Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Biochemical and structural studies of

*Mycobacterium tuberculosis* isocitrate lyase

Ram Prasad Bhusal

A thesis submitted in fulfilment of the requirements for the degree of

**Doctor of Philosophy in Chemistry**

School of Chemical Sciences

The University of Auckland

2018
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Extent of contribution by PhD candidate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrote the Thesis Chapter</td>
<td>90</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td>review of publication and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td>review of publication and proofreader of the Thesis Chapter</td>
</tr>
<tr>
<td>Ghader Bashiri</td>
<td>Helped on publication</td>
</tr>
<tr>
<td>Brooke Kwai</td>
<td>Helped on publication</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Ghader Bashiri</td>
<td></td>
<td>09/03/2018</td>
</tr>
<tr>
<td>Brooke Kwai</td>
<td></td>
<td>09/03/2018</td>
</tr>
</tbody>
</table>

Last updated: 28 November 2017
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

**Chapter 2: Recombinant production and purification of Mycobacterium tuberculosis isocitrate lyase**

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Planned and conducted experiments, collected and analysed experimental data, and wrote the Thesis Chapter</th>
</tr>
</thead>
</table>

| Extent of contribution by PhD candidate (%) | 90 |

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td>Supervision of research and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td>Supervision of research and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Ghader Bashiri</td>
<td>Helped with production of ICL2</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:
- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Ghader Bashiri</td>
<td></td>
<td>09/03/2018</td>
</tr>
</tbody>
</table>

Last updated: 28 November 2017
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3: Kinetics and inhibition study of Isocitrate lyase 1 (ICL1) Published on MedChemComm 2017, 8, 2155–2163. The work done by other co-authors in the publication has not been included in this Chapter.

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Planned and conducted experiments, collected and analysed experimental data, and wrote the Thesis Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>90</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td>Supervision of research, review of publication and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td>Supervision of research, review of publication and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Anne Swartjes</td>
<td>Synthesised (2S,3R)-2-methylisocitrate</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Anne Swartjes</td>
<td></td>
<td>27/02/2018</td>
</tr>
</tbody>
</table>
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. Please include one copy of this form for each co-authored work. Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4: Structural and biochemical studies of M. tuberculosis isocitrate lyase 2 (ICL2)

<table>
<thead>
<tr>
<th>Co-Author</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td>Supervision of research and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td>Supervision of research and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Ghader Bashiri</td>
<td>Helped with solving the structure of ICL2 and conducted SAXS experiments</td>
</tr>
<tr>
<td>Jóhannes Reynisson</td>
<td>Conducted Molecular dynamics experiment</td>
</tr>
<tr>
<td>Anne Swartjes</td>
<td>Synthesised (2S,3R)-2-methylisocitrate</td>
</tr>
</tbody>
</table>

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ivanhoe Leung</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Jonathan Sperry</td>
<td></td>
<td>12/03/2018</td>
</tr>
<tr>
<td>Ghader Bashiri</td>
<td></td>
<td>09/03/2018</td>
</tr>
<tr>
<td>Jóhannes Reynisson</td>
<td></td>
<td>09/03/2018</td>
</tr>
<tr>
<td>Anne Swartjes</td>
<td></td>
<td>27/02/2018</td>
</tr>
</tbody>
</table>
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

**Chapter 5: Itaconic acid as a green and sustainable substrate for pharmaceuticals. Published on Green Chem. 2016, 18 (8), 2453-2459**

| Nature of contribution by PhD candidate | Planned and conducted experiments, collected and analysed experimental data, and wrote the Thesis Chapter |
| Extent of contribution by PhD candidate (%) | 90 |

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jonathan Sperry</td>
<td>Supervision of research and review of publication and proofread of the Thesis Chapter</td>
</tr>
<tr>
<td>Ivanhoe Leung</td>
<td>Proofread of the Thesis Chapter</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jonathan Sperry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ivanhoe Leung</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Last updated: 28 November 2017
Abstract

The work described in this Thesis has focused on the study of *Mycobacterium tuberculosis* isocitrate lyase (ICL), a key enzyme of the glyoxylate cycle. The work included recombinant production and purification of the different ICL isoforms, the development of new biophysical assays to study ICL inhibitors, biochemical and structural studies of ICL isoform 2, and the sustainable synthesis of pharmaceutically important nitrogen-heterocycles using itaconic acid, a weak inhibitor of ICL, as a starting material.

The use of the *Escherichia coli* and *Mycobacterium smegmatis* expression systems to produce the four isoforms of ICL (isoforms 1, 2, 2a and 2b) was explored. Soluble ICL1, ICL2 and ICL2b were successfully produced and purified. Recombinant ICL2, widely regarded as unstable was found to be a stable enzyme that can be stored in buffers without imidazole.

A combined nuclear magnetic resonance (NMR) and thermal shift assay strategy was developed to study ICL inhibitors using ICL1 as a model system. The NMR-based activity assay enables the detection of substrate (isocitrate) consumption and product (succinate) formation in real time, allowing enzyme kinetics and enzyme inhibition to be followed and quantified. It is informative but relatively low throughput. In contrast, the thermal shift assay measures protein stability upon ligand binding. It is an indirect assay but allows the screening of ICL ligands in a high-throughput manner. The effectiveness of this combined strategy was demonstrated by using existing ICL1 inhibitors.

Protein X-ray crystallography was used for the structural and biochemical characterisation of ICL2, which was found to be a tetramer. Each monomeric unit contains two distinct domains. Structural and kinetic studies revealed that the C-terminal domain is an acetyl-coenzyme A binding domain. Binding of acetyl-CoA leads to a global conformational change, which leads to the activation of the isocitrate lyase and methylisocitrate lyase activities of ICL2.

The synthesis of pharmaceutically important nitrogen-heterocycles using itaconic acid, a weak inhibitor of ICL, was also investigated. A library of diverse N-heterocycles including dihydroindolone, indole, dihydroindolizinone, 2-pyridone, pyrido[1,2-a]indole dihydrocarbazolone and carbazole have been prepared from dimethyl itaconate and pyrrole, two compounds that are readily attainable from biomass.
Acknowledgements

I would like to take this opportunity to thank my supervisors A/Prof. Jonathan Sperry and Dr Ivanhoe Leung for the help and guidance in the last three and half years. I would like to thank Jon for accepting me as a PhD student in his group and giving me an opportunity to work with freedom. I would like to thank Ivan for introducing me to new techniques in chemical biology including molecular biology and NMR. I always feel grateful to learn techniques directly from him. Without his support and encouragement, it would have been hard to understand research in depth. Both Ivan and Jon have managed to pass a passion and enthusiasm for research in chemical biology that I hope I can continue to pursue.

I would like to thank my co-workers and collaborators for their contribution. I would like to thank Dr Ghader Bashiri for supervision of ICL2 project. I learned lots and enjoyed working with him. I want to thank Brooke Kwai for working on the same project and giving hands on conducting various experiments. I would also like to thank Dr Jóhannes Reynisson and Krunal Patel for conducting computational works. I would also like to thank Anne Swartjes for synthesising (2R,3S)-2-methylisocitrate.

I would like to thank all group members of Ivan. In particular, I would like to thank Danilo for answering all my questions on molecular biology techniques and valuable discussions. I would also like to thank Dona for continuous help and being so good for last two years. I also want to thank Praveen, Nabangshu, Mike, Yu, Oi Wei, Naasson, Jinal, Eva and Aimee.

I would like to thank all group members of Jon. Especially, I would like to thank Lachlan Blair for guidance on synthesis. I would like to thank him for helping me to settle in Jon’s lab and being good friend and encouraging me throughout this three and half years. I would also like to thank Ashley, Andrew, Emma, Josh, Matthew, Emily and Allen.
I would also like to thank Nepalese friends, who makes my living easier in Auckland. Especially, I would like to thank Prabhat and Sarala, Amal and Pankaja, Sanjay and Apsara, Balram and Sarada, Roshan and Anu, Nabin, Achyut and Daya.

Thanks now go to my family. I would like to thank for my wife (Rachana) for all the continuous support, encouragement and understanding. Without her presence it would have been hard to finish PhD on time. I am grateful to have a daughter (Ritvi) during this period. I would like to thank my mom and mother-in-law for coming to New Zealand and take care of my baby. It would have been difficult to finish PhD without their support. I would also like to thank my brother (Shiva), sister-in-law (Kamala), sister (Babi) and brother-in law (Pawan) for continuous support. I would also like to thank Rachana’s family (father-in-law, brother-in-laws (Laxman and Abhisek), sister in laws (Susmita and Sony) for their understanding and supports. I would also like to thank Prof. Jin Bok Hee from South Korea for her inspiration. Finally, I would like to dedicate this thesis to my father, who would have been a happiest person on completing my PhD.
# Abbreviations and acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_{280}$</td>
<td>Absorbance</td>
</tr>
<tr>
<td>AAs</td>
<td>Amino Acids</td>
</tr>
<tr>
<td>Acetyl-CoA</td>
<td>Acetyl-coenzyme A</td>
</tr>
<tr>
<td>An-ICL</td>
<td><em>Aspergillus nidulans</em> isocitrate lyase</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulphate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>Carbon 13</td>
</tr>
<tr>
<td>Cbz</td>
<td>Carboxybenzyl</td>
</tr>
<tr>
<td>cryo-EM</td>
<td>Cryo-electron microscopy</td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
</tr>
<tr>
<td>dCTP</td>
<td>Deoxycytidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>Deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>DI</td>
<td>Deionised</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>4-FBA</td>
<td>4-Fluorobenzylamine</td>
</tr>
<tr>
<td>GATB</td>
<td>Global Alliance for TB Drug Development</td>
</tr>
</tbody>
</table>
GNAT  Gcn5-related N-acetyl transferase
GSK  GlaxoSmithKline
'H  Proton
HEPES  4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV  Human immunodeficiency virus
HMF  5-Hydroxymethylfurfural
HTS  High throughput screening
IA  Itaconic acid
ICL  Isocitrate lyase
IC₅₀  Half maximal inhibitory concentration
IDH  Isocitrate dehydrogenase
IMAC  Immobilised metal affinity chromatography
IPTG  Isopropyl β-D-1-thiogalactopyranoside
kₖₑₐₜ  Catalytic constant
Kᵢ  Inhibitory constant
Kᵢₐₙ₃₅  Interaction constant
Kₔₐ  Michaelis constant
LAC  (1R, 5S)-anhydroisosaccharino-δ-lactone
LB  Luria-Bertani media
LDH  Lactate dehydrogenase
LGO  (-)-Levoglucosenone
LHMDS  Lithium bis(trimethylsilyl)amide
LIC  Ligase independent cloning
MAD  Multiwavelength anomalous diffraction
β-ME  β-Mercaptoethanol
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCPBA</td>
<td>Meta-Chloroperoxybenzoic acid</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular replacement</td>
</tr>
<tr>
<td><em>M. smegmatis</em></td>
<td><em>Mycobacterium smegmatis</em></td>
</tr>
<tr>
<td><em>M. Tuberculosis</em></td>
<td><em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>MTS</td>
<td>Tetrazolium</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular weight cut off</td>
</tr>
<tr>
<td>NAD</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>4-NMM</td>
<td>N-Methylmorpholine</td>
</tr>
<tr>
<td>NMP</td>
<td>N-Methyl-2-pyrrolidone</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Optical density</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>Pd/C</td>
<td>Palladium on carbon</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PIPES</td>
<td>Piperazine-&lt;i&gt;N&lt;/i&gt;,&lt;i&gt;N&lt;/i&gt;'-bis(2-ethanesulfonic acid)</td>
</tr>
<tr>
<td>PPA</td>
<td>Polyphosphoric acid</td>
</tr>
<tr>
<td>Propionyl-CoA</td>
<td>Propionyl-Coenzyme A</td>
</tr>
<tr>
<td>rt</td>
<td>Room temperature</td>
</tr>
<tr>
<td>[S]</td>
<td>Substrate concentration</td>
</tr>
<tr>
<td>SAR</td>
<td>Structure-activity relationship</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small-angle X-ray scattering</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size-exclusion chromatography</td>
</tr>
<tr>
<td>SOC</td>
<td>Super optimal broth with catabolite repression</td>
</tr>
<tr>
<td>Succinyl-CoA</td>
<td>Succinyl-coenzyme A</td>
</tr>
<tr>
<td>TAACF</td>
<td>Tuberculosis Antimicrobial Acquisition and Coordinating Facility</td>
</tr>
<tr>
<td>(−)-TAMP</td>
<td>(1R, 2R)-trans-2-Aminomethylcyclopropanecarboxylic acid</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic acid cycle</td>
</tr>
<tr>
<td>TCC-A</td>
<td>Thermo-catalysis and ammonisation</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris acetate-EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
</tr>
<tr>
<td>TEA</td>
<td>Triethylamine</td>
</tr>
<tr>
<td>TEMED</td>
<td>Tetramethylthlenediamine</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch Virus</td>
</tr>
<tr>
<td>THF</td>
<td>Tetrahydrofuran</td>
</tr>
<tr>
<td>Tₘ</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra violet</td>
</tr>
<tr>
<td>UV/vis</td>
<td>Ultraviolet/Visible</td>
</tr>
<tr>
<td>V</td>
<td>Velocity</td>
</tr>
<tr>
<td>Vₘ₉₉</td>
<td>Velocity maximum</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
# Table of Contents

Abstract........................................................................................................................................... I
Acknowledgements............................................................................................................................. II
Abbreviations and acronyms................................................................................................................ IV

**Chapter 1: Introduction** .................................................................................................................. 1

1.1 Tuberculosis (TB)......................................................................................................................... 1
1.2 Energy metabolism of *M. tuberculosis* ...................................................................................... 1
1.2.1 Energy sources ........................................................................................................................ 1
1.2.2 Tricarboxylic acid cycle........................................................................................................... 2
1.2.3 Glyoxylate cycle ....................................................................................................................... 3
1.2.4 Methylcitrate cycle.................................................................................................................... 4
1.3 Isocitrate lyase (ICL) ................................................................................................................... 5
1.3.1 Isocitrate lyase in *M. tuberculosis* ......................................................................................... 5
1.3.2 ICL in *M. tuberculosis* glyoxylate cycle ............................................................................. 7
1.3.3 ICL in *M. tuberculosis* methylcitrate cycle ......................................................................... 8
1.3.4 Additional role of ICL in *M. tuberculosis* for antibiotic resistance ...................................... 9
1.4 Structural and mechanistic aspects of ICL .................................................................................. 10
1.4.1 Catalytic mechanism ............................................................................................................... 10
1.4.2 Role of divalent cation in catalytic mechanism of ICL ........................................................... 12
1.4.3 Structural aspects of ICL ....................................................................................................... 12
1.5 Inhibition of mycobacterial ICL as potential anti-TB therapy ............................................... 16
1.5.1 Inhibitors of ICL .................................................................................................................... 16
1.6 Overview of this Thesis .............................................................................................................. 27
1.6.1 General overview .................................................................................................................. 27
1.6.2 Specific overview .................................................................................................................. 27
Chapter 2: Recombinant production and purification of *Mycobacterium tuberculosis* isocitrate lyase

2.1 Introduction ................................................................................. 28
2.2 Objectives ............................................................................... 31
2.3 Production and purification of ICL1 ........................................... 31
  2.3.1 Cloning ............................................................................... 31
  2.3.2 Protein expression trials ....................................................... 35
  2.3.4 Large-scale production and purification ............................... 36
    2.3.4.1 Immobilized metal affinity chromatography (IMAC) ....... 36
    2.3.4.2 Size exclusion chromatography (SEC) ....................... 37
  2.4 Production and Purification of ICL2 ........................................... 39
    2.4.1 Cloning ........................................................................... 39
    2.4.2 Protein expression trials .................................................... 45
    2.4.3 Large-scale production and purification ............................ 46
  2.4.4 Optimisation for stability of ICL2 ......................................... 47
  2.5 Production and purification of ICL2a and ICL2b ....................... 49
    2.5.1 Design of synthetic gene, cloning and transformation ....... 49
    2.5.2 Expression trial of ICL2a .................................................. 50
    2.5.3 Expression trial of ICL2b .................................................. 56
    2.5.4 Production and purification of ICL2b ................................. 56
  2.6 Conclusion .............................................................................. 58

Chapter 3: Kinetics and inhibition study of Isocitrate lyase 1 (ICL1) ............... 59
  3.1 Introduction ............................................................................. 59
  3.1.1 Assays for ICL .................................................................... 59
  3.1.2 $^1$H nuclear magnetic resonance spectroscopy for kinetics and inhibition studies .......... 60
3.1.3 Thermal shift assay for high-throughput screening of protein-ligand interactions........61
3.2 Objective .........................................................................................................................62
3.3 ICL1 kinetics by \(^1\)H NMR ..........................................................................................62
3.3.1 Isocitrate lyase activity by \(^1\)H NMR ........................................................................62
3.3.2 Role of Mg (II) in the activity of ICL1 .......................................................................64
3.3.3 Kinetic parameters for ICL1 with \(\text{DL-}\)isocitrate by \(^1\)H NMR ..........................65
3.3.4 Kinetic parameters of ICL1 with (2\(R\),3\(S\))-2-methylisocitrate by \(^1\)H NMR ........67
3.3.4.1 Enzyme inhibition study of (2\(R\),3\(S\))-2-methylisocitrate .................................70
3.4 \(^1\)H NMR and thermal shift based method for inhibition studies of ICL1 .................71
3.4.1 \(^1\)H NMR as a tool for inhibition studies of ICL1 ................................................71
3.4.1.1 Single concentration inhibition assay ..................................................................72
3.4.1.2 Quantification of inhibition assay by measuring half maximal inhibitory concentration (IC\(_{50}\)) ........................................................................................................73
3.4.2 Thermal shift assay to compliment \(^1\)H NMR-based inhibition assay .....................78
3.5 Conclusion .......................................................................................................................81

