I-II evolution</td>
<td>L561S</td>
<td>S</td>
<td>0.51</td>
<td>4.35</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>4GUS1</td>
<td>L R</td>
<td>ND</td>
<td>3.31</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>5GUS12</td>
<td>Y C S G L R</td>
<td>ND</td>
<td>5.60</td>
<td>7.15</td>
</tr>
<tr>
<td></td>
<td>7GUS7</td>
<td>T S R</td>
<td>0.54</td>
<td>3.89</td>
<td>7.15</td>
</tr>
<tr>
<td></td>
<td>7GUS10</td>
<td>L S</td>
<td>0.83</td>
<td>3.58</td>
<td>4.30</td>
</tr>
<tr>
<td>Lp6 revisit</td>
<td>7GUS13</td>
<td>T S</td>
<td>0.62</td>
<td>3.45</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td>7GUS14</td>
<td>S S</td>
<td>0.56</td>
<td>3.52</td>
<td>6.25</td>
</tr>
<tr>
<td></td>
<td>7GUS24</td>
<td>S R</td>
<td>0.34</td>
<td>2.41</td>
<td>6.88</td>
</tr>
<tr>
<td>Rational design</td>
<td>T509A</td>
<td>A</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

#### Table 3-4: Kinetic parameters in the presence of 20% t-BuOH and in 50 mM NaPi, pH 7.4 buffer (reflects physiological conditions).

<table>
<thead>
<tr>
<th>Variants</th>
<th>(k_{cat}) (s(^{-1}))</th>
<th>(K_m) (mM)</th>
<th>(k_{cat}/K_m) (mM(^{-1})s(^{-1}))</th>
<th>(k_{cat}) (s(^{-1}))</th>
<th>(K_m) (mM)</th>
<th>(k_{cat}/K_m) (mM(^{-1})s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>5.24 ± 0.56</td>
<td>0.14 ± 0.017</td>
<td>38.25 ± 4.114</td>
<td>76.57 ± 2.46</td>
<td>0.14 ± 0.019</td>
<td>530.9 ± 71.9</td>
</tr>
<tr>
<td>L561S</td>
<td>41.52 ± 0.71</td>
<td>0.52 ± 0.084</td>
<td>80.34 ± 13.388</td>
<td>40.05 ± 1.00</td>
<td>0.14 ± 0.015</td>
<td>279.7 ± 29.4</td>
</tr>
<tr>
<td>4GUS1</td>
<td>18.56 ± 0.6</td>
<td>0.22 ± 0.012</td>
<td>62.35 ± 18.810</td>
<td>4.31 ± 0.16</td>
<td>0.031 ± 0.007</td>
<td>140.6 ± 32.9</td>
</tr>
<tr>
<td>5GUS12</td>
<td>35.65 ± 0.5</td>
<td>0.40 ± 0.081</td>
<td>89.44 ± 18.691</td>
<td>67.38 ± 3.12</td>
<td>0.17 ± 0.030</td>
<td>402.3 ± 74.5</td>
</tr>
<tr>
<td>7GUS7</td>
<td>19.96 ± 0.5</td>
<td>0.57 ± 0.049</td>
<td>35.14 ± 3.176</td>
<td>16.69 ± 0.74</td>
<td>0.051 ± 0.012</td>
<td>329.8 ± 78.4</td>
</tr>
<tr>
<td>7GUS10</td>
<td>14.95 ± 0.7</td>
<td>0.46 ± 0.094</td>
<td>32.21 ± 6.733</td>
<td>6.54 ± 0.18</td>
<td>0.047 ± 0.007</td>
<td>138.9 ± 20.6</td>
</tr>
<tr>
<td>7GUS13</td>
<td>15.31 ± 0.6</td>
<td>0.43 ± 0.068</td>
<td>35.26 ± 5.759</td>
<td>41.09 ± 0.72</td>
<td>0.14 ± 0.010</td>
<td>299.0 ± 21.8</td>
</tr>
<tr>
<td>7GUS14</td>
<td>34.17 ± 1.2</td>
<td>0.60 ± 0.076</td>
<td>56.53 ± 7.358</td>
<td>37.70 ± 0.40</td>
<td>0.10 ± 0.005</td>
<td>364.2 ± 17.2</td>
</tr>
<tr>
<td>7GUS24</td>
<td>23.67 ± 1.2</td>
<td>0.47 ± 0.086</td>
<td>50.01 ± 9.477</td>
<td>26.80 ± 0.23</td>
<td>0.067 ± 0.003</td>
<td>401.8 ± 17.2</td>
</tr>
<tr>
<td>T509A</td>
<td>6.76 ± 0.21</td>
<td>0.31 ± 0.034</td>
<td>21.75 ± 2.445</td>
<td>46.25 ± 2.04</td>
<td>0.10 ± 0.019</td>
<td>486.0 ± 100.6</td>
</tr>
</tbody>
</table>
The kinetic behavior of the wild-type enzyme in 50 mM phosphate buffer (NaPi), pH 7.4 is consistent with the values reported in various literature\(^5\).\(^7\).\(^9\). In buffer only, none of the variants were better than wild-type. In t-BuOH, all the variants chosen from library screening were better than wild-type. However, rational design variant, T509A, was not better than wild-type in t-BuOH. This would explain its absence in our evolution program.

The final variant (5GUS12) carries eight mutations and it was the most catalytically efficient variant. Amongst these mutations, the mutation L561S carried the dominant effect. The presence of single mutation L561S conferred 8-fold improvement in its turnover rate (\(k_{\text{cat}}\)) value when compared to WT (Figure 3-8) in t-BuOH.

![Figure 3-7: The relative activities of the variants. Top: Comparison of the activity of variants under physiological condition and in 20% t-BuOH. Bottom: Comparison of the t-BuOH stability of the variants as determined by the midpoint of a melting curve, [t-BuOH]\(^{50}\) and their activities under physiological condition.](image)

The final variant (5GUS12) carries eight mutations and it was the most catalytically efficient variant. Amongst these mutations, the mutation L561S carried the dominant effect. The presence of single mutation L561S conferred 8-fold improvement in its turnover rate (\(k_{\text{cat}}\)) value when compared to WT (Figure 3-8) in t-BuOH.
Consistent with our selection criteria during screening, this improvement was sacrificed slightly for better \( t \)-BuOH stability in 4GUS1. The final variant 5GUS12 recovered most of the activity that was lost by the addition of \( t \)-BuOH and had a higher catalytic efficiency compared to L561S and 4GUS1. Although the \( k_{cat} \) improved by approximately 8-fold in the presence of \( t \)-BuOH when L561S was introduced, the catalytic efficiency \( (k_{cat}/K_m) \) in the presence of \( t \)-BuOH only doubled (Figure 3-8). This was sustained by variants obtained in subsequent generations, which was interesting because we did not expect 4GUS1 to be as efficient as L561S due to its lower \( k_{cat} \). This may be explained by the reduced \( K_m \) in 4GUS1 and can be rationalized from its structure. The substitution in K495R would decrease the distance between residue K495 and the C\( \alpha \) backbone of D454 that interacts with the catalytic residue E413. Consequently, this would lead to increased rigidity of the network formed (Figure 3-8, insets) and alter the binding behavior in the active site. The increased rigidity may also explain the higher thermal stability of the variant (Figure 3-9). However, the mutations that were added in the subsequent round reversed the stability effect. 5GUS12 was the least thermostable compared to L561S and 4GUS1, but the mutations in 5GUS12 had benefitted its catalytic activity in both conditions—presence and absence of \( t \)-BuOH. The observations above showed that it was difficult to co-evolve for improved catalytic properties in different conditions—absence and presence of \( t \)-BuOH—and increased stability simultaneously.

Figure 3-8: The trend in evolutionary progress of \( \beta \)-GUS activity in 20 % \( t \)-BuOH. Insets show the interactions of K495(WT) and R495 (mutant) with D454 and the network of 495-454-416 that may modulate the position of base catalyst residue, E413.
Most of the variants that had improved activity in $t$-BuOH also had improved stability. This was with the exception of the variant 7GUS7 (I560T/L561S), 7GUS13 (I560T/L561S/K596R), and the most efficient variant 5GUS12. These had reduced thermostability compared to the wild-type. Hence, it may appear that the mutations at Lp6 with the exception of I560T appeared to be beneficial for thermostability even though they only conferred slight improvement in activity in the presence of $t$-BuOH. Lower thermostability in 5GUS12 may have been due to the other mutations from the other regions. Notably, 7GUS10 (S557L/L561S) was 7 °C more stable than the wild-type even though it was only three times faster than the wild-type. However, there was an exception – the double mutant I560T/L561S. The neighbouring mutations may have interacted negatively to cause a reduction in thermostability. Some of the stability was recovered by K596R. When the mutation I560T was removed from the triple mutation combination I560T/L561S/K596R, the lost stability in double mutant I560T/L561S was recovered.

Figure 3-9: Thermal stability values of β-GUS variants. Most mutations at Lp6 (L561S, S557L, F593S and K596R) improved the thermostability of β-GUS; the exception was I560T, that appeared detrimental as a double mutant (I560T/L561S). The additional mutations accumulated in the final round of Lp2 evolution (5GUS12) reduced the thermostability of β-GUS.
3.3.5. Stability study in various denaturants

The use of organic solvents are sometimes needed in chemoenzymatic reactions\textsuperscript{11–14} and in some cases have been shown to be able to improve enzyme performance\textsuperscript{15}. We compared the activity of the variants towards \( t \)-BuOH and two alcoholic solvents (isopropyl alcohol and ethanol) that are often used in the industry. We also determined their resistance towards glutaraldehyde that is used in the industry as a cross-linker to stabilize enzymes in industrial process\textsuperscript{16}. Mutations that confer greater solvent tolerance to \( \beta \)-GUS may be advantageous for biocatalysis by the glucuronylsynthase. A greater resistance towards glutaraldehyde also offers the possibility of obtaining cross-linked enzyme crystals that can be used for structural studies or in the enzymatic reactions. In addition, \( \beta \)-GUS is a common reporter enzyme\textsuperscript{17,18} but suffers from function loss during tissue fixation due to poor stability, particularly in glutaraldehyde or formaldehyde\textsuperscript{19}. Hence, it would be interesting to know if greater \( t \)-BuOH stability also translated to greater stability in other solvents that may be useful for other \( \beta \)-GUS applications.

The tolerance in different solvents was determined using a residual activity assay. The concentrations at which the enzymes were 50\% inactivated, [solvent]\textsuperscript{50} were determined. Relative stabilities of the variants compared to the wild-type \( \beta \)-GUS were visualized in a heat map generated using the online analytics tool Plotly (Montreal, Quebec). The heat map allowed for direct comparisons of all the data in a two-dimensional matrix where lighter colors represented lower solvent stabilities and darker colors represented higher solvent stabilities (Figure 3-10).

Generally, variants that were more tolerant of \( t \)-BuOH were also more tolerant of other alcoholic solvents compared to \( \beta \)-GUS. However, in the other alcoholic solvents, the enhancement seen was less impressive than in \( t \)-BuOH. The most interesting variant was the 7GUS14 (L561/F593S) that had higher resistance towards both \( t \)-BuOH and glutaraldehyde. It displayed the most consistency in stability across all solvents. When compared to the parameters obtained in the previous section, this variant also showed good thermostability and was nearly as efficient as L561S. Therefore, this variant was taken as an enzyme with the most enhanced biophysical properties, while the L561S variant offers the greatest benefits in turnover rate.
3.4. Discussion

3.4.1. Structural rationalization of L561S and other variants

The change from leucine to serine at residue 561 is the key mutation in the evolution towards increased activity in the presence of t-BuOH organic solvent. The fact that it was found in two parallel evolutions, Lp2 and then Lp6, establishes the pivotal role of this mutation. We attempted to rationalize the structural basis for the enhancement conferred by L561S. In the crystal structure obtained by Wallace et al., the residue L561 is found on a subunit interface (Figure 3-11a). The Leu-Leu distance is around 3.3 Å, which allows weak non-polar interaction between the Leu from adjacent subunits. Leu→Ser substitution increases this distance to 7.5 Å, which is not possible for molecular interaction. However, this creates a space for solvent molecules to potentially dock. We investigated this possibility by subjecting the wild-type and L561S variant to docking analysis using FTMap (Figure 3-11 (b)-(d)). FTMap is a server that identifies hot spots on macromolecules that bind small organic molecules or probes.
Locations where the probes are predicted to dock are known as solvent clusters. One of the solvent clusters predicted was on the Leu-Leu interface in the wild-type (Figure 3-11b). However, mutation to Ser at this position removes this potential docking site (Figure 3-11c).

Figure 3-11: Positions of the mutations in the crystal structure 3K4D. (a) Positions of the active site, the disordered loop and the L561 residues on two subunits. L561 is at the interface of the subunit but is unlikely to interact. It may interact with the disordered loop to partially obstruct the active site entrance. (b) L561 dimeric interface with solvent clusters (dots) generated by FTMap. (c) FTMap revealed that the mutation to serine displaces the solvent cluster (d) Mutations that conferred stability were generally found at solvent-exposed loops.

Hence, from structural examination, we deduce that the mutation Leu→Ser reduced subunit interactions by disrupting the interaction between the 561 residues of adjacent subunits. This increases the flexibility of the loop harbouring residue 561 (Lp6). This loop and the N-terminal portion of the disordered loop (residue 361-362), which can also interact with the leaving group of the substrate, are within accessible distance of each other. The two loops can partially obstruct the entrance to the active site (Figure 3-11a). This inevitably affects the dynamics involved in substrate binding and product release. Substitution to the smaller amino acid (Leu→Ser) opens up the aperture and consequently facilitates these events. Therefore, the pivotal mutation
L561S affects turnover rates more than stability. This is supported by the characteristics of the other Lp6 variants, which saw additional mutations to the L561S single mutant increase the thermostability while sacrificing some of the gain in $k_{cat}$ by the single mutation L561S. In most cases, the mutations (K596R and F593S) occurred on solvent exposed positions, but they are not implicated in any loop motion that is likely to benefit activity (Figure 3-11d). The trend in activity-stability sacrifice seen with L561S is congruent with observations from other directed evolution experiments\textsuperscript{21,22}.

### 3.4.2. Implications of screening and evolutionary strategies that were adopted

In the first round of experiments, the activity of the native enzyme was high and the screening method failed to identify mutants with improved activity. There may have been mutants with enhanced activity, but the change in activity would have been within the systemic errors inherent in library screening. This made it difficult to identify any clones with improved activity. In the second set of experiments, the $t$-BuOH greatly reduced the activity of the native enzyme. In this case, the screening method could identify mutants that had increased activity because the increment was above the error levels of the screens. Therefore, it appears that it is easier to identify mutants with enhanced activity when starting with a template that has low activity.

Although the pooling method is a powerful screening approach, it may not have been appropriate for the first stage evolution. Even though we had taken into account the various factors that would affect the discrimination of a well containing better clones compared to the rest of the screening plate, our reasoning had failed because we overestimated the improvement that we expected to see. We had chosen to pool four cfu/well based on accounts from the literature that proposed that pooling ten colonies could increase chances of detecting super-mutants (clones with five-fold improvement)\textsuperscript{3,23,24}. We reduced the number of colonies pooled according to our expectation that we were unlikely to detect any super-mutants, but we had assumed that we would see a mutant that was at least two-fold better.

The assumption that we had adopted in the pooling strategy may have been valid if we were screening to enhance a different function \textit{e.g.} to increase one of its promiscuous activity such as the $\beta$-glucosidase activity. However, our aim was to
increase its native activity in the presence of excessive native substrate. This was with the hope that we could isolate a variant that has a more rapid turnover at high substrate load, while concomitantly raising its $K_m$ (Figure 3-12). This evolutionary pressure was very subtle and was almost akin to enhancing the $\beta$-GUS native activity. As such, a more reasonable expectation should have been 1.1–1.2 fold improvement that would have been well within our screening error range. In this case, even without pooling, discriminating these hits over the noise level of library screening would have been difficult. Employing the pooling strategy regardless of the number of colonies pooled had simply made matter worse. All these suggest that the first approach may not be as easy as it had seemed.

We had attempted evolution by region to increase screening efficiency. This had worked to a certain extent because we could search a sequence space very thoroughly. This enabled the identification of single mutations that have pivotal roles, in this case L561S. This was isolated in parallel evolution – the Stage I-II evolution – and revisited Lp6 evolution. However, not many mutations that added benefits to L561S were found in these regions subsequently. The prevailing proof for this was in the fifth round, where most mutations were accumulated outside of the TIM-barrel region (Lp2 and Lp6 targeted prior to the fifth round). This suggests that too many mutations introduced simultaneously into the active site or catalytic domain (Lp2 and Lp6) would be detrimental to the enzyme. A better strategy may be to screen a few regions in parallel and select the best progenitor from one of the regions as a template for a subsequent round which targets a different region. This concept would be similar to the iterative saturation mutagenesis (ISM) that was developed in the Reetz’s laboratory.25 The
difference is that this involves the recombination of fragments rather than single sites. This approach had worked in the final round, where we isolated 5GUS12, which had regained the catalytic efficiency that was sacrificed in the previous round. This variant also outperformed L561S in its catalytic efficiency in physiological condition, making it a variant that can function well under both conditions.

3.5. Concluding remarks

As part of the two-step strategy towards an improved glucuronylsynthase, this step yielded three potential candidates for the second step of the evolution – L561S, 5GUS12 and L561S/F593S – where the mutation E504G could be introduced site-specifically. The mutation L561S had dominated the evolution program when t-BuOH was added and was determined to be important for improving turnover rate in t-BuOH. Hence, we took the L561S as the first variant to test in the two-step approach. This will be elaborated on in the next chapter.
3.6. References

20. Kozakov, D. et al. The FTMap family of web servers for determining and


4. Engineering the glucuronylsynthase

4.1. Preamble

In Chapter 3, we identified a mutation, L561S, which was more stable and performed better than β-GUS in the presence of 20% t-BuOH. We hypothesized that this enhancement would be translated to higher glucuronidation activity when we introduce E→G mutation into the variant at residue 504 to produce the glucuronylsynthase (Syn) variant. Thus, we introduced this mutation into L561S using site directed mutagenesis. Glucuronylsynthase (Syn) and recombinant Syn L561S were then purified as described in Section 2.10. Purified enzymes were tested by the McLeod group in synthetic reactions using testosterone as the acceptor substrate. Reaction mixtures were incubated at 30 °C for three days and the products were analysed using NMR. However, we could not detect the testosterone glucuronide product for the Syn L561S reaction even though Syn showed approximately 65% conversion for the glucuronidation of testosterone (data not presented). Hence, we decided to reprioritize our approach and attempted instead to engineer the glucuronylsynthase (Syn)\(^1\) using the linear approach (Figure 4-1, right side).

![Figure 4-1: Two engineering approaches used in attempts to improve the Syn system. The two-part (indirect) approach was the original strategy that involves β-GUS evolution first before incorporating E504G mutation. This chapter deals with the direct approach. In this approach, a recombinant that already has E504G mutation incorporated was directly engineered instead.](image)

Before commencing the one-step approach, we devoted some effort to the development and optimization of a library screening method as we lacked a suitable
assay for Syn library screening. We also had to consider better library design – smaller but more effective library – because the assays that would potentially be applied for Syn library screening would be more demanding than the β-GUS assay. To proceed with assay development, we first had to contemplate the selection pressure for the screens. We will be describing these in this chapter.

4.2. Setting the selection parameters for library screens

There were three parameters that had to be set for the engineering endeavor: choice of acceptor substrate, co-solvent and glycosyl donor:acceptor ratio. All of these will create the selection pressure for the evolution effort.

(A) Substrates

Like its parent glycosidase (β-GUS), the glucuronylsynthase is specific for the glycosyl moiety but is expected to show a wide range of activity for different steroids. As our aim was to develop a variant that would be more efficient in synthesizing glucuronides of drug components that contain steroid acceptors, we narrowed our substrate selection to steroid based substrates. We were presented with three choices of steroid acceptor substrates: CMO-DHEA, DHEA and testosterone. They contain steroid scaffolds and are mostly hydrophobic. However, CMO-DHEA has an oxime group that aids solubility. The activity of Syn with CMO-DHEA was the most understood as this was used in most of the process development work in the synthetic laboratory. Hence, we chose CMO-DHEA as the substrate for most of our evolution work.

(B) Co-solvent

The addition of t-BuOH would be necessary because the acceptor substrates have poor solubilities\(^2,3\). In the optimization process by the McLeod group, 10% of t-BuOH was found to be optimal for their Syn chemoenzymatic reactions, due predominantly to increased steroid (acceptor substrate) solubility. Furthermore, the use of t-BuOH (between 5–20%) does not form t-butyl glycoside, and thus would not pose as a competing reaction. From these past observations, we expected that the addition of
more $t$-BuOH could improve the catalytic activity of the glucuronylsynthase. To induce sufficient pressure for the evolution, we doubled the amount of $t$-BuOH to 20%. This was with the hope that we would be able to isolate a more efficient and more stable variant compared to Syn.

(C) Glycosyl donor (1FGlcA): Acceptor ratio

In previous investigations, it was found that a reaction using 1:1 molar equivalent of 1FGlcA halted after 14 h\(^4\). This was caused by decreased availability of the 1FGlcA substrate over the entire course of the reaction (36 h). Hence, a higher molar equivalence of 1FGlcA (5:1) was used to achieve near-complete conversion rates. This also reduced product inhibition. We would use a ratio of 3:1 (donor:acceptor) in our screening procedure.

4.3. Method Development

4.3.1. General materials and methods

All buffers were prepared as described in Chapter 2. Oligonucleotide primers for the incorporation of E504G and library randomization were purchased from Integrated DNA Technologies (IDT, Coralville, USA). The sequences of the primers are listed in Appendix D. The host and expression system used were BW25142 (Chapter 2.3) and pJ401 (Chapter 2.4). Purified enzyme solutions were obtained using manual HisTrap purification as described in Section 2.10.4. \(\alpha\)-D-glucuronyl fluoride (1FGlcA), CMO-DHEA and testosterone were prepared by the McLeod research group (Research School of Chemistry, ANU). Buffers and indicators that were used for the development of an indicator displacement assay were also prepared by the McLeod group. Mobile phases were prepared using chromatography-grade MeOH and Milli-Q water filtered through 0.22 \(\mu\)m filters. Chromatography-grade MeOH was purchased from Merck. Milli-Q water (mQH\(_2\)O) was collected from Millipore water purification system. The same solvent system was used to quench the reactions for library screening.
4.3.2. Tailoring the reaction time and amount of crude lysates for screening

The workflow for growth and expression was similar to the procedures described in Section 2.7.1.1 with one subtle difference. The volume for growth and expression was increased to 500 μL so that we could maximize the amount of crude lysates that could be obtained from growth and expression in 96-well plates. We anticipated that this volume would yield approximately 0.025–0.05 mg/mL of overexpressed enzyme, which would be enough for the library screening assays. This was based on the following observations. First, based on Bradford assay of crude lysate, 5 mL of pJ401/Syn/BW25142 grown overnight can yield 70 mg of the protein per L of cell culture. Second, SDS-PAGE analysis typically showed that Syn made up 50% of the total protein expressed. Therefore, 500 μL culture would give us approximately 0.035 mg/mL of glucuronylsynthase enzymes. Third, in typical synthetic reactions using 1–2 mg/mL of purified Syn, a three-day reaction course would yield ~95% conversion of CMO-DHEA. Hence, allowing the screening reaction to run overnight (12–24 h) should yield approximately 9–10% conversion of CMO-DHEA, which should yield sufficient CMO-DHEA glucuronide product to be detected and to differentiate between improved variants and Syn like variants.

We set up the assays for library screening using 500 μL of overnight expressed cultures induced by 0.5 mM IPTG at 30 °C. Cells were clarified by centrifugation and resuspended in 150 μL of 50 mM NH₄OAc buffer. Lysis was done at room temperature by adding 50 μL of 4 U/μL of rLysozyme™ and agitating the mixture at high-speed for 15 min. Lysates were then cleared by centrifuging at 3, 700 rpm for 15 min. 150 μL of cleared lysate was used in 400 μL of reaction containing 2 mM (~0.6 mg/mL) of CMO-DHEA and 6 mM 1FGlcA (1:3). Reactions were incubated at 30 °C overnight for 12–16 h. 120 μL overnight reaction was then quenched by addition of MeOH to a final concentration of 50% (v/v) MeOH and assayed.

4.3.3. Assay designs for glucuronylsynthase screens

The major challenge for Syn engineering was the development of a suitable assay for library screening. The most reliable and straightforward assay for detecting hits would be by monitoring glucuronidated products using HPLC or LCMS. However, to
use either method for screening an instrument with 96-well plates automation capability would be required. We did not have this facility at the start of the program. Instead, we attempted to optimize an indicator displacement assay that would allow us to monitor the amount of fluoride liberated.

**4.3.3.1. Fluoride-selective indicator displacement assay**

As glycosynthetic reactions generally use an α-sugar fluoride donor, the displacement of fluoride \( F^- \) is a common feature across all glycosynthetic reactions. Therefore, a \( F^- \) indicator displacement assay can be applied to a wide range of glycosynthases, including the glucuronylsynthase system. Using this method, we would be able to assay libraries of \( 10^3 \)-\( 10^4 \) clones.

The \( F^- \) indicator displacement assay utilizes a tetravalent zirconium (IV)-EDTA complex and a 2’-sulfoflavanol\(^5\). Mixing the two compounds generates a Zr-EDTA-flavanol complex \textit{in situ} and emits blue fluorescence (\( \lambda_{\text{em}} = 460 \) nm). In the presence of free \( F^- \), 2’-sulfoflavanol is displaced and fluorescence will be quenched (Figure 4-2). The reduction in fluorescence intensity corresponds to the amount of \( F^- \) present that can be used to monitor the extent of the glycosynthase reaction.

![Figure 4-2: The reaction involved in an indicator displacement assay. The \( F^- \) ions generated during the glucuronylsynthase reaction displaces the 2’-sulfoflavanol from the Zr-EDTA-sulfoflavanol complex](image)

We tested the limit of detection of this screening assay using 0–100 \( \mu \)M of sodium fluoride, NaF. This was done by plotting the response factor, \( F/F_r \) against variable NaF concentrations, where \( F_r = \) response for 0 \( \mu \)M NaF and \( F = \) response at different NaF concentrations. The value at 0 \( \mu \)M of NaF solution would be an indication of the noise level in this assay and the \( F/F_r \) would account for the signal-to-noise level of this
system. Using this system, we could discriminate down to 5 μM F⁻ (Figure 4-3a). However, when we tested the system with crude cell lysates and 1× BugBuster™ to simulate the conditions of library screening, the discrimination power decreased to 50 μM (Figure 4-3b).

Figure 4-3: Quantifying the limit of detection in the fluorescence indicator displacement assay. (a) The sensitivity of the assay in the absence of cell matrices in library screens can reach as low as 5 μM. Inset shows the response factors at [NaF] lower than 20 μM. (b) The sensitivity of the assay is compromised in the presence of cell matrices. There was no discriminatory power when the [NaF] was lower than 50 μM.

Two reasons may account for the reduced performance in the library screening simulation. First, low-levels of non-enzymatic hydrolysis of the α-D-glucuronyl fluoride donor could have interfered with the assay at lower F⁻ concentrations. Second, the additional components that would be present under screening conditions could have caused the interference.

We found that the presence of cell lysate raised the pH of the reaction by nearly one unit (Figure 4-4). The ligand-exchange reaction is known to be sensitive to pH with an optimal around 5.8. If the presence of other cell components causes such differences to the pH, then this assay is unsuitable for high-throughput screening of the glucuronylsynthase libraries as the increase in pH would reduce the sensitivity of the assay. Furthermore, the use of t-BuOH in the reactions might complicate the situation further. Hence, this method would be limiting and another assay was needed.

4.3.3.2. Developing the LCMS method for library screening

Most of the problems faced in the previous section could be circumvented with an LCMS fitted with auto plate screening capability. The LCMS method would be more
straightforward and would be capable of detecting glucuronide drugs in derivatized urine samples at concentrations below 10 nM\(^6,7\). When the plate LCMS became available, we developed two analytical procedures: one for the analysis of enzyme kinetics and another for rapid analysis of libraries.

To resolve the peaks, we used an Agilent Poroshell C-18 column of 30 mm length and 1.8 mm diameter packed with 2.7 \(\mu\)m beads attached to a 5 mm guard column. Temperature of the column was maintained at 30 °C. The mobile phase system consisted of 10\% MeOH/90\% 10 mM NH\(_4\)OAc (mobile phase A) and 90\% MeOH/10\% 10 mM NH\(_4\)OAc (mobile phase B), and the flow rate was set at 0.2 mL/min.

A mock reaction solution containing 6 mM 1FGlcA, 2 mM CMO-DHEA, and 0.8 mM of CMO-DHEA glucuronide was prepared. This was then diluted two-fold in neat MeOH and further diluted 20-fold in MeOH/water (50:50). The signal strength for the glucuronide was relatively weak. Hence, the fragmentation power and capillary voltage (\(V_{cap}\)) were optimized to increase the signal to noise ratio for CMO-DHEA glucuronide. Table 4-1 lists the different conditions tried, with the final condition underlined. Using the fixed parameter, different ratios of the mobile phase solvents A and B were tested to optimize the separation of CMO-DHEA and CMO-DHEA glucuronide (Table 4-2). We also made further modifications to the method in order to achieve shorter analysis time without losing the resolution of the two peaks. This was achieved using 50:50 ratio of A/B. A calibration curve (Figure 4-5) was then obtained for CMO-DHEA glucuronide. The LCMS chromatograms are shown in Appendix 4-1 (Figure A4-1).
The optimized analytical method was then used to test enzymatic reactions with crude lysates. Overnight glucuronylsynthase reaction assays were then set up to exemplify screening conditions described in Section 4.3.2. We simulated a reaction that would mimic a potential hit with two times the activity of Syn by doubling the amount of crude lysate used in the reaction. Four 400 µL reactions for each Syn and hypothetical “hit” were set up. Peak areas for the hypothetical hits’s reactions were two times higher than the Syn (Table 4-3). This meant that we were able to discriminate improved variants with at least 2-fold improvement. The experiment validated the screening process and the analytical method.

Table 4-1: The different detector parameters optimized. Bolded and underlined are the finalized parameters.

<table>
<thead>
<tr>
<th>Capillary voltage (V&lt;sub&gt;cap&lt;/sub&gt;)</th>
<th>Fragmentation voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500</td>
<td>100, 175, 200, 225, 250, 275</td>
</tr>
<tr>
<td><strong>3000</strong></td>
<td>80, 100, 175, <strong>200</strong>, 225, 250</td>
</tr>
<tr>
<td>3500</td>
<td>80, 100, 175, 200, 225, 250</td>
</tr>
<tr>
<td>4000</td>
<td>80, 100, 175, 200, 225, 250</td>
</tr>
</tbody>
</table>

To further reduce the analysis time, we performed the chromatography using only a 5 mm guard column. Both analytes (steroid substrate and glucuronide product) eluted about 0.5 minutes quicker. We lost some resolution between the peaks but the discrimination of the peak areas between the Syn and the hypothetical hit was not compromised. Hence, we proceeded with this method for library screening as we could decrease the analysis time to less than 0.9 minutes per sample (Appendix 4-2, Figure A4-2).

In summary, we have developed, optimized and validated an analytical method that would allow us to screen the synthase libraries. We have also optimized an analytical method that would be suitable for kinetic characterization of the screening hits. The method developed for screening allowed us to detect Syn hits with two-fold improved catalytic rate. Approximately 400 clones were screened per night using this method.
4.4. Results and Discussion

4.4.1. Glucuronylsynthase rational engineering

Directed evolution represents an elegant way of engineering enzyme, but the success of an evolution project depends on the size and the quality of the library. The larger the screening capacity, the higher the chances of isolating improved variants.
This poses a challenge for directed evolution experiments that are performed without the convenience of high-throughput screening facilities. This challenge can be overcome using directed evolution approaches that engineer enzymes based on structure-guided knowledge (semi-rational engineering). Structure-guided evolution approaches aim to create smarter libraries, i.e. small but effective libraries, and involve screening libraries targeted at a specific site (site-saturation libraries) or at a particular region. We employed semi-rational design for directed evolution of the Syn so that the throughput of the libraries could be managed.

Targets for semi-rational engineering were chosen based on the consideration that the substrate system of the parent β-GUS and the β-GUS/E504G variant, Syn, is different. The β-GUS binds glucuronide, while the Syn binds glucuronyl fluoride of a different anomeric configuration and releases glucuronide. Therefore, the substrate system of the glucuronylsynthase is different from its parent. This may contribute to suboptimal binding of the glucuronylsynthase substrate and result in modest performance. Furthermore, the E504G mutation prevents the enzyme from forming a glycosyl-enzyme covalent bond, the reaction intermediate. This prevents the formation of the intermediate state that in the parent enzyme is highly stabilizing. In several studies, it was found that the intermediate in β-glucosidase offers a transition state stabilization up to 10 kcal/mol through substrate-enzyme interaction at the glycosyl C2 position. In one directed evolution experiment of a glycosynthase, it was postulated that the mutations accumulated in the improved glycosynthase mutant was imparted by conformational changes that resulted in the restoration of the transition state interactions at C2. Based on this, we rationalize that choosing sites that are on the flexible loops close to the Syn substrate binding pocket may yield useful mutations that would favour the new substrate system and subsequently lead to enhanced glucuronylsynthase activity. Following this rationale, we chose eight different sites for single-site saturation mutagenesis and performed random mutagenesis on a disordered loop in the active site region. Each of the experiment is discussed separately in the following sections.

4.4.1.1. Site saturation mutagenesis (SSM)

Our targeted sites for site saturation mutagenesis study (SSM) were: (i) 160, 162, and 164, which we refer to as the Domain 1 cluster, (ii) 361 and 362, which will be referred to as the disordered loop sites and (iii) N412, M447, and D508. These sites are
not directly involved in substrate binding but they reside on flexible loops that harbor other residues that are involved in substrate interaction (Figure 4-6). As such, they may alter the active site plasticity to accommodate a two-substrate system or stabilize the active site conformation in favour of the glucuronylsynthase substrates.

We used degenerate NNK codons to randomize each site separately (where N=a/c/g/t and K=g/t). This would encode all 20 amino acids with a degeneracy of 32 codons per 20 amino acids. The library size for NNK randomization was calculated according to the algorithm used in CASter\textsuperscript{12,13} (Appendix 4-3). We screened 84 clones in each library using 2 mM of CMO-DHEA and 3 molar equivalent of α-glucuronyl-fluoride (1FGlcA) donor. Using the algorithm described in Appendix 4-3, this would achieve approximately 85% coverage of the libraries. Table 4-4 summarizes the quality of the library, the quality of the screens and the activity distribution in each library.

![Figure 4-6: Sites (turquoise and blue) targeted for single-site saturation mutagenesis are labeled. Active site residue E413 and a glucaro-δ-lactam inhibitor in the crystal structure (PDB ID: 3K4D) are shown with carbons in white. The inactivated nucleophilic residue (E504G) is shown as dots and labeled as G504.

The coefficient of variation (% CV) was higher than the ideal 20%\textsuperscript{14}. This was because we had less control over the evaporation of t-BuOH co-solvent from 96-well plates during the course of a 12–16 h reaction. As a result, the problem of edge effects\textsuperscript{15,16} was more pronounced than what we encountered in the previous chapter. We attempted to minimize this effect by using higher volumes of reaction volume (400 μL as opposed to the typical 200 μL reaction volume used in previous chapter) and by double sealing the assay plates. This reduced the % CV, but still did not eliminate the risks of false-positives and false-negatives. To minimize our chances of missing potential hits, it was thus necessary to conduct a secondary screen.
Table 4-4: Summary of Syn SSM library screening. % of wells that were at least half as active as Syn (> 0.5x) and % of wells that had less than 20% glucuronylsynthase activity compared to Syn. Sites 160, 162, 164, 361 and 362 are less susceptible to mutations compared to sites 412, 447 and 508. % CV and mutation frequency are the indicators of the screening quality and library quality.

<table>
<thead>
<tr>
<th>Sites</th>
<th>&gt; 0.5x Syn activity (% well)</th>
<th>&lt; 0.2x Syn activity (% well)</th>
<th>Inter-plate % CV</th>
<th>Library quality</th>
<th>Mutation frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>100</td>
<td>0.00</td>
<td>26.0</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>162</td>
<td>84.4</td>
<td>2.08</td>
<td>28.0</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>164</td>
<td>100</td>
<td>0.00</td>
<td>33.0</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>361</td>
<td>85.4</td>
<td>2.08</td>
<td>16.0</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>362</td>
<td>89.6</td>
<td>4.17</td>
<td>16.0</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>412</td>
<td>36.5</td>
<td>61.5</td>
<td>33.0</td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>447</td>
<td>26.0</td>
<td>39.6</td>
<td>35.0</td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>508</td>
<td>27.1</td>
<td>61.5</td>
<td>32.0</td>
<td></td>
<td>62</td>
</tr>
</tbody>
</table>

Secondary screen was conducted on the top 10% wells from libraries that were neutral. Some of the clones isolated in the primary screening were identified as false positives as they did not perform as well as they did in the primary screens. Clones that showed as much glucuronylated product as Syn in the secondary screen were also sequenced. A few false positive clones from library 361 were also picked for sequencing. These were identified as L361A and L361V. Table 4-5 lists the relative activities of the variants in the secondary screen. The clones that were confirmed by the secondary screen to have the best improvement in their respective libraries, Y160G and H162Q, were purified and characterized. Their kinetic properties will be discussed in Section 4.4.2.

Table 4-5: Relative improvement of various mutants in secondary screen. Mutants were confirmed as false or true hits in the secondary screen. Standard deviations were taken from three replicates.

<table>
<thead>
<tr>
<th>Mutants</th>
<th>Relative improvement over SynWT (Average)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn</td>
<td>1.00 ± 0.07</td>
<td>Parent (Control)</td>
</tr>
<tr>
<td>Syn Y160G</td>
<td>1.59 ± 0.24</td>
<td>Potential positive hit</td>
</tr>
<tr>
<td>Syn Y160V</td>
<td>0.84 ± 0.19</td>
<td>Neutral mutation</td>
</tr>
<tr>
<td>Syn H162A</td>
<td>1.50 ± 0.09</td>
<td>Potential positive hit</td>
</tr>
<tr>
<td>Syn H162F</td>
<td>1.44 ± 0.21</td>
<td>Potential positive hit</td>
</tr>
<tr>
<td>Syn H162P</td>
<td>1.06 ± 0.08</td>
<td>Neutral mutation</td>
</tr>
<tr>
<td>Syn H162Q</td>
<td>1.77 ± 0.21</td>
<td>Potential positive hit</td>
</tr>
<tr>
<td>Syn H162R</td>
<td>1.14 ± 0.12</td>
<td>Neutral mutation</td>
</tr>
<tr>
<td>Syn L361A</td>
<td>0.42 ± 0.01</td>
<td>False positive</td>
</tr>
<tr>
<td>Syn L361V</td>
<td>0.92 ± 0.12</td>
<td>False positive</td>
</tr>
</tbody>
</table>
4.4.1.2. Disordered loop targeted random mutagenesis

We targeted a disordered loop (360-370) for random mutagenesis. A random library focused on residues 286–420 was generated such that it included the disordered loop, an N-terminal flanking region of 73 residues and a C-terminal flanking region of 50 residues. This region is also approximately half of the TIM-barrel (active site domain) and is positioned spatially opposite the inactivated nucleophilic residue E504G (Figure 4-7a). Engineering of this region may lead to the alteration of active site preference to accommodate the Syn substrates.

Ten clones were randomly picked to determine the average mutation rate ($\mu$). The $\mu$ determined was 1.0 amino acid change per gene (standard deviation, $\sigma = 0.67$). The standard deviation differed slightly from the theoretical standard deviation ($\sigma_{\text{theoretical}} = 1.0$) for the calculated $\mu$ assuming Poisson distribution. The difference reflected a deviation from the Poisson distribution that could have been caused by the small sampling size used in the $\mu$ determination. Approximately 550 clones in six 96-well plates were screened. Based on the empirically determined $\mu$, the screening size covered about 15% of the theoretical library size ($\binom{150}{1} \times 20$ possible amino acids in each position, where the number of residues randomized = 150). 40% of the library was nearly as active as Syn, and indicated the neutrality of the region towards mutations (Figure 4-7b).

Figure 4-7: Screening of the targeted random mutagenesis library. (a) The region that was subjected to random mutagenesis (blue). This region encompasses the disordered loop (red) and makes up the part of the TIM-barrel that is opposite the E504G (shown as dots). (b) Distribution of the synthase library relative activities (ranked in order) in six plates. Inset shows the inter-plate variation in each plate.
Several clones appeared to be better than the Syn. However, when eight of these were sequenced, six of them were confirmed as Syn. False identification was caused by a slightly higher inter-plate variation in one of the plates (~ 35%). When we closely examined the distribution of activity within the library (Figure 4-7), even though the sequenced mutants appeared to have more than 1.5-fold improvement in activity, there were other control wells (red bars in Figure 4-7) that had as high as 2-fold improvement. Consequently, this plate was more susceptible to false positives. Of the eight clones that were sequenced, the other two contained single mutations P371T and K372E respectively. These were purified and characterized to confirm their catalytic properties.

4.4.2. Kinetic properties of glucuronylsynthase and its variants

Kinetic parameters for glucuronylsynthase variants H162Q, Y160G, P371T and K372E in the presence of 20% t-BuOH were measured. We also recombined the two mutations P371T and K372E with H162Q and characterized them. Table 4-6 lists the variants that were characterized and the mutations that they harbor.

Table 4-6: Mutations in the variants that were kinetically characterized. Kinetic parameters for Syn and six Syn variants were obtained.

<table>
<thead>
<tr>
<th>Variants</th>
<th>Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn</td>
<td>E504G</td>
</tr>
<tr>
<td>Syn Y160G</td>
<td>E504G + Y160G</td>
</tr>
<tr>
<td>Syn H162Q</td>
<td>E504G + H162Q</td>
</tr>
<tr>
<td>R2V1</td>
<td>E504G + P371T</td>
</tr>
<tr>
<td>R2V2</td>
<td>E504G + K372E</td>
</tr>
<tr>
<td>R3V1</td>
<td>E504G + H162Q + P371T</td>
</tr>
<tr>
<td>R4V2</td>
<td>E504G + H162Q + P371T + K372E</td>
</tr>
</tbody>
</table>

Kinetic characterization was first performed using CMO-DHEA concentrations between 0 mM and 4 mM, but when we attempted to fit the data to the Michaelis-Menten (MM) model, these concentrations appeared to be significantly lower than the $K_m$. We repeated some the experiment with higher concentrations of CMO-DHEA i.e. 8, 10 and 12.5 mM. However, even at these concentrations, the $V_{max}$ was not approaching saturation (Figure 4-8) and signifies that we were still operating at concentrations below the $K_m$ value for the CMO-DHEA acceptor substrate. Consequently, the parametric values $k_{cat}$ and $K_m$ determined from these plots may still
be inaccurate. Therefore, we determined the $k_{cat}/K_m$ by extrapolating the linear slope at low substrate concentrations in the rate plot (Figure 4-8, dotted lines).

At low substrate concentrations \textit{i.e.} concentrations lower than the $K_m$ ($K_m \ll [S]$), the MM equation can be simplified to $v_0 = \frac{V_{max}[S]}{K_m}$ as the $[S]$ term in the denominator of the MM equation approaches zero\textsuperscript{17}. Hence, at low substrate concentrations, $v_0$ exhibits linear dependence with $[S]$, and the $V_{max}/K_m$ \textit{i.e.} $k_{cat}/K_m$ can be determined from the linear slope extrapolated from the low $[S]$ concentrations in the rate plot\textsuperscript{17}. This would give a more accurate estimate of the enzyme’s catalytic efficiency than from using the $k_{cat}$ and $K_m$ parameters obtained from the MM curve-fit obtained under non-saturating substrate concentrations, as is the case for the glucuronylsynthase and its variants.

Figure 4-8: Plots of initial rates versus increasing CMO-DHEA concentration for Syn (blue) and Syn H162Q (black) and the curve-fit parameters from applying the Michaelis-Menten model. Data was taken from three measurements, $n = 3$. Saturation could not be obtained even though the regression values indicated a good fit of data. Curve-fit parameters may be inaccurate. The $k_{cat}/K_m$ parameters were determined from the extrapolated slope (dotted lines) of the plot.

To compare the improvement in catalytic activity, we compared the turnover rate of the variants at sub-saturating CMO-DHEA concentration (4 mM). Under this condition, single mutant Syn H162Q was the most efficient with a rate of $0.80 \times 10^{-2}$ s$^{-1}$, followed by Syn Y160G that has a turnover rate of $0.25 \times 10^{-2}$ s$^{-1}$ (Figure 4-9a). Hence, the single mutant, Syn H162Q catalyzed the synthetic reaction by approximately 4-
times quicker than Syn, while Syn Y160G only improved the glucuronilation activity by 1.5-times with 4 mM CMO-DHEA. Values of $k_{cat}/K_m$ (obtained from initial $v_0$) also point towards the variant Syn H162Q to be approximately 4-times more efficient than Syn (Figure 4-9b). Hence the mutations Y160G and H162Q improve both the turnover rate and enzyme efficiency simultaneously by 2 and 4-times respectively.