Chapter 4: Structural and biochemical studies of \textit{M. tuberculosis} isocitrate lyase 2 (ICL2) ..................................................................................................................................................82
4.1 Introduction .....................................................................................................................82
4.1.1 \textit{M. tuberculosis} ICL2 ............................................................................................82
4.1.2 Structure determination using X-ray crystallography ..............................................82
4.2 Objectives .......................................................................................................................84
4.3 Crystallography of ICL2 ................................................................................................84
4.4 Structure of ICL2 ..........................................................................................................89
4.5 Kinetics studies of ICL2 ...............................................................................................96
4.5.1 Kinetics of ICL2 in the presence of acetyl-CoA using the substrate \(\text{DL-}\)isocitrate ....96
4.5.2 Kinetics parameters of ICL2 in the presence of propionyl-CoA, succinyl-CoA and CoA using the substrate dl-isocitrate ............................................................................................................................................. 100
4.5.3 Kinetics of ICL2 using (2R,3S)-2-methylisocitrate as substrate ............................................................................................................. 104
4.6 Thermal shift assay of ICL2 in the presence of acetyl-CoA ........................................................................................................... 110
4.7 Crystallisation of ICL2 in the presence of acetyl-CoA ..................................................................................................................... 111
4.8 Active site of ICL2 .............................................................................................................................................................................. 120
4.9 Conclusion ....................................................................................................................................................................................... 125

Chapter 5: Itaconic acid as a green and sustainable substrate for pharmaceuticals ................................................................. 126
5.1 Introduction ..................................................................................................................................................................................... 126
5.1.1 Green chemistry ........................................................................................................................................................................ 127
5.1.2 Green pharmaceutical chemistry ............................................................................................................................................. 127
5.1.3 Biomass for synthesis of fine chemicals .................................................................................................................................. 131
5.1.3.1 Synthesis of Ranitidine and δ-aminolevulinic acid from cellulose-derived 5-(chloromethyl)furfural ................................................................. 131
5.1.3.2 Synthesis of enantiopure cyclopropyl esters from (−)-levoglucosenone ......................................................................................... 133
5.1.3.3 Synthesis of muscarine-like compounds from (1R, 5S)-anhydroisaccharino-δ-lactone (LAC) ........................................................................................................................................ 135
5.1.4 Importance of nitrogen-heterocycles ........................................................................................................................................ 136
5.1.5 Nitrogen heterocycles from biomass ........................................................................................................................................ 137
5.1.5.1 Conversion of cellulose to N-heterocycles ................................................................................................................................. 137
5.1.5.2 Conversion of biomass-derived building blocks to N-heterocycles ........................................................................................... 138
5.1.5.3 Synthesis of N-heterocycles from biomass derived building blocks ........................................................................................... 139
5.1.6 Itaconic acid (IA) ............................................................................................................................................................................ 142
5.1.7 Pyrrole ....................................................................................................................................................................................... 143
5.2 Objective ....................................................................................................................................................................................... 145
6.1.7 Cloning in the presence of ligase .......................................................................................... 178
  6.1.7.1 Preparation of insert and vector .................................................................................. 178
  6.1.7.2 Ligation ...................................................................................................................... 179
6.1.8 Transformation ................................................................................................................ 179
  6.1.8.1 Heat shock general procedure .................................................................................... 179
  6.1.8.2 Electroporation general procedure ............................................................................ 180
  6.1.8.3 Transformation of recombinant pNIC28-Bsa4 vector to E. coli XL10-Gold cells.... 181
  6.1.8.4 Transformation of recombinant pYUB28b vector to E. coli Top10 cells................. 181
6.1.9 Plasmid preparation and DNA sequencing ................................................................. 181
  6.1.10 Small scale optimisation and large scale protein production of ICL1, ICL2a and ICL2b using BL21 (DE3)............................................................................................................. 181
    6.1.10.1 Transformation of recombinant pNIC28-Bsa4 vector to BL21 (DE3) cells........ 181
    6.1.10.2 Starter culture........................................................................................................ 181
    6.1.10.3 Small scale growth optimisation............................................................................. 182
    6.1.10.4 Large scale growth of ICL1 and ICL2b ............................................................... 183
    6.1.10.5 Cell lysis .............................................................................................................. 183
  6.1.11 Small scale optimisation and large scale protein production of ICL2 and ICL2a using BL21 LOBSTR and M. smegmatis ........................................................................................................ 184
    6.1.11.1 Transformation of recombinant pYUB28b vector to plasmid pGro7 co-transformed BL21 LOBSTR ...................................................................................................................... 184
    6.1.11.2 Transformation of recombinant pYUB28b vector to M. smegmatis MC24517 ...... 184
    6.1.11.3 Starter culture........................................................................................................ 184
    6.1.11.4 Small scale growth optimisation......................................................................... 185
    6.1.11.5 Large scale growth of ICL2 ................................................................................ 187
    6.1.11.5 Cell lysis.............................................................................................................. 188
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.12</td>
<td>Chromatography</td>
<td>188</td>
</tr>
<tr>
<td>6.1.12.1</td>
<td>ICL1 and ICL2b</td>
<td>188</td>
</tr>
<tr>
<td>6.1.12.2</td>
<td>ICL2</td>
<td>191</td>
</tr>
<tr>
<td>6.1.13</td>
<td>Digestion of polyhistidine tag by tobacco etch virus (TEV) protease</td>
<td>192</td>
</tr>
<tr>
<td>6.2</td>
<td>NMR Methods</td>
<td>193</td>
</tr>
<tr>
<td>6.2.1</td>
<td>Materials</td>
<td>193</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Buffer for NMR experiment</td>
<td>193</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Sample preparation</td>
<td>193</td>
</tr>
<tr>
<td>6.2.3.1</td>
<td>ICL1</td>
<td>193</td>
</tr>
<tr>
<td>6.2.3.2</td>
<td>ICL2</td>
<td>194</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Sample acquisition and processing</td>
<td>195</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Data treatment</td>
<td>195</td>
</tr>
<tr>
<td>6.3</td>
<td>Thermal shift assay</td>
<td>195</td>
</tr>
<tr>
<td>6.4</td>
<td>X-ray crystallography</td>
<td>196</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Acetyl-CoA free ICL2</td>
<td>196</td>
</tr>
<tr>
<td>6.4.1.1</td>
<td>Sample preparation</td>
<td>196</td>
</tr>
<tr>
<td>6.4.1.2</td>
<td>Initial Screening</td>
<td>196</td>
</tr>
<tr>
<td>6.4.1.3</td>
<td>Fine Screening</td>
<td>197</td>
</tr>
<tr>
<td>6.4.1.4</td>
<td>Additive screening</td>
<td>197</td>
</tr>
<tr>
<td>6.4.1.5</td>
<td>X-ray Data collection and processing</td>
<td>197</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Acetyl-CoA bound structure</td>
<td>198</td>
</tr>
<tr>
<td>6.4.2.1</td>
<td>Sample preparation</td>
<td>198</td>
</tr>
<tr>
<td>6.4.2.2</td>
<td>Screening</td>
<td>198</td>
</tr>
<tr>
<td>6.4.2.3</td>
<td>X-ray Data collection and processing</td>
<td>199</td>
</tr>
<tr>
<td>6.5</td>
<td>Synthesis of N-heterocycles</td>
<td>200</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>6.5.1</td>
<td>General</td>
<td>200</td>
</tr>
<tr>
<td>6.5.2</td>
<td>General Procedure A; C-acylation</td>
<td>201</td>
</tr>
<tr>
<td>6.5.3</td>
<td>General Procedure B; N-acylation</td>
<td>201</td>
</tr>
<tr>
<td>6.5.4</td>
<td>General Procedure C; Dehydrogenation</td>
<td>201</td>
</tr>
<tr>
<td>6.5.5</td>
<td>Synthesis</td>
<td>202</td>
</tr>
</tbody>
</table>

References | 213 |
Chapter 1: Introduction

1.1 Tuberculosis (TB)

Tuberculosis (TB) is an infectious disease that is caused by the bacterium Mycobacterium tuberculosis.1, 2 There are two forms of TB infection, active TB infection and latent TB infection. Active TB refers to the condition in which M. tuberculosis are actively dividing. Active TB is contagious and the infected individual typically displays signs and symptoms that include (but not limited to) productive cough, fever, weight loss, energy loss, appetite loss and night sweat. However, most M. tuberculosis infections do not lead to any symptoms. This condition is known as latent TB infection. When M. tuberculosis enters the body of an individual, the bacteria are phagocytised by macrophages. M. tuberculosis have a thick, waxy mycolic acid capsule, which enables them to survive inside the macrophages and prevent the conversion of phagosomes to phagolysomes,3, 4 thus leading to latent TB infection. Whilst most latent TB infections do not progress to active TB infection, if the infected individual becomes immunocompromised,5, 6 for example during advanced HIV infection,7 the probability of developing active TB may increase significantly.

1.2 Energy metabolism of M. tuberculosis

1.2.1 Energy sources

The life cycle of M. tuberculosis means that the bacteria may spend a significant amount of time (years or even decades) persisting inside macrophages of the infected host. The environment inside macrophages is harsh. Typically, it is low in external nutrients, relatively hypoxic and has an acidic pH.8 M. tuberculosis have therefore evolved highly specialised

---

1Part of the work from this chapter was published as described in:
energy metabolism system to survive in this environment. Unlike other pathogens that tend to consume one carbon source before adapting to the next source, *M. tuberculosis* can catabolise multiple carbon sources simultaneously. These carbon sources include fatty acids and cholesterols, which are present in relative abundance inside macrophages. Interestingly, fatty acids and lipids may even be the preferred carbon sources of *M. tuberculosis* during both infection and dormancy.

1.2.2 Tricarboxylic acid cycle

The tricarboxylic acid (TCA) cycle is the main aerobic metabolic pathway to produce energy in almost all living organisms. The starting point of the TCA cycle is acetyl-coenzyme A (acetyl-CoA), which is produced by glycolysis or through the fatty acid spiral (Figure 1.1). The first step of the TCA cycle is the citrate synthase-catalysed reaction of oxaloacetate and acetyl-CoA to give citrate. Citrate is then isomerised by aconitase to give isocitrate via the intermediate cis-aconitate. The next step is the oxidative decarboxylation of isocitrate into 2-oxoglutarate (also known as α-ketoglutarate) by isocitrate dehydrogenase, which is then followed by another oxidative decarboxylation reaction, catalysed by α-ketoglutarate dehydrogenase, to give succinyl-coenzyme A (succinyl-CoA). Succinyl-CoA is then broken down to succinate, a reaction that is catalysed by succinyl-CoA synthetase. Succinate then undergoes further modifications. First, succinate is catalysed by succinate dehydrogenase into fumarate. Then, fumarate undergoes catalytic hydration by fumarase to give malate. Finally, malate dehydrogenase catalyses the oxidation of malate to generate oxaloacetate. The TCA cycle is the most important link to the electron transport chain in which adenosine triphosphate (ATP) is produced, but the cycle wastes two carbons in the form of carbon dioxide (Figure 1.1). Thus, this cycle is not ideal for energy production when external nutrients are scarce.
Figure 1.1: TCA cycle (green), glyoxylate cycle (blue) and methylcitrate cycle (red). The early steps of the glyoxylate cycle resembles the TCA cycle, in which acetyl-CoA is converted into D-isocitrate.

1.2.3 Glyoxylate cycle

The glyoxylate cycle is an alternative pathway to the TCA cycle.\textsuperscript{17, 18} The early steps of the glyoxylate cycle resemble those in the TCA cycle, in which acetyl-CoA is converted into isocitrate via citrate and \textit{cis}-aconitate. The major point of difference between these two cycles is the role of isocitrate.\textsuperscript{19} In the glyoxylate cycle, isocitrate is cleaved by isocitrate lyase to give succinate and glyoxylate. The second enzyme in the glyoxylate cycle, malate synthase, then
catalyses the formation of malate from glyoxylate and acetyl-CoA. Other TCA cycle intermediates, including oxaloacetate, are replenished by anaplerotic reactions from succinate and malate, products of the glyoxylate cycle. The advantage of bypassing the TCA cycle is that it helps to utilise all carbons available in nutrient deprived conditions by bypassing two decarboxylation steps of the TCA cycle (Figure 1.1).10

1.2.4 Methylcitrate cycle

During infection, the pools of carbon that M. tuberculosis can utilise include fatty acids from the host as well as the bacteria’s internal lipid reserves.20 Although natural animal fatty acids are composed of an even number of carbons, bacteria including mycobacteria possess the ability to synthesise odd-chain fatty acids.21 However, β-oxidation of odd-chain fatty acids is potentially harmful to the bacteria, because the process generates propionyl-coenzyme A (propionyl-CoA) and propionate,22-24 both of which are toxic.25-27 In addition to fatty acids, mycobacteria can also utilise cholesterol as a carbon source,28,29 the metabolism of which also yields propionyl-CoA and propionate as by-products.22-24

The methylcitrate cycle is an important pathway that bacteria could employ to transform toxic propionate and propionyl-CoA into non-toxic substances, succinate and pyruvate.30 The first step of the methylcitrate cycle is the methylcitrate synthase-catalysed conversion of propionyl-CoA and oxaloacetate to give methylcitrate. Methylcitrate is then isomerised to form methylisocitrate, a step that is catalysed by methylcitrate dehydratase. The final step of the methylcitrate cycle is the methylisocitrate lyase-catalysed conversion of methylisocitrate to give succinate and pyruvate (Figure 1.1).
1.3 Isocitrate lyase (ICL)

1.3.1 Isocitrate lyase in *M. tuberculosis*

Isocitrate lyase (threo-2R-isocitrate-glyoxylate lyase, E.C. 4.1.3.1, also known as isocitratase, isocitritase, isocitrase or isocitric lyase) is the first enzyme of glyoxylate cycle (for review, see reference 31). ICL is an Mg$^{2+}$-dependent enzyme that catalyses the reversible lysis of a C-C bond of d-isocitrate to form glyoxylate and succinate (Figure 1.2). ICL is present in bacteria (including mycobacteria), fungi and plants, but not in humans or animals. In *M. tuberculosis*, there are two known isoforms of ICL; ICL1 (428 amino acids (AAs)) and ICL2 (766 AAs), which are encoded by the genes *icl1* and *aceA* (also known as *icl2*), respectively. Exceptions can be found in some mycobacterial species including *M. tuberculosis* H37Rv, one of the most studied laboratory strains, in which the gene *aceA* is split into two open reading frames, *aceAa* and *aceAb*. It is not known whether these two genes (*aceAa* and *aceAb*) encode stable proteins (denoted as ICL2a and ICL2b respectively) and/or whether the proteins that these genes encode possess ICL activities. Sequence analyses indicate that although ICL1 and ICL2 share 27% sequence identity, ICL1 is a prokaryotic-like ICL isoform, whilst ICL2 is a eukaryotic-like ICL isoform. To date, the most studied isoform of *M. tuberculosis* ICLs is ICL1, as recombinant ICL2 was found to be unstable *in vitro*. ICL1 has a stronger affinity to the substrate d-isocitrate than ICL2 and is more active than ICL2 *in vitro*.

ICL plays an important role in the fatty acid metabolism of actively replicating *M. tuberculosis*. Genetic studies showed that the expression level of both *icl1* and *aceA* increased when *Mycobacteria* were grown in media containing palmitate or acetate (i.e. metabolites of fatty acids). *M. tuberculosis* mutants lacking both the *icl1* and *aceA* genes were found unable to grow and survive in media with fatty acid as the sole carbon source and a carbon mixture that contained fatty acids. However, the growth of the same mutant was not affected in media with carbohydrates as the carbon source. Interestingly, complementation of the same mutant with
plasmids carrying *icll* or *aceA* gene restored growth on fatty acids. These experiments suggest that ICL1 and ICL2 are essential for fatty acid metabolism of *M. tuberculosis* during the actively replicating state, although there may be some functional redundancy amongst these two enzymes in their fatty acid metabolism roles.

ICL also plays a functional role in the survival of *M. tuberculosis* during non-replicating persistent/latent infection. The gene that encodes ICL was found to be upregulated when *M. tuberculosis* was grown under hypoxic conditions and in growth-arrested bacilli in mice (to mimic persistent infection). ICL was also found to be essential for the survival of non-replicating *M. tuberculosis* in nutrient deprived (but oxygen-rich) media, as demonstrated by experiments using an *icll* deletion mutant of *M. tuberculosis* H37Rv, which also lacks the *aceA* gene. These experiments showed that ICL activity is essential for the survival of *M. tuberculosis* in the non-replicating dormant phase.