![Kinetic characterization of Syn and Syn variants.](image)

Figure 4-9: Kinetic characterization of Syn and Syn variants. (a) The initial $v_0$ of enzymes using 4 mM of CMO-DHEA acceptor substrate. The standard deviations were taken from at least three independent replicates, $n = 3$. (b) The trends of CMO-DHEA $k_{cat}/K_m$ for Syn and the variants determined from the slope of $v_0$ against [S]. Error bars indicate the standard deviation from three data sets, $n = 3$.

We recombined Syn H162Q with P371T and K372E. In the absence of H162Q mutation, the mutations P371T and K372E were worse than the Syn. This trend is different from the screening data where both mutants were identified as potential hits.
with at least two-times the activity of Syn. The kinetics experiment here therefore confirm that these were also false positives. Nevertheless, when the mutations were recombined with H162Q, the variants R3V1 and R4V2 appeared to be 2-times more efficient than Syn. Nonetheless, both recombined variants were not as good as single mutant Syn H162Q. The observations provide confirmation that the H162Q mutation is indeed advantageous, but both P371T and K372E mutations are in fact slightly deleterious. The mutations H162Q, P371T and K372E also appeared to be interacting in an additive manner i.e. the effects of the mutations combine additively.

The difficulty in reaching kinetic saturation that we faced was also reported by Wilkinson et al.\textsuperscript{2}, and is a reflection of the high $K_m$ for the acceptor substrate. This is expected since the wild-type (β-GUS) from which Syn was derived is known to be glycosyl specific, but a generalist for the acceptor moiety. This feature is advantageous to us for the purpose of developing a system that can be applied to a broad range of acceptor substrates. To investigate the potential broad range applicability of the Syn variants, we compared the relative performance of Syn, Syn Y160G and Syn H162Q in the presence of sub-saturating levels of acceptor substrates – 4 mM testosterone and 4 mM CMO-DHEA (Figure 4-10). Both mutations conferred similar enhancement in glucuronylation activity for both steroidal substrates. This suggested that the variants would be capable of faster glucuronylation with a broad range of acceptor substrates.

![Figure 4-10](image_url)

Figure 4-10: Investigating the relative performance of Syn variants with different steroid acceptor substrates. The steroid acceptor substrates tested were: (a) CMO-DHEA, and (b) testosterone. (c) The relative performance of Syn variants tested against the two acceptor substrates at a fixed substrate concentration of 4 mM.
4.4.3. Stability of glucuronylsynthase and its variants

4.4.3.1. Half-lives in t-BuOH and thermostability

Acceptor substrates of the glucuronylsynthase system that will be of practical interest are typically hydrophobic. Examples include codeine and steroidal compounds, such as CMO-DHEA and testosterone. The use of t-BuOH as a solvent overcomes the solubility issues of hydrophobic acceptor substrates and improves the reaction rate. This necessitates the addition of t-BuOH in the glucuronylsynthase reaction but its stability in solvents had not been determined. Furthermore, the duration to achieve full conversion of CMO-DHEA is considerably long (2–3 days). It would be interesting to know if the Syn and its variants remain active throughout the course of the reaction. Hence, we measured the half-lives of the Syn and single mutants (Syn H162Q and Syn Y160G) in 20% t-BuOH. Enzyme solutions were incubated with 20% t-BuOH at 30 ºC between 0 h and 12 h. Incubated solutions were sampled at 1, 2, 4, 7 and 12 h. The residual activities of the enzymes sampled at different time points were plotted as a function of incubated time. The half-lives (Table 4-7) were then determined from the sigmoidal function described in Section 2.11.6.

We observed half-lives (t_{1/2}) of less than 2 h for Syn and less than 30 min for the glucuronylsynthase variants (Table 4-7). These values were considerably shorter than the usual reaction duration of 24–36 h, and were incompatible with our observation from monitoring glucuronide production during the synthetic reactions. The glucuronide peak area continued to increase even after 12 hours. This discrepancy could be reconciled with the fact that we had performed the incubation for stability study without a ligand or substrate. We suspected that the presence of 1FGlcA substrate stabilized the glucuronylsynthase enzyme in the reactions. To verify this, we monitored the unfolding of the enzyme in the presence and absence of 1FGlcA (Figure 4-11) using circular dichroism (CD). Without 1FGlcA, alteration to the secondary structure of the enzyme was observed within the day (Figure 4-11(a–b)). However, in the presence of 1FGlcA, the secondary structures of mutant enzymes

<table>
<thead>
<tr>
<th>t_{1/2} (hr)</th>
<th>Syn</th>
<th>Syn Y160G</th>
<th>Syn H162Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7</td>
<td>0.73</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>1.9</td>
<td>0.72</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>1.8</td>
<td>0.75</td>
<td>1.0</td>
</tr>
<tr>
<td>Average</td>
<td>1.8</td>
<td>0.73</td>
<td>1.0</td>
</tr>
<tr>
<td>SD</td>
<td>(± 0.095)</td>
<td>(± 0.014)</td>
<td>(± 0.0049)</td>
</tr>
</tbody>
</table>
remained unchanged (as determined by CD) after overnight incubation in t-BuOH (Figure 4-11(c–d)).

The stabilization effect of 1FGlcA was also observed in a thermal shift assay experiment. We obtained the $T_m$ of the enzymes using the procedures described in Section 2.12.1 and observed that the $T_m$ of the enzymes increased significantly with the addition of 1FGlcA (Figure 4-12). This observation is similar to other experiments that have described the stabilizing effects of substrates and ligands. We also compared the degree of stabilization by 1FGlcA on the variants by estimating the percentage difference in the $T_m$ with 1FGlcA and without 1FGlcA (Figure 4-12a). Thus, the larger
the percentage difference, the larger is the ligand stabilization effect. The increase in the T_m induced by the presence of 1FGlcA was the most significant for Syn H162Q, suggesting that the variant is more stabilized by the substrate than the parent Syn. Increased ligand stabilization effect by the mutation H162Q can be explained by the inter-domain interaction between the loop in Domain 1 and the active site (Figure 4-12b).

Figure 4-12: Effects of 1FGlcA on T_m. (a) The stabilization afforded by 1FGlcA for each variant. The values of T_m with and without 1FGlcA for the three variants are listed in the inset table below the plot. (b) The inter-domain interaction between Domain I and the active site in the crystal structure with a glucaroo-δ-lactam (GDL) inhibitor. D163 is a substrate-binding site approximately 3.2 Å away from the carboxylate group of GDL. (c) Thermal dissociation curves with and without 1FGlcA for (i) Syn, (ii) Syn Y160G (iii) Syn H162Q. Syn appears to be more thermostable compared to the variants.

In addition, we also observed a shoulder during dissociation (in black) in the thermal denaturation profiles of enzymes incubated with 1FGlcA (Figure 4-12c). This is comparable to the biphasic transition that was observed in the endotherms of other enzymatic systems at subsaturating ligand concentration^{21,22}. Our observations can be
accounted for by considering the simplest model for the denaturation of an enzyme: Folded (F) ↔ Unfolded (U). During denaturation, the two forms are in equilibrium and at T_m they exist in 1:1 molar ratio. In the presence of ligand, the folded form binds a ligand molecule that upon unfolding, is released. At subsaturating ligand concentration level, the released ligand can be taken up by a transient intact enzyme molecule^{21,22}. Consequently, this gives rise to the biphasic transition that is seen during denaturation of enzymes with subsaturating ligand concentration. In our case, the presence of the shoulder in the thermal dissociation curves suggests that the event of 1FGlcA exchange between the unfolded and folded forms occurs during denaturation of the Syn variants when 1FGlcA is present as a ligand.

The different degree of ligand stabilization afforded by the presence of 1FGlcA for the three variants suggests that the interaction between Domain I and the active site has changed and this may have affected the substrate affinity of this loop. The loop that harbours residue H162 is involved in ligand and substrate binding through the neighbouring residue D163 (Figure 4-12b). When the ligand 1FGlcA is bound in the active site, the interaction between this residue and the ligand locks the loop in place, thereby increasing the rigidity of the enzyme and hence, the stability. The mutation H162Q may change the affinity of this loop for the 1FGlcA ligand and thus, result in higher degree of 1FGlcA stabilization for the Syn H162Q variant.

### 4.4.3.2. Inactivation of glucuronylsynthase in the presence of t-BuOH and urea

The activities of Syn and variants, Syn Y160G and Syn H162Q, at different t-BuOH concentrations were determined. Enzyme solutions were incubated in different t-BuOH concentrations for two hours at 30 °C. Their activities after incubation were determined using reaction assays containing 4 mM of steroid CMO-DHEA and 3 mM glycosyl fluoride. Residual activity was calculated using the activity of the samples incubated with 0% t-BuOH as the reference. To monitor urea-induced denaturation, enzyme solutions were incubated in urea for 30 min at 30 °C. Urea induced denaturation was observed using intrinsic fluorescence as described in Section 2.12.2. Unfolding was monitored in the presence of 2–8 M urea. The concentrations of t-BuOH, ([t-BuOH]^{50}) and urea ([Urea]^{50}) at which the enzymes lost 50% of their initial activity
or are 50% unfolded (Table 4-8) were determined from sigmoidal curve fit values (Figure 4-13).

Table 4-8: The concentration of \( t\)-BuOH at which 50% of the activity is lost for Syn and the variants and the concentration of urea at which 50% of the enzyme samples were unfolded.

<table>
<thead>
<tr>
<th>Urea (M)</th>
<th>t-BuOH (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn</td>
<td>4.62 (± 0.088) 21.8 (± 0.66)</td>
</tr>
<tr>
<td>Syn Y160G</td>
<td>2.38 (± 0.020) 15.7 (± 1.50)</td>
</tr>
<tr>
<td>Syn H162Q</td>
<td>2.04 (± 0.032) 19.5 (± 0.83)</td>
</tr>
</tbody>
</table>

Figure 4-13: Inactivation of the enzyme solutions in \( t\)-BuOH and urea. (a) Residual activities of the enzymes as measured using reactions containing 3 mM donor substrate and 4 mM acceptor substrate. (b) Unfolding of the enzymes in urea was monitored using intrinsic fluorescence and by comparing the ellipticity measurements at 325 nm and 345 nm.

Unfolding of the enzymes in urea was observed with a red-shift in fluorescence. The wild-type glucuronolysynthase could tolerate higher concentrations of urea before it was fully unfolded. This and the earlier data on thermostability indicate that the variants have traded-off some biophysical properties in exchange for improved catalytic activity. This trade-off effect\(^{23-25}\), is a common occurrence in evolution when attempting to evolve for enhanced catalytic property because activity and stability are governed by different features. It is widely believed that stability benefits from an increase in structural rigidity while some degree of structural flexibility is required to modulate the activity. We would not have been able to increase both rigidity and flexibility at the
same time with only a single mutation (Y160G or H162Q). This would have caused the trade-off that was seen in our case.

Intriguingly, although we had evolved Syn H162Q and Syn Y160G to have higher catalytic activity in the presence of 20% \( t \)-BuOH, we have not managed to increase its \([t\text{-BuOH}]\)\textsuperscript{50}, its \( t \)-BuOH tolerance. When this result was compared alongside the thermostability profiles and the half-lives without 1FGlcA obtained in the earlier section, Syn consistently showed higher stability compared to its variants. All the results taken together indicate that we have increased the dependence of the variants on the 1FGlcA to stabilize them.

Overall, although we have increased the synthetic activity of the glucuronylsynthase variants, Syn Y160G and Syn H162Q, we have made small sacrifices in their biophysical properties. We have likely increased their dependence on 1FGlcA to stabilize the enzymes. This corroborates earlier discussions that suggested that the intrusion of the loop from Domain I into the active site in Domain III plays a significant role in substrate binding and that the role of this loop in substrate binding is also an important factor for the stabilization of the glucuronylsynthase active site.

### 4.4.4. The trend in glycosyl binding properties of Syn and variants

Isothermal calorimetry titration (ITC)\textsuperscript{26} experiments are used to determine binding parameters by measuring the heat absorbed or heat liberated during the binding event. ITC does not establish kinetic parameters but its thermodynamic parameter (\( K_a \) or \( K_d \)) is often taken as an approximation of an enzyme’s \( K_m \) values\textsuperscript{27,28}. ITC can therefore be used to estimate the binding affinities of the three variants for the 1FGlcA substrate. The disadvantage of using this method to quantify the binding affinity of the glucuronylsynthase with the 1FGlcA is that, measurements may be complicated by background hydrolysis of 1FGlcA. However, analysis of the binding isotherm (Appendix 4-4, Figure A4-4) did not indicate the presence of background hydrolysis, or the background hydrolysis is not significant enough to completely swamp the heat liberated from the binding event. In addition, 1FGlcA is used as a substrate in enzymatic reactions that were run over 3–5 days, which implies that the background hydrolysis, if present, should be on a longer timescale than the ITC experiment that
takes approximately two hours. The results from ITC can therefore give an approximation of the trends in the binding affinity among the different variants.

The ITC experiments were carried out as detailed in **Section 2.12.4**. Independent measurements (n = 2) to estimate the dissociation constant ($K_d$) were performed using 100 µM of freshly prepared enzyme and 4–5 mM of freshly prepared ligand. The ligand used was 1FGlcA prepared in 50 mM phosphate (NaPi) buffer, pH 7.4. We titrated 1.25 µL of 1FGlcA into the enzyme, with an equilibration time of 300 s between each injection. A non-linear regression global curve fit was used to calculate the binding parameters. The binding was exothermic ($\Delta H < 0$) and a best fit with stoichiometry $n = 1$ was obtained from an independent model using the curve-fitting function on NanoAnalyze (TA Instruments). The affinities of the glucuronysynthase variants for 1FGlcA were significantly higher than Syn. Dissociation constants ($K_d$) for Syn, Syn Y160G and Syn H162Q were determined to be in the order of $10^{-4}$ M (Syn), $10^{-5}$ M (Syn Y160G) and $10^{-6}$ M (Syn H162Q) (Figure 4-14). Representative binding isotherms for the three samples are shown in **Appendix 4-4**.

The standard deviations for replicates of ITC measurements on Syn were large and mirrors previous attempts in the McLeod group where their binding parameters were also estimated with large standard errors⁴. Although this indicated some degree of inaccuracy in the determination, it also served as an indication of its weak affinity for
1FGlcA. In both instances, the suboptimal binding of Syn with 1FGlcA would cause low values of the critical parameter, $c$, which determines the shape of its binding isotherm. This is determined by the equation\(^{26,29}\): $c = E_t \times K_a$, where $E_t$ = total enzyme concentration, $K_a$ = determined association constant. The critical parameter, $c$, is related to the steepness and the resolution of the titration curve. A lower $c$-value indicates that the binding is too poor and the titration curve is too shallow for accurate determination of the $K_a$ parameter. For accurate determination of the binding constant, the value of $c$ should fall between 10 and 100\(^{29}\). Since the binding of Syn with 1FGlcA is weak, the $c$ parameter would have been too small for its $K_a$ to be determined with high accuracy (Table 4-9). However, the measurements for Syn Y160G and Syn H162Q were more accurate as the requisite in the $c$ parameter was easier to satisfy because they both had better affinity for 1FGlcA than the Syn.

Table 4-9: The values of the $c$-parameter in Syn and the variants. The enzyme concentration used was 100 µM. The required total enzyme ($E_t$) to achieve $c$-parameter of at least 10 is given in the last column. For Syn H162Q, the required $E_t$ is less than what was used. Using 100 µM of the enzyme solutions, we achieved the required $10 < c < 100$ for Syn H162Q.

<table>
<thead>
<tr>
<th>E(_t) used</th>
<th>$c$-parameter</th>
<th>Required $E_t^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn</td>
<td>100 µM</td>
<td>0.28</td>
</tr>
<tr>
<td>Syn Y160G</td>
<td>100 µM</td>
<td>7.24</td>
</tr>
<tr>
<td>Syn H162Q</td>
<td>100 µM</td>
<td>14.1</td>
</tr>
</tbody>
</table>

To summarize, the glycosyl substrate binding affinities of Syn and the two variants can be ranked as such: Syn H162Q > Syn Y160G > Syn. The $K_d$ determined for Syn with 1FGlcA (355 µM) also agrees well with the $K_m$ of the Syn with 1FGlcA (200 µM) that was determined by Wilkinson et al.\(^2\) When we compared the $K_d$ values with the kinetic parameter $k_{cat}/K_m$, the trend reflects a correlation between increased affinity for 1FGlcA of the variants, Syn Y160G and Syn H162Q, and their catalytic efficiencies (Figure 4-15). This also agrees with our earlier experiment in the thermal shift assay in which Syn H162Q and Syn Y160G both saw larger stabilization effects by the 1FGlcA substrate and that suggests that the loop conformation change induced by the mutations increased substrate affinity.
Figure 4-15: Trends in $K_d$ and $k_{cat}/K_m$ of the Syn variants. Improvement in catalytic efficiency seemed to be correlated with higher affinity.

4.5. Structural deduction of mutations Y160G and H162Q

The consequence of Y160G mutation can be rationalized from the crystal structure of β-GUS (PDB ID: 3K4D)\(^{30}\). Y160 forms a H-bond with S557 (Figure 4-16) that will be disrupted when tyrosine is replaced by glycine. Consequently, the mobility of both loops increases. In the case of residue 160, an increase in the dynamics of this loop changed the substrate binding property, thus changing its glycosyl affinity. However, disruption to the Y160-S557 bond also caused the flexible loop harbouring S557 to be less rigid. Since the S557 loop is involved in subunit interactions, decreasing its rigidity would result in lower stability. As a result, Syn Y160G exhibited increased activity but lower stability.

Figure 4-16: Interaction between Y160 and S557.
The mutation H162Q does not appear to be directly involved in glycosyl binding. In the crystal structure, the side chain of H162 points away from the active site (Figure 4-16). The only possible interaction that a glutamine at position 162 may have is a hydrogen bond between its carbonyl group on the side chain with its backbone amino group. His→Gln mutation may rigidify the loop to facilitate better interactions between its neighboring residue, D163 with the non-native α-glucuronyl substrate (1FGlcA). The large % change in $T_m$ observed upon the addition of 1FGlcA suggests a high degree of stabilization by the presence of the ligand, which supports this line of reasoning. The new loop conformation induced by Q162 may restore some of the stabilizing effects of the transition state in the native hydrolytic enzyme that would have been lost with the substitution at the nucleophilic residue. As proposed in the work by Kim et al., restoration of the glycosyl C2-hydroxyl group interaction with the enzyme was one main reason that accounted for the improvement brought about by their mutations\textsuperscript{11}. Along a similar line, we postulate that the mutation H162Q changes the dynamics of the loop that alters the interaction of its neighbor, D163, with the active site pocket. Consequently, this adjusts the active site conformation to favour the glucuronylsynthase donor substrate of an inverted anomeric configuration.

Figure 4-17: Position and spatial orientation of residue 162 and 163 in the crystal structure. His→Gln mutation at position 162 was performed using the mutagenesis function on PyMol with an RMS value 0.027. The side chain of residue 162 points away from the active site and does not interact with the GDL inhibitor (substrate analogue), but is capable of hydrogen bond interaction with its backbone amino group. Altering this residue may have changed the position of the loop so that the binding affinity between D163 and the inhibitor (GDL) also changed.
4.6. Conclusion and future directions

Our experiment shows that targeting sites that can modulate loops that affect glycosyl binding has been a reasonable strategy for engineering a faster glucuronylsynthase. Glucuronylsynthase single site-saturation libraries Y160 and H162 each produced a variant (Syn Y160G and Syn H162Q) that were two and four-times more efficient than Syn. Our thermal shift assay experiment suggests that there may be a causal relationship between 1FGlcA binding affinity and the stabilization of the active site pocket through better substrate-enzyme interaction. X-ray crystal structures of the glucuronylsynthase soaked with a 2’deoxy-2,2-difluoro-α-glucuronyl fluoride as a substrate analogue may provide structural evidence to support this. In addition, since both mutations appear to reduce the stability of the enzyme, it may be worthwhile to consider an evolution for increased stability before recombining the mutations.

Surprisingly, random mutagenesis of the disordered loop that is located in the active site space did not yield any beneficial mutants. This loop is poised for loop reconfiguration that can affect glycosyl-enzyme interactions. The best mutations identified (P371T and K372E) from screening this library were confirmed to be slightly deleterious when combined with the Syn or with Syn H162Q. It seems that targeting this region is disadvantageous for improving the glucuronylsynthase activity. Nevertheless, it must be noted that the library was generated with low mutation rate (µ = 1), which could produce a neutral library. Furthermore, the screening data for this library was associated with large errors. Consequently, there may be minute improvements or neutral mutations that were not detected. Since the accumulation of neutral mutations can act synergistically to produce an overall improvement, evolution of this loop and nearby regions may warrant a second investigation.

The outcome from incorporating E504G into L561S showed that our initial strategy was not effective. We thought that this was due to increased competing hydrolytic activity (Figure 4-16). However, we were not sure how this could have arisen. We made two inferences. First, the serine mutation (L561S) reactivated the hydrolytic activity. This would be possible if the introduction of mutation E504G has not completely abolished the parental activity. In the seminal work on glycosynthases done by the Withers group, they quantified that their glycosynthase retained
(8 \times 10^{-6})\% of the glycosidase activity. This is very low but it suggests the possibility that a variant with improved glycosidase activity could have higher background hydrolysis when E504G is incorporated.

Second, there was wild-type contamination. Since the glucuronylsynthase activity is very low and the \( \beta \)-GUS activity is very high, slight contamination of the \( \beta \)-GUS species may reverse the accumulation of glucuronide product from the glucuronylsynthase reaction, thus inhibiting the synthetic reaction. However, we could not pinpoint the source for \( \beta \)-GUS contamination. It could be caused by human error or from using the wrong expression system. Yet, a third possible contamination source may have been translational misincorporation. This had been quoted in the literature\textsuperscript{32}, but had never been systematically investigated. We will describe this in the next chapter.
## 4.7. References


5. β-GUS contaminant and translational misincorporation in the glucuronylsynthase system

5.1. Introduction

In Chapter 4, we attributed the absence of glucuronylsynthase activity in Syn L561S to wild-type contamination. We suggested several possibilities of contaminating sources. One of these was translational misincorporation. In this chapter, we will show that wild-type contamination from translational misincorporation can give rise to wild-type contamination that can affect the performance of the glucuronylsynthase system.

Translational misincorporation is a biological event that arises from errors during protein synthesis. This can be caused by inaccurate base transcription or codon translation. Resulting from this is the introduction of trace amounts of contaminating species, which can reduce the quality and yield of desired protein. The rate of translational misincorporation varies depending on the protein and the expression system that is used. Production of recombinant protein under unnatural environment and high metabolic burden can increase the rate of misincorporation. Nevertheless, the rates of misincorporation are typically quoted as 1:10^6-8 for base transcription in *E. coli*, and 1:10^3-4 for codon translation in *E. coli* and mammalian expression systems.

Generally, the rates of misincorporation are too low for it to be detected or for it to cause adverse effects. However, occasionally, misincorporation may lead to observable altered function or protein misfolding. As such, translational misincorporation can have far-reaching consequences for protein structure and function studies. It also represents a niche issue for the preparation of therapeutic proteins as minor contaminants can induce immunogenic responses. However, its occurrence barely has significant consequences in biocatalysis, including that for glycosynthases, but the protein purity for a pharmaceutical company marketing a recombinant protein may be of utmost concern and knowing the factors that affect translational misincorporation may be extremely useful. Unfortunately, the techniques that have been used to detect misincorporation are not ideal for rapid and easy detection under numerous conditions.
There are several ways to detect translational misincorporation. This includes studies using isoelectric focusing\(^\text{11}\), Edman sequencing\(^\text{12}\), two-dimensional gels\(^\text{13}\) and mass spectrometry coupled with peptide mapping analysis\(^\text{9,14–16}\). Alternatively, we can deduce the effects of translational misincorporation by investigating altered properties, for example reduced activity\(^\text{5,17,18}\) or misfolding\(^\text{7,19}\) of a protein. We will explore the translational misincorporation in glucuronylsynthase by monitoring the reacquisition of the native (hydrolytic) activity in the wild-type, β-GUS.

The investigation on wild-type contamination and translational misincorporation will be pursued in three parts. First, we will investigate the catalytic action of nucleophilic catalyst E504G. Then, we will systematically eliminate exogenous sources of wild-type contamination. Lastly, we will observe the effects of translational misincorporation using codon usage, expression conditions and different glycoside substrates. The hydrolytic activity of the β-GUS is extremely high so that reversions of E504G to the wild-type (E504) can be easily detected with an activity assay that can be monitored with crude lysates. Hence, some idea of the factors that affect misincorporation can also be easily detected over a variety of conditions.

5.2. **Results**

5.2.1. **The catalytic nucleophile of *E. coli* β-GUS E504**

The role of residue E504 in the *E. coli* β-GUS has been reported previously\(^\text{20}\). When sequence alignment was performed with other sequences that are related to *E. coli* β-GUS (human β-GUS, mouse β-GUS, and *E. coli* β-GAL), the position E504 is fully conserved and aligns with the nucleophilic residues in the human β-GUS (E540)\(^\text{21}\) and *E. coli* β-GAL (E537)\(^\text{22}\) (Figure 5-1a). In addition, we studied the effects of small nucleophilic molecule formate on the Syn hydrolytic activity, where we titrated 0 M, 2 M and 4 M formate into a reaction containing pNPGlcA and Syn. Formate acts as a small nucleophilic molecule that occupies the space created by the EÆG substitution in the active site pocket of Syn, thereby reactivating the hydrolytic activity\(^\text{23,24}\).
We observed a 30-fold recovery in hydrolytic activity (Figure 5-1b) when formate was present. Doubling the concentration of the activator molecule also doubled the recovered activity. The reaction progress curves also revealed a lag time that became more pronounced with increasing formate concentration. This may be an indication of slow equilibration of the formate, which took the place of the glutamate as the nucleophile, in the active site. We did not characterise the α-glucuronyl formate product but the formation of the yellow product that was monitored under UV/Vis absorbance of 405 nm indicated the liberation of pNP (Figure 5-1c). Hence, all observations thus far is compatible with the assignment of the E504 as the nucleophilic residue\textsuperscript{20,25–28}. The reacquisition of hydrolytic activity in the presence of small nucleophilic molecules is also in line with the other glycosynthases\textsuperscript{23,24,29–32} and indicates that the mechanism of glucuronylsynthase is likely to be similar to the proposed mechanisms for other glycosynthases.

5.2.2. Eliminating β-GUS contamination

Several experiments were set-up to ascertain that we had not introduced β-GUS contamination through the growth and expression system, or through the media components that were used. First, we created a double knocked-out mutant by substituting glutamate with glycine and alanine at the second catalytic residue, E413 that is responsible for base hydrolysis. If there were wild-type species contamination from external or internal factors, then cultures of glucuronylsynthase and the double knocked-out variant would exhibit similar background activities. The hydrolytic activities of both Syn E413A (E413A/E504G) and Syn E413G (E413G/E504G), and Syn (E504G) crude lysates, were observed qualitatively over a period of one week using 1 mM of pNP\textsubscript{GlcA} as substrate. The growth of BW25142 strain, used as the host cell for this experiment, and the overexpression of arylsulfatase (AtaS) were used as negative controls. The AtaS sulfatase catalyzes the hydrolysis of sulfate esters and does not hydrolyse glucuronides. It is also nearly the same size (60 kDa) as the β-GUS (69 kDa). Thus, it is a good negative control to establish that the overexpression of an enzyme in BW25142 host does not induce the synthesis of contaminating β-GUS. The overexpression of wild-type β-GUS served as the positive control. Whole cell lysates

--- 120 ---
Figure 5-1: Confirming E504G as the nucleophilic residue. (a) Sequence alignment of four glycosyl hydrolases at position E413 and E504 (E. coli β-GUS). (b) Recovery of hydrolytic activity in the presence of small nucleophile is indicated by the relative activity of reaction mixtures with formate compared with reaction mixtures without formate. 2 M and 4 M HCOONa were titrated into the enzyme incubation buffer containing 50 μM of Synthase enzyme. Turnover rate, $k_{cat}$, of the hydrolytic rate in 0 M formate was obtained from $k = 7625$ M$^{-1}$cm$^{-1}$. (c) Schematic diagram of the chemical rescue reaction.

were obtained from 10 mL of overnight cell cultures and lysed using rLysozyme™ (50 kU/g of cell mass). Overexpression of gene products was visualised on SDS-PAGE.

There was no β-GUS activity (Figure 5-2a) detected in the negative controls, BW25142 (empty cell line) and AtaS. In the SDS-PAGE analyses (Figure 5-2b), the overexpressed gene products for AtaS corresponded with the size of the sulfatase used, and no overexpression of gene products was detected for the empty cell line BW25142. The positive control, β-GUS turned over the pNPGlcA almost instantaneously (within a minute). The glucuronylsynthase exhibited low but noticeable β-glucuronidase activity.
It turned over the pNPGlcA overnight. However, the reaction assay containing double mutants of E504G recombined with E413A or E413G remained colourless even after one week. This indicated that the presence of hydrolytic activity by the double mutants was too low to be observed and was thus considered negligible.

The above results inferred two points. Firstly, the drastic loss of hydrolytic activity caused by mutating E413 implied that the glucuronylsynthase with a standalone E413 was capable of hydrolysis, albeit slowly and inefficiently. Secondly, the presence of contaminating wild-type that gave rise to observable β-GUS activity was unlikely to have arisen from the constituents of the growth media or from the expression system used. Hence, it could only be caused by an intrinsic factor such as inaccurate transcriptional and translational events during the course of growth and expression.

<table>
<thead>
<tr>
<th>(a)</th>
<th>Syn E413A</th>
<th>Syn E413G</th>
<th>Syn</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BW25142</td>
<td>AtaS</td>
<td>WT</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5-2: Eliminating external sources as potential contaminant. (a) Qualitative analysis of hydrolytic activities. The intensity of the yellow colour gives a crude indication of the rate of pNP liberation (b) Expression profiles of triplicate expressions of Syn (E504G) and Syn variants (E413A/E504G and E413G/E504G), host cell BW25142 and negative control represented by AtaS, and wild-type β-GUS used in this study.

5.2.3. Codon usage of mutation E504G affects the extent of hydrolytic by glucuronylsynthase

The genetic code is degenerate and leads to redundancy of codon usage. This gives rise to bias in codon usage that results in a particular set of codons being favored
by an organism for optimal translational process\textsuperscript{31}. Favored codons occur more frequently in an organism and are called common codons. We utilized this bias to investigate inaccurate translational events.

The misincorporation of glycine to glutamate in the \textit{E. coli} system was observed using mass spectrometry by Huang \textit{et al.}\textsuperscript{15} recently. Drawing on this finding, we incorporated two sets of glycine codon at residue E504. Misincorporation at this residue from G\textsuperscript{Æ}E would therefore introduce the revertant G504E that is essentially the wild-type (\(\beta\)-GUS). Hence, the extent of mistranslational events could be monitored through a simple \(\beta\)-GUS activity assay \textit{i.e.} the hydrolysis of \(\rho\)NPGlcA. The four degenerate codons that encode glycine are ggt (2.8), ggc (3.0), gga (0.7) and ggg (0.9)\textsuperscript{*}. We encoded “ggt” and “gga” to produce the common and the rare E504G mutants. If the effects of inaccurate translational events are small, they both should have comparable hydrolytic (\(\beta\)-GUS) and synthetic (glucuronylsynthase, \(\text{Syn}\)) activities.

Using His-Trap enzymes purified from His-Trap columns dedicated to each glucuronylsynthase (\(\text{Syn}\)) variants that were expressed from the two codon sets, we measured their hydrolytic activity with \(\rho\)NPGlcA. Their activities were measured from three replicates and averaged (Figure 5-3a). The hydrolytic activity of the \(\text{Syn}\) encoded with rare codons was approximately \(10^3\) fold higher than the \(\text{Syn}\) encoded with common codons (Figure 5-3b). However, there was only a slight difference between the \(K_m\) values of both the glucuronylsynthases compared to the wild-type (Figure 5-3b). From an unpaired \(t\)-test analysis, the difference was insignificant (\(p > 0.05\)), \(t(4) = -1.32, p = 0.25\) for rare \(\text{Syn}\), and \(t(4) = 1.59, p = 0.20\) for the common \(\text{Syn}\), and was likely due to systematic errors, as reflected in the large standard deviations of the \(K_m\) values of the glucuronylsynthases. Given that the \(K_m\) values are similar, we infer that the hydrolytic activity that were observed in the glucuronylsynthase samples were caused by the same species and the contaminant was most likely contaminating wild-type.

Both versions of the glucuronylsynthase were also tested for their synthetic activity in glucuronylsynthase enzymatic reactions. We detected glucuronylated products (CMO-DHEA glucuronide) in the reactions performed using \(\text{Syn}\) encoded by common codon but could not do so with the \(\text{Syn}\) encoded by rare codon. Comparison of

\textsuperscript{*} Values in parentheses denote the average frequency of the codon per 100 codons in \textit{E. coli}. 

— 123 —
the synthetic activity $k_{cat}$ against the hydrolytic $k_{cat}$ revealed that the synthetic rate of the Syn encoded with common E504G was 10$^2$-fold faster than the rate of its hydrolytic activity (Table 5-1). Thus, this would still result in a net production of β-glucuronide products. By contrast, the Syn encoded with rare E504G had a hydrolytic $k_{cat}$ value 10$^3$-fold higher than the hydrolytic $k_{cat}$ value of the common E504G Syn. Although we were not able to quantify the CMO-DHEA glucuronide production by the rare E504G Syn, we expected both versions of the Syn to have the same synthetic ability. Therefore, based on this assumption and the comparisons of the hydrolytic $k_{cat}$ values of both mutants, the hydrolysis rate of the rare E504G Syn will outpace its production of β-glucuronide. Consequently, this would result in no net production of glucuronylated products (Figure 5-4).

Figure 5-3: Kinetics of the hydrolytic activity exhibited by glucuronylsynthase encoded with rare and common codons. (a) Michaelis-Menten curve-fit of the β-GUS activity in enzymatic assays containing E504G encoded by rare codon (red) and common codon (black), where rare codon = gga and common codon = ggt. Curve-fit was done using the average of three replicates obtained from two different batches of expressed enzymes. (b) The relative activities of WT, Syn (rare) and Syn (common) in logarithmic values. The activity exhibited by Syn (rare) is 1000-fold higher than Syn (common). WT activity is more than 10$^5$ higher than the Syn (common). Embedded table compares the kinetic parameters of the β-glucuronidase activity obtained for wild type and the two versions of glucuronylsynthase.

Table 5-1: The turnover rates $k_{cat}$ of purified β-GUS and glucuronylsynthase activities by wild-type, Syn (rare) and Syn (common). $^{(a)}$ could not be determined because no products could be detected. $^{(b)}$ synthetic activity determined using 4 mM CMO-DHEA and 6 mM glucuronyl fluoride as in Chapter 4.

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>Syn (rare)</th>
<th>Syn (common)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{cat}$ [s$^{-1}$] {hydrolytic}</td>
<td>76.0 (± 2.6)</td>
<td>0.460 (± 0.26)</td>
<td>5.26 x 10$^4$ (± 0.35 x 10$^4$)</td>
</tr>
<tr>
<td>$k_{cat}$ [s$^{-1}$] {synthetic}</td>
<td>ND$^{(a)}$</td>
<td>ND$^{(a)}$</td>
<td>1.67 x 10$^{2.0(b)}$ (± 3.2 x 10$^{-3}$)</td>
</tr>
</tbody>
</table>
Overall, the results showed that the choice of codon affects the amount of β-GUS contaminants, which is an indication of different levels of mistranslational events. The choice of codon appears to have a real and observable effect on the efficiency of the glucuronylsynthase system.

5.2.4. Minimizing hydrolysis using different expression conditions

Overexpression is one of the factors that can increase translational inaccuracy that leads to translational misincorporation. Overexpression subjects the host organism to an unnatural environment that increases metabolic stress, which in turn increases error rates that can introduce impurities. The level of metabolic stress that is induced upon the cells depends on different overexpression conditions and this can subsequently affect the rate of translational inaccuracy.

In the glucuronylsynthase system, we can investigate the effects of different overexpression conditions by monitoring the G→E reversion at residue 504. We set up an experiment for this purpose. This experiment served two purposes. First, it served to provide additional support that translational inaccuracy events were accountable for the β-GUS activity seen. Second, through this experiment we would be able to determine conditions that can minimize the effects of translational inaccuracy.
We overexpressed glucuronylsynthase in BW25142 host cell for different lengths of time at three temperatures (18, 30 and 37 °C). Whole cells were lysed using a detergent based lysis method (BugBuster™). The overexpressed products were then visualized on SDS-PAGE (Appendix 5-1, Figure A5-1) and the total amount of crude lysates were determined using Bradford assay. Their activities were then measured using pNPGlcA assay and standardized according to the amount of total lysate. Standardized activities of the crude lysates were then normalized relative to the average activity observed from samples grown and expressed at 18 °C for 4 h (Figure 5-5).

Regardless of expression time, the samples expressed at 18 °C exhibited very low levels of β-GUS activity (0.022 – 0.020 nM/min/mg total enzyme) (Appendix 5-2, Table A5-1). Comparing this to the $k_{\text{cat}}$ value obtained in Table 5-1, this is approximately 0.020% of wild-type activity. The level of β-glucuronidase activity determined from samples expressed for 4 h doubled when the induction temperature was increased from 18 °C to 30 °C, and from 30 °C to 37 °C. The trend was different for samples obtained from 8 h and overnight expression. In these cases, the increase in the level of β-glucuronidase activity observed was larger when the induction temperature...
was increased from 18 °C to 30 °C compared to when the induction temperature was raised from 30 °C to 37 °C. In the case of overnight expression, the increase in β-glucuronidase activity was nearly 20-fold when the induction temperature was increased from 18 °C to 30 °C but only doubled when the induction temperature was raised from 30 to 37 °C. At 30 °C and 37 °C, we also observed an exponential increase in the β-glucuronidase activity after 8 h induction. In addition, it may be worth noting that the errors in the overnight expression sample at 37 °C was caused by large differences between duplicates on different days, and was likely caused by a difference in the incubation period (~ 20 h on Day 1 and ~ 16 h on Day 2). Using available data on the growth curve of BW25142 cell line harbouring constitutive expression systems (pJWL1030/gusA/BW25142) and of an inducible system harbouring the GUS gene (pET28a/GUS/GMS407) in GMS407 (DE3) cell line, the samples on Day 1 may have already entered the stationary phase, resulting in higher levels of misincorporation. Both growth curves showed two stages of growth phase where the growth slowed down after 4 hours, and enters into the stationary phase after 20 hours (Appendix 5-3, Figure A5-2).

From all the data, we infer that the rate of misincorporation increases exponentially as it transitions to a different growth phase but remains largely consistent during a particular phase. Prolonged induction at 30 °C and 37 °C increases the levels of misincorporation significantly, thus introducing more β-GUS contaminants. Altogether, the data suggests that starvation of the cells at the latter stages of overnight growth and expression is a major contributing factor of translational misincorporation of the glucuronylsynthase. This data also points towards the nature of the occurrence of β-GUS activity as one being related to an intrinsic factor of growth and expression rather than an external factor.

5.2.5. Hydrolytic activity of the glucuronylsynthase on other glycosides

Next, we tested the glucuronylsynthase system for glycosidase activity using purified glucuronylsynthase and purified wild-type on three glycoside substrates: β-glucuronide (native), β-glucoside and β-galactoside (Figure 5-6). β-glucoside is very similar to native β-glucuronide. It only differs at the C5 position, where the –COOH group is replaced by a CH₂OH. By contrast, β-galactoside is comparatively different
from the native β-glucuronide with C4 bearing a different configuration, and C5 a different functional group.

β-GUS exhibits promiscuous activity with β-glucoside and β-galactoside. The E504 residue is involved in substrate binding through the formation of a glycosyl-enzyme covalent bond. Thus, mutation here will change the glycosyl specificity. If the E504G mutant (glucuronylsynthase) exhibits basal levels of hydrolysis, then it may show different substrate specificity than the wild-type. Therefore, observable changes in the $K_m$ for different glycosyl substrates would indicate the presence of genuine hydrolysis by the glucuronylsynthase. Conversely, the $K_m$ with the glycoside substrates will be the same for both purified wild-type and glucuronylsynthase if the hydrolytic activity by the glucuronylsynthase is due to the presence of contaminating wild-type in purified enzyme sample.

Figure 5-6: The glycosyl rings of the three glycoside substrates studied. In parentheses are the substrates that harbour the glycosyl rings. β-D-glucopyranuronic acid is the native glycosyl ring of β-GUS native substrate, β-glucuronide.

The $k_{cat}$ values for the glucuronylsynthase were very low, approximately five orders of magnitude lower than the wild-type, for each of the glycosides (Figure 5-7). This might reduce the accuracy of the hydrolytic kinetics parameter that was determined for the glucuronylsynthase, but still serve as a good estimate. The substrate preferences of the parent β-GUS and glucuronylsynthase for the β-glucuronide and β-glucoside substrates were similar (Figure 5-7). However, there was a pronounced difference in the $K_m$ values exhibited by Syn and the wild-type for the β-galactopyranose substrate. The mutant displayed a lower binding affinity. This suggested that the β-galactosidase activity observed in the glucuronylsynthase was not caused solely by β-GUS contamination and may reflect the mutant’s catalytic aptitude with this substrate. This also suggests that the E413 is capable of low levels of base catalysis that gives rise to detectable water-assisted hydrolysis.
Water-assisted hydrolysis is extremely inefficient. Wang et al. reported an overall $k_{\text{cat}}$ of $10^{-6}$ s$^{-1}$ for their glycosynthase and $10^{-7}$ s$^{-1}$ as the true value for its parent glycosidase activity\textsuperscript{23}. This meant that 10% of the glycosidase activity seen in their synthase could be attributed to genuine hydrolysis caused by glucuronylsynthase. If we assume the same case, then genuine basal hydrolysis caused by the rare E504G contributes insignificantly to the hydrolysis, and in the case of common E504G would have been infinitesimal. Unfortunately, such low level of activity would neither be useful as a starting point for evolving the galactosidase activity nor for glucuronylsynthase evolution. Hence, it would be better to tailor growth and expression factors to reduce the misincorporation rates.

Figure 5-7: Hydrolytic kinetic parameters of the glucuronylsynthase and the wild type with different glycosides. The trend in $K_m$ values of purified Syn and $\beta$-GUS with $\beta$-glucuronide, $\beta$-glucoside and $\beta$-galactoside substrates. Inset: Enlarged view for $\beta$-glucuronide substrates. The turnover rates of purified $\beta$-GUS and Syn with three glycoside substrates.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrates</th>
<th>$K_m$ (mM)</th>
<th>$k_{\text{cat}}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn ($x10^6$)</td>
<td>$\beta$-glucuronide</td>
<td>526 ($\pm$ 35)</td>
<td>3.94 ($\pm$ 0.74)</td>
</tr>
<tr>
<td></td>
<td>$\beta$-glucoside</td>
<td>76 ($\pm$ 2.6)</td>
<td>0.134 ($\pm$ 0.040)</td>
</tr>
<tr>
<td></td>
<td>$\beta$-galactoside</td>
<td>1.42 ($\pm$ 0.14)</td>
<td>0.0197 ($\pm$ 0.00067)</td>
</tr>
<tr>
<td>$\beta$-GUS</td>
<td>$\beta$-glucuronide</td>
<td>76 ($\pm$ 2.6)</td>
<td>0.134 ($\pm$ 0.040)</td>
</tr>
<tr>
<td></td>
<td>$\beta$-glucoside</td>
<td>526 ($\pm$ 35)</td>
<td>3.94 ($\pm$ 0.74)</td>
</tr>
<tr>
<td></td>
<td>$\beta$-galactoside</td>
<td>1.42 ($\pm$ 0.14)</td>
<td>0.0197 ($\pm$ 0.00067)</td>
</tr>
</tbody>
</table>
5.3. Discussion

The use of glucuronylsynthase system to detect translational misincorporation is a chance event

Our observations on misincorporation events came under unusual circumstance. In many cases, translational misincorporation is probed by monitoring its detrimental effects on the function, specificity and stability of a protein. By contrast, our study involved monitoring the reacquisition of the native activity (β-GUS) in a knocked out system (glucuronylsynthase) whose function is reversed of its parent. Essentially, we were monitoring the effect of amino acid reversion at a catalytic residue that is critical for function. This was only possible because the wild-type activity was considerably higher than the mutant E504G’s activity. This meant that the slightest incidence of the reversion (~ 0.001%) was observable. Consequentially, this provided a useful alternative for us to estimate the misincorporation events.