Unlike the functional redundancy of ICL1 and ICL2 in fatty acid metabolism during the actively replicating state of *M. tuberculosis*, the roles of ICL1 and ICL2 in the survival of *M. tuberculosis* in the non-replicating dormant phase are not well-understood. An early study using an *icll* deletion mutant of *M. tuberculosis* showed the mutant is unable to survive long-term within macrophages or cause virulence in mice, although the role of *aceA* was not addressed in that study. A follow-up study by the same authors found that both ICL1 and ICL2 were required for the survival of *M. tuberculosis* in macrophages and in mice by using a mutant with both *icll* and *aceA* deleted. The results were further supported by studies supplementing wild-type (WT) *M. tuberculosis* with an inhibitor that blocks both ICL1 and ICL2 (nitropropionate, see Section 1.5), in which the bacteria showed similar growth pattern in fatty acids to that of the *icll* and *aceA* double deletion mutant. In contrast to the earlier findings, single deletion of *icll* or *aceA* showed little effect on bacterial growth in macrophages and in mice. It is likely that during the persistent phase of infection, *M. tuberculosis*...
tuberculosis exists in a mixture of different metabolic states and both ICL1 and ICL2 are important for growth and survival. Further studies, for example using chemical probes that selectively inhibit ICL1 and ICL2, would be an informative strategy to elucidate the roles of ICL1 and ICL2 in latent M. tuberculosis at different metabolic states during persistent infection.

1.3.2 ICL in the M. tuberculosis glyoxylate cycle

ICL is the first enzyme of the glyoxylate cycle. As M. tuberculosis utilises fatty acids as a carbon source during both the actively replicating and the non-replicating latent phases, ICL and the glyoxylate cycle may be essential to the growth and survival of the bacteria. The upregulation of the both isocitrate lyase and malate synthase during persistent infection inside macrophages is well documented for many other bacteria. However, in M. tuberculosis, while the expression of icl1 was induced, the level of glcB (the gene that encodes malate synthase) was not significantly altered in the active state when grown in media containing fatty acids as the carbon source. The level of glcB was even found to be downregulated in the dormant state. This lack of coordination in the regulation of isocitrate lyase and malate synthase in the glyoxylate cycle is unique to M. tuberculosis and has not been found in other bacteria. There are proposals that succinate and glyoxylate formed as a result of ICL catalysis may have other roles rather than their canonical roles in fatty acid metabolism and gluconeogenesis. There may exist a glyoxylate-to-glycine shunt, the proposed main function of which is to replenish the pool of nicotinamide adenine dinucleotide (NAD) to support the minimal hypoxic metabolism in the bacterium’s latent state. However, such a pathway has not been fully characterised. It has also been shown that under hypoxic conditions, succinate, another product of ICL catalysis, may be the preferred respiratory electron donor to sustain succinate dehydrogenase activity and facilitates oxidative synthesis of ATP. Interestingly, a recent report showed that isocitrate dehydrogenase (IDH), the enzyme that competes with ICL for the substrate d-
isocitrate, may be regulated by glyoxylate in mycobacteria, as opposed to reversible phosphorylation that is commonly found in other bacteria such as E. coli. This finding is interesting because it suggests that there may exist a cross-regulation mechanism between the TCA and glyoxylate cycles in mycobacteria.

1.3.3 ICL in the M. tuberculosis methylcitrate cycle

Three enzymes, including methylcitrate synthase, methylcitrate dehydratase and methylisocitrate lyase are involved in the bacterial methylcitrate cycle. However, bioinformatics studies suggested that the genome of M. tuberculosis only encodes methylcitrate synthase and methylcitrate dehydratase, but not methylisocitrate lyase, even though M. tuberculosis can survive in a high propionate environment. In order to identify the enzyme that is responsible for the lysis of methylisocitrate in M. tuberculosis, its growth in media containing propionate was studied. Interestingly, ICL was found to be essential for the growth of M. tuberculosis in propionate, as the icl1 deletion mutant of M. tuberculosis cannot grow in propionate or cholesterol-containing media, despite the presence of saturated amounts of carbohydrates and fatty acids. A M. tuberculosis mutant that lacks both the icl1 and aceA genes cannot grow in media containing propionate as the sole carbon source. However, complementing the double deletion mutant with plasmids encoding icl1 and aceA may restore mycobacterial growth in propionate media. These experiments showed that ICLs likely play a role in both the glyoxylate and methylcitrate cycles in M. tuberculosis.

In addition, in vitro experiments with purified ICL1 and ICL2 proteins showed that both enzymes could use isocitrate and methylisocitrate as substrates (Figure 1.2a), although isocitrate is the preferred substrate over methylisocitrate for both ICL1 and ICL2. In order to understand the substrate selectivity of ICL, crystallisation of the C191S variant of ICL1, in
which the catalytic cysteine residue in the active site was mutated to a serine residue in order to slow down the catalytic activity of the enzyme, was attempted in the presence of methylisocitrate.\textsuperscript{36} Interestingly, in the resulting crystal structure the reaction products, pyruvate and succinate, were observed inside the enzyme active site instead of methylisocitrate, indicating the mutant was still active during crystallisation. Nonetheless, the results showed that ICL1 could accommodate the methyl group of methylisocitrate in the hydrophobic pocket of the enzyme, further supporting the proposal that ICL may play a dual role in the survival of \textit{M. tuberculosis} in a high fatty acid content environment that is characteristic of both active and persistent infections.

\textbf{1.3.4 Additional role of ICL in \textit{M. tuberculosis} for antibiotic resistance}

Recently, ICL was linked to the development of antibiotic resistance in \textit{M. tuberculosis}. By using both metabolomics and gene expression analyses, it was demonstrated that ICL was activated when \textit{M. tuberculosis} were subjected to sub-lethal doses of three different anti-TB drugs; rifampicin, streptomycin or isoniazid.\textsuperscript{48} \textit{M. tuberculosis} with an \textit{icl1} deletion showed a 100-fold increase in sensitivity to these antibiotics.\textsuperscript{48} Furthermore, studies that used a \textit{M. tuberculosis} mutant complemented with a K189E variant of ICL1, whose mutation kept the enzyme in an open conformation and hence reduced its catalytic activity, showed increased sensitivity towards rifampicin and streptomycin, but not to isoniazid.\textsuperscript{49} These experiments suggest that the catalytic activity of ICL may play a role in mycobacterial defence against (at least some) antibiotics, although the exact molecular mechanism is not clear. Interestingly, when the \textit{icl1} deletion mutant was co-incubated with antioxidants including thiourea or 2,2,6,6-tetramethylpiperidin-1-yl)oxyl (Tempol), the heightened susceptibility towards the antibiotics was reduced.\textsuperscript{48} The results suggest that ICL may play a role in the metabolic defence against antibiotic-induced oxidative stress. This discovery is significant because long-term treatments
and increasing antibiotic resistance are two of the biggest challenges preventing the eradication of TB. These recent studies increase the appeal of ICL as a target in anti-TB drug discovery programmes.

1.4 Structural and mechanistic aspects of ICL

1.4.1 Catalytic mechanism

The exact mechanism by which ICL converts isocitrate into glyoxylate and succinate is not fully understood, but a retro Claisen-type condensation pathway has been proposed (Figure 1.2b). The first step involves deprotonation of the isocitrate hydroxyl group followed by fragmentation of the isocitrate to form glyoxylate and succinate. Mutagenesis and bioinformatics studies showed that the highly conserved KKCGH sequence motif at the enzymes’ active site (residues 189-193 in ICL1, also found in ICLs from other organisms) is essential for the catalytic activity of the enzyme. Three residues are proposed to be involved in the catalytic cycle, including the cysteine and histidine residues in the conserved KKCGH sequence motif, along with an unidentified residue that is close to the substrate at the active site. The catalytic cysteine residue (Cys191 in M. tuberculosis ICL1) serves as a general acid to aid the formation of the succinate co-product, whilst the histidine residue (His193 in ICL1) interacts with the catalytic cysteine, decreasing its pKₐ and aiding the formation of the aci-carboxylate intermediate that forms upon C(2)-C(3) bond cleavage. The unidentified residue presumably acts as a general base to deprotonate the C(2) hydroxyl group of the isocitrate substrate. In ICL1, this residue is proposed to be either an arginine (Arg228), tyrosine (Tyr89) or histidine (His180).
Figure 1.2: (a) ICL catalyses the reversible conversion of isocitrate or methylisocitrate to succinate and glyoxylate or pyruvate; (b) Consensus mechanism for ICL catalysis. An unidentified residue acts as a general base to deprotonate the hydroxyl group of isocitrate. In ICL1, this residue was proposed to be Arg-228, Tyr-89 or His-180. The catalytic cysteine (Cys-191) serves as a general acid to aid the formation of the succinate co-product. The histidine residue (His-193) interacts with the catalytic cysteine to aid the formation of the aci-carboxylate intermediate.
1.4.2 Role of a divalent cation in the catalytic mechanism of ICL

Divalent metal ions play an important role in the catalytic activity of both ICL1 and ICL2.32 The substrate d-isocitrate binds to ICL by chelating the active site metal ion via the C(1) carboxylic acid group and C(2) hydroxyl group. Mg$^{2+}$ is required for optimal catalytic activity, whereas replacement of Mg$^{2+}$ by Mn$^{2+}$ results in a 60% loss of catalytic activity. Other divalent cations like Co$^{2+}$, Fe$^{2+}$, Ca$^{2+}$, Ba$^{2+}$, Ni$^{2+}$, Cd$^{2+}$, Zn$^{2+}$, Cu$^{2+}$ and Hg$^{2+}$ are not effective substitutes for Mg$^{2+}$, and negligible activities were detected in the absence of divalent cations.

1.4.3 Structural aspects of ICL

To date, only the crystal structure of *M. tuberculosis* ICL isoform 1 was solved [PDB IDs: 1F61 (apo-ICL), 1F8I (nitropropionate and glyoxylate complex), 1F8M (bromopyruvate complex)]$^{50}$ and 5DQL (2-vinyl isocitrate complex)$^{54}$. ICL2 was found to be unstable and no structural information is available.$^{32,55}$ ICL1 crystallised as either a dimer or tetramer, with each monomer subunit containing 14 $\alpha$-helices and 14 $\beta$-strands. The core of ICL1 contains 8 $\alpha$-helices and 8 $\beta$-strands forming a $\alpha/\beta$-barrel-like structure. The highly conserved KKCGH sequence motif extends from $\beta$5 of the $\alpha/\beta$-barrel forming a loop structure (residues 185-196) and connects to a small $\beta$-sheet domain that consists of five short $\beta$-strands ($\beta$6, $\beta$7, $\beta$9, $\beta$10, $\beta$11). At the C-terminal, $\alpha$12-$\alpha$14 project away from the barrel and interact with nearby subunit (Figure 1.3a).

When comparing the crystal structure of apo-ICL1 and the holo-forms with glyoxylate, nitropropionate (a mimic of the product succinate), or bromopyruvate (an inhibitor of ICL), conformational changes are observed (Figure 1.3b).$^{50}$ In particular, large changes are apparent around the active site loop region that includes the catalytically-essential sequence motif KKCGH (residues 185-196 of ICL1), and the C-terminal end (residues 411-428 of ICL1) of the adjacent subunit. In the open conformation, the catalytic cysteine (Cys191) is relatively far
away from other active site residues, thus allowing the substrate (and other solvent molecules) to access the binding site. Upon binding of the substrate, the enzyme adopts a closed conformation and access to the active site is then blocked by the active site loop (residues 185-196). The C-terminal end of the adjacent subunit then moves to the space that was opened up due to the closure of the active site loop, forming a lid. Electrostatic interactions between positively charged lysine residues at the active site loop (Lys189 and Lys190) and the negatively charged glutamate residues at the C-terminal lid (Glu423 and Glu424) are important to keep the enzyme in its catalytically active conformation. Disrupting these interactions may lead to a reduction in the enzyme’s catalytic activity.

Interestingly, His180, one of the residues that was proposed to act as a general base during the catalytic cycle of ICL1, is also an important residue for the enzyme’s catalytic activity and oligomerisation state. Mutagenesis studies showed that ICL1 H180A is a monomer that is catalytically inactive. Molecular dynamics calculations showed that loss of interactions between His180 and Tyr89 (another residue that was proposed to be important for ICL1 activity) may cause changes in the orientation of α6 and the C-terminal helices (α12-α14), thus affecting the protein’s ability to interact with other monomer subunits to form an oligomer. The loss of such interactions also affects the stability of ICL1’s catalytically-active conformation, as the active site loop is no longer being held together by the lid of the adjacent subunit in the tetramer.
**Figure 1.3:** (a) Crystal structure of apo-ICL1 (PDB id: 1f61), which crystallised as a dimer (subunit A: blue and subunit B: orange). Four α-helices are involved in the dimerisation between the two subunits; (b) Conformational change was observed when ICL1 was bound to glyoxylate and nitropropionate (PDB ID: 1F8I; subunit A: red; subunit B: wheat; only two subunits shown). The active site loop was found to be in the close conformation in the presence of glyoxylate and nitropropionate (insert: blue to red), and the C-terminal of the adjacent subunit move over to the space that was previously occupied by the active site loop (insert: orange to wheat).
It has long been proposed that bacterial and plant ICLs may be switched on and off by reversible acetylation and succinylation respectively,\textsuperscript{57-59} although the exact molecular and structural basis of such control is not known. Recently, two post-translational modifications were found on \textit{M. tuberculosis} ICL, including lysine acetylation and lysine succinylation.\textsuperscript{49, 60, 61} Interestingly, one of the succinylated lysine residues (Lys189) is located in the catalytically-essential conserved sequence motif (KKCGH).\textsuperscript{49} Although it is not possible to recombinantly produce (at least in large quantity) succinylated-ICL for structural and mechanistic studies, the effect of lysine succinylation may be mimicked by a lysine to glutamate mutation. Structural and kinetic studies using K189E ICL showed that replacing the positively-charged lysine with the negatively-charged glutamate (a similar effect to lysine succinylation) may keep the enzyme in the open form, thus rendering the enzyme catalytically inactive. The regulation of the TCA and glyoxylate cycles are important for the survival of \textit{M. tuberculosis} because both cycles take \textit{d}-isocitrate as an intermediate.\textsuperscript{19} The presence of a potential feedback mechanism involving reversible succinylation of the active site lysine in ICL1 to activate and deactivate the glyoxylate cycle is not surprising given the tight control that is required to regulate these two closely related cycles. This finding, together with another recent study that showed glyoxylate may positively control the activity of \textit{M. tuberculosis} isocitrate dehydrogenase (IDH),\textsuperscript{62} provide insights into the regulation between these two cycles and may enable the development of new anti-TB drugs that can selectively activate and deactivate ICL and IDH at different stages of infection.

In addition to succinylation, Bi and co-workers reported that lysine acetylation may regulate the activities of \textit{M. tuberculosis} ICL1 by using recombinant ICL1 that was acetylated at K392 (K392Ace ICL1) and K332 (K332Ace ICL2).\textsuperscript{63} A two-fold increase in isocitrate lyase activity was observed with K392Ace ICL1 when compared to the WT, although K392 acetylation did not alter methylisocitrate lyase activity. K322 acetylation was found to be deactivating, with a
2-fold decrease in ICL and methylisocitrate lyase activities observed. To demonstrate the physiological relevance of this finding, the authors conducted cell-based experiments to compare the level of expression and activity of non-acetylated ICL1, K392Ace ICL1 and K332Ace ICL1 in cells that were grown in media containing glucose, acetate or propionate at different concentrations. In agreement with previous studies, the level of expression of non-acetylated ICL1, K322Ace ICL1 and K392Ace ICL1 increased when the cells were grown in acetate or propionate. It was found that the level of K322 acetylation positively correlated to the concentration of acetate or propionate that was being used. However, the expression level of the K392Ace ICL1 was stable in all concentrations of acetate and propionate. Interestingly, over an extended period of time (55 days), the expression level of non-acetylated ICL1 and K332Ace ICL1 dropped within 10 days, whilst the level of K392Ace remained stable throughout the period. Side by side, the authors showed that there was a 1.7 fold increase in ICL activity of the cell extracts from cells that were grown in propionate than those that were grown in glucose. Overall, it was postulated that the post translational acetylation of K392 may increases the activity and abundance of ICL1 in the stationary phase of *M. tuberculosis*.