Deducing translational misincorporation

When we first noticed that the glucuronylsynthase had higher than expected β-glucuronidase activity, we instinctively assumed that this was caused by wild-type β-GUS contamination due to an oversight in the experimental design. We thought this could be abolished through extra precautions. When this had failed to overcome the contamination issue problem, only two other explanations could have accounted for the presence of hydrolytic activity in the glucuronylsynthase system. The first was associated with the catalytic function of E504 and the mechanism of the glucuronylsynthase. We were opposed to this idea because the role of E504G is well established and our chemical rescue experiment supported this. Even though the stand-alone E413 may be capable of low levels of hydrolysis through water-assisted catalysis, this would be very inefficient and cannot account for the relatively rapid turnover of pNPGlcA. Secondly, intrinsic factors related to the growth and expression machinery in bacteria introduced a contaminating wild-type species through the misincorporation of G→E.
Bacterial cell replication and gene expression involve multitude biological components, which act together to create a seamless operation. However, each step in the process is prone to very low levels of errors either during transcription ($10^{-5}$-($-6$)35, translation ($10^{-3}$-($4$)36 or DNA replication ($10^{-7}$-$(-8$)$1,37. Although all three types of errors are intricately tied in that one can affect the other two, the errors arising from translation inaccuracy is known to be more frequent. Translation errors can be caused by ribosome errors during mRNA decoding ($10^4$), amino misacylation ($10^5$), and amino acid misrecognition by tRNA ($10^5$)35. Our experiments with codon usage and expression conditions provided strong circumstantial evidence that these erroneous events were likely the main contributors of the presence of impurities that gave rise to β-glucuronidase activity in the glucuronylsynthase system. In addition, this theory is supported by a recent novel observation of the G$\rightarrow$E misincorporation in a recombinant peptide in E. coli, which suggests that this amino acid misincorporation is possibly a prevalent occurrence in E. coli.

It is known that overexpression conditions can significantly affect the misincorporation rate. Its occurrence is more prevalent in high-yield expression systems due to metabolic stress. When stressed, the frequency of ribosome errors and amino acid misacylations increase, causing codon mistranslation. We have found that overexpressing the glucuronylsynthase system under high metabolic stress, which E. coli. is not optimized for affected the levels of β-GUS activity. Even though stronger overexpression conditions i.e. higher temperature and prolonged expression period for the glucuronylsynthase system yielded more enzymes, this also increased the incidence of translational errors. This gave rise to a higher proportion of mistranslated G504E products, which are essentially wild-type (β-GUS) impurities.

The proof for translational misincorporation was further strengthened by experiments with different codons. The genetic code is redundant, whereby the 20 canonical amino acids can be encoded by 61 codon sets. The usage of these codon sets is optimized differently in different in organisms according to their cellular tRNA makeup. This had been optimized so that translational inaccuracy can be minimized, causing the display of codon usage bias. In E. coli, its preferred codon sets for glycine are the ggt and ggc, while gga is a rare codon. Using the rare codon gga to encode the glycine mutation at E504 would have caused sub-optimal codon usage.
due to lower abundance of the rare codon cellular tRNA and result in higher rates of misincorporation\textsuperscript{40,41,42}.

The high native hydrolytic rate seen for β-GUS also extended to the promiscuous hydrolytic activities. This had allowed us to characterize these activities in the glucuronylsynthase system and confirmed that inefficient base catalysis can occur. Unfortunately we could not quantify this, as we did not use an analytical method such as the mass spectrophotometry (MS) or gel electrophoresis to quantify the amounts of misincorporated product. Even if we had attempted to use this, we might not have been able to quantify such low levels of misincorporated products since the application of analytical methods to detect misincorporated products that are orders of magnitude less than the error-free species is often challenging\textsuperscript{34}. Furthermore, this is subjected to the availability of a sensitive and reliable MS instrument. Hence, detecting misincorporation events through crucial active sites, particularly in a fragile system such as the glucuronylsynthase, can be a useful alternative for translational fidelity studies when an analytical method is unavailable. Nonetheless, this method is subject to limitations as it requires that there is an available detection method sensitive enough to detect the activity of the mistranslated product.

**Implications of translational misincorporation on the glucuronylsynthase system and biocatalyst engineering**

The presence of hydrolytic activity in the glucuronylsynthase system can undermine its performance because of the nature of this system. Nevertheless, if the hydrolysis occurs at a rate that is much slower than the synthetic rate, this is not a problem.

We have determined that two factors contribute to hydrolytic activity in the glucuronylsynthase system. The first is the basal hydrolytic activity caused by the base catalyst E413. In the scheme of things, this is of little consequence to the glucuronylsynthase system. In comparison, changes in codon usage and expression conditions together can reduce the rates of misincorporation by factors up to $10^3$. This recovery rate is actually higher than what we achieved in the chemical rescue and illustrates the importance of optimal expression systems.
A stronger expression system or overexpression condition is often deemed helpful for enzyme and protein production, but as we have shown in this exercise, it may not always be the case. High expression levels increase the amount of the products, which gives the appearance of a more productive manufacturing process. However, this increases the metabolic burden on the expression system and promotes translational inaccuracy. In the case of glucuronylsynthase, the cost of using strong expression systems or conditions to increase the yields of the enzyme can be significant. Hence, the cost-benefit of using high-yield expression systems or conditions should be carefully weighed. Whilst we have not conducted a study using different expression systems, our study suggests that it might be worthwhile to consider codon optimization and bioprocessing optimization, which are often regarded as trivial, to reduce translational misincorporation while engineering biocatalysts.

5.4. Conclusion

For biocatalyst production, translational misincorporation can cause reduced activity or substrate specificity and result in undesired side activity. However, to our knowledge, this has not been known to inhibit the reaction in which the biocatalyst is used such as what we saw with the synthetic activity of the glucuronylsynthase encoded with rare codon. Our case was unique, and it brought to light the significance of ensuring high translational accuracy in the production of biocatalysts. Although we are unlikely to abolish β-GUS contamination, we can reduce the incidence of misincorporation by tailoring the expression systems and conditions. More importantly, our observation in this chapter stresses upon the need to evolve a faster glucuronylsynthase enzyme.
5.5. References


6. Mutational tolerance of the *Escherichia coli* β-GUS

6.1. Introduction

Mutations, generally classified as deleterious, neutral and beneficial, create diversity in the genetic pool. They are the driving force behind evolution because they can alter protein structure\(^1,2\), properties\(^3,4\) and function\(^5\). These changes result in variation to their traits or phenotypes, which is also known as phenotype variation. Examples of phenotype variation include the inactivation of protein function, the improvement of a promiscuous activity (adaptation), or the acquisition of a different biophysical property\(^6\).

Mutational tolerance describes the ability of a protein in resisting changes to its traits when subjected to random mutations. At the molecular level, mutational tolerance involves looking at the properties of a molecule as a function of mutation rate. One can generate large libraries of mutated genes and then examine the activity, or some other property such as the stability of the variants. If the gene encodes a protein that is essential for the survival of the organism, then one can look at the survival rates as a function of mutation rate to gain some idea of mutational tolerance\(^7\)–\(^11\). These experiments generate a distribution of fitness data that is then used to compute mutational tolerance. Mutational tolerance can be computed by plotting the distribution of fitness against mutation rate\(^9\) or by calculating a library’s inactivation probability based on the fitness data\(^7,8,10,11,12\). An enzyme that has higher mutational tolerance is able to withstand higher mutation rate.

There appears to be many ways to quantify mutational tolerance. For example, studies by Guo *et al.* in Loeb’s group includes the development of a substitutability index, the \(x\)-factor, to quantify the mutational tolerance of the AAG\(^10\), and the DNA Pol I\(^11\). In other their studies, the parameters \(\alpha\) and \(\beta\) that were developed by Charlesworth\(^13\) were used to quantify mutational tolerance and negative epistasis\(^9,14\), and the \(\alpha/\beta\) ratio was used to describe the degree of epistasis\(^9\).
It had been shown that mutational tolerance can facilitate adaptation\textsuperscript{15,16}. Hence, better understanding of the factors that impact mutational tolerance may lead to better understanding of the driving forces of evolution or adaptation that may benefit laboratory evolution experiments. There are various ways to study mutational tolerance. Structure-function analyses through experiments that investigate mutation effects can provide useful insights into what drives mutational tolerance. In studies involving the DNA repair enzyme 3’-methyladenine DNA glycosylase (AAG)\textsuperscript{10} and DNA Polymerase I (DNA Pol I)\textsuperscript{11} performed in Loeb’s group, the group has demonstrated that there is a relationship between the mutational tolerance of the residues in the enzymes and their evolutionary conservation score, structural motifs and solvent accessibility.

Other factors that affect mutational tolerance are the population size\textsuperscript{8,17} and thermostability\textsuperscript{7,9,18}. In one experiment, Bloom \textit{et al.} subjected a P450 variant to three parallel evolution experiments such that each experiment would have a different rate of accumulation in diversity\textsuperscript{8}. They found that high diversity yielded a population that has higher mutational tolerance and better thermostability due to a surplus of stabilizing mutations that are present in diverse population. It has also been shown in various experiments that evolution is caused by adaptive mutations that also tend to be destabilizing\textsuperscript{19,20,21}. As a result, proteins tend to accumulate mutations with compensatory roles with improved biophysical properties\textsuperscript{22} that can compensate for these destabilizing effects of adaptive mutations. As such, there is a strong association between stability and mutational tolerance, which also promotes the ability of a protein to evolve for new function (evolvability)\textsuperscript{23,24,25}.

Analyses such as those described above leads to better understanding of mutational tolerance that can facilitate the development of small and smart directed evolution libraries through structure-based library design. However, when multiple mutations are involved, each mutation may change the network of interactions in a macromolecule. Consequently, the net effects of mutations may not always be equal to the sum of their individual influence and this complicates structure-function analysis or library design. This effect, also called epistasis, is caused by non-additive interactions between mutations and can impact the mutational tolerance of an enzyme. When an epistatic mutation arises, the mutational tolerance of an enzyme, which is expected to
fall off in a linear manner, may decrease in an exponential manner as the effects of the mutation can overshadow the effects of other mutations. One study that was performed through computational modelling predicted that epistasis can enhance mutational tolerance. This was later verified in an experimental study by Bershtein et al.\textsuperscript{9} that concluded that lower mutational tolerance was associated with weaker epistasis and that the apparent robustness of an enzyme increases with the presence of thermostable mutations. The study not only provides experimental evidence on the impact of epistasis on mutational tolerance, but it also provides additional insights into how enzyme stability can affect mutational tolerance.

In this chapter, we will first study the mutational tolerance of the \textit{E. coli} β-GUS. This enzyme is a multimer that consists of multiple domains. These features are expected to result in increased structural complexity, which is likely to cause the enzyme to exhibit a different mutational tolerance from structurally simpler forms of enzyme. One simple monomeric enzyme that have had its mutational tolerance study comprehensively is the TEM-1 by Bershtein et al.\textsuperscript{9} Their study, aimed at providing insights into the effects of epistasis caused by the accumulation of deleterious mutations on the protein’s ability to tolerate mutations, had computed various $\alpha$-values of the TEM-1 for quantifying the mutational tolerance of TEM-1 under different selection pressures. The $\alpha$ parameter of the TEM-1 can therefore serve as a primary reference to facilitate direct comparisons of the β-GUS mutational tolerance against an enzyme that is structurally less complicated. Secondly, we will investigate the mutational tolerance of eight individual active site residues. Some of these residues are implicated in substrate binding and interacts with the glycosyl C6 position. The libraries will be screened against two different substrates that differ at the C6 position- the native substrate, β-glucuronide (pNPGlcA) and a promiscuous substrate, β-glucoside (pNPGlc). We will associate the mutational tolerance of the amino acids with their evolvability. From these experiments, we aim to understand the factors that govern the mutational tolerance of β-GUS. Insights gained from the experiments can be useful for understanding its evolvability and for guiding future designs of laboratory evolution, particularly for the β-GUS enzyme.
6.2. Results

6.2.1 Random mutagenesis on whole gene and selected regions in enzyme

We subjected the whole length of *gusA* gene to random mutagenesis with six MnCl$_2$ concentrations. This created six sub-libraries of *gusA* random libraries, each with different mutation rate ($\mu$). Appendix 6-1 describes the method used to determine the accuracy of the $\mu$ determination. Mutation rates were determined by randomly sequencing eight colonies from each sub-library (Appendix 6-1, Table A6-1 and Figure A6-1). The sub-libraries were screened using the X-gluc blue-white screening method described in Section 2.7.2 and the pNPGlcA solution based assay described in Section 2.7.3. The percentage of active mutants in a sub-library was taken as the measure of the sub-library’s fitness, $W_n$.

$W_n$ was determined by taking the percentage of a library screened against pNPGlcA that retained 99% of the activity (activity cut-off 1%) or by enumerating the fraction of colonies that turned blue after overnight incubation on a colony agar plate consisting 0.8 mg/mL X-gluc. We plotted this against the mutation rate and found that the sub-library’s fitness declined exponentially with the mutation rate. This fitness decline was taken as a measure of the mutational tolerance of β-GUS. Since our experiment is analogous to a random mutational drift experiment in the absence of purifying selection, the exponential function that describes our fitness decline can be written in the form shown in Equation 6-1:

$$W_n = K \exp (-\alpha \mu)$$

(6-1)

where $W_n$ is the fitness of the sub-libraries, $\mu$ is the mutation rates or the mutational burden and $\alpha$ is the rate of fitness decline.

Higher values of $\alpha$ indicate that a library is less tolerant of high mutation rates and imply that it is less tolerant. $K$ is a constant that should be 100 (100% activity when $\mu = 0$) in an ideal case. Since the value of $K$ represents an ideal case, the experimentally determined value of $K$ also gives an indication of how accurate the assessment is. The rate of inactivation ($\alpha$) for β-GUS is 0.46 with a $K$ value of 104 (Figure 6-1).
We compared the $\alpha$-value of $\beta$-GUS to the one that was obtained for TEM-1$^9$ under random mutational drift as this selection pressure is the one that closely reflects our experimental design. We also compared their solvent accessibility surface areas (Table 6-1). The average SASA value was calculated from the Solvent Accessibility Surface Areas (SASA) function in PyMol and then averaged by the number of residues. The $\alpha$-value that was used for comparison from the TEM-1 study was the one obtained from the library that was allowed to drift under highest fitness level (Lib0)$^9$. The $\beta$-GUS appears to be slightly less tolerant than the TEM-1 with $\alpha$-value 0.46. In addition, the SASA of $\beta$-GUS is approximately 10 Å$^2$ less than the TEM-1. This is unsurprising since we expect the functional quaternary structure of the macromolecule assembly to be roughly globular, and for globular shapes, the surface:volume ratio decreases with increasing size. The decrease in average SASA calculated for the $\beta$-GUS as its oligomeric number increases also suggests that there are more buried residues in the $\beta$-GUS.

6.2.2 Tolerance of active site residues towards mutation

We extended our study to focus on the mutation tolerance at several active sites. The active site residues chosen are listed in Table 6-2. The table also includes the key feature associated with each residue. The residues that were chosen can be classified into four distinct category: (i) glycosyl binding residues (ii) non-binding residues (iii) leaving group and (iv) catalytic loop flip residues. The crystal structure of glucar-δ-lactam (GDL) bound $\beta$-GUS obtained by Wallace et al.$^{27}$ (PDB ID: 3K4D) was used to
Table 6-1: Solvent accessibility areas. The total solvent accessibility area (SASA) and average SASA per residue of a reference enzyme (TEM-1) and the different structural compositions in β-GUS.\(^{(a)}\) Only the monomer was considered for TEM-1 because it is a monomeric protein\(^{(a)}\). \(^{(b)}\) the two subunits in the crystal structure of β-GUS had slightly different solvent accessibility areas \(^{(c)}\) the references for α-value and crystal structure of TEM-1.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Structural assembly</th>
<th>No. of AA</th>
<th>Total SASA</th>
<th>Average SASA</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEM-1 (reference)(^{(c)})</td>
<td>Monomer (^{(a)})</td>
<td>263</td>
<td>11984.71</td>
<td>45.57</td>
<td>0.37</td>
</tr>
<tr>
<td>β-GUS</td>
<td>Monomer A (^{(b)})</td>
<td>603</td>
<td>24246.36</td>
<td>40.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monomer B (^{(b)})</td>
<td>603</td>
<td>24647.54</td>
<td>40.87</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Dimer</td>
<td>1206</td>
<td>45738.56</td>
<td>37.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tetramer</td>
<td>2412</td>
<td>85096.22</td>
<td>35.28</td>
<td></td>
</tr>
</tbody>
</table>

shortlist glycosyl and non-glycosyl binding residues (Figure 6-2a). Substrate \(p\)NPGlcA was docked into the active site by pair-fitting \(p\)NPGlcA with the inhibitor GDL. The pair-fit RMSD value was 0.19 (Figure 6-2b).

Of the shortlisted glycosyl binding residues, we identified D163, Y472 and R562 as targets for single-site saturation mutagenesis. These are within 4 Å of the carboxylate group at the C6 position in \(p\)NPGlcA and would interact with the glycosyl group of \(p\)NPGlcA at this position. We chose these residues so that we could compare the mutational tolerance when screened using \(p\)NPGlcA against the mutational tolerance when screened with \(p\)NPGlc. The \(p\)NPGlc is a promiscuous substrate of β-GUS and bears a different C6 functionality \(i.e.\) a –CH\(_2\)OH group at C6 in \(p\)NPGlc (a β-glucoside) in contrast to a –COOH in \(p\)NPGlcA (a β-glucuronide) (Figure 6-2c). We would use the comparison between the two screening data sets to investigate the relationship between mutational tolerance and evolvability.

Four other residues N412, M447, Y469 and T509 that are not implicated in glycosyl binding were also investigated as these were in locations that may modulate the catalytic residues (E413 and E504) and the glycosyl binding residues. Amino acid N412 is a fully conserved residue that does not interact directly with the substrate, but interacts with the E413 catalytic residue \(via\) a bridging water molecule (Figure 6-3a). It is interesting that this residue is fully conserved and we were interested to know if
Table 6-2: The sites chosen for randomization and their roles. The sites are classified according to the classification column. The sites are listed according to the order in which they were investigated.

<table>
<thead>
<tr>
<th>Sites</th>
<th>Comments</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>D163</td>
<td>Glycosyl C6 interaction</td>
<td>Glycosyl binding residue</td>
</tr>
<tr>
<td>R562</td>
<td>Glycosyl C6 interaction</td>
<td>Glycosyl binding residue</td>
</tr>
<tr>
<td>N412</td>
<td>glutamate E413 that participates in general acid/base hydrolysis</td>
<td>Non-binding residue</td>
</tr>
<tr>
<td>M447</td>
<td>Leaving group interaction</td>
<td>Aglycone binding residue</td>
</tr>
<tr>
<td>T509</td>
<td>Important site based on literature not involved in binding</td>
<td>Non-binding residue</td>
</tr>
<tr>
<td>Y468</td>
<td>Flipped residue resulting from loop conformation change</td>
<td>Catalytic loop flip residues</td>
</tr>
<tr>
<td>Y469</td>
<td>Flipped residue resulting from loop conformation change</td>
<td>Catalytic loop flip residues</td>
</tr>
<tr>
<td>Y472</td>
<td>Glycosyl C6 interaction upon loop conformation change</td>
<td>Catalytic loop flip residues/Glycosyl binding residue</td>
</tr>
</tbody>
</table>

Figure 6-2: Interaction between substrate-binding residues and pNPGlcA. (a) Key interactions between E. coli β-GUS and docked substrate pNPGlcA. pNPGlcA was pair-fitted to the inhibitor glucaro-δ-lactam (GDL) in the inhibitor bound β-GUS crystal structure obtained by Wallace and co-workers. Interaction distances were derived from the superimposed pNPG substrate and the residues. (b) The superimposition of pNPGlcA (white sticks) over GDL (represented as blue lines) in the crystal structure. (c) Glycosyl rings of β-glucoside (top) and β-glucuronide (bottom).

randomizing this amino acid would also yield interesting changes in enzyme activity. M447 was also interesting because it does not interact directly with the glycosyl ring, but interacts with the p-nitrophenolate (pNP) leaving group. This residue is also found to occlude the cleft of the active-site (Figure 6-3b) and may control product release. M447 also interacts with Y469 in a sulfur-π-ring interaction (Appendix 6-2, Figure A6-2a). The Y469 residue is involved in a loop conformation change involving itself, Y468 and Y472 (Appendix 6-2, Fig A6-2) that happens upon substrate binding. This
conformational change appears to be important for substrate interaction as Y472 flips into the active site to interact with the glycosyl ring. Hence, these residues were also investigated.

Figure 6-3: Interactions of several active site residues in β-GUS. (a) Interaction between E413 and N412 in the active site of β-GUS with a GDL inhibitor is mediated by a bridging water molecule (white dot). (b) The loop that bears M447, Y468 and Y469 (represented as surface in cyan) can occlude the exit of hydrolysis products. Docked active sites is pNPGlcA (yellow) using the pair-fitting function that is available in PyMol.

**6.2.3. Developing a statistical approach to study the mutational tolerance of single-site saturation mutagenesis libraries**

The mutational tolerance of an amino acid residue can be denoted by the fraction of the library that appears active. However, for the solution-based assays such as the pNPGlcA assay, an inactivity threshold has to be assigned to classify inactive variants. For example, in a library whose threshold value is designated 10%, any variant exhibiting less than 10% activity will be considered inactive. The fraction of the library that remains active above this value is taken as a statistical representation of the fitness. However, this can be ambiguous because the threshold value is a subjective value. It is also less suitable for single site-saturation mutagenesis libraries because it does not account for the bias that is imparted by the presence of wild-type. The mutation frequency of site-saturation libraries is the percentage of variants in the library that has been successfully mutated. This depends on the efficiency of the randomization at a particular site. The typical mutation frequency achieved by the NNK codon randomization method lies between 60 and 95% (values taken from literature\textsuperscript{28,29} and our experience). A library with lower efficiency randomization rate has more wild-type encoded clones that will lead to overestimation of its mutational tolerance.
We designed a method based on the activity distribution of an individual library to derive a measure for the mutational tolerance of our site-saturation libraries. The mutants in a site-saturation library will display a range of activities. When ranked in order, the activity distribution fits a sigmoidal function. The gradient that describes the rate at which activity is lost in the library is then taken as the statistical measure of the mutational tolerance at each site. This is represented by \( m_3 \), the inflection point of the sigmoid curve that fits Equation 6-3. Figure 6-4 illustrates the method using three hypothetical libraries.

\[
y = y_{\min} + \frac{y_{\max} - y_{\min}}{1 + \left(\frac{x}{x_{1/2}}\right)^S}
\]

where \( y_{\max} \) (m1) and \( y_{\min} \) (m2) are the asymptotes of the sigmoidal curve, \( S_{1/2} = \) slope at \( x_{1/2} \) and \( x_{1/2} = \) inflection point of the sigmoidal curve. A more negative \( S_{1/2} \) corresponds to higher inactivation rate and hence, reflects lower mutational tolerance. This number is independent of the screening size (Appendix 6-3, Figure A6-3) or the mutation frequency, and is not affected by the ambiguity in assigning a threshold value.

6.2.4. Quantifying the mutational tolerance of the single-site saturation mutagenesis libraries

Figure 6-5 summarizes the mutational tolerance of the eight single-site saturation mutagenesis libraries. The mutational tolerance of each site towards the substrates \( pNPGlcA \) and \( pNPGLc \) was compared. For each site, two independent screening experiments for both substrates were performed. We found that the non-glycosyl binding residue M447 was the most tolerant followed by T509 and N412. The M447 library also appears to have similar mutational tolerance with both substrates, suggesting that it is not evolvable. T509 appeared to be most evolvable with several screening data points that was above 1.0 (improved activity). It also appeared to have a lowest inactivation factor (most tolerant) with the non-native substrate, again indicating better evolvability.

We randomly picked clones in the screened libraries to sequence. When the activity distribution curves (Appendix 6-4, Figure A6-4) and sequencing data (Appendix 6-4, Table A6-2) were analysed, we found that there were various mutants in
CHAPTER 6

Figure 6-4: Simulation of activity distribution profile following data sorting of activity levels relative to wild-type in a library (hypothetical). Figures (a)–(c) each represent an intolerant library, moderately tolerant library and tolerant library with clones that are more active than wild-type. In each of the curves, the sigmoidal curve-fit parameters are listed and their residual errors, where $m_1$ = maximum asymptote, $m_2$ = lower asymptote, $m_3$ = inflection point and $m_4$ = the gradient at the inflection point. Curve-fitting was performed using Kaleidagraph 4.5. Table in the corner (top right) summarizes these parameters and compares the gradient $m_4$ to the threshold method using three discrete bins i.e. the 0.1x WT activity, 0.1–1x WT activity and > 1x WT activity.

the M447 library that retained partial activity, while there were three mutants in library T509 (T509A/G/S) and one mutant in library N412 (N412D) that appeared partially active in the $\beta$NPGlcA screen. This indicates that N412 is not tolerant of mutations and imply that its role in co-ordinating the interaction between E413 and the substrate through the bridging water is an important one. We purified M447S/T and T509A, which had appeared frequently in our sequencing data. M447T was also chosen for further confirmation because it appeared to have lost its partial activity with the $\beta$NPGlc substrate.

The library R562 stimulated some interest. Although this library appears to be totally inactivated in the $\beta$NPGlcA assay, it was not so in the $\beta$NPGlc assay. This
indicates that some of the R562 variants are trading their β-GUS function to acquire some glucosidase functionality. We sequenced several R562 clones that appeared to have glucosidase activity but not β-GUS activity, and found mutations R562H/P/S/T. Of these, the R562S and R562T mutants were isolated twice. They were purified to further confirm their activity.

Figure 6-5: Mutational tolerance of sites chosen for site-saturation mutagenesis study. E413, which should be completely inactivated is a control to represent full library inactivation. Thus, any values more negative than E413 signifies inactivation. Error bar represents standard deviation of the inactivation factor from two independent screening exercises.

6.2.5. Verification of activity in mutants

We purified T509A, R562S, R562T, M447G and M447S by metal affinity chromatography as described in Chapter 2. Kinetic parameters were obtained and are given in Figure 6-6. However, we were not able to achieve substrate saturation for the glucoside substrate with both M447G/S mutants (Appendix 6-5, Figure A6-7). We did not attempt to repeat the kinetics using higher substrate concentrations due to solubility concerns. This reflects an increase in $K_m$ of the M447 mutants for the glucoside substrate, which is interesting as M447 is not known as a substrate binding residue. Its position at the entrance of the active-site (Figure 6-3) may contribute to this effect. It
suggests that the dynamics at the active-site entrance controls substrate uptake and product release and consequently affect substrate binding.

The enzyme activities of R562 and M447 variants measured at a particular substrate concentration were compared with the WT. The substrate concentration was chosen such that it would be similar to the one used in the screening experiment and was 1.6 mM for the β-GUS assay (1 mM in screening), and 12.8 mM for glucosidase assay (9 mM in screening). The variant R562S and R562T were not completely inactive in pNPGlcA even though they only retained < 1% of β-GUS activity. M447G retained partial β-GUS activity and partial glucosidase activity but M447S retained partial β-GUS activity and was almost inactive as a glucosidase (< 1%). In general, the trends observed using purified enzymes agreed well with the observations from screening data.

We also obtained the kinetic parameters for T509A (Table 6-3) and determined its thermostability using the thermal shift assay that was described in Section 2.12.1 (Figure 6-7). T509A is reported in various literatures as a mutation that can improve the xylosidase and β-galactosidase activity of the β-GUS. We found that the β-glucosidase activity (pNPGlc activity) of T509A is three-times more efficient.
compared to the parent. This, along with the other reports, would suggest that this mutation is important for accepting non-native glycosyl rings.

Although the T509A demonstrated an enhancement in the promiscuous activity with pNPGlc, we did not observe any dramatic loss in its native activity (activity against pNPGlcl). This is reasonable since it is not directly involved in catalysis, but it is located on the same loop, a few residues away from the catalytic residue. Consistent with other reports on other similar type of mutations in other enzymes, the enhancement of the promiscuous activity by T509A mutation is likely due to its ability to enable the accommodation of different substrates through conformational changes. This mutation has also been reported in a thermostability study, but it had appeared together with other mutations. Its effect on stability as a single mutant was not reported. We found that, its effect as a single mutant was not important, and is in fact, slightly disadvantageous for improving thermostability. This was likely offset by the other mutations that had appeared concurrently with T509A in the thermostable variant found in the report. Therefore, the T509A mutation appears to be important for function, but not for stability. This exemplifies the activity-stability trade-off trend that has already been extensively studied.

6.2.6. Limitations of the statistical method used for studying mutational tolerance of single site-saturation mutagenesis library

Overall, the statistical method that we have developed to describe the mutational tolerance of our site-saturation libraries tallied well with the activity distribution of the
libraries. However, there may be limitations to the method. Of great concern, would be the utility of the method when applied to larger screening data and screening data that involves mutants with huge activity improvement. Both in our method development and in our application, the largest activity improvement we saw was a modest 2.5-fold in pNPGlc screen of the T509 library. The method used does not properly account for the positive effects of the improved mutants on the mutational tolerance of the site. Further, although the data in both cases could still be fitted to the sigmoidal function, the fit becomes less ideal. The R-value in the hypothetical library (Figure 6-4c) decreased to 0.985 compared with the R-values of > 0.999 in the other two hypothetical libraries. Additional parameters to incorporate the effects from neutral and significantly improved mutations would increase the utility of the model.

It is also important to note that the method is extremely sensitive to the effects of screening variation. The replicates of the pNPGlcA screen of T509X library for example, showed a high standard deviation. We replotted and recalculated the inactivation factor of T509 using only the data of the sequenced mutants (Appendix 6-6, Figure A6-6) and found that the inactivation factor for one of the plate was underestimated. This is caused by large screening errors (% CV in the plate = 30%), which resulted in a huge distribution of the wild-type activity data. A better way to implement the method would be to apply the method on data sets whose sequence is already verified (Appendix 6-6, Figure A6-6).

6.3. Discussion and concluding remarks

We have used the $\alpha$-value of TEM-1 as a reference for evaluating the mutational tolerance of $\beta$-GUS. Based on their $\alpha$-values, it seemed that the $\beta$-GUS is slightly less tolerant than the TEM-1. However the difference is smaller than expected and may not be significantly different. In fact, when we conducted an unpaired t-test analysis, it appears that the difference between the two values is not statistically significant at confidence interval level above 99% even though they are different below 95% confidence level (Appendix 6-7, Table A6-3 and Figure A6-7). Nevertheless, it must be noted that the $\alpha$-value is a subjective quantitative measure as it depends on the criteria that is used to plot the y-axis ($W_n$) in Figure 6-1. Criteria that are less stringent would
result in a higher apparent mutational tolerance. Our criteria is different from the one used in the TEM-1 study. In our study, two assays were used – the \( p\text{NPGlcA} \) assay and the blue-white screening assay with saturating amounts of X-gluc. The X-gluc assay represents the lowest possible stringency level. For the former, the level of stringency in our study depended on the activity cut-off that was set in the \( p\text{NPGlcA} \) assay (1%). This cut-off seemed comparable to the X-gluc assay. Hence, both our criteria represent the least stringent criteria. Comparing it to the unselected library at the highest fitness level (highest antibiotic concentration, \( \alpha = 0.37 \)) may not be the most suitable comparison. A comparison against the unselected library screened with lower antibiotic concentration (lower stringency) may be a better reference for us. The \( \alpha \)-value of the TEM-1 approaches 0.1 at lower ampicillin concentration, and is substantially higher than the \( \alpha \)-value of \( \beta \)-GUS.

It is counter-intuitive that the \( \beta \)-GUS appears less robust or to be only as robust as the TEM-1. One expects the larger \( \beta \)-GUS to have more network of interactions that can buffer the effects of undesirable mutations. One possible explanation for this is that the \( \beta \)-GUS consists of several domains and exists as a tetramer, resulting in regions that are more sensitive to mutations. The other is that tetramerization of \( \beta \)-GUS gave rise to a higher proportion of buried residues, which are generally known to be less tolerant\(^{10}\), in its quaternary structure. Mutations at the interfacial regions may give rise to deleterious mutations that disrupt the interaction required to maintain its quaternary structure. Therefore, fewer mutations are needed to destabilize the enzyme and henceforth inactivating it.

To test the effects of mutational interactions on the mutational tolerance of \( \beta \)-GUS, we fitted the data in Figure 6-1 to a fitness decline model with an epistatic parameter (\( \beta \)) (Equation 6-2).

\[
W_n = K \exp (-\alpha \mu - \beta \mu^2) \tag{6-2}
\]

where \( W_n \) is the fitness of the sub-libraries, \( \mu \) is the mutation rates or the mutational burden and \( \alpha \) is the rate of fitness decline, and \( \beta \) is the epistasis parameter. The \( \beta/\alpha \) value indicates the strength of directional epistasis with a lower value corresponding to weaker epistasis. When directional epistasis is taken into account, the \( \beta \)-GUS has a \( \beta/\alpha \) value of 0.076, and an \( \alpha \)-value of 0.33 (Appendix 6-7, Figure A6-8). This \( \alpha \)-value is lower than the value of 0.46 that was obtained from Equation 6-1 and is much closer to
the reported TEM-1 $\alpha$-value. This indicates again that the mutational tolerance of β-GUS may not be significantly different from the mutational tolerance of TEM-1.

We projected the $\alpha$ and $\alpha/\beta$ values of the β-GUS obtained from Equation 6-2 onto the plot that correlates epistasis and mutational tolerance (plot of $\beta/\alpha$ versus $\alpha$) in Figure 2b of reference (9). The β-GUS appears to be under stronger epistatic influence compared to the TEM-1. At $\alpha \sim 0.36$, the $\beta/\alpha$ value of the TEM-1 approaches zero ($\sim 0.01$), whereas for β-GUS, its $\beta/\alpha$ value of 0.076 is much greater than zero. This supports our inference that interactions within the macromolecule that contributes to a proper fold and structure of the β-GUS is a major determining factor of the mutational tolerance of β-GUS. This is further supported by an examination of the tetrameric β-GUS structure, generated from interposing the dimeric crystal structure, that indeed revealed several distinct examples of subunit and inter-domain interactions (Appendix 6-8, Figure A6-9), including a solvent exposed disordered region that is involved in domain swapping (PDB ID: 3LPF, 3LPG).27

One of the residues at the interface regions that can totally inactivate an enzyme upon mutation is the R562 residue. This residue appears on a mobile loop at the interface of two subunits. The intolerance of this residue may be associated with destabilization that results from the disruption of the subunit-subunit interaction when this residue is mutated. However, this residue also seems to have a better tolerance in the presence of $p$NPGlc (non-native substrate), suggesting that it is capable of adaptability. This also suggests that the mutation at this position does not completely disrupt the interaction between the two subunits and a weaker interaction between the subunits may have given rise to conformation changes that allow the macromolecule to accommodate other substrates. The fact that R562 may be capable of adaptability is not surprising since R562 is also a glycosyl binding site and hence, mutation on this residue may optimize it towards better acceptance of a non-native glycosyl ring. In addition, this residue is on a loop and loops in general have been known to be excellent targets for modulating enzyme activity20,37,38.

We had also studied the mutational tolerance of several other substrate binding residues that are located on loops using site-saturation mutagenesis and compared their evolvability. However, none had showed the same ability to adapt as R562. Conversely,
the residue T509, which does not participate in substrate binding, seemed the most evolvable. This residue was also one of the more tolerant residues, but it showed reduced thermostability. This does not agree with the generalization that increased mutational tolerance leads to increased stability, which would then increase evolvability. Further, M447, that had the highest mutational tolerance, also did not appear to be the most evolvable. One of the mutations that had retained partial activity with the native substrate was inactive when assayed with the promiscuous substrate. These suggest that the general observation - that mutational tolerance promotes evolvability - do not constitute as “hard and fast rules”. Interestingly, the tolerant residues, M447 and T509, and the residues that seem more pliable (R562 and T509) are located on surface areas. This is agreeable with observations thus far that solvent accessibility promotes mutational tolerance. It appears that both mutational tolerance and solvent accessibility can affect evolvability but, the main driving force of the adaptation process of β-GUS depends strongly on mutations that will alter the intermolecular interaction within the enzyme and the dynamics of the loop motion so that it can adopt a structural conformation that can better accommodate non-native substrates\[33. In other words, these would be the major factors that drive the promiscuity of β-GUS.

Taken together, solvent accessibility does increase the mutational tolerance of an amino acid residue, but tolerable residues are not always more evolvable. In considering library design, protein dynamics such as loop motions that are involved in conformational changes may be the more important factors for consideration. This may be an even more important factor in larger macromolecules since the inter-molecular forces (subunit and inter-domain interactions) within an enzyme will inevitably influence its dynamics\[39. However, with increasing complexity in the structure, it will become more difficult to predict the influences of the participating interaction within the molecular assemblies of an enzyme. The protein dynamics of large molecules such as the β-GUS is more complex due to more possibilities of mutations at the interfacial regions that can be sensitive to mutations. This suggests that in directed evolution studies of large enzymes, it may be more fruitful to mutate at low mutation rates as there will be a higher risk of introducing a deleterious mutation that can totally negate the effects of other positive and neutral mutations that may appear together with it.
6.4 References


7. Conclusion and future directions

This thesis was initiated with the purpose to explore different approaches to obtain a glucuronylsynthase for the synthesis of O-glucuronides with enhanced catalytic rate. Even though the applicability of the glucuronylsynthase was successfully demonstrated by the McLeod group, the $k_{cat}$ of the glucuronylsynthase remains low, which left it considerable room for improvement.

The original intent of this thesis was to improve the $k_{cat}$ of the glucuronylsynthase a directed evolution methods and a two-stage strategy outlined in Scheme I of Figure 1-4 (Chapter 1). The two-step approach was designed to facilitate screening processes because the screening method for the evolution of glucuronylsynthase would have been too involved. This was an unusual approach because most engineering programs will choose to engineer the variant that already has the synthetic capability\textsuperscript{1,2}. If this had worked, it could have represented a novel way of screening. However, working with the wild-type that already has a high activity meant that any improvement was also subjugated by the noise in the screens. Subsequently, we added $t$-BuOH because this was found to be advantageous for the chemoenzymatic reactions of the glucuronylsynthase. This partially inactivated the wild-type and essentially meant that we also introduced an additional selection pressure in the evolution. Consequently, it was easier to identify improved variants. Out of this second exercise, we isolated mutation L561S that had eight-times higher hydrolytic activity compared to wild-type in the presence of $t$-BuOH. However, conversion of the L561S variant to glucuronylsynthase (SynL561S) did not lead to a functional glucuronylsynthase. This had then prompted us to focus the evolution on the linear one-step glucuronylsynthase engineering approach.

We successfully applied site-saturation mutagenesis to evolve a glucuronylsynthase variant that had higher synthetic activity. The best variant contained single mutation H162Q that had conferred more than four-fold improvement. Considering the limitations that were faced in terms of screening capability, this was a decent enhancement. Part of this success was due to the implementation of the strategy
that had considered stabilization of the active site conformation as the key factor. Our success with this strategy suggests that future semi-rational glucuronylsynthase engineering endeavours may target sites that are involved directly or indirectly in the active site configuration.

The fold-improvement we obtained compares well with the median fold improvement of 5.4, as assessed by Nannemann et al. and with other directed evolution experiments of glycosynthases that were performed with libraries larger than ours (fold improvement 5, 7 and 27). This paints an encouraging prospect for future attempts to further improve the glucuronylsynthase enzyme using directed evolution approach. It suggests that the more straightforward direct path approach is more effective than the two-step approach i.e. sacrificing screening capacity would have been more worthwhile than exercising the first approach.

Our success at isolating the H162Q mutation reopens another possible screening route for the future of glucuronylsynthase enzyme engineering. This may also be a suitable starting point for a second attempt to optimize the fluoride-selective assay discussed at the beginning of Chapter 4 as we may be able to overcome the limits of detection. We could not use this assay because of low sensitivity but we may be able to optimize the assay to work for continuing the evolution from an improved variant. This can replace the slower LCMS method as the main method for primary screening.

The unsuccessful attempt in converting the L561S to a functional glucuronylsynthase was against our expectation, which prompted us to ask several questions. Was our reasoning that by increasing the base catalysis we would also increase the synthetic activity flawed or was there a flaw in the glucuronylsynthase system itself that we had overlooked? Prior to this study, translational misincorporation had been reported for the glycosynthase system but was not known to be a major issue. We considered this a trivial issue and did not pay much attention to this detail. Our experience with L561S suggested otherwise. We set-up a systematic study to study the translational misincorporation of the glucuronylsynthase.

The experiments in Chapter 5 indicated that the intrinsic errors during protein translation increased the hydrolytic process significantly, indicating the presence of
wild-type β-GUS contaminants. In turn, this adversely affected the glucuronylsynthase system. The rate of the hydrolysis process could be reduced by at least $10^3$ times by manipulating codon bias usage and overexpression conditions. We have found that the misincorporation rates increased by a factor of 40 when we increased the expression temperature and induction time. Data from cell growth study suggests that at longer induction times, the cell growth enters a different phase where cells may start competing for food. Assimilating both data, it shows strong evidence that increased metabolic stress increases translational misincorporation. In addition, the rate of G504E mistranslation when we used a rare codon was $10^3$ times higher than when a common codon was used. These two factors could have contributed to the failure of converting L561S into a glucuronylsynthase. The E504G/L561S was produced using the rare codon gga and at high temperatures for prolonged expression duration. E504G/L561S may have to face increased competing hydrolytic activity caused by mistranslated products (G504E/L561S), which would explain why it did not work.

The results from translational misincorporation suggest that adopting the linear approach would have been the more prudent option. Our current standing shows that the linear approach would decrease the risk of introducing a contaminant with higher hydrolytic activity that would have resulted from misincorporation. Therefore, despite the difficulty in screening method, the linear approach is a more elegant approach and may offer a higher chance of selecting better variants.

Regardless of the approach that future endeavours would like to explore, our study on the mutational tolerance of the multi-domain, tetrameric β-GUS may provide insightful details that may be of use. Higher mutational tolerance can promote evolvability because this would mean that the enzyme is less prone to significant destabilization, which may cause complete inactivation of the enzyme. The ability of enzymes to tolerate mutations is therefore advantageous if we were trying to evolve for a new function.

From the mutational tolerance study, we deduced that the intramolecular interaction within a subunit and the interactions at the tetramerization interface are important in influencing the mutational tolerance of the β-GUS. Our idea that the inter-domain interaction has significant influence on the mutational tolerance and hence the
evolvability of the enzyme is supported by the H162Q mutation found in the glucuronylsynthase enzyme engineering program. The mutation that had conferred higher glucuronylation activity, H162Q is found on the loop from Domain I that participates in inter-domain interaction. Specifically, this loop extrudes into the active site space in the TIM-barrel domain. The mutation is located adjacent to D163 that had been established as an important substrate binding site. The H162Q mutation had increased the binding affinity of the loop in domain I, presumably through its neighbour, with the 1FGleA substrate that is accommodated in the active site cavity in the TIM-barrel. The presence of 1FGleA had offered the variant significantly more stabilization effect than for its parent glucuronylsynthase suggesting that there was an increased dependence on the enzyme-glucuronyl donor interaction. Hence, the loop where the H162Q is located is involved in the binding of the sugar donor substrate through inter-domain interaction. This interaction is important for stabilizing the enzyme active site conformation and subsequently promoted its ability to adapt glucuronylsynthase function.