1.5 Inhibition of mycobacterial ICL as potential anti-TB therapy

1.5.1 Inhibitors of ICL

Given the central role ICL plays in the glyoxylate and methylisocitrate cycles, ICL is a current inhibition target for antimicrobial applications including (but not limited to) latent TB. However, despite considerable effort has been spent in both academia and industry, no compounds have progressed through to the clinical trial stage. There are three major challenges in targeting ICLs, including; (1) the polar nature of the ICL binding pocket; (2) the small size of the natural substrates, and; (3) the need to target both ICL1 and ICL2.
The polar nature of the ICL binding pocket favours small, polar molecules. It is therefore not surprising that early inhibitor discovery work that involved plant, fungal and bacterial ICLs have all resulted in compounds that are analogues of the substrate isocitrate, succinate or glyoxylate. These included itaconate\(^{64}\), nitropropionate\(^{65}\), and bromopyruvate\(^{66}\). Both itaconate and nitropropionate are succinate analogues and non-covalent inhibitors of ICL,\(^{64, 65}\) but bromopyruvate, a glyoxylate analogue, was found to covalently bind to ICL via the catalytic cysteine residue at the active site.\(^{66}\) The first inhibition study of \(M.\) \(tuberculosis\) ICL1 and ICL2 was conducted by Bentrup \textit{et al.}\(^{32}\) In that study, the inhibition constant \((K_i)\) of several ICL inhibitors (itaconate 1, nitropropionate 2 and bromopyruvate 3) against both ICL1 and ICL2 were determined (Table 1.1). However, while these compounds are relatively potent inhibitors against \(M.\) \(tuberculosis\) ICL, they are non-selective and possess high toxicity, rendering them unsuitable as potential drug candidates.\(^{67-71}\) For example, nitropropionate also inhibits succinate dehydrogenase,\(^{72}\) a TCA cycle enzyme that is essential for the survival of humans.\(^{45, 46}\)

Interestingly, some ICL inhibitors were found to be isoform selective. These include itaconic anhydride 4, which was found to be a stronger inhibitor of ICL1 \((K_i\) value = 190 \(\mu\)m) than ICL2 \((480 \mu\)M).\(^{32}\) Moreover, oxalate and malate were also found to be isoform selective inhibitors, which inhibited 50% of ICL1 at 5 mill molar (mM) concentration and did not inhibit ICL2.\(^{32}\) These experiments suggested that the active site of ICL1 and ICL2 may not be entirely identical, or there may exist allosteric inhibition sites on ICL1 and ICL2.

High throughput screening (HTS) has been applied to develop ICL inhibitors.\(^{55, 73-78}\) The most notable HTS studies were conducted by GlaxoSmithKline (GSK) and the Global Alliance for TB Drug Development (GATB), who screened 900,000 compounds, and the Tuberculosis Antimicrobial Acquisition and Coordinating Facility (TAACF), who screened just under 101,000 compounds. The TAACF screen was moderately successful.\(^{76}\) For instance, 4-(2-
pyridyl)-2-thiazolamines were identified as a potential new scaffolds for ICL inhibitors. In addition, heterocyclic benzopyranones were also identified as ICL inhibitors. As there are very few previous reports of compounds containing these scaffolds as anti-TB drug candidates, they constitute interesting starting points for future medicinal chemistry programmes. Recently, a library of compounds was screened and found methyl-4-(4-methoxyphenyl)-4-oxobut-2-enoate 5 inhibits ICL1 with $K_i$ of 1.85 μM. However, this compound shows cytotoxicity to mouse liver and kidney cells. In another study, Ji and co-workers screened several Mannich base compounds. In that study, 3-((1-(4-chlorophenyl)-5-(6-methoxynaphthalen-2-yl)-3-oxopentyl)amino)benzoic acid 6 was found to be a strong ICL1 inhibitor with a half maximal inhibitory concentration (IC$_{50}$) of 0.0535 μM. The challenge for HTS is the relatively small size of the ICL1 binding pocket as the number of compounds that can fit the binding pocket is limited, which also restricts the scope of any structure-activity relationship study. None of the compounds from HTS were tested against ICL2. Besides HTS, several academic research groups have tried to develop ICL1 inhibitors using other approaches, but none of the molecules were tested against ICL2. Vinšová and co-workers have focused on synthesising derivatives of salicylanilides, including salicylanilide benzoates, salicylanilide pyrazinoates, salicylanilide diethyl phosphates, salicylanilide $N,N$-substituted carbamates and thiocarbamates. These compounds were tested for inhibition of ICL1. At 100 μM concentration, 5-chloro-2-hydroxy-$N$-(4-(trifluoromethyl)phenyl)benzamide 7 and 4-chloro-2-((4-(trifluoromethyl)phenyl)carbamoyl)phenyl pyrazine-2-carboxylate 8 showed ~59% inhibition of ICL1. At 10 μM concentration, 4-bromo-2-((4-(trifluoromethyl)phenyl)carbamoyl)phenyl acetyl-L-phenylalaninate 9 and (S)-4-bromo-2-((4-fluorophenyl)carbamoyl)phenyl diphenylcarbamate 10 showed ~22% inhibition of ICL1. However, these compounds were cytotoxic and were therefore not suitable to progress as drug candidates. The same group also explored 2-methoxybenzanilide, 2-methoxy-2′-
hydroxybenzanilide derivatives and their thioxo analogues as potential inhibitors against ICL1. At 10 μM concentration, 4-chloro-N-(2-hydroxy-4-(trifluoromethyl)phenyl)-2-methoxybenzothioamide showed ~23% inhibition against ICL1. In another study, the group of Sriram synthesised four classes of molecule such as phthalazinyl derivatives, 5-nitro-2-furoic acid hydrazones, 5-nitro-2,6-dioxohexahydro-4-pyrimidinocarboxamides and 3-nitropropionamides. These compounds were then tested against ICL1. A phthalazinyl derivative 2-(3-(4-bromo-2-fluorobenzyl)-4-oxo-3,4-dihydropthalazin-1-yl)-N’-(3-nitrobenzylidene)acetohydrazide showed ~62% inhibition of ICL1 at 100 μM. 5-Nitro-2-furoic acid hydrazones derivatives, including 5-nitro-N’-((3-nitrofuran-2-yl)methylene)furan-2-carbohydrazide and N’-((1-(4-bromophenyl)ethylidene)-5-nitrofuran-2-carbohydrazide, showed 89 and 73% ICL1 inhibition respectively. Several 3-nitropropionamide derivatives also showed inhibitory action against ICL1. These include 4-cyclopropyl-6-(3,5-dimethyl-4-(3-nitropropanoyl)piperazin-1-yl)-7-fluoro-5-methoxy-1-oxo-1,4-dihydronaphthalene-2-carboxylic acid (IC<sub>50</sub> = 0.10 μM) and 4-cyclopropyl-7-fluoro-5-methoxy-6-(4-(3-nitropropanoyl)piperazin-1-yl)-1-oxo-1,4-dihydronaphthalene-2-carboxylic acid (IC<sub>50</sub> = 0.12 μM). The flavonoid, quercetin, showed ICL1 inhibition with an IC<sub>50</sub> of 3.57 μM. However, quercetin is not being progressed due to poor penetration of the mycobacterial cell wall as these compounds were highly polar with multiple hydroxyl groups present. In addition to small molecule inhibitors, Yin et al. and Liu et al. used molecular docking to guide the design of peptide-based ICL1 inhibitors. This is a potential method to bypass the polar active site as peptides target the oligomerisation interface of ICL1, which is catalytically inactive in the monomeric form. Although the best reported peptidyl inhibitor, a heptapeptide, was only moderately potent in vitro (IC<sub>50</sub> = 126 μM), this demonstrates that the oligomerisation interface could be a potential inhibition target.
Recently, Pham et al developed an isocitrate analogue inhibitor, 2-vinyl-$\alpha$- isocitrate 19, which is a so-called mechanism-based inhibitor. The molecule was found to be isoform selective, which inhibits ICL1 strongly (interaction constant, $K_{\text{inact}} = 22 \mu\text{M}$) than ICL2 ($K_{\text{inact}} = 420 \mu\text{M}$). The mechanism of inhibition relies on the catalytic conversion of isocitrate to succinate and glyoxylate. The nucleophile Cys191 attacks the methylene group of 2-vinyl glyoxylate to form a homopyruvate adduct and thus inactivates the enzyme. The advantages of this approach are fourfold. Firstly, it forms a irreversible or slow reversible covalent adduct at the active site. Secondly, it reacts with Cys191 at the substrate binding site that is less prone to mutation thus lead to resistance. Thirdly, it inhibits both ICL1 and ICL2. Fourthly, it did not show cytotoxicity levels at above 400 $\mu\text{M}$. However, it inhibits isocitrate dehydrogenase with $IC_{50}$ of 10 $\mu\text{M}$ and also its inhibition activity decreased by glyoxylate cycle intermediates $\alpha$-malate, glyoxylate and succinate.
<table>
<thead>
<tr>
<th>Name of inhibitors</th>
<th>Structure</th>
<th>$K_i$/$K_{\text{inact}}$/$IC_{50}$</th>
<th>% inhibition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itaconate</td>
<td><img src="image1.png" alt="Structure" /></td>
<td>$K_i = 120 \mu M$</td>
<td>$K_i = 220 \mu M$</td>
</tr>
<tr>
<td>Nitropropionate</td>
<td><img src="image2.png" alt="Structure" /></td>
<td>$K_i = 3 \mu M$</td>
<td>$K_i = 110 \mu M$</td>
</tr>
<tr>
<td>Bromopyruvate</td>
<td><img src="image3.png" alt="Structure" /></td>
<td>$K_i = 120 \mu M$</td>
<td>$K_i = 710 \mu M$</td>
</tr>
<tr>
<td>Itaconic anhydride</td>
<td><img src="image4.png" alt="Structure" /></td>
<td>$K_i = 190 \mu M$</td>
<td>$K_i = 480 \mu M$</td>
</tr>
</tbody>
</table>

Table 1.1: Inhibitors of ICL
<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>![Chemical Structure]</th>
<th>IC$_{50}$</th>
<th>% inhibition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methyl-4-(4-methoxyphenyl)-4-oxobut-2-enoate</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-((1-(4-Chlorophenyl)-5-(6-methoxynaphthalen-2-yl)-3-oxopentyl)amino)benzoic acid</td>
<td>6</td>
<td>IC$_{50}$ = 0.0535 µM</td>
<td></td>
</tr>
<tr>
<td>5-Chloro-2-hydroxy- N-[4--(trifluoromethyl)phenyl]benzamide</td>
<td>7</td>
<td>% inhibition = 59 (100 µM)</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1.1: Inhibitors of ICL*
<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Structure</th>
<th>% Inhibition</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-chloro-2-((4-(trifluoromethyl)phenyl)carbamoyl)phenyl pyrazine-2-carboxylate</td>
<td><img src="image" alt="Structure 8" /></td>
<td>59 (100 µM)</td>
</tr>
<tr>
<td>(S)-4-bromo-2-[4-(trifluoromethyl)phenylcarbamoyl]phenyl 2-acetamido-3-phenylpropanoate</td>
<td><img src="image" alt="Structure 9" /></td>
<td>22 (10 µM)</td>
</tr>
<tr>
<td>4-bromo-2-((4-fluorophenyl)carbamoyl)phenyldiphenylcarbamate</td>
<td><img src="image" alt="Structure 10" /></td>
<td>22 (10 µM)</td>
</tr>
</tbody>
</table>

Table 1.1: Inhibitors of ICL
<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>% Inhibition</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-chloro-2-methoxy-N-(4-(trifluoromethyl)phenyl)benzothioamide</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>23% (10 µM)</td>
</tr>
<tr>
<td>2-(3-(4-bromo-2-fluorobenzyl)-4-oxo-3,4-dihydrophthalazin-1-yl)-N'-(3-nitrobenzylidene)acetohydrazide</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>62% (100 µM)</td>
</tr>
<tr>
<td>5-nitro-N'-(3-nitrofuran-2-yl)methylene)furan-2-carbohydrazide</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>89% (100 µM)</td>
</tr>
</tbody>
</table>

*Table 1.1: Inhibitors of ICL*
<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Chemical Structure</th>
<th>IC₅₀ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N'-(1-(4-bromophenyl)ethylidene)-5-nitrofuran-2-carbohydrazide</td>
<td><img src="image1" alt="14" /></td>
<td>% inhibition = 73 (100 µM)</td>
</tr>
<tr>
<td>4-Cyclopropyl-6-(3,5-dimethyl-4-(3-nitropropanoyl)piperazin-1-yl)-7-fluoro-5-methoxy-1-oxo-1,4-dihyronaphthalene-2-carboxylic acid</td>
<td><img src="image2" alt="15" /></td>
<td>IC₅₀ = 0.1 µM</td>
</tr>
<tr>
<td>4-Cyclopropyl-7-fluoro-5-methoxy-6-(4-(3-nitropropanoyl)piperazin-1-yl)-1-oxo-1,4-dihyronaphthalene-2-carboxylic acid</td>
<td><img src="image3" alt="16" /></td>
<td>IC₅₀ = 0.12 µM</td>
</tr>
<tr>
<td>Quercetin</td>
<td><img src="image4" alt="17" /></td>
<td>IC₅₀ = 3.57 µM</td>
</tr>
</tbody>
</table>

Table 1.1: Inhibitors of ICL
<table>
<thead>
<tr>
<th>Heptapeptide</th>
<th>18: His-Tyr-Ser-Ser-His-Met-Arg</th>
<th>IC$_{50}$ = 126 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Vinyl-δ-isocitrate</td>
<td><img src="image" alt="Chemical Structure" /></td>
<td>$K_{inact} = 22$ µM</td>
</tr>
</tbody>
</table>

Table 1.1: Inhibitors of ICL
1.6 Overview of this Thesis

1.6.1 General overview

The main objective of the work reported in this thesis was to study the role of ICL in the fatty acid metabolism of *M. tuberculosis*, specifically its role in the glyoxylate and methylcitrate cycles. A particular focus will be on ICL2, the lesser studied isoform. A second objective of this work was to synthesise valuable nitrogen-heterocycles from itaconic acid, a weak inhibitor of ICL.

1.6.2 Specific overview

**Chapter 2:** Molecular biology techniques were applied for the production and purification of four isoforms of isocitrate lyase; ICL1, ICL2, ICL2a and ICl2b.

**Chapter 3:** An NMR-based method was developed to study the kinetics and inhibition of ICL by using ICL1 as a model system. The use of the thermal shift method for high-throughput screening of potential ICL1 ligands was also described.

**Chapter 4:** Protein X-ray crystallography was used for structure determination of ICL2. The activity of ICL2 was investigated by NMR.

**Chapter 5:** Itaconic acid was investigated as a green starting material for organic and medicinal chemistry.

**Chapter 6:** Experimental procedures are described.
Chapter 2: Recombinant production and purification of *Mycobacterium tuberculosis* isocitrate lyase

2.1 Introduction

There are two known isoforms of ICL in *M. tuberculosis* (see Chapter 1). These are ICL1 (428 AAs; for protein sequence see Table 2.1) and ICL2 (766 AAs; Table 2.2). In *M. tuberculosis* H37Rv, an open reading frame mutation was found in the gene *aceA*, which encodes ICL2. This splits the gene into two open reading frames, *aceAa* and *aceAb*, which encodes ICL2a (367 AAs) and ICL2b (398 AAs) respectively (Table 2.3). It is not known whether ICL2a or ICL2b are stable proteins and/or whether these proteins possess ICL activities.

| MSVVGTPKSAEQIQQEWDNTPRWKDVTRTYSAEDVVALQGSVVEEQHTLARRGAEVLWEQQLHDLEWVNALG |
| ALTGNMAVQQRAGLKIYLSGWQVAGDALSNGHTYDQSLYPANSVFQVVRINNALQRADQIAKIGD |
| TSVENLAPIVADGEAGFGGALNYELQKALIAAGVAGSHEDQLASEKCGHLGKVLIPTQHIRTLET |
| SARLAADVADVPTVIIARTDAEAATLITSDVDERDQPFIETGRTREGFYRTKNGIEPCIARAKAYAPFAD |
| LIWMTGTPDLEAAQFSEAIVKAEYPDQMLAYNCSPSFNWKHLDDATIAKFKELAAMGFKFQFITLAG |
| FHALNYSMFDLAYGVAQNSYAGHELQEREFAEEERGYTAKTHQREVAGYFDRIATTVPNSSTTALTG |
| STEEGQFH |

Table 2.1: Translated amino acid sequence of ICL1.
Table 2.2: Translated amino acid sequence of ICL2.

| a) | MAIAETDTEVHTPFEQDFEKDV AATQRYFDSSRFAGIIRLYTARQVVEQRG TIPVDHVAREAAAGAFYERLRELFAARKSITTGFYPSPGQAVSMKR MGEAIYLGWWATS A KGSSTE DPGDLAS YPLSQVPDDAAVLVRALLTADRNQHYLRQSMERQRAATP AYDFRPIIADACTGHGDPHVRNLIRFQVEVGVPYHIEDQGPGTKKCGHQGKGVL VPSDEQIKRLNAARPQLDIMR VQVI IVARTDAEANLIDS RADERQPDFFQLGATKL DVPYSYKSCFLAMVRRFTNWASRSSMVIFS MLATASTRPA VGLSAKAFSAWSPTRSTRGRTASSRSTAFFSTRSSRGWRFGRTTRA |

| b) | MTYGEADVLEFGQSEGEPI MAPEEWRFAARSLHAARAKA KELGDPP WDCELAKTEGYYQIRGGIPYAIASKLAAAPFADILW METKTADLADARQF AEAIHEEPQMLAYNLSPF NWDRTTGMTEEM RFP EELKMGVF NHYGHQIDGVAEEAFRALQDGMLARL QRMRLVESPYRTFQTL GGFRSDAAL AASSGR TATMKMGKSTHQHQLVQTEVPRLLEEELAMLWSGH YQLDKLVRQLRPQAGSEVLELHG |

Table 2.3: Translated amino acid sequence of a) ICL2a and; b) ICL2b.

*M. tuberculosis* ICL1 and ICL2 have been produced recombinantly in *E. coli* before. In 1999, Bentrup and co-workers reported the production of stable and active recombinant mycobacterial ICL1 with a N-terminal polyhistidine tag based on the sequence of *M.*
*tuberculosis* CSU93. The authors amplified the gene using polymerase chain reaction (PCR) and ligated the gene that encodes ICL1 to the pET based p6HisF-11d vector and transformed to *E. coli* HB101 for recombinant protein production. Liu and co-workers reported the production of ICL1 with a N-terminal polyhistidine tag by using *E. coli* BL21 (DE3). The gene was cloned into pET28a. In both cases, the protein was purified by using metal affinity chromatography. The groups of Sacchettini and Meek also reported the production of ICL1 without a purification tag by using *E. coli* BL21 (DE3). In both cases, ion exchange chromatography and size exclusion chromatography were used for protein purification.

The first production of the mycobacterial ICL2 enzyme with a N-terminal polyhistidine tag by using *E. coli* as a host was reported by Bentrup and co-workers. The ICL1 expression condition was applied to produce ICL2. The authors reported ICL2 as an unstable protein. Protein denaturation was observed when the protein was stored at -80 °C in a piperazine-$N,N'$-bis(2-ethanesulfonic acid (PIVES)-based buffer at pH 8.0 in the presence of 100 mM NaCl and 300 mM imidazole. In 2006, Sacchettini and co-workers reported the production ICL2 without a purification tag by using *E. coli* BL21 (DE3). The protein was purified by ion exchange chromatography and size exclusion chromatography. The protein was stored in a tris(hydroxymethyl)aminomethane (Tris)-based buffer at pH 7.5 in the presence of 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT) and 100 mM NaCl. The authors did not comment on the stability of the protein. In 2017, Meek and co-workers documented the production of stable truncated ICL2 (residues 1-605) with a C-terminal polyhistidine tag by using *E. coli* BL21 (DE3) as a host. The protein was purified by affinity chromatography.
2.2 Objectives

Given the importance of ICL in the fatty acid metabolism of *M. tuberculosis*, the production of highly purified, stable and active ICL at high yield will enable medicinal chemistry efforts to develop inhibitors that target ICL. The objective of the work described in this Chapter is therefore to find the optimal conditions for the production, purification and storage of ICL1, ICL2, ICL2a and ICL2b.

2.3 Production and purification of ICL1

2.3.1 Cloning

The recombinant production of ICL1 was first optimised. The procedure largely followed the method that reported by Savitsky *et al.*\(^6\) and Gileadi *et al.*\(^7\) In order to aid the production and purification of the recombinant protein, the pNIC28-Bsa4 vector (Figure 2.1a) was chosen so that the protein will include a N-terminal polyhistidine tag for affinity purification and a tobacco etch virus (TEV) protease cleavage site for the removal of the purification tag.

Next, a synthetic DNA fragment (gBlocks from Integrated DNA Technologies) encoding *M. tuberculosis* ICL1 was designed. Due to the high repeat density near the 3’ end of ICL1 (Table 2.4), the DNA sequence was codon optimised for the ease of DNA synthesis.\(^8\) In order to clone it onto the pNIC28-Bsa4 vector,\(^6,7\) the tacttccaatccatg sequence was added to the 5’ end and cagtaaaggtggata was added to the 3’ end of the codon-optimised DNA sequence encoding *M. tuberculosis* ICL1 (Table 2.4).

The cloning method for pNIC28-Bsa4 is ligase independent cloning (LIC), which does not require DNA ligase (Figure 2.1b). The pNIC28-Bsa4 vector was first linearised by using the BsaI restriction enzyme, which recognises the sequence GGTCTC(N), in which N is any nucleotide. The uncut vector possessed a *Bacillus subtilis* sacB gene (Figure 2.1a) that encodes the enzyme levansucrase. The expression of levansucrase in the presence of sucrose is lethal to
E. coli, which can be used as a negative selection marker. Next, complementary cohesive overhangs were created by treating the opened vector with T4 DNA polymerase in the presence of deoxyguanosine triphosphate (dGTP). The overhangs were created using the 3’→5’ exonuclease activity of T4 DNA polymerase. The same procedure, this time with deoxycytidine triphosphate (dCTP), was used to generate the complementary cohesive ends with the synthetic DNA fragment (Figure 2.1b). The recombinant plasmid was then transformed to XL10-Gold competent cells. The negative selection of the uncut vector was achieved by adding 5% sucrose to the culture media. Following a plasmid minipreparation, the sequence of the plasmid was confirmed by DNA sequencing (DNA Sequencing Centre, The University of Auckland). The correct plasmid was then used to transform BL21 (DE3) competent cells for protein expression.
Table 2.4: The synthetic gene fragment of *icl1* used in this study. Sequences *taacttcaatcc* and *cagtaaaggtggata* were added to the 5’ and 3’ ends respectively for cloning to the vector pNIC28-Bsa4.
Figure 2.1: a) Plasmid map of the vector pNIC28-Bsa4; b) Ligase-independent cloning strategy of the synthetic DNA fragment that encodes *icl1* to the vector pNIC28-Bsa4. Red represents the 3' end comprising of amino acid sequences.
2.3.2 Protein expression trials

The expression levels of the proteins in the transformed BL21 (DE3) cells were then evaluated. As the plasmid has a T7-promotor, isopropyl β-D-1-thiogalactopyranoside (IPTG) could be used to promote the production of the recombinant protein. Three different IPTG concentrations (0.2, 0.5 and 1.0 mM), three different incubation temperatures after induction (18, 28 and 37 °C) and two different incubation times (three hours and 16 h) were tested. Using denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 2.2), ICL1 was found to be expressed and soluble under all the tested conditions. The optimal temperature selected for large-scale production of ICL1 was 18 °C and the optimal IPTG concentration was 0.2 mM. The optimal incubation time was 16 h.

![SDS-PAGE showing the expression trial of ICL1 (MW = 49.64 kDa) at 18 °C with three different IPTG concentrations (0.2, 0.5 and 1.0 mM) and 16 h incubation. First lane represents a protein ladder. Lanes 2 and 3 represent whole cell and cell lysate at 0.2 mM IPTG concentration. Lanes 4 and 5 represent whole cell and cell lysate at 0.5 mM IPTG concentration. Lanes 6 and 7 represent whole cell and cell lysate at 1.0 mM IPTG concentration.](image-url)

**Figure 2.2:** SDS-PAGE showing the expression trial of ICL1 (MW = 49.64 kDa) at 18 °C with three different IPTG concentrations (0.2, 0.5 and 1.0 mM) and 16 h incubation. First lane represents a protein ladder. Lanes 2 and 3 represent whole cell and cell lysate at 0.2 mM IPTG concentration. Lanes 4 and 5 represent whole cell and cell lysate at 0.5 mM IPTG concentration. Lanes 6 and 7 represent whole cell and cell lysate at 1.0 mM IPTG concentration.
2.3.4 Large-scale production and purification

The purification of the recombinant ICL1 was then optimised. Two column chromatography methods, immobilised metal affinity chromatography (IMAC) and size-exclusion chromatography (SEC), were tested.

2.3.4.1 Immobilized metal affinity chromatography (IMAC)

The concept of IMAC was introduced in 1975 by utilising the interaction between transition metals such as copper and zinc and histidine and cysteine residues of proteins.\textsuperscript{11} The interaction of fused polyhistidine affinity tag to immobilised nickel in a column has been exploited since the 1980s for the purification of recombinant proteins.\textsuperscript{12} As ICL1 was produced with an N-terminal polyhistidine tag, IMAC using Ni(II) was used for purification. The UV\textsubscript{280} chromatogram showed ICL1 was eluted as a single peak. The typical yield of ICL1 that was purified by IMAC was around 70 mg per five litres of media.

Next, the removal of the polyhistidin tag by using TEV protease was tested. First, the protein was incubated with TEV protease at 4 °C in 1:50 molar ratio for 24 hours. Unfortunately, the condition found to be unsatisfactory, as only about 10% of the proteins was found to be cleaved as judged by SDS-PAGE. The concentration of TEV protease was therefore increased to a 1:10 molar ratio, and the incubation time was increased to 48 hours. However, even with the increased concentration of TEV protease and increased incubation time, TEV protease did not cleave more than 10% of the protein, which was demonstrated by a thin band at 47 kDa in SDS-PAGE (Figure 2.3). Control experiments using another fusion protein showed that the TEV protease was active (data not shown). It was therefore reasoned that the unsuccessful attempts were likely due to the structure of the protein. It is postulated that the TEV protease binding site may be sterically hindered due to the compact structure of ICL1. As previous studies showed that the catalytic activity of ICL1 was similar with or without a N-terminal
polyhistidine tag,\textsuperscript{1,2,4,5} it was therefore decided that ICL1 with the polyhistidine tag would be used for further experiments.

**Figure 2.3:** SDS-PAGE displaying unsuccessful attempts at polyhistidine tag cleavage from ICL1 (MW = 49.64 kDa) to give ICL without purification tag (MW = 47.086). Lane 1 represents protein ladder. Lane 2 represents ICL1 before treatment with TEV protease. Lane 3 represents ICL1 treated with TEV protease (1:10).

### 2.3.4.2 Size exclusion chromatography (SEC)

Size exclusion chromatography (SEC) is a well-known chromatographic technique for the purification of peptides and proteins, in which the separation of macromolecules is based on molecular weight.\textsuperscript{13} Typically, size exclusion columns are made of cross linked polymer beads such as agarose (Sepharose), dextran-epichlorohydrin (Sephadex) and dextran-polyacrylamide (Sephacryl) with a controlled pore size. When a solution containing macromolecules of different molecular weight is passed through the column, small macromolecules are trapped inside the pores of the beads in the column and will therefore take a longer time to elute. Large molecules, in contrast, will not fit inside the pores of the beads in the column and will therefore
elute faster. SEC was used as the final purification step to obtain pure ICL1. The protein was eluted as a single peak in the chromatogram (Figure 2.4a). The purity of the protein was confirmed by SDS-PAGE (Figure 2.4b) and the folding of the ICL1 was confirmed by thermofluor data (Figure 2.4c). Overall, 45 mg of highly purified ICL1 was obtained from 8 L of culture.
**Figure 2.4:** a) SEC purification chromatogram of ICL1; b) SDS-PAGE showing the purity of ICL1 (MW = 49.64 kDa) after SEC. Sample run was composed of fractions 19 to 26. Lane 1 represents the protein ladder and lane 2 presents ICL1; c) Protein melt curve showing an unfolding transition of ICL1. The sample contained 20 μM ICL1 in 50 mM Tris-HCl pH 7.5. The temperature was increased from 25 to 95 °C at 1 °C increment every 60 seconds.

### 2.4 Production and Purification of ICL2

#### 2.4.1 Cloning

The recombinant production of ICL2 was then optimised. The DNA sequence (Table 2.5) that encodes ICL2 from *M. tuberculosis* CDC1551 possesses 2301 nucleotides. The sequence was long for the commercially-available synthetic gene fragment that was used for ICL1; a full synthetic gene was therefore required (plasmid pUCIDT; obtained from Integrated DNA Technologies).
### Table 2.5: DNA sequence of AceA

| ATGGCCATCGCCGAAACGGACACCAGTGCCACACACCCGTTTCGAGCAGGACTTTTGAGAAAGACGTAAGCG |
| CCACCTCAGCGATACTTCAGCAGCTGCGCTTTGCTGGATGACTCGCTACACGCCGCCCCGCCAAGGTCG |
| GGAACAGCGCGCAGACATCCGCTGAGCAGGAGGCGACGGGCGGCGCCCTTTCGAGGCTG |
| CTCGGCGAACTTCTGGCAGGGCGCCAGAGAGAATTTGCAACAGGCTCTGGTTG |
| GCAAGAGCGGAGTGGATCAGGAGGAGCTACACTCAGGAGGTCCACACACCGTTCGAGCAGGACTTTGAGAAAAGACGTAGCCG |
| CCACTCAGCGATACTTCGACAGCTCGCGCTTTGCTGGGATCATTCGGCTC |
| GCCGCTCTGCTACCCGCGCCAGACGCCAACCAAGCTATCTACGAGCATGAGCGAGGTCCACACACCGTTCGAGCAGGACTTTGAGAAAAGACGTAGCCG |
| GACACCCGTTGACATCCGCTGAGCAGGAGGCGACGGGCGGCGCCCTTTCGAGGCTG |
| GTACGAACACCTGATCCGCTTTCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG |
| GCACCAAGAAGATCGCCGCCACCAAGGCGGCGCAAGATCCTGTGCTGCCAGCAGCAGATCAAGCGGCCTCAA |
| CGCGCCGCCATTCACGCTGCATTCGGCAGGGCGGACTCGCCGCGACCCGGGAGG |
Although ICL2 had been produced recombinantly using *E.coli* previously, it was reported to be an unstable protein.\(^1\) It is possible that *E. coli* may lack all the molecular machineries that are required for the folding and post-translational modification of ICL2, thus leading to unstable proteins.\(^{14}\) Mycobacterial hosts, in contrast, should possess all the necessary molecular chaperones and modifying enzymes, and are therefore better choice to get stable proteins.\(^{15}\) Non-pathogenic mycobacterial species such as *Mycobacterium smegmatis* (*M. smegmatis*) and *Mycobacterium vaccae* were previously used for the production of mycobacterial proteins.\(^{16},\ 17\) The transformable mutant *M. smegmatis* strain (mc\(^2\) 155) was first isolated in 1990 from wild-type ATCC 607.\(^{18}\) Since then, it has been used as the system of choice in the expression of mycobacterial proteins. Another strain *M. smegmatis* (mc\(^2\) 4517) was also generated. It allows the expression of proteins using the T7 promoter system in a similar manner to *E. coli*.\(^{19}\) It was therefore decided to test the production of ICL2 in both *E. coli* and *M. smegmatis*. The pYUB28b vector,\(^{20}\) which was designed from the *M. smegmatis* expression vector pYUB1049, was chosen as it can be used for both *E. coli* and *M. smegmatis*.

The polymerase chain reaction (PCR), a method that is commonly used to amplify particular DNA sequences, was then used to amplify the DNA sequence that encodes ICL2 from the synthetic gene. As the pYUB28b vector contains the NdeI and HindIII restriction sites, two primers, one forward and one backward, were designed based on the NdeI and HindIII restriction sites sequence and melting temperatures (T\(_m\)) (Table 2.6). The T\(_m\) difference between the forward and the reverse primer was kept within 2 °C and in between 60-64 °C, which is recommended for the PCR experiment.\(^{21}\) The PCR product was analysed and confirmed by agarose gel electrophoresis (Figure 2.5). It was then cloned into the pYUB28b vector (Figure 2.6). Both the PCR product and the vector were double digested with NdeI and HindIII. The digested inserts and vectors were incubated with ligase to get a recombinant pYUB28b-AceA plasmid.
Table 2.6: Forward and reverse primers that were used for PCR reactions to amplify aceA from the synthetic gene. *aactagaatcagcat* is added to the 5’ end of the forward primer and *aaatcaagcctt* is added to the 5’ end of the reverse primer so that the amplified PCR product can be cloned into the pYUB28b vector using the NdeI and HindIII restriction sites.
Figure 2.5: Agarose gel electrophoresis (1%) showing the amplified PCR product that encodes AceA. The observed length of PCR product was similar to the calculated length (2328 bp).
Figure 2.6: a) Plasmid map of the vector pYUB28b; b) Cloning strategy of the PCR-amplified DNA fragment that encodes AceA to the vector pYUB28b. The red colour represents:

5' CATATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCCGAATTCGAGCTCCGTCGACAAGCTT3'
3' GTATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGGCTTAAGCTCGAGGCAGCTGTTCGAA5'
The recombinant plasmid was then transformed to *E. coli* Top10 competent cells to get multiple copies of the recombinant plasmid. The plasmid was confirmed by agarose gel electrophoresis and DNA sequencing (DNA Sequencing Centre, The University of Auckland). The correct plasmid was then transformed into *E. coli* BL21 (DE3) LOBSTR competent cells and *Mycobacterium smegmatis* mc²4517.

### 2.4.2 Protein expression trials

Small-scale expression trials were then conducted with both *E. coli* BL21 (DE3) LOBSTR and *M. smegmatis* mc²4517. *E. coli* BL21 (DE3) LOBSTR was used because it is low in histidine-rich protein.²² This will allow efficient IMAC purification even if the expression level of the polyhistidine-tagged recombinant protein is low. In addition, the *E. coli* BL21 (DE3) LOBSTR is co-transformed with the plasmid pGro7, which encodes the GroEL-GroES chaperone. The GroEL-GroES chaperone may facilitate the folding of its client proteins to their native state by stabilising the unfolded or partially folded proteins, thus providing enough time for proper folding to occur.²³,²⁴

Interestingly, soluble ICL2 was found under all the tested conditions (Table 2.7). Based on the yield of soluble ICL2 production that was judged by denaturing SDS-PAGE, BL21 (DE3) LOBSTR with leaky expression at 37 °C was selected for large-scale production of ICL2. BL21 (DE3) LOBSTR was selected over *M. smegmatis* for scale up production because *M. smegmatis* is slow-growing and relatively difficult to handle when compared to *E. coli*.

<table>
<thead>
<tr>
<th>S.N.</th>
<th>Organism</th>
<th>Temperature/°C</th>
<th>Condition</th>
<th>Soluble ICL2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BL21 LOBSTR</td>
<td>18</td>
<td>leaky</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>autoinduction</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5 mM IPTG</td>
<td>√</td>
</tr>
<tr>
<td>2</td>
<td><em>M. smegmatis</em></td>
<td>37</td>
<td>leaky</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>autoinduction</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5 mM IPTG</td>
<td>√</td>
</tr>
</tbody>
</table>

**Table 2.7:** Different conditions that were tested for ICL2 expression.
2.4.3 Large-scale production and purification

Firstly, transformed BL21 (DE3) LOBSTR was grown at 37 °C and used as a starter culture for ICL2 production. The starter culture was then diluted with fresh media containing 0.1% arabinose, which was then incubated for 18 hrs at 37 °C before it was harvested by using centrifugation. Arabinose (0.1 %) helps to induce the production of GroEL-GroES as pGro7 is an *araB* promoter.\textsuperscript{25}  \textsuperscript{26} The cell pellet was resuspended in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 7.5) and lysed using a cell disrupter to get soluble ICL2. The pYUB28b vector added a N-terminal fusion polyhistidine tag to the protein,\textsuperscript{20} which enables ICL2 to be purified by IMAC and SEC in a similar manner to ICL1 (Figure 2.7a). The folding of the ICL2 was confirmed by thermofluor data. Overall, 30 mg of the purified protein was obtained from 5 L culture.

![Figure 2.7: a) SDS-PAGE showing the purity of ICL2 (MW = 85.32 kDa) after SEC. Lane 1 represents the protein ladder and lane 2 presents ICL2. Fraction 12 was run on the gel (see Figure 2.8 b, page 47); b) Protein melt curve showing an unfolding transition of ICL2. The sample contained 20 μM ICL2 in 50 mM Tris-HCl pH 7.5. The temperature was increased from 25 to 95 °C at 1 °C increment every 60 seconds.](image-url)
2.4.4 Optimisation for stability of ICL2

ICL2 was reported to be unstable when stored in vitro in PIPES buffer at pH 8.0 (20 mM PIPES, 50 mM NaH₂PO₄, 100 mM NaCl, 300 mM imidazole). However, the results obtained in this study showed that freshly produced and affinity chromatography-purified ICL2 in HEPES buffer at pH 7.5 containing sodium chloride (NaCl), β-mercaptoethanol (β-ME) with and without imidazole (Figure 2.8) was stable. However, it was also found that freezing the protein for long-term storage in HEPES buffer containing imidazole may lead to protein precipitation (Figure 2.9a). After freeze-thawing the affinity chromatography-purified ICL2 (in the presence of imidazole) at -80 °C, about 50 % of ICL2 was found to have aggregated, as reflected by the chromatogram of SEC, which showed that these aggregated proteins eluted in the void volume of SEC (Figure 2.9a). Freezing and thawing of proteins may induce protein aggregation. This is a known phenomenon and it could be influenced by factors including low temperature and additives in the buffer. Interestingly, when ICL2 was stored at -80 °C in HEPES buffer without imidazole, the SEC chromatogram showed only a peak that corresponded to the folded protein (Figure 2.9b). Therefore, it is likely that the aggregation of ICL2 was induced by imidazole in the buffer rather than by freeze-thawing. Therefore, affinity column purification should be followed immediately by buffer exchange to remove the imidazole.
Figure 2.8: Chromatograms of ICL2 purification. a) IMAC purification and; b) SEC purification. The SEC purification was conducted immediately after IMAC purification without freeze-thawing the protein sample.

Figure 2.9: SEC chromatograms of ICL2 purification after the partially-purified protein (by IMAC) was at -80 °C a) buffer with imidazole; b) buffer free of imidazole.
2.5 Production and purification of ICL2a and ICL2b

In *M. tuberculosis* H37Rv, the gene that encodes ICL2 is split into two open reading frames, *aceAa* and *aceAb*, which encode ICL2a and ICL2b respectively. It is not clear whether ICL2a and ICL2b are soluble or functional proteins.

2.5.1 Design of synthetic gene, cloning and transformation

Synthetic DNA fragments (gBlocks) encoding *M. tuberculosis* ICL2a and ICL2b were first designed. In a similar manner to the production of ICL1, the *taccttcaatccatg* sequence was added to the 5’ end and *taacagtaaaggtggata* was added to the 3’ end for cloning to the pNIC28-Bsa4 vector (Tables 2.8 and 2.9). The cloning procedures followed those as described for ICL1. The sequence of the plasmid was confirmed by DNA sequencing (DNA Sequencing Centre, The University of Auckland).
Table 2.8: The synthetic AceAa fragment used in this study. Sequences \textit{taccttccaatcc} and \textit{cagtaaaggtgata} were added to the 5' and 3' ends respectively for cloning into the vector pNIC28-Bsa4.

<table>
<thead>
<tr>
<th>TACTTCCAATCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATGACCTACGGAGGCGCTGGACGTGCTCGAATTCGGTCAGAGCGAGGCGAACCCATGTTGCATGG</td>
</tr>
<tr>
<td>CTCCCGAGGATGCGCGGGGTCTGCGCCGGTGCATGCTGCCGCCGGGCAACGGCATTCATGG</td>
</tr>
<tr>
<td>GGGCGAATGCGCTGCCAGCTGAGCGCCAGGGCGCATGCTGGCGCTGGTTCGACGCGAAGATG</td>
</tr>
<tr>
<td>ATACCATGCGGTCGCAACCTGCTGGCCAGGCGGCAAGACCTGGCGCTGGGCTGCTGG</td>
</tr>
<tr>
<td>TACTTCCAATCC</td>
</tr>
</tbody>
</table>

Table 2.9: The synthetic AceAb fragment used in this study. Sequences \textit{taccttccaatcc} and \textit{cagtaaaggtgata} were added to the 5' and 3' ends respectively for cloning into the vector pNIC28-Bsa4.

2.5.2 Expression trial of ICL2a

The recombinant plasmid (pNIC28-Bsa4) encoding ICL2a was transformed in \textit{E. coli} BL21 (DE3). Expression trials were then conducted using three different incubation temperatures (18, 28 and 37 °C) and IPTG concentrations (0.2, 0.5 and 1.0 mM). Unfortunately, none of these conditions led to soluble ICL2a production (Figure 2.10).
E. coli Arctic Express (DE3) and Origami (DE3) were then tested. E. coli Arctic Express (DE3) allows the production of a protein at low temperature (8-12 °C).\textsuperscript{28, 29} Low temperatures increase the probability of the proper folding of a protein.\textsuperscript{30} E. coli Arctic Express strains are engineered with cold-adapted chaperonins Cpn10 and Cpn60 from the psychrophilic bacterium, *Oleispira Antarctica*, which helps the proper folding of the protein at low temperature.\textsuperscript{29} Origami (DE3) possesses an oxidative cytoplasmic environment that helps to produce the soluble protein by allowing the formation of disulphide bonds in a protein.\textsuperscript{31} However, none of these conditions gave soluble ICL2a (Table 2.10).

The *AceAa* gene was also cloned to pNH-TrxT vector, which adds a fusion peptide thioredoxin, of 128 AAs, together with a polyhistidine tag and TEV cleavage site to ICL2a. This is because thioredoxin may aid to the proper folding of the protein.\textsuperscript{32} The plasmid was transformed into BL21 (DE3), and expression trials were done at three different temperature (18, 28 and 37 °C) and three different IPTG concentrations (0.2, 0.5 and 1.0 mM). However, the protein was not expressed in soluble form and accumulated as insoluble aggregates.

As *E. coli* may lack the proper machinery to fold mycobacterial proteins, expression trials of ICL2a in *M. smegmatis* and *E. coli* BL21 (DE3) LOBSTR with extra copies of chaperones were then tested. Similarly to *AceA*, the *AceAa* gene was amplified using PCR (for the primer sequences, see Table 2.11) and analysed and confirmed by agarose gel electrophoresis (Figure 2.11). It was then cloned into the pYUB28b vector.

The recombinant plasmid was then used to transform competent BL21 (DE3) LOBSTR cells containing the plasmid pGRo7 and *M. smegmatis* mc\textsuperscript{2}4517. However, ICL2a was accumulated as insoluble inclusion bodies under all conditions used (Table 2.12).

Finally, the insoluble aggregates of ICL2a were then tested for refolding *in vitro*. The cells were harvested and resuspended in denaturing buffer containing 6 M guanidine. The denatured protein was purified using IMAC and the purified protein was refolded by rapid dilution.
However, the protein was precipitated instead of refolding. It is therefore concluded ICL2a could be not expressed as a soluble protein in both *E.coli* and mycobacterial hosts.

**Figure 2.10:** SDS-PAGE showing expression trial of ICL2a (MW = 43.042) in BL21 (DE3) cells. a) 18 °C, b) 28 °C and c) 37°C. First lane represents protein ladder. Lanes 2 and 3 represent whole cell and cell lysate in 0.2 mM IPTG. Lanes 4 and 5 represent whole cell and cell lysate in 0.5 mM IPTG. Lane 6 and 7 represent whole cell and cell lysate of 1.0 mM IPTG.
<table>
<thead>
<tr>
<th>S.N.</th>
<th>Organism</th>
<th>temperature/°C</th>
<th>IPTG/mM</th>
<th>protein expression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insoluble</td>
</tr>
<tr>
<td>1</td>
<td><em>E. coli</em> [BL21 (DE3)]</td>
<td>18</td>
<td>0.2</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td><em>E. coli</em> [BL21 (DE3)]</td>
<td>28</td>
<td>0.2</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td><em>E. coli</em> [BL21 (DE3)]</td>
<td>37</td>
<td>0.2</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td><em>E. coli</em> [BL21 (DE3)]</td>
<td>12</td>
<td>0.2</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td><em>E. coli</em> [arctic express (DE3)]</td>
<td>12</td>
<td>0.2</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td><em>E. coli</em> [origami (DE3)]</td>
<td>18</td>
<td>0.2</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td><em>E. coli</em> [origami (DE3)]</td>
<td>28</td>
<td>0.2</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>X</td>
</tr>
</tbody>
</table>
Table 2.10: Additional conditions used for the expression trial of ICL2a using different strains of *E. coli*. The conditions used varied from 12 to 37 °C and 0.2 to 1.0 mM IPTG concentration.

<table>
<thead>
<tr>
<th></th>
<th>8</th>
<th>E.coli [origami (DE3)]</th>
<th>1</th>
<th>0.2</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>37</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

**Design of forward primer based on NdeI restriction site**

**NdeI restriction site**

5’...CA\textsubscript{T}ATG...3’
3’...GT\textsubscript{T}ATAC...5’

*A section of forward sequence with additional aactagaatcgacat*  
5’aactagaatcgacat\textsubscript{ATG} gcacatgcgcgaaacggacaccgaggtcaca 3’

**Forward Primer;**  
5’aactgaatcgacat\textsubscript{ATG} GCC ATC GCC GAA ACG GAC 3’  
\(T_m = 60.8^\circ C\)

**Design of reverse primer based on HINDIII restriction site**

**HindIII restriction site**

5’...A\textsubscript{T}AGCTT...3’
3’...TTGCA\textsubscript{A}...5’

*A section of reverse complement with additional aaatcaagctt*  
5’ aaatcaagctt cacgccgctgctgccctgccgcaacgacccgc 3’

**Reverse Primer;**  
5’ aaatcaagctt TCA GGC CGT CGT C 3’  
\(T_m = 60.4^\circ C\)

Table 2.11: Forward and reverse primers that were used for PCR to amplify aceAa from the Genomic DNA. *aactagaatcgacat* is added to the 5’ end of the forward primer and *aaatcaagctt* is added to the 5’ end of the reverse primer so that the amplified PCR product can be cloned into the pYUB28b vector using the NdeI and HindIII restriction sites.
Figure 2.11: 1% Agarose gel electrophoresis visualising the PCR product \( AceAa \). The observed length was similar to the calculated length 1131 bp.

<table>
<thead>
<tr>
<th>S.N.</th>
<th>Organism</th>
<th>Temperature/°C</th>
<th>Condition</th>
<th>Soluble ICL2a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BL21 LOBSTR</td>
<td>18</td>
<td>leaky</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>autoinduction</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5 mM IPTG</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>M. smegmatis</td>
<td>37</td>
<td>leaky</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>autoinduction</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5 mM IPTG</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.12: Results of expression trial of ICl2a in BL21 (DE3) LOBSTR and \( M.\ \text{smegmatis} \).
2.5.3 Expression trial of ICL2b

ICL2b recombinant plasmid transformed BL21 (DE3) cells were grown at 37 °C. Expression trials were conducted by using three different temperatures (18, 28 and 37 °C) and IPTG concentrations (0.2, 0.5 and 1.0 mM). ICL2b was found to be overexpressed under all the tested conditions. However, at 28 and 37 °C, ICL2b was mainly found in insoluble inclusion bodies, instead of forming soluble proteins. The optimal conditions for soluble ICL2b production were 18 °C with two different IPTG concentrations (0.5 and 1.0 mM). A higher yield was found in the presence of 1.0 mM IPTG. These conditions were therefore used for large-scale production (Figure 2.12).

![Figure 2.12: SDS-PAGE showing the expression trials of ICL2b (MW = 47.13 kDa) at 18 °C.](image)

The first lane represents protein ladder. Lanes 2 and 3 represent whole cell and cell lysate with 0.2 mM IPTG. Lanes 4 and 5 represent whole cell and cell lysate with 0.5 mM IPTG. Lanes 6 and 7 represent whole cell and cell lysate with 1.0 mM IPTG.

2.5.4 Production and purification of ICL2b

Similarly to ICL1, ICL2b was purified by IMAC and SEC. The affinity column purified ICL2b contains an N-terminal polyhistidine tag and a TEV protease cleavage site. In order to cleave
the N-terminal polyhistidine peptide, the purified protein was incubated with TEV protease at 4 °C in a 1:50 molar ratio for 18 hours. The polyhistidine tag was found to be cleaved under these conditions (Figure 2.13a). The tag-free ICL2b was further purified by IMAC and SEC consecutively, to get pure ICL2b (26 mg from 8 L culture) (Figure 2.13a).

**Figure 2.13:** a) SDS-PAGE showing ICL2b with (MW = 47.13 kDa) and without (MW = 44.58 kDa) a polyhistidine tag. Lane 1 represents the protein ladder. Lane 2 represents ICL2b with a polyhistidine tag. Lane 3 represents ICL2b without a polyhistidine tag b) SDS-PAGE showing purity of ICL2b after SEC.
2.6 Conclusion

The work described in this Chapter summarises the effort in the production of all four isoforms of *M. tuberculosis* ICL (ICL1, ICL2, ICL2a and ICL2b). Soluble ICL1, the most studied isoform, was produced recombinantly using an *E. coli* expression system. The cleavage of the N-terminal polyhistidine tag of ICL1 was found to be problematic, which is likely due to the structure of the protein. Nonetheless, highly purified ICL1 was produced following IMAC and SEC purifications. Soluble ICL2 was also produced recombinantly using both the *E. coli* and *M. smegmatis* expression systems. Soluble proteins were obtained under all tested conditions. Highly purified ICL2 was obtained by IMAC and SEC. In contrast to previous reports, ICL2 was found to be stable in buffers without imidazole. The production of soluble ICL2a and ICL2b were also attempted. Despite the different expression conditions and expression systems (*E. coli* and *M. smegmatis*) that were tested, it was not possible to obtain soluble ICL2a. Refolding of ICL2a from inclusion bodies was also unsuccessful. In contrast, soluble ICL2b was obtained using expression conditions that were similar to ICL1. These experiments provided us the conditions to obtain soluble, stable and highly purified ICL1, ICL2 and ICL2b, which could then be used for structural and kinetic experiments (see Chapters 3 to 6).
Chapter 3: Kinetics and inhibition studies of Isocitrate lyase 1 (ICL)¹

3.1 Introduction

3.1.1 Assays for ICL

Several assays were developed to study the kinetics and inhibition of ICLs. The most common assays are based on ultraviolet/visible (UV/vis) spectrophotometry, which rely on the quantification of glyoxylate, one of the products of the enzymatic reaction catalysed by ICL. However, as glyoxylate cannot absorb UV or visible light, it needs to be chemically or enzymatically modified.¹

The chemical-coupled method was developed in 1959 by Dixon et al.² In this method, glyoxylate is chemically modified by phenylhydrazine to form a hydrazone that absorbs UV light at 390 nm. Thus, the conversion of isocitrate to succinate and glyoxylate by ICL is correlated with the amount of hydrazone that is formed (and hence UV/vis absorption). However, the disadvantages of using this method are three-fold. Firstly, this is an indirect measurement of the product glyoxylate. The accuracy of the experiment is dependent on the chemical coupling step. For instance, the rate of glyoxylate-phenylhydrazone complex formation is pH dependent. In particular, the rate slows significantly above pH 7, which may impact on the accuracy of the enzyme kinetic measurement. Secondly, phenylhydrazine is unstable at pH 7 and above. This leads to a slow increase in the UV absorption even in the absence of glyoxylate.³ Thirdly, the accuracy of the experiment may be affected by other components of the reaction mixture such as ICL inhibitors that if aromatic, may absorb UV/vis light at similar wavelengths.

¹The work that is described in this chapter was conducted in collaboration with Anne Swartjes, who synthesised (2R,3S)-2-methylisocitrate.

Part of the work from this chapter was published as described in:

An alternative to the chemical coupled method to determine ICL kinetics is to use the enzyme lactate dehydrogenase (LDH). LDH catalyses the reduction of glyoxylate to form glycolate. This reaction is coupled with the oxidation of NADH, which is used as a cofactor of LDH, to NAD⁺. The decrease of NADH can be measured by the addition of the tetrazolium (MTS) dye. When NADH is oxidised, MTS is reduced to soluble formazan, which absorbs UV at 490 nm. There are several limitations with this method. Firstly, it is indirect and is reliant on both the activity of LDH as well as the reduction of MTS to form formazan. Secondly, autoxidation of NADH to NAD⁺ may lead to inaccurate results over time. Thirdly, this method is not suitable to measure methylisocitrate lyase activity because LDH cannot accept pyruvate as a substrate.

In addition to kinetic assays, binding assays have also been developed. The first assay was based on native non-denaturing mass spectroscopy, which can be applied to study covalent ligand binding (and in theory, non-covalent binding, provided the protein-ligand interaction is strong enough to survive the electrospray ionisation process). The second method was based on intrinsic protein fluorescence. In ICL1, Tyr89 and Trp93 are found in close proximity to the active site. Therefore, upon ligand binding, the intrinsic fluorescence (\(\lambda_{\text{ex}} = 290\) nm and \(\lambda_{\text{em}} = 320\) nm) of the protein may change as a result of changes in environment and/or conformation.

### 3.1.2 \(^1\)H nuclear magnetic resonance spectroscopy for kinetics and inhibition studies

The application of NMR spectroscopy to the study of biological systems started in the early 1950s. The first publication reporting the use of NMR spectroscopy to study proteins was disclosed in 1968. The authors applied \(^1\)H NMR to monitor the hydrolysis of acetylcholine by horse cholinesterase. Since then, NMR spectroscopy has become one of the most important biophysical techniques in enzymology. For example, NMR spectroscopy is now commonly applied in applications that include structure determination of proteins and enzymes, \textit{in situ}
characterisation of enzyme reactions and products, the studies of protein dynamics and the measurement of protein-ligand interactions.\textsuperscript{11}

The use of NMR spectroscopy to study enzyme kinetics has several advantages. Firstly, all reactants and products (as long as they contain non-exchangeable hydrogens) can be monitored in the same spectrum, provided there is no overlap of signals. Secondly, the amount of substrate consumption and product formation can be accurately quantified by simply following changes in the peak area of the resonances associated with the substrate and/or reaction product(s). Thus, this method does not rely on the accuracy of chemical or enzymatic coupling reactions.

3.1.3 Thermal shift assay for high-throughput screening of protein-ligand interactions

The thermal shift assay was developed in the early 2000s by Pantoliano \textit{et al.} as a high-throughput technique to screen for protein-ligand binding.\textsuperscript{12} This method is based on the phenomenon that ligand binding may stabilise\textsuperscript{13-15} or destabilise\textsuperscript{16} the protein to thermal denaturation. Thus, ligand binding may be reflected by changes in the protein melting temperature ($T_m$).

Protein $T_m$ may be measured by the use of a fluorescent dye. When the protein unfolds (for example, due to thermal denaturation), it exposes its hydrophobic core. By using a fluorescent dye that binds hydrophobic amino acids and is sensitive to changes in hydrophobic environment,\textsuperscript{17} the protein $T_m$ can be measured by following changes in the fluorescence that is emitted due to binding of the dye to the open hydrophobic core of the protein. The assay is simple. It can be applied in an automated fashion using multi-well plate format, thus it is widely applied as a high-throughput method to screen for new protein ligands.\textsuperscript{12, 18, 19}
3.2 Objective

The objective of the work described in this chapter is to develop a NMR-based method for the studies of ICL kinetics and inhibition using ICL1 as a model system. In addition to $^1$H NMR spectroscopy, a thermal shift assay was also to be investigated for screening of ICL inhibitors.

3.3 ICL1 kinetics by $^1$H NMR

3.3.1 Isocitrate lyase activity by $^1$H NMR

$^1$H NMR was first applied to monitor the ICL1-catalysed conversion of isocitrate to succinate and glyoxylate. DL-Isocitrate, which is commercially available, was used as the substrate. MgCl$_2$ was added to the reaction mixture as Mg$^{2+}$ was previously shown to be important for ICL1 activity. $^1$H spectra were recorded at ~1.3 minute intervals. The experiments were conducted in 90% H$_2$O and 10% D$_2$O (Figure 3.1a). The large water peak was suppressed using the excitation sculpting method so that the resonances of the substrate and product could be visualised and quantified accurately. Upon addition of the enzyme, the peaks that corresponded to isocitrate dropped in intensity, which was accompanied by the appearance of a new singlet peak at 2.3 ppm, corresponding to succinate (Figure 3.1b). Glyoxylate cannot be observed because it contains only exchangeable hydrogen atoms. Integration of the isocitrate and succinate peaks showed that the reaction appeared to slow down when ~50% of the isocitrate was consumed (Figure 3.1c). As the isocitrate was a racemic mixture, this result infers that ICL1 has a preference for one enantiomer, which is in agreement with a previous study that showed D-isocitrate is the preferred substrate of the enzyme.
a) DL-isocitric acid is converted into succinic acid by isocitrate lyase in the presence of Mg²⁺.

b) HPLC chromatograms showing the conversion of isocitrate to succinate over time:
- 10.5 minutes: 100% isocitrate
- 9.2 minutes: 75% isocitrate, 25% succinate
- 7.9 minutes: 50% isocitrate, 50% succinate
- 6.6 minutes: 25% isocitrate, 75% succinate
- 5.3 minutes: 12.5% isocitrate, 87.5% succinate
- 4 minutes: 100% succinate

C) Graph showing the concentration of succinate over time:
- [Succinate] (µM) vs. Time (minutes)
Figure 3.1: (a) Isocitrate lyase catalyses the conversion of isocitrate to glyoxylate and succinate; (b) ¹H NMR spectroscopy to monitor ICL1-catalysed turnover of isocitrate into succinate (c) Corresponding plot of the isocitrate turnover data. The curve was added to aid visualisation. The sample contained 190 nM ICL1, 1 mM DL-isocitrate, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The hashtag (#) indicates Tris/Tris-D11 peak. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).

### 3.3.2 Role of Mg (II) in the activity of ICL1

Divalent metals play important role in the activity of ICL1. Previous studies showed that Mg²⁺ (and to a lesser extent, Mn²⁺) are required for optimal ICL1 activity.⁴ ²² In order to confirm the concentration of divalent magnesium that is required for optimal activity of the enzyme, the reaction was run using different concentrations of MgCl₂. Under the tested reaction conditions, it was found that 500–5000 μM of MgCl₂ was required for the optimal activity, and this concentration was therefore used in all subsequent kinetic and inhibition assays (Figure 3.2). The enzyme showed negligible activity without Mg²⁺. Mg²⁺ is absolutely required for the reaction because the substrate was proposed to bind the active site of ICL via chelation of a magnesium ion.²² ²³ However, any role of Mn²⁺ cannot be measured by ¹H NMR spectroscopy. Mn²⁺ is paramagnetic, which enhances the longitudinal and transverse relaxation rates of NMR resonances. Thus, the isocitrate and succinate signals will be severely broadened and hence accurate measurements cannot be performed.
Figure 3.2: Mg\(^{2+}\) is required for optimal ICL1 activity. Sample contained 190 nM ICL1, 1 mM DL-isocitrate, varying concentration of MgCl\(_2\) and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H\(_2\)O and 10% D\(_2\)O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).

3.3.3 Kinetic parameters for ICL1 with DL-isocitrate by \(^1\)H NMR

Next, the kinetic parameters of ICL1 were measured. This was conducted using a fixed concentration of ICL1 (190 nM) and Mg\(^{2+}\) (5 mM), and varying concentration of DL-isocitrate. The initial rate was calculated using the turnover of isocitrate to succinate and glyoxylate over four minutes. The kinetic parameters were calculated using Hanes-Woolf plot, which involved linearisation of the Michaelis-Menten equation for the analysis of kinetic data (Equation 3.1).\(^{24}\)
Equation 3.1: Hanes-Woolf equation

\[
\frac{[S]}{V} = \frac{K_M}{V_{\text{max}}} + \frac{[S]}{V_{\text{max}}}
\]

The Michaelis constant \((K_M)\) was found to be \(290 \pm 10 \, \mu\text{M}\) and the catalytic constant \((k_{\text{cat}})\) was determined to be \(4.3 \pm 0.1 \, \text{s}^{-1}\) (Figure 3.3). These values were similar to those obtained by Gould et al. using the aforementioned LDH assay, which were \(190 \, \mu\text{M}\) and \(5.24 \, \text{s}^{-1}\) respectively (Table 3.1).\(^{25}\) The slight discrepancy between the two measurements is likely due to differences in the reaction conditions. Overall, this validated the accuracy of our \(^1\text{H} \, \text{NMR}\) assay to study ICL1 kinetics.

Figure 3.3: A Hanes-Woolf plot of ICL1. Sample contained 190 nM ICL1, varying concentration of \(\text{DL}-\text{isocitrate}\), 5 mM \(\text{MgCl}_2\) and 50 mM Tris/Tris-D11 (pH 7.5) in 90\% H\(_2\)O and 10\% D\(_2\)O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).
Table 3.1: Kinetic parameters of ICL1 with DL-isocitrate as substrate. Measurements were made using samples containing 190 nM ICL1, varying concentrations of DL-isocitrate, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).

<table>
<thead>
<tr>
<th></th>
<th>$K_M$ / μM</th>
<th>$k_{cat}$ / s⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>290 ± 10</td>
<td>4.3 ± 0.1</td>
</tr>
<tr>
<td>Gould et al.²⁵</td>
<td>190</td>
<td>5.24</td>
</tr>
</tbody>
</table>

3.3.4 Kinetic parameters of ICL1 with (2R,3S)-2-methylisocitrate by ¹H NMR

It was reported that mycobacterial ICL1 may catalyse the conversion of (2R,3S)-2-methylisocitrate to pyruvate and succinate in presence of Mg²⁺ (Figure 3.4a). In order to test this proposal, (2R,3S)-2-methylisocitrate was synthesised according to the procedure reported by Darley et al (by Anne Swartjes),²⁶ and the reaction was monitored by the ¹H NMR assay at ~1.3 minute intervals. Upon addition of the enzyme, the peaks corresponding to (2R,3S)-2-methylisocitrate dropped in intensity, a change which was accompanied by the appearance of two new singlet peaks at 2.3 and 2.26 ppm corresponding to succinate and pyruvate (Figure 3.4b). The keto-enol tautomerism of pyruvate (Figure 3.5) under the reaction condition directly effects the intensity of –CH₃ peak of pyruvate,²⁷ so we did not use the –CH₃ peak of pyruvate for quantification of enzyme kinetics. The kinetic parameters were calculated using Hanes-Wolf plot. The Michaelis constant ($K_M$) was found to be 560 ± 28 μM and the catalytic constant ($k_{cat}$) was determined to be 0.5 ± 0.1 s⁻¹ (Figure 3.6). These values were not much different to those obtained by Gould et al. using the aforementioned LDH assay, which were 720 μM and 1.25 s⁻¹ respectively (Table 3.2).²⁵
Figure 3.4: (a) Isocitrate lyase catalyses the conversion of (2R,3S)-2-methylisocitrate to succinate and pyruvate; (b) \(^1\)H NMR spectroscopy to monitor ICL1-catalysed turnover of (2R,3S)-2-methylisocitrate into succinate and pyruvate. The sample contained 2 µM ICL1, 1 mM (2R,3S)-2-methylisocitrate, 5 mM MgCl\(_2\) and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H\(_2\)O
and 10% D$_2$O. The hashtag (#) indicates Tris/Tris-D11 peak and asterisk (*) indicates impurities in the reaction mixture.

**Figure 3.5:** Keto-enol tautomerism of pyruvate under the aqueous condition.

![Diagram of keto-enol tautomerism of pyruvate](image)

**Figure 3.6:** A Hanes-Wolf plot of ICL1 with (2R,3S)-2-methylisocitrate. The sample contained 2 µM ICL1, varying concentrations of (2R,3S)-2-methylisocitrate, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).
<table>
<thead>
<tr>
<th></th>
<th>$K_M$ / $\mu$M</th>
<th>$k_{cat}$ / s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>560 ± 28</td>
<td>0.5 ± 0.01</td>
</tr>
<tr>
<td>Gould et al.$^{25}$</td>
<td>720</td>
<td>1.25</td>
</tr>
</tbody>
</table>

**Table 3.2:** Kinetic parameters of ICL1 with (2R,3S)-2-methylisocitrate as substrate.

Measurements were made using samples containing 2 $\mu$M ICL1, varying concentrations of (2R,3S)-2-methylisocitrate, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. The errors shown are the standard deviation from three separate measurements.

### 3.3.4.1 Enzyme inhibition study of (2R,3S)-2-methylisocitrate

In contrast to $d$-isocitrate, a reduction of ICL1 activity was observed at high (2R,3S)-2-methylisocitrate concentrations (above 1000 $\mu$M) and did not follow the Michaelis-Menten equation (Figure 3.7). The phenomenon is likely due to substrate inhibition.$^{28, 29}$ There are several reasons behind substrate inhibition. These include uncompetitive and non-competitive substrate inhibition,$^{30}$ multiple binding sites,$^{31}$ and an allosteric mechanism$^{32}$. A plausible mechanism of ICL1 inhibition by (2R,3S)-2-methylisocitrate at high concentration is that, instead of the active orientation, (2R,3S)-2-methylisocitrate may bind ICL1 in a different orientation in the catalytic pocket leaving the enzyme non-productive. Substrate inhibition of other enzymes at high concentrations by forming non-productive enzyme-substrate complexes has been reported previously.$^{33, 34}$
Figure 3.7: ICL1 kinetics with (2R,3S)-2-methylisocitrate showing enzyme inhibition at higher concentrations of (2R,3S)-2-methylisocitrate. Measurements were made using samples containing 2 μM ICL1, varying concentrations of (2R,3S)-2-methylisocitrate, 5 mM MgCl\(_2\) and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H\(_2\)O and 10% D\(_2\)O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).

3.4 \(^1\)H NMR and a thermal shift based method for inhibition studies of ICL1

3.4.1 \(^1\)H NMR as a tool for inhibition studies of ICL1

The new NMR-based assay was also applied to study ICL1 inhibition. First, the method was evaluated by using four known ICL inhibitors. These include three so-called “first generation inhibitors”, itaconate 1, nitropropionate 2 and bromopyruvate 3. Methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate 20, an inhibitor that was discovered recently by Liu et al.\(^{35}\) using high-throughput screening, was also used.
3.4.1.1 Single concentration inhibition assay

The single concentration inhibition experiment was designed for a preliminary screening to find if a molecule inhibits the enzyme. In addition, the method is appropriate to compare the strength of inhibition between inhibitors. To validate the method, the single concentration inhibition experiment of four known inhibitors were first conducted (Figure 3.8). In accordance with literature results, our \(^1\)H NMR assay showed that nitropropionate and bromopyruvate were the most potent inhibitors, followed by itaconate. Methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate was the weakest of the four tested.

![Image](https://example.com/image.png)

**Figure 3.8:** Single concentration inhibition data of ICL1 inhibitors. The sample contained 190 nM ICL1, 1 mM \(DL\)-isocitrate, 100 \(\mu\)M inhibitor (if applicable), 5 mM MgCl\(_2\) and 50 mM Tris/Tris-D11 (pH 7.5) in 90\% H\(_2\)O and 10\% D\(_2\)O. The curves were added to aid visualisation. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).
3.4.1.2 Quantification of inhibition assay by measuring the half maximal inhibitory concentration (IC$_{50}$)

IC$_{50}$ is the concentration required of an inhibitor to reduce the activity by 50 percent. Under certain conditions it can used to express the affinity of the enzyme inhibitor. It is widely used in drug discovery process to find the suitability of a molecule as an inhibitor. Any inhibition assay method should be appropriate for measuring IC$_{50}$. To check the appropriateness of our $^1$H NMR method, we obtained quantitative inhibition information (IC$_{50}$) of 4 known inhibitors. (Table 3.3 and Figures 3.9–3.12). The IC$_{50}$ values of nitropropionate and bromopyruvate were found to be 14.7 ± 1.8 μM and 17.5 ± 1.0 μM, respectively. The IC$_{50}$’s for itaconate and methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate were 29.4 ± 4.1 μM and 250 ± 7 μM, respectively.

Overall, these results show that $^1$H NMR is a useful tool to study ICL1 inhibition in vitro, enabling a rapid evaluation of inhibitor strength as well as providing accurate quantitative information such as IC$_{50}$. 
<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Structure</th>
<th>IC$_{50}$ / μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitropropionate</td>
<td><img src="image1" alt="Structure" /></td>
<td>14.7 ± 1.8</td>
</tr>
<tr>
<td>Bromopyruvate</td>
<td><img src="image2" alt="Structure" /></td>
<td>17.5 ± 1.0</td>
</tr>
<tr>
<td>Itaconate</td>
<td><img src="image3" alt="Structure" /></td>
<td>29.4 ± 4.1</td>
</tr>
<tr>
<td>Methyl-4-(4-methoxyphenyl)-4-oxobut-2-enoate</td>
<td><img src="image4" alt="Structure" /></td>
<td>250 ± 7</td>
</tr>
</tbody>
</table>

**Table 3.3:** IC$_{50}$ values of ICL1 inhibitors with DL-isocitrate as substrate. Measurements were made using samples containing 190 nM ICL1, 1 mM DL-isocitrate, 5 mM MgCl$_2$, varying concentrations of inhibitor and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. The errors shown are the standard deviation from three separate measurements. (see Figures 3.9–3.12).
Figure 3.9: IC$_{50}$ measurement for nitropropionate. Samples contained 190 nM ICL1, 1 mM DL-isocitrate, varying concentrations of inhibitor, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). The IC$_{50}$ was 14.7 ± 1.8 μM.
Figure 3.10: IC₅₀ measurement for bromopyruvate. Samples contained 190 nM ICL1, 1 mM DL-isocitrate, varying concentrations of inhibitor, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). The IC₅₀ was 17.5 ± 1.0 μM.
Figure 3.11: IC$_{50}$ measurement for itaconate. Samples contained 190 nM ICL1, 1 mM DL-isocitrate, varying concentrations of inhibitor, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). The IC$_{50}$ was 29.4 ± 4.1 μM.
**Figure 3.12**: IC\textsubscript{50} measurement for methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate. Samples contained 190 nM ICL1, 1 mM \textit{dl}-isocitrate, varying concentrations of inhibitor, 5 mM MgCl\textsubscript{2} and 50 mM Tris/Tris-D\textsubscript{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). The IC\textsubscript{50} was 250 ± 7 μM.

3.4.2 Thermal shift assay to compliment \textsuperscript{1}H NMR-based inhibition assay

Although \textsuperscript{1}H NMR spectroscopy was found to be a useful method to study ICL1 inhibition, it is relatively low throughput and labour intensive. A high throughput assay is needed to facilitate the efficient screening and development of new ICL inhibitors. The use of a thermal shift assay to compliment the \textsuperscript{1}H NMR-based kinetic assay was therefore evaluated.

First, the melting temperature of ICL1 was measured. As MgCl\textsubscript{2} is important for the activity of the enzyme, a saturating concentration of 1 mM was used. The melting temperature of ICL1 in the presence of MgCl\textsubscript{2} was found to be 43.0 °C. Next, the melting temperatures of ICL1 in
the presence of a saturating concentration (1 mM) of the aforementioned inhibitors and MgCl₂ were measured. Addition of bromopyruvate or itaconate were found to stabilise ICL1, with a positive shift to melting temperature of 52.5 °C and 53.3 °C respectively. Interestingly, nitropropionate and methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate were found to destabilise the protein, with a negative thermal shift to 40.9 °C and 37.6 °C respectively (Figure 3.13).

A negative thermal shift upon ligand binding has been previously observed for other protein systems. A positive thermal shift may be observed if the ligand induces the protein to adapt a more stable ‘closed’ conformation, whilst a negative thermal shift may be observed if the ligand keep the protein in a less stable ‘open’ conformation. Previous structural studies by Sharma et al. showed that ICL1 may undergo a two-step conformational change upon substrate binding. Indeed, a crystal structure of ICL1 in the presence of both nitropropionate and glyoxylate was found to adapt a ‘closed’ conformation (PDB id: 1F8I). Nitropropionate is a structural analogue of succinate. We reasoned that the binding of nitropropionate on its own may keep ICL1 in an open conformation in order to allow glyoxylate to bind. However, in the presence of both nitropropionate and glyoxylate, the protein can then undergo a conformational change to the ‘closed’ conformation, as suggested by the crystal structure, to catalyse the reverse reaction. This proposal is consistent with the mechanism suggested by Sharma et al. It should also be noted that binding of nitropropionate or methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate may induce a change in the oligomeric state of ICL1, which exists as a tetramer in solution. However, based on evidence from X-ray crystallography and molecular docking, both compounds are not known to bind at the oligomerisation interface between the ICL1 monomers, or interfere with the residues that were previously identified as important for the protein’s oligomerisation.
Figure 3.13: Protein melt curve for ICL1 inhibitors. The sample contained 20 μM ICL1, 1 mM compounds (where applicable) and 1 mM MgCl₂ in 50 mM Tris-HCl pH 7.5. The temperature was increased from 25 to 95 °C at 1 °C increment every 60 seconds. The melting temperature of ICL1 in the presence of MgCl₂ was 43.0 °C. The melting temperatures of ICL1 in the presence of bromopyruvate, itaconate, nitropropionate and methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate were 52.5 °C, 53.3 °C, 40.9 °C and 37.6 °C respectively.
3.5 Conclusion

By using ICL1 as a model system, the general applicability of a combined $^1$H NMR and thermal shift assays to screen for and evaluate ICL inhibitors was demonstrated. Both methods are relatively simple to carry out. In contrast to current fluorescence-based assays that rely on enzyme or chemically coupled reactions, the NMR assay enables a direct observation of substrate consumption and product formation and is therefore less prone to errors. One minor drawback of the NMR assay is the low throughput associated with monitoring reactions in real time. Typically, around 15 to 20 minutes of measurement time (six to eleven $^1$H experiments) was needed to obtain an initial rate. This equates to around seven hours of total measurement time to obtain a full kinetic analysis or a complete IC$_{50}$ curve with, for example, six concentration points in triplicate. In contrast, the thermal shift assay is a high-throughput method enabling semi-automatic measurements using multi-well plates. Ligand binding to ICL1 was easily identified through a change in the protein’s melting temperature. Interestingly, we observed both positive (i.e. stabilising) and negative (i.e. destabilising) thermal shifts of ICL1 with four known inhibitors, likely due to the inhibitors keeping ICL1 in either the open or closed conformations. A very informative, but low throughput NMR method and a high throughput, but less informative T$_m$ shift assays complement each other when carrying out ICL inhibition studies.
Chapter 4: Structural and biochemical studies of *M. tuberculosis* isocitrate lyase 2 (ICL2)

4.1 Introduction

4.1.1 *M. tuberculosis* ICL2

The physiological role of ICL2 in *M. tuberculosis* lipid and fatty acid metabolism is not clear. There appears to be a discrepancy between *in vitro* experiments, which showed that ICL2 possesses poor isocitrate lyase activity and no methylisocitrate lyase activity, and cell-based *in vivo* experiments, which showed that ICL1 and ICL2 are jointly required for the growth and virulence of *M. tuberculosis*. The function of an enzyme is determined by its three dimensional structure. Sequence analyses indicated that ICL2 contains an N-terminal catalytic domain and a C-terminal domain of unknown function. Homology modelling by Munoz-Elias and co-workers showed high sequence and structure similarities between the catalytic domains of ICL1 and ICL2. The atomic details of full length ICL2, and in particular the C-terminal domain, are therefore of particular importance to understand the biochemical and physiological function of ICL2.

4.1.2 Structure determination using X-ray crystallography

There are a few techniques to study the structure of biological macromolecules. These include X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (cryo-EM). X-ray crystallography is the most common method for solving protein structures. Protein NMR spectroscopy is not applicable to ICL2 due to its size, whilst cryo-EM is a promising technique but requires specialist equipment and image processing methods.

---

1 The work that is described in this chapter was conducted in collaboration with Dr Ghader Bashiri, Dr Jóhannes Reynisson and Anne Swartjes. G.B. provided supervision and technical support in the determination of the ICL2 structures. He also performed SAXS experiments of ICL2. J.R. conducted molecular dynamics experiments. A.S. synthesised (2R,3S)-2-methylisocitrate
The first protein structure that was solved by X-ray crystallography was myoglobin. It was solved by Kendrew et al in 1958. In 1960, the structure of haemoglobin was reported by Perutz et al. Although both reported structures were of low resolution (especially when compared to modern standards), it was a significant breakthrough in structural biology and both Kendrew and Perutz were awarded the Nobel Prize in chemistry in 1962 for this work. The first high resolution image of a protein (hen egg-white lysozyme) was published in 1965. To date, over 111800 protein structures have been solved by X-ray crystallography, which accounts for over 90% of the total protein structures that have been deposited in the protein data bank (PDB).

The first crucial step of protein crystallography is to obtain high-quality diffracting protein crystals, which typically relies on the purity and stability of the sample. The principle of crystallisation is to bring the sample to a state of supersaturation and finally nucleation. The major factors that influence crystal nucleus formation are physicochemical parameters, such as pH, temperature and ionic strength, and the concentration of precipitating agents, such as salt (e.g. ammonium sulfate), organic solvent (e.g. ethanol) and highly soluble polymers (e.g. polyethylene glycol). Optimising the physicochemical parameters may decrease the solubility of a protein, thus enabling the protein to reach the supersaturation state and subsequently precipitation or crystallisation. The addition of precipitating agents may also facilitate protein crystallisation, as the precipitating agents will compete with the protein for water. The most common methods for protein crystallisation are vapour diffusion (hanging or sitting drop), batch crystallisation, dialysis, free interface diffusion and microfluidics. After obtaining a high quality protein crystal, the next step involves X-ray diffraction. Sources of X-ray beam ranges from commercial laboratory-based X-ray source (e.g. rotating anode) or synchrotron X-ray source. The X-ray diffraction pattern can then be processed to determine protein structures using either molecular replacement (MR) or experimental phasing methods.
such as multiple isomorphous replacement (MIR)\textsuperscript{10} and multiwavelength anomalous diffraction (MAD)\textsuperscript{18}.

4.2 Objectives
The objectives of the work described in this chapter were to solve the crystal structure of ICL2 and investigate its biochemical function. These results in turn provide a fundamental understanding on the role of ICL2 in \textit{M. tuberculosis} metabolism.

4.3 Crystallography of ICL2
After successful purification of stable ICL2 (see Chapter 2), initial screening of ICL2 crystallisation was carried out by sitting drop vapour diffusion method using three commercial screens JCSG-\textit{plus},\textsuperscript{19} Morpheus\textsuperscript{20} and PACT \textit{premier}\textsuperscript{21}. These experiments were performed at a protein concentration of 13 mg/mL in the buffer containing 20 mM HEPES, 150 mM NaCl and 1 mM \(\beta\)-mercaptoethanol. None of the tested conditions facilitated ICL2 crystal growth, with protein precipitation observed in majority of the drops. The concentration of protein sample and the presence of a ligand (that may stabilise the protein) are two important factors that may facilitate crystallisation.\textsuperscript{22} Keeping these two factors in mind, co-crystallisation of ICL2 with succinate, one of the products of the ICL2-catalysed reaction, was attempted. As protein precipitation was observed at 13 mg/mL, the protein concentration was reduced to 8 mg/mL.

After 15 days at 18 \(^\circ\)C, rod shaped crystals were observed with the Morpheus screen conditions \textbf{D2} [10\% \textit{w/v} PEG 8K, 20\% \textit{v/v} ethylene glycol 0.02 M of each alcohol (0.02 M 1,6-hexanediol, 0.02 M 1-butanol, 0.02 M \((R,S)\)-1,2-propanediol, 0.02 M 2-propanol, 0.02 M 1,4-butanediol, 0.02 M 1,3-propanediol) and 0.1 M MES/imidazole pH 6.5 (Figure 4.1a)] and \textbf{H3} [(10\% \textit{w/v} PEG 4K, 20\% \textit{v/v} glycerol, 0.02 M amino acids [0.02 M sodium L-glutamate, 0.02 M D L-}
alanine, 0.02 M glycine, 0.02 M DL-lysine-HCl, 0.02 M DL-serine) and 0.1 M MES/imidazole pH 6.5 (Figure 4.1b)].

**Figure 4.1:** ICL2 crystals were observed after 15 days at 18 °C in the Morpheus screen conditions D2 (a) and H3 (b).

A fine screen was subsequently prepared based on the successful crystallisation conditions (Table 4.1). Crystals were grown in several conditions after 15 days at 18 °C. Ten crystals were harvested, flash-frozen in liquid nitrogen and sent to the Australian Synchrotron for X-ray diffraction. The best diffracting crystal was grown in the condition composed of 9% w/v PEG 4K, 18% v/v glycerol, 0.02 M amino acids, 0.1 M MES/imidazole pH 6.9. A full dataset was collected at 2.9 Å resolution, which was then processed and used to solve the initial ICL2 structure by MR using the structure of *Aspergillus nidulans* isocitrate lyase (An-ICL, PDB code: 1DQU) as a search model. An-ICL shows 35% amino acid sequence identity with ICL2, covering only the N-terminal part of the latter.

To increase the diffraction resolution, the above-mentioned condition from the fine screen was used as a base condition to screen the Morpheus additive screen. The drops were set up in a 96-well crystal tray, in which the additive screen was added to the crystallisation conditions to make up the 10% of the total volume. Cubic crystals were grown in various conditions, which
were tested at the Australian Synchrotron. A crystal grown in the presence of sodium tungstate as the additive diffracted to 1.8 Å resolution, which was used to solve the structure by MR using the previously solved ICL2. A possible reason of the increase in the resolution is the enhancement of crystal packing and stability in the presence of sodium tungstate. The enhancement of crystal packing and stability by crosslinking monomers in the presence of polyoxometalate including polyoxotungstate via electrostatic interactions between negatively charged metal and positively charged surface of the protein as well as hydrogen bonding has been reported previously.24

Since ICL2 possesses a C-terminal part that is not present in the original search model (An-ICL), with no apparent sequence homology to known sequences in the databank, the resulting MR structure contained only the N-terminal domain of the protein. There was, however, clear electron density present for the missing C-terminal domain. A combination of automated (Phenix) and manual (Coot) model building was used to trace the C-terminal domain, resulting in a complete model. The final structure was refined to 1.8 Å resolution, with crystallographic \( R_{\text{value}} \) of 23.41% and \( R_{\text{free}} \) of 25.55%. Ramachandran plots produced by MolProbity25 indicated that 98.9% of the residues are within favoured regions with no outliers.

All data processing, structure determination and structure analyses were performed with Dr Ghader Bashiri.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>PEG4K/16%</td>
<td>PEG8K/18%</td>
<td>PEG8K/20%</td>
<td>PEG8K/22%</td>
<td>PEG8K/24%</td>
<td>PEG8K/16%</td>
<td>PEG8K/18%</td>
<td>PEG8K/20%</td>
<td>PEG8K/22%</td>
<td>PEG8K/24%</td>
</tr>
<tr>
<td></td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
</tr>
<tr>
<td></td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
</tr>
<tr>
<td></td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
</tr>
<tr>
<td></td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
</tr>
<tr>
<td>B</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>PEG4K/16%</td>
<td>PEG8K/18%</td>
<td>PEG8K/20%</td>
<td>PEG8K/22%</td>
<td>PEG8K/24%</td>
<td>PEG8K/16%</td>
<td>PEG8K/18%</td>
<td>PEG8K/20%</td>
<td>PEG8K/22%</td>
<td>PEG8K/24%</td>
</tr>
<tr>
<td></td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
</tr>
<tr>
<td></td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
</tr>
<tr>
<td></td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
</tr>
<tr>
<td></td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
</tr>
<tr>
<td>C</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>PEG4K/16%</td>
<td>PEG8K/18%</td>
<td>PEG8K/20%</td>
<td>PEG8K/22%</td>
<td>PEG8K/24%</td>
<td>PEG8K/16%</td>
<td>PEG8K/18%</td>
<td>PEG8K/20%</td>
<td>PEG8K/22%</td>
<td>PEG8K/24%</td>
</tr>
<tr>
<td></td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.03M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
<td>EG, 0.02M</td>
</tr>
<tr>
<td></td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>EG, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
</tr>
<tr>
<td></td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
</tr>
<tr>
<td></td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
<td>pH 6.9</td>
</tr>
<tr>
<td>D</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>PEG4K/16%</td>
<td>PEG4K/18%</td>
<td>PEG4K/20%</td>
<td>PEG4K/22%</td>
<td>PEG4K/24%</td>
<td>PEG4K/16%</td>
<td>PEG4K/18%</td>
<td>PEG4K/20%</td>
<td>PEG4K/22%</td>
<td>PEG4K/24%</td>
</tr>
<tr>
<td></td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
</tr>
<tr>
<td></td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
</tr>
<tr>
<td></td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
<td>pH 6.1</td>
</tr>
<tr>
<td>E</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
<td>8%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>PEG4K/16%</td>
<td>PEG4K/18%</td>
<td>PEG4K/20%</td>
<td>PEG4K/22%</td>
<td>PEG4K/24%</td>
<td>PEG4K/16%</td>
<td>PEG4K/18%</td>
<td>PEG4K/20%</td>
<td>PEG4K/22%</td>
<td>PEG4K/24%</td>
</tr>
<tr>
<td></td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>GLY, 0.02M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
<td>AA, 0.1M</td>
</tr>
<tr>
<td></td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>ALC, 0.