Other observations that had been pointed out in the mutational tolerance studies may also facilitate future enzyme engineering endeavours of the β-GUS enzyme. When we performed site saturation mutagenesis on the sites that are implicated in the binding of the C6 functional group and tested the tolerance of the enzyme in the presence of the β-glucuronide (native) and β-glucoside (non-native) substrates, we found that generally the mutational tolerance exhibited in the presence of both substrates were the same. In contrast, when we subjected a site (T509) that was not directly involved in substrate binding, but that had the potential to modulate the β-GUS catalytic residue E504, mutations (T509A and T509S) that conferred higher promiscuous activity were isolated. Interestingly, the mutation that had conferred the glucuronylsynthase with higher activity is also not an important binding site residue but lies adjacent to the important substrate binding residue D163. These suggest that targeting residues adjacent to important binding sites may be fruitful endeavours. This notion is not new as it had been found that neighbouring residues often coevolve with conserved positions that have functional importance to compensate for destabilizing adaptive mutations. Hence, targeting sites that are in the vicinity of an important binding site might also result in significant changes in the activity of an enzyme as this would can modulate the flexibility of loops where binding sites are found.
Overall, this project had been an ambitious one and was complicated due to the size of the enzyme, the initial target set for the two-step approach and the unavailability of a quick glucuronylsynthase screening method. Interestingly, the initial approach that was meant to make it easy had an inverted effect- made the effort more cumbersome and complicated. Our experience suggests that it is more efficient to attempt to increase very low levels of the non-native synthetic activity. We had made very small progress compared to what we would have liked to in terms of fulfilling the original objective. However, we did isolate a glucuronylsynthase variant that was better, which can be used as the template to continue the glucuronylsynthase evolution and that can be used to develop quicker screening method for this future development. More importantly, the work undertaken to answer the main aim of this thesis had led to a better understanding of translational misincorporation that would apply not only to the glucuronylsynthase system but also to the production of biocatalysts in general. Furthermore, the penultimate chapter that had explored the mutational tolerance of the enzyme might initiate more of such studies. These studies can lead to further consensus that can further our understanding of mutational tolerance and evolvability of multi-domain oligomeric enzymes.
7.1. References


Appendix

Engineering the glycoside hydrolase β-glucuronidase (β-GUS) to improve a non-native activity for the synthesis of O-glucuronides
### Appendix A: Equipments and Consumable Suppliers

<table>
<thead>
<tr>
<th>Equipments and consumables</th>
<th>Manufacturers and suppliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>5415 Microcentrifuge</td>
<td>Eppendorf</td>
</tr>
<tr>
<td>5804 Centrifuge</td>
<td>Eppendorf</td>
</tr>
<tr>
<td>AKTA FPLC Sysyem</td>
<td>G.E. Healthcare</td>
</tr>
<tr>
<td>Amicon Ultra-15 Centrifugal Filter Unit</td>
<td>Millipore</td>
</tr>
<tr>
<td>ASB270BT Autoclave</td>
<td>Astell Scientific</td>
</tr>
<tr>
<td>Bioline II Isolate Plasmid kit</td>
<td>Bioline</td>
</tr>
<tr>
<td>C1000 Thermal Cycler</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Chirascan Circular Dichroism</td>
<td>Applied Photophysics</td>
</tr>
<tr>
<td>CO8000 Cell Density Meter</td>
<td>WPA Biowave</td>
</tr>
<tr>
<td>GeneAmp PCR System 9700</td>
<td>SLM Instruments</td>
</tr>
<tr>
<td>HiTrap Desalting Column</td>
<td>G.E. Healthcare</td>
</tr>
<tr>
<td>I-Cycler</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Liquid Chromatography/Mass Spectrometry</td>
<td>Agilent Technologies</td>
</tr>
<tr>
<td>MicroPulser Electroporator</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>MJ Mini Personal Thermal Cycler</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Multiskan Ascent</td>
<td>Thermo Fischer Scientific</td>
</tr>
<tr>
<td>NanoDrop® ND-1000 Spectrophotometer</td>
<td>NanoDrop</td>
</tr>
<tr>
<td>NanoITC</td>
<td>Texas Instruments</td>
</tr>
<tr>
<td>Orion ROSS Combination pH Electrode</td>
<td>Orion Pacific Pty Ltd</td>
</tr>
<tr>
<td>QIAprep Spin Miniprep kit</td>
<td>Qiagen</td>
</tr>
<tr>
<td>R20A2 and R9A Rotors for VX22G centrifuge</td>
<td>Hitachi</td>
</tr>
<tr>
<td>Rotorfix 32 Centrifuge</td>
<td>Hettich</td>
</tr>
<tr>
<td>SE250 Mighty Small II Mini Vertical Electrophoresis Unit</td>
<td>Hoefer</td>
</tr>
<tr>
<td>SpectraMax® M2/M2e Microplate Reader</td>
<td>Molecular Devides</td>
</tr>
<tr>
<td>Superdex 200 Size Exclusion Column</td>
<td>Amersham Biosciences</td>
</tr>
<tr>
<td>UV transilluminator 254/312 nm with UVItec camera</td>
<td>Hanimax Statesman</td>
</tr>
<tr>
<td>Veriti 96-Well Thermal Cycler</td>
<td>Applied Biosystems</td>
</tr>
<tr>
<td>VX22G High Speed Centrifuge</td>
<td>VWR</td>
</tr>
<tr>
<td>Wide Mini Sub Cell Electrophoresis Tank</td>
<td>Bio-Rad</td>
</tr>
<tr>
<td>Wizard® SV Gel and PCR Clean-Up System</td>
<td>Promega</td>
</tr>
</tbody>
</table>

---
Appendix B: Chemical suppliers

Chemicals
1 kb DNA Marker
2-propanol (IPA)
4-(2-aminoethyl) benzene sulfonyl fluoride hydrochloride
Acetone
Acetonitrile
Acryl/Bis 37.5:1 40% (w/v)
Agar
Agarose
Ammonium chloride
Ampicillin
BigDye® Terminator and Sanger Sequencing Buffer

BioTaq™ Polymerase and PCR buffer
Bromophenol Blue
BugBuster® Protein Extraction Reagent
Calf Intestinal Alkaline Phosphatase and Buffer
Dimethyl Sulfoxide
dNTPs
DpnI and Buffer
EcoRI and Buffer
Ethylendiaminetetraacetic acid
Ethanol
Ethidium Bromide
Ethyl acetate
Glycerol
Glycine
Hydrochloric acid
Imidazole
Isopropl β-D-1-thiogalactopyranoside
Kanamycin
Low Range SDS-PAGE Marker
Methanol
Magnesium Chloride
Magnesium Chloride, PCR buffer
Manganese Chloride
N,N,N’,N”-Tetramethylethlenediamine
NdeI and Buffer
Nutrient Broth
Orthophosphoric acid – 85%
*p-nitrophenol
*p-nitrophenyl galatoside
*p-nitrophenyl glucoside
*p-nitrophenyl glucuronide
Pfu Polymerase and Buffer

Manufacturers and suppliers
New England Biolabs
Merck
Sigma Corporation
Amresco
Sigma Corporation
Sigma Corporation
AnalaR
Amresco
Biomolecular Resource Facility, JCSMR
Bioline
Sigma Corporation
Novagen
New England Biolabs
Sigma Corporation
Roche
New England Biolabs
New England Biolabs
Ajax Chemicals
Merck
Sigma Corporation
Merk
Amresco
Ajax Chemicals
Sigma Corporation
Astral
AG Scientific Ltd
Bio-Rad
Merk
Sigma Corporation
Roche
Ajax Chemicals
Sigma Corporation
New England Biolabs
Difco
AnalaR®
Sigma Corporation
Sigma Corporation/Chem-Impex International Inc
Sigma Corporation/Chem-Impex International Inc
Acros Organics
Thermo Scientific/Agilent
Phenylmethylsulfonyl fluoride  
Phusion polymerase and Buffer  

Potassium phosphate, monobasic  
Potassium phosphate, dibasic  
RedSafe™ DNA Stain  
rLysozyme™  
Sodium Chloride  
Sodium Dodecyl Sulfate  
Sodium Hydroxide  
Sodium phosphate, monobasic  
Sodium phosphate, dibasic  
T4 DNA Ligase  

Taq Polymerase  
tert-butanol  
Tris  
Tryptone  
Yeast Extract  
β-mercaptoethanol  

Technologies  
Sigma Corporation  
Finnzyme/New England Biolabs  
Ajax Chemicals  
Chembio  
Novagen  
Chem Supply  
Amresco  
Ajax Chemicals  
Ajax Chemicals  
Ajax Chemicals  
Thermo Scientific/Fermentas/New England Biolabs  
Roche  
Sigma Corporation  
Amresco  
Difco  
Difco  
Sigma Corporation
Appendix C: Recipes of media and buffers

Recipes for growth media

M9 media
Na$_2$HPO$_4$  -----  6.4% (w/v)
K$_2$HPO$_4$  -----  1.5% (w/v)
NaCl  -----  0.5% (w/v)
NH$_4$Cl  -----  0.25% (w/v)
Top up to desired volume with MilliQ water (mQH$_2$O)
For agar plates preparation: add 1.5% (w/v) agar
Sterilized by autoclaving

LB media
Yeast extract  -----  0.5% (w/v)
Tryptone  -----  1% (w/v)
Sodium chloride  -----  1% (w/v)
Top up to desired volume with MilliQ water (mQH$_2$O)
For agar plates preparation: add 1.5% (w/v) agar
Sterilized by autoclaving

2XYT
Yeast extract  -----  1.0% (w/v)
Tryptone  -----  1.6% (w/v)
Sodium chloride  -----  0.5% (w/v)
Top up to desired volume with MilliQ water (mQH$_2$O)
Sterilized by autoclaving

YENB
Yeast extract  -----  0.75% (w/v)
Nutrient broth  -----  0.8% (w/v)
Top up to desired volume with MilliQ water (mQH$_2$O)
Sterilized by autoclaving
**Recipes for biochemical solutions and buffers**

**Sodium phosphate buffer (NaPi) components**
Disodium hydrogen phosphate (Na$_2$HPO$_4$), 200 mM ----- 28.39 g in 1 L MilliQ water
Sodium dihydrogen phosphate (NaH$_2$PO$_4$), 200 mM ----- 27.58 g in 1 L MilliQ water

**Sodium phosphate buffer (NaPi), 200 mM**
200 mM Na$_2$HPO$_4$ ----- ~750 mL
200 mM NaH$_2$PO$_4$ ----- ~250 mL (titrant)
Adjust pH to pH 7.4 with NaOH and H$_3$PO$_4$

**Sodium phosphate buffer (NaPi), 50 mM**
Dilute 200 mM NaPi 1:4 v/v in MilliQ water (mQH$_2$O), final pH 7.4

**Assay buffer with t-BuOH**
Mix 200 mM NaPi pH 7.4 and neat t-BuOH to desired concentrations
Top up with mQH$_2$O

**Ni$^{2+}$ affinity column purification (HisTrap) binding buffer (Buffer A)**
NaPi, 200 mM ----- 250 mL (50 mM)
Sodium chloride (NaCl) ----- 29.22 g (0.5 M)
Adjusted to pH 7.4 with NaOH and H$_3$PO$_4$
Top up to 1 L with MilliQ water (mQH$_2$O)

**Ni$^{2+}$ affinity column purification (HisTrap) elution buffer (Buffer B)**
NaPi, 200 mM ----- 250 mL (50 mM)
Sodium chloride (NaCl) ----- 29.22 g (0.5 M)
Imidazole ----- 34.04 (0.5 M)
Adjusted to pH 7.4 with NaOH and H$_3$PO$_4$
Top up to 1 L with MilliQ water (mQH$_2$O)

**10X SB buffer (Sodium Hydroxide-borate) for DNA gel electrophoresis**
Sodium hydroxide ----- 8 g
Top up with 1 L mQH$_2$O
Boric acid (to adjust pH to 8.0) ----- ~48 g
To make 1X SB buffer:
Dilute 100 mL 10X SB buffer in 900 mL mQH$_2$O
### Bromophenol blue (BPB) loading buffer

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bromophenol blue</td>
<td>0.05%</td>
</tr>
<tr>
<td>Glycerol</td>
<td>75%</td>
</tr>
<tr>
<td>Tris buffer, pH 8.0</td>
<td>2.5 mM</td>
</tr>
</tbody>
</table>

Made up to desired volume with mQH$_2$O

### SDS-PAGE running buffer (1X)

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>14.4 g</td>
</tr>
<tr>
<td>Tris</td>
<td>3.0 g</td>
</tr>
<tr>
<td>SDS</td>
<td>1.0 g</td>
</tr>
</tbody>
</table>

Made up to desired volume with mQH$_2$O

### SDS-PAGE loading buffer

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris buffer pH 6.8</td>
<td>0.31 M</td>
</tr>
<tr>
<td>Sodium dodecyl sulfate (SDS)</td>
<td>10%</td>
</tr>
<tr>
<td>Glycerol</td>
<td>50%</td>
</tr>
<tr>
<td>β-mercaptoethanol (βME)</td>
<td>25%</td>
</tr>
<tr>
<td>Bromophenol blue</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

Made up to desired volume with mQH$_2$O
### Appendix D: Oligonucleotide Primers

Table A-1: Table listing oligonucleotide primers used. F: forward primer; R: reverse primer; nnk, mnn denotes degenerate codons.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequences (5’–3’)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUS 164X F</td>
<td>act tcc atg atn nk tta act atg ceg g</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS 562QC R</td>
<td>aga tce ctt tct tgt tgc cgc cca cmn nca ata tgc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS D163 QC F</td>
<td>agt taa aga ann nat gga agt aag act gct ttt tct tgc cg</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS D163X</td>
<td>gtc tta ctt cca tnn ntt ctt taa cta tgc ceg aat cc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS D399-D415 F</td>
<td>ata gcg cgt gac aaa aac cac cc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS D399-D415 R</td>
<td>aat aat ccc tgc acc ttg cgg</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS D508X R</td>
<td>aat acg ggc tgn nka cgt tag cc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS E413A/G F</td>
<td>att gcc aac gst ceg gat acc</td>
<td>Site saturation mutagenesis at E413</td>
</tr>
<tr>
<td>GUS E413X R</td>
<td>tat tgc caa c nn kcc gga tac ceg tcc</td>
<td>Site saturation mutagenesis at E413</td>
</tr>
<tr>
<td>GUS E504A R</td>
<td>acg ceg tat ggc gtg atg atc atc ggc</td>
<td>Site specific mutation E504A Codon optimization at E504G</td>
</tr>
<tr>
<td>GUS E504G (common) R</td>
<td>acg ceg taa ceg gtg atg atc atc ggc</td>
<td>Codon optimization at E504G</td>
</tr>
<tr>
<td>GUS E504G (rare) R</td>
<td>acg ceg tat ceg gtg atg atc atc ggc</td>
<td>Site specific mutation E504S Fragment library construction</td>
</tr>
<tr>
<td>GUS E504S R</td>
<td>acg ceg taa ctg gtg atg atc atg ggc</td>
<td>Site specific mutation E504S</td>
</tr>
<tr>
<td>GUS E505-S522 F</td>
<td>atc acc gaa tac ggc gtg g</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>GUS E505-S522 R</td>
<td>gcc atg cac act gat act ctt</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>GUS F161X</td>
<td>age agt ctt can nkc atg att tct tta act atg c</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS F162X</td>
<td>age agt ctt act tcn nkg att tc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS G362X</td>
<td>tct ctt tan nka tgt gtt tgg aag c</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS K370 R</td>
<td>aag ceg gca can nkc ceg gga aag aac tg</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>Primers</td>
<td>Sequences (5' - 3')</td>
<td>Comments</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>GUS K372E</td>
<td>aac aag ccg gaa gaa ctg tac agc</td>
<td>Site specific mutation K372E</td>
</tr>
<tr>
<td>GUS L361X</td>
<td>taa cct ctc tnn kgg cat tgg ttt cg</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS N412</td>
<td>tat ccg gtt cnn nng cca tac tcc aca tca cc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS N412</td>
<td>gag tgc cnn nge acc gga tac ceg tcc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS M447</td>
<td>gcg tca atg tan nnt tct gcg acg ctc aca ccg</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS M447</td>
<td>age gtc gca gaa nnn tac att gac gca ggt gat cgg</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS P371T</td>
<td>aag ccg gca aca aga cga aag acc</td>
<td>Site specific mutation P371T</td>
</tr>
<tr>
<td>GUS R562X</td>
<td>aag cca tat tgn nk gtt gcg gta aca aga aag g</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS R562X</td>
<td>agg cat att gnn ngt tgg cgg taa cca gaa agg g</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS T509A</td>
<td>ccc gcg taa cgc atc cac gcg gta tt</td>
<td>Site specific mutation T509A</td>
</tr>
<tr>
<td>GUS T509X</td>
<td>gcg tgg atn nnt tag ceg gcg tgc act c</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS Y160X</td>
<td>agc acg tct nnk ttc cat gat ttc ttt aac</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS Y469X</td>
<td>gaa ccc tta tnn kgg atg gta ttt cca aag c</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS Y472X</td>
<td>tta cgg atg gnn kgt cca aag cgg</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>GUS Y478X</td>
<td>cct gcg tnn kta ceg tgg atg gta tgt cc</td>
<td>Site saturation mutagenesis</td>
</tr>
<tr>
<td>gusA 1016F</td>
<td>act gcg cag atg aac atg gc</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1166R</td>
<td>aag tgc gct tgc tga gtt tcc c</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1289F</td>
<td>aag cca cgg gta aac tcc acc cc</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1289R</td>
<td>gtc gag ttt acg cgt tgc ttc gc</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1339F</td>
<td>atg ttc tgc gac gct cac acc gat a</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1339R</td>
<td>gta ceg gtg tga cgc tgc cag aac</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>Primers</td>
<td>Sequences (5' - 3')</td>
<td>Comments</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>gusA 1364 R'</td>
<td>aga gat cgc tga tgg</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1448 R</td>
<td>ttc tct gec tgt tcc aaa tgg</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1511 F</td>
<td>tat cat cac cga ata cgg cgt gga taa g</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 1511 R</td>
<td>tat cca cgc cgt att cgg tga tga taa tgc</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 367nt F</td>
<td>att gcc ggg aaa agt gta cg</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 5' int</td>
<td>ata acc ttc acc cgg</td>
<td>Sequencing</td>
</tr>
<tr>
<td>730 R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gusA 544 R</td>
<td>tgt tcg ggc tgt tgt aga gc</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 857 F</td>
<td>taa cca caa acc gtt cta ctt tac tgc</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>gusA 857 R</td>
<td>aac ggt tgg tta atc agg aac tgt tgc</td>
<td>Fragment library construction</td>
</tr>
<tr>
<td>pET28a/GUS</td>
<td>ggt ggt gga att ctc att gtt tgc c</td>
<td>Construction of pET28a/GUS WT</td>
</tr>
<tr>
<td>EcorI R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pET28a/GUS</td>
<td>cgc ggc agc cat atg tta cgt cct gta gaa acc</td>
<td>Construction of pET28a/GUS</td>
</tr>
<tr>
<td>V2L F</td>
<td></td>
<td>WT</td>
</tr>
<tr>
<td>pJe01 F</td>
<td>cac aat tcc aca acg gtt tc</td>
<td>Library construction</td>
</tr>
<tr>
<td>pJe01 R</td>
<td>cga ccg get cag tgc aaa g</td>
<td>Library construction</td>
</tr>
<tr>
<td>pJe02 F</td>
<td>att cca cca cgg tgt ccc</td>
<td>Library construction</td>
</tr>
<tr>
<td>pJexp seq F'</td>
<td>ctc gaa aat aat aaa ggg aaa atc ag</td>
<td>Sequencing</td>
</tr>
</tbody>
</table>
Figure A2-1: The standard curves used to calculate the extinction coefficient values for pNP in 50 mM NaPi, pH 7.4. The molar coefficient was taken from the average of four curves.

Figure A2-2: The standard curves used to calculate the extinction coefficient values for pNP in 50 mM NaPi, pH 7.4 containing 20% t-BuOH, The molar coefficient was taken from the average of four curves.
Appendix 3-1

Table A3-1: Library sizes and the oversampling size required to achieve 95% library coverage in directed evolution experiments. They are dependent on the peptide length of an enzyme. Oversampling library sizes were calculated according to the to equation described by Patrick et. al.\(^1\)

<table>
<thead>
<tr>
<th>Length of protein or fragment (no. of AA residues)</th>
<th>( f_{\text{mutation}} )</th>
<th>Theoretical library size (no. of variants)</th>
<th>95% coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>6.00E+02</td>
<td>1.80E+03</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1.78E+05</td>
<td>5.35E+05</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3.49E+07</td>
<td>1.05E+08</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.80E+03</td>
<td>5.40E+03</td>
</tr>
<tr>
<td>300</td>
<td>2</td>
<td>1.61E+06</td>
<td>4.84E+06</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>9.62E+08</td>
<td>2.89E+09</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>3.60E+03</td>
<td>1.08E+04</td>
</tr>
<tr>
<td>600</td>
<td>2</td>
<td>6.47E+06</td>
<td>1.94E+07</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>7.74E+09</td>
<td>2.32E+10</td>
</tr>
</tbody>
</table>

Theoretical library size calculated by:

\[ nC_r \]

where,

- \( n \) = length of fragment
- \( r \) = average mutation frequency

The oversampling factor that is required to achieve 95% library coverage is approximately three.

This is calculated by:

\[ L = -V\ln(1-F) \]

where,

- \( V \) = total number of possible mutant sequences
- \( F = 1 - P(0) \)
- \( F \) represents the probability of all variants occurring at least once
- Hence, \( F = 1 - P(0) \); where \( P(0) \) = probability that a sequence does not contain any mutation (0 substitution)

References:
Table A3-2: Mutations accumulated in β-GUS evolution

| Lib  | D31 | R71 | F74 | D196 | S37 | S173 | I271 | K277 | H313 | I346 | F357 | S376 | N381 | A387 | V486 | A426 | A431 | I441 | H452 | K453 | L463 | V484 | K495 | K487 | A542 | G545 | L561 |
|------|-----|-----|-----|------|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
|      |     |     | D   | G    | L   | A    | R    | A    | A    | A    | V    | A    | A    | R    | V    | T    | I    | L    | S    | E    | D    | D    | R    | G    |     |     |     |     |     |
| L1   | 1   | 1   |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|      | Low |     |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|      | Lp2 |     |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| L1.  | 1   | 2   |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|      | High|     |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|      |     |     |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| L3   | 2a  |     |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|      | Lp2 |     |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|      | 2   |     |     |      |     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

Legend: Error-prone PCR, Error-prone + recombination (ep-STEP)
Table A3-2 (cont): Mutations accumulated in β-GUS evolution. Highlighted in blue denotes the point when t-BuOH was added.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L4</td>
<td>2bLp</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L5</td>
<td>1Lp4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L6</td>
<td>1Lp6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L7</td>
<td>3Lp</td>
<td>2+6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Table A3-2: Mutations accumulated in β-GUS evolution. Highlighted in blue denotes the point when t-BuOH was added.
Table A3-2 (cont): Mutations accumulated in β-GUS evolution. Highlighted in blue denotes the point when \( t \)-BuOH was added.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L8</td>
<td>4Lp</td>
<td>2+6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L9</td>
<td>5-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GUS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4Lp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2+6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>V</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>C</td>
<td>S</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>C</td>
<td>S</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Library type:
- S: Recombination (StEP)
- S: Error-prone PCR
Table A3-3: Relative activity of 4GUS mutants. All mutants were more than 3x more active than wild-type and could not be discriminated. Activities were compared with the wells that had single mutation L561S in the same plate instead.

<table>
<thead>
<tr>
<th>Mutant(s)</th>
<th>No. of occurrences</th>
<th>Mutations</th>
<th>Relative WT activity</th>
<th>Relative L561S activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>4dGus.1</td>
<td>1</td>
<td>H313L/K495R/L561S</td>
<td>&gt; 3</td>
<td>0.76</td>
</tr>
<tr>
<td>4dGus.3</td>
<td>1</td>
<td>K277Q/L561S</td>
<td>&gt; 3</td>
<td>0.70</td>
</tr>
<tr>
<td>4dGus.12</td>
<td>1</td>
<td>K277Q/A431V/L561S</td>
<td>&gt; 3</td>
<td>0.74</td>
</tr>
<tr>
<td>4GUS21</td>
<td>2</td>
<td>K495R/L561S</td>
<td>&gt; 3</td>
<td>1.11</td>
</tr>
<tr>
<td>4GUS21</td>
<td>5</td>
<td>A431V/L561S</td>
<td>&gt; 3</td>
<td>0.63</td>
</tr>
<tr>
<td>L561S</td>
<td>18</td>
<td>L561S</td>
<td>&gt; 3</td>
<td>1.00</td>
</tr>
<tr>
<td>4GUS11</td>
<td>2</td>
<td>K277Q</td>
<td>&gt; 3</td>
<td>0.70</td>
</tr>
<tr>
<td>4GUS31</td>
<td>1</td>
<td>H313L</td>
<td>&gt; 3</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Appendix 3-3

Table A3-4: Mutations accumulated in revisited Lp6 libraries after the addition of t-BuOH (Rd 1). Rel Act (bfe): relative activity over wild-type in buffer, Rel Act (t-BuOH): relative activity over wild-type in t-BuOH.

|         | Rel Act (bfe) | Rel Act (t-BuOH) | Rel stab | G599 | F537 | F568 | F596 | F560 | V548 | F564 | F537 | F593 | G598 | F582 | A542 | V533 | A511 | Q547 | K576 | N566 | G594 | S557 |
|---------|---------------|------------------|----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1ltp6.1 | 1.33          | 1.6209           | 0.84     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.2 | 1.32          | 0.99             | 0.75     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.3 | 1.07          | 1.6001           | 1.49     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.4 | 1.00          | 0.9416           | 0.94     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.5 | 0.34          | 0.6402           | 1.87     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.6 | 1.16          | 0.8893           | 0.77     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.7 | 1.28          | 0.8984           | 0.70     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.8 | 1.25          | 0.8035           | 0.64     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.9 | 1.24          | 0.9777           | 0.79     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.10| 0.61          | 1.6219           | 2.66     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.11| 0.42          | 3.4126           | 8.04     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.12| 0.12          | 0.6628           | 5.65     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.13| 0.28          | 1.0551           | 3.75     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.14| 0.49          | 3.9321           | 8.08     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.15| 1.20          | 1.5435           | 1.29     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.16| 0.24          | 0.3888           | 1.60     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.17| 0.13          | 0.5634           | 4.21     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.18| 0.17          | 0.4173           | 2.51     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.19| 0.47          | 1.1469           | 2.42     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.20| 0.37          | 0.8875           | 2.40     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.21| 0.75          | 1.88             | 2.52     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.22| 1.25          | 1.1775           | 0.94     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.23| 0.04          | 0.1751           | 3.93     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.24| 0.33          | 0.5456           | 1.66     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.25| 0.47          | 0.7265           | 1.54     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.26| 0.18          | 0.4115           | 2.25     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.27| 0.24          | 0.7326           | 3.08     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.28| 2.26          | 1.5493           | 0.68     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.29| 1.99          | 1.5345           | 0.77     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.30| 0.60          | 2.6              | 4.31     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 1ltp6.31| 1.08          | 4.06             | 3.75     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
Table A3-5: Mutations accumulated in revisited Lp6 libraries after the addition of t-BuOH (Rd 2). Bfe: relative activity over wild-type in buffer, Res: Residual activity when 20 % t-BuOH was included.

<table>
<thead>
<tr>
<th>2Lp6 clones</th>
<th>Other designations</th>
<th>Bfe</th>
<th>t-BuOH</th>
<th>Res</th>
<th>543</th>
<th>557</th>
<th>560</th>
<th>561</th>
<th>594</th>
<th>598</th>
<th>596</th>
<th>593</th>
<th>597</th>
<th>592</th>
<th>596</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>L561S</td>
<td>0.51</td>
<td>4.353</td>
<td>8.54</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>1.53</td>
<td>3.785</td>
<td>2.47</td>
<td>T</td>
<td>S</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>1.11</td>
<td>4.805</td>
<td>4.32</td>
<td>T</td>
<td>S</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>1.21</td>
<td>4.68</td>
<td>3.87</td>
<td>S</td>
<td></td>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>0.71</td>
<td>3.265</td>
<td>4.61</td>
<td>I</td>
<td>T</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>7GUS10</td>
<td>0.83</td>
<td>3.579</td>
<td>4.30</td>
<td>L</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>7GUS13</td>
<td>0.62</td>
<td>3.45</td>
<td>5.56</td>
<td>T</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>7GUS14</td>
<td>0.56</td>
<td>3.52</td>
<td>6.25</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-</td>
<td>1.02</td>
<td>4.99</td>
<td>4.88</td>
<td>I</td>
<td>T</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>0.80</td>
<td>5.15</td>
<td>6.44</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>7GUS24</td>
<td>0.35</td>
<td>2.41</td>
<td>6.88</td>
<td>S</td>
<td></td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>0.40</td>
<td>2.95</td>
<td>7.39</td>
<td>R</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>-</td>
<td>0.77</td>
<td>3.64</td>
<td>4.72</td>
<td>T</td>
<td>S</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>-</td>
<td>0.37</td>
<td>0.87</td>
<td>2.74</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>-</td>
<td>1.57</td>
<td>3.15</td>
<td>8.45</td>
<td>I</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>1.06</td>
<td>5.56</td>
<td>3.54</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>-</td>
<td>0.51</td>
<td>6.30</td>
<td>5.95</td>
<td>V</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>-</td>
<td>0.98</td>
<td>3.33</td>
<td>7.73</td>
<td>I</td>
<td>S</td>
<td>C</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>-</td>
<td>0.77</td>
<td>4.55</td>
<td>5.87</td>
<td>I</td>
<td>S</td>
<td></td>
<td></td>
<td>S</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>-</td>
<td>0.17</td>
<td>1.67</td>
<td>10</td>
<td>L</td>
<td>S</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>7GUS7</td>
<td>0.54</td>
<td>3.89</td>
<td>7.15</td>
<td>S</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>-</td>
<td>1.05</td>
<td>5.89</td>
<td>5.58</td>
<td>S</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4-1

(a) Variable fragmentation voltage

(b) Variable capillary voltage

Figure A4-1: Optimization of the LC/MS ionization parameters. (a) The fragmentation voltage of the LC/MS is varied between 100–250 V. (b) The capillary voltage of the LCMS is varied between 2000–4500 V.
Appendix 4-2

(a) 1X lysate load

Free CMO-DHEA

CMO-DHEA glucuronide

Area 19028.1 RFU^2

(b) 2X lysate load

Free CMO-DHEA

CMO-DHEA glucuronide

Area 32399.3 RFU^2

Figure A4-2: Resolution of the free CMO-DHEA and CMO-DHEA glucuronide peaks using only the guard column at two difference lysate loads. (a) Resolution of the peaks and the peak area of CMO-DHEA glucuronide with 1x lysate load. (b) Resolution of the peaks and the peak area of CMO-DHEA glucuronide with 2x lysate load. The method can discriminate variants that are two-fold improved.
Appendix 4-3

Table A4-1: Library sizes for various site-saturation libraries.

<table>
<thead>
<tr>
<th>No. of AA position</th>
<th>Codon degeneracy</th>
<th>$f_{\text{mutation}}$</th>
<th>Theoretical library size (no. of variants)</th>
<th>95% coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDT</td>
<td>12</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>1</td>
<td>NNK</td>
<td>32</td>
<td>32</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>NNN</td>
<td>64</td>
<td>64</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>NDT</td>
<td>12</td>
<td>144</td>
<td>432</td>
</tr>
<tr>
<td>2</td>
<td>NNK</td>
<td>32</td>
<td>1024</td>
<td>3072</td>
</tr>
<tr>
<td></td>
<td>NNN</td>
<td>64</td>
<td>4096</td>
<td>12288</td>
</tr>
<tr>
<td></td>
<td>NDT</td>
<td>12</td>
<td>1728</td>
<td>5184</td>
</tr>
<tr>
<td>3</td>
<td>NNK</td>
<td>32</td>
<td>3.28E+04</td>
<td>9.83E+04</td>
</tr>
<tr>
<td></td>
<td>NNN</td>
<td>64</td>
<td>2.62E+05</td>
<td>7.86E+05</td>
</tr>
<tr>
<td></td>
<td>NDT</td>
<td>12</td>
<td>2.07E+04</td>
<td>6.22E+04</td>
</tr>
<tr>
<td>4</td>
<td>NNK</td>
<td>32</td>
<td>1.05E+06</td>
<td>3.15E+06</td>
</tr>
<tr>
<td></td>
<td>NNN</td>
<td>64</td>
<td>1.68E+07</td>
<td>5.03E+07</td>
</tr>
</tbody>
</table>

The oversampling factor that is required to achieve 95% library coverage is approximately three.

This is calculated by:

$$L = -V \ln(1-F)$$

where

- $V$ = total number of possible mutant sequences
- $F = 1 - P(0)$
- $F$ represents the probability of all variants occurring at least once

Hence, $F = 1 - P(0)$; where $P(0)$ = probability that a sequence does not contain any mutations (0 substitution)

Reference:

Appendix 4-4

Figure A4-3: ITC binding isotherms of glucuronylsynthase and its variants (a) SynWT (b) SynH162Q (c) Syn Y160G.
Figure A4-4: Reaction scheme for the glucuronidation of CMO-DHEA (k₁) and the competing hydrolytic reaction (k₋₁).
Appendix 5-1

Figure A5-1: SDS-Page analysis of the expression levels of Syn expressed at different temperature and length of time on two different days (D1 and D2). Bradford quantification of protein amounts was also compared to the overexpression levels indicated on these SDS-pages.
Table A5-1: Activity of the crude lysates measured from samples expressed on two different days (intra-day replicate). On each day, duplicate of each samples were grown and expressed. D1 = Day 1, D2 = Day 2. Numbers behind D1 or D2 denotes the duplicate on each day.

<table>
<thead>
<tr>
<th>Induction temperature</th>
<th>Induction time</th>
<th>Activity (nM/min/ug) x 10^{-2}</th>
<th>D1-1</th>
<th>D1-2</th>
<th>D2-1</th>
<th>D2-2</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 °C</td>
<td>4 h</td>
<td></td>
<td>2.24</td>
<td>2.34</td>
<td>2.13</td>
<td>2.13</td>
<td>2.21 ± 0.10</td>
</tr>
<tr>
<td></td>
<td>8 h</td>
<td></td>
<td>2.01</td>
<td>1.37</td>
<td>2.32</td>
<td>2.21</td>
<td>1.98 ± 0.43</td>
</tr>
<tr>
<td></td>
<td>overnight</td>
<td></td>
<td>2.21</td>
<td>2.16</td>
<td>1.79</td>
<td>2.42</td>
<td>2.14 ± 0.26</td>
</tr>
<tr>
<td>30 °C</td>
<td>4 h</td>
<td></td>
<td>5.86</td>
<td>6.02</td>
<td>2.86</td>
<td>2.75</td>
<td>4.37 ± 1.82</td>
</tr>
<tr>
<td></td>
<td>8 h</td>
<td></td>
<td>6.53</td>
<td>5.24</td>
<td>10.1</td>
<td>12.0</td>
<td>8.48 ± 3.14</td>
</tr>
<tr>
<td></td>
<td>overnight</td>
<td></td>
<td>33.0</td>
<td>43.1</td>
<td>41.7</td>
<td>37.5</td>
<td>38.8 ± 4.54</td>
</tr>
<tr>
<td>37 °C</td>
<td>4 h</td>
<td></td>
<td>9.85</td>
<td>12.07</td>
<td>5.59</td>
<td>6.63</td>
<td>8.54 ± 2.97</td>
</tr>
<tr>
<td></td>
<td>8 h</td>
<td></td>
<td>25.7</td>
<td>23.3</td>
<td>6.55</td>
<td>10.6</td>
<td>16.5 ± 9.40</td>
</tr>
<tr>
<td></td>
<td>overnight</td>
<td></td>
<td>79.8</td>
<td>93.3</td>
<td>33.1</td>
<td>24.6</td>
<td>57.7 ± 34.0</td>
</tr>
</tbody>
</table>
Figure A5-2: Growth curve of constitutive and inducible host-vector expression system. (a) Growth curve of IPTG inducible expression system harbouring gusA gene in GMS407 (DE3) cell line, pET28a/gusA/GMS407(DE3). (b) Growth curve of constitutive expression system harbouring gusA gene in BW25142 cell line, pJWL1030/gusA/BW25142. The growth curve of inducible expression system harbouring gusA gene in BW25142 (pJ401/gusA:E504G/BW25142) is expected to be similar to both the growth curves.
Appendix 6-1

A6-1  Mutation rate determination for the random mutagenesis of the whole gene

We evaluated the μ computation by first comparing the empirical standard deviation (SD) and the theoretical SD (\(\sqrt{\mu}\)) were calculated (Table A6-1). As the library distribution is expected to follow Poisson distribution, the empirical SD should fall close to the \(\sqrt{\mu}\). At higher μ, the SD becomes larger. This is likely caused by small sampling size that resulted in insufficient coverage of broader distribution. This method may not be the best way to evaluate the μ that was determined. Hence, we also projected the trend of μ in response to increasing MnCl\(_2\) concentration (Figure A6-1). The mutation rate increases proportionally with MnCl\(_2\) concentration with an expected rate of 11 amino acids per mM MnCl\(_2\) and a projected mutation rate of 0.7 AA per gene when the MnCl\(_2\) is 0 mM (y-intercept). The y-intercept is slightly higher than expected as one would expect the mutation rate to be closer to 0 when the MnCl\(_2\) is 0 mM. We forced the trendline through 0, which resulted in a slightly higher rate of increase in mutation rate (~14 AA per mM MnCl\(_2\)) (Figure A6-1, dotted line). This may underestimate the empirical μ determined at higher MnCl\(_2\) above 0.30 mM and can result in slightly lower mutational tolerance computation. Nevertheless, the difference between the two expected values is small and we do not expect it to cause significant difference.

<table>
<thead>
<tr>
<th>Empirical μ</th>
<th>Empirical SD</th>
<th>Theoretical SD</th>
<th>% deviation of SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.13</td>
<td>0.835</td>
<td>1.061</td>
<td>21.3</td>
</tr>
<tr>
<td>2.50</td>
<td>1.77</td>
<td>1.58</td>
<td>10.7</td>
</tr>
<tr>
<td>1.88</td>
<td>2.10</td>
<td>1.37</td>
<td>53.4</td>
</tr>
<tr>
<td>5.00</td>
<td>1.20</td>
<td>2.24</td>
<td>46.3</td>
</tr>
<tr>
<td>4.38</td>
<td>3.00</td>
<td>2.09</td>
<td>43.3</td>
</tr>
<tr>
<td>7.00</td>
<td>6.28</td>
<td>2.65</td>
<td>137</td>
</tr>
</tbody>
</table>

Table A6-1: Deviation of empirical (experimentally determined) and theoretical SD, and the quartiles of the mutation rates in each library.
Figure A6-1: Mutation rate increased in a linear manner in response to the amount of MnCl₂. Error bars represent the deviation from the median values. Dotted line represents the line fit when forced through the origin.
Figure A6-2: Location of residues targeted for site-saturation mutagenesis in the E. coli β-GUS crystal structure (PDB ID: 3K46). (a) Interaction between M447 and Y469. (b) Position of Y468-Y472 and M447 relative to superimposed pNPGlcA (yellow) in the crystal structure. (c) Superimposition of loop N466-Y472 of E. coli β-GUS crystal structure without GDL inhibitor bound (apo) over the crystal structure with GDL inhibitor bound (PDB ID: 3K4D). (d)-(f) Position and movement of residues Y468 (d), Y469 (e) and Y472 (f) in the presence of GDL. The residues Y472 and Y468 flipped in the opposite direction. Y472 flipped into the active site and forms interaction with the carboxylate group at C6, while Y469 flipped out. Adapted from Wallace et al.1
Appendix 6-3

Figure A6-3: Comparison of profile slopes (m4) with different number of data points.
### Activity distribution curves of site saturation libraries (pNPGlcA screening data)

<table>
<thead>
<tr>
<th>Clones</th>
<th>D163</th>
<th>D163</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1" alt="" /></td>
<td><img src="image2" alt="" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clones</th>
<th>N412</th>
<th>N412</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image3" alt="" /></td>
<td><img src="image4" alt="" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clones</th>
<th>E413 (control for full inactivation)</th>
<th>Not determined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image5" alt="" /></td>
<td></td>
</tr>
</tbody>
</table>

Continued on the next page
Figure A6-4: Activity distribution curves of site-saturation libraries from screening data with pNPGlcA substrate (left) and pNPGlc substrate (right).
Table A6-2: Summary of sequenced mutations from site saturation libraries. I = Inactive; P = retained partial activity (5–80%); A = as active as WT (80–120%); O = Better than wild type (>120% WT activity). *nd not determined

<table>
<thead>
<tr>
<th>Library / Mutations</th>
<th>Activity (pNPGA)</th>
<th>Activity (pNPG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D163X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>E</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>M</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>P</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Q</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>R</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Y</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>R562X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>F</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>H</td>
<td>I</td>
<td>P</td>
</tr>
<tr>
<td>N</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>P</td>
<td>I</td>
<td>P</td>
</tr>
<tr>
<td>Q</td>
<td>I</td>
<td>P</td>
</tr>
<tr>
<td>D</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>S</td>
<td>I</td>
<td>P</td>
</tr>
<tr>
<td>T</td>
<td>I</td>
<td>P</td>
</tr>
<tr>
<td>V</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>W</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>M447X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>P</td>
<td>I</td>
</tr>
<tr>
<td>D</td>
<td>P</td>
<td>nd</td>
</tr>
<tr>
<td>E</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>F</td>
<td>P</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>P</td>
<td>I</td>
</tr>
<tr>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>P</td>
<td>P</td>
<td>I</td>
</tr>
<tr>
<td>R</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>S</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>T</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>V</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>N412X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>D</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>E</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Y468X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>D</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>H</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>K</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>M</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>N</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>P</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>S</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>T</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Y469X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>C</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>D</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>H</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>K</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>M</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>N</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>P</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>S</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>T</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Y472X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>K</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>M</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>N</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>R</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>S</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>T</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>V</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>T509X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>O</td>
</tr>
<tr>
<td>D</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>P/I</td>
<td>P/I</td>
</tr>
<tr>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>K</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>N</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>N412X</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>(cont.)</td>
<td>P</td>
<td>I</td>
</tr>
<tr>
<td>Q</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>R</td>
<td>I</td>
<td>I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>T509X</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(cont.)</td>
<td>S</td>
<td>A</td>
<td>O</td>
</tr>
<tr>
<td>V</td>
<td>I</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>P</td>
<td>P</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 6-5

Kinetics curve for purified enzyme characterization of glucuronidase activity (left) and glucosidase activity (right)

M447G

- **Kinetics Curve for M447G**

  - **Glucuronidase Activity**
    - **k_{cat}** (M447G) (s⁻¹) vs. [pNPG] (mM)
    - **k_{cat}** (M447G) (s⁻¹) vs. [pNPG] (mM)

M447S

- **Kinetics Curve for M447S**

  - **Glucosidase Activity**
    - **k_{cat}** (M447S) (s⁻¹) vs. [pNPG] (mM)
    - **k_{cat}** (M447S) (s⁻¹) vs. [pNPG] (mM)

(continued on the next page)
APPENDIX: CHAPTER 6

T509A

R562S

R562T

Continued on the next page
Figure A6-5: Michalis-Menten kinetic curves of purified enzymes (WT, M447G/S, R562S/T and T509A) from with $p$NPGA substrate (left) and $p$NP substrate (right).
Appendix 6-6

Figure A6-6: Comparison of activity distribution data between raw screening data and a subset of screening data. (a) Activity distribution of T509A replotted based on sequencing data and three control wells (Inactivation factor = -8.96) (b) Activity distribution of raw T509A screening data (Inactivation factor = -3.48 and -12.0 for each replicate).

Appendix 6-7

Figure A6-7: Curve-fitting data to obtain input values for unpaired t-test analysis. (a) Fitness decay of TEM-1 at the highest fitness level. The data used to plot TEM-1 was taken from reference 2, supplementary information, Table 1 (1P0 to 8P0). The mutation rate was estimated from Figure 1 of reference 2 and may deviate by ±0.2. The α-value compares very well with the reported value of 0.37. (b) Linear transformation of decay plot in panel a according to Equation 6-2 for TEM-1. (c) Linear transformation of decay plot, Figure 6-1, according to Equation 6-2 for β-GUS.
Appendix 6-7

Table A6-3: Unpaired two-tailed t-test analysis of TEM-1 and β-GUS α-values. Computing the significance of the difference between the α-value of TEM-1 and β-GUS. α-value of TEM-1 was obtained from Figure A6-7a The standard error (S.E.) of β-GUS and TEM-1 were obtained from the \textsc{LINEST} function of the linear regression in Figure A6-7b.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>α</th>
<th>S.E.</th>
<th>dF</th>
<th>( t_{\text{crit}} )</th>
<th>p-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEM-1</td>
<td>0.33</td>
<td>0.053</td>
<td>7</td>
<td></td>
<td>2.86</td>
<td>0.014 Significant at p &lt; 0.05; not significant at p &lt; 0.01</td>
</tr>
<tr>
<td>β-GUS</td>
<td>0.46</td>
<td>0.1</td>
<td>5</td>
<td></td>
<td>2.86</td>
<td>0.014 Significant at p &lt; 0.05; not significant at p &lt; 0.01</td>
</tr>
</tbody>
</table>

Figure A6-8: The fitness decline of β-GUS fitted with an epistasis parameter (β). This was fitted to Equation 6-2. The α-value is 0.33 and the \( \beta/\alpha \) value is 0.078. When fitted to this equation, the mutational tolerance of β-GUS appears to be the same as TEM-1 and has a \( \beta/\alpha \) value higher than the TEM-1. These values were compared with TEM-1’s value on the plot of \( \beta/\alpha \) versus α in reference 2. β-GUS appears to be under larger epistastic influence.
Figure A6-9: Interactions between domains and between subunits in β-GUS. (a) Structure of the active site domain (TIM-barrel) of β-GUS. Part of Domain I that intrudes into the active site space in Domain III is shown in “salmon red”. In magenta at the bottom of the barrel is a hairpin loop from Domain II, which is located at the base of Domain III. The two catalytic residues are highlighted in green and the disordered region is colored red. The disordered loop is known to extrude into the active site of an
adjacent subunit. This is involved not only in domain-domain interaction, but also in subunit-subunit interaction. (b) The disordered loop in is solvent exposed. Red regions represent amino acid residues between 361-376 that is disordered in the native structure and the structure with GDL bound, but that appears to interact with the active site of an adjacent subunit in the presence of two inhibitors. (c) Subunit interaction between mobile loops at the C-terminal of the active site where R562 is located. The region (Lp6) in the different subunits (A and D) are colored in yellow and brown.

References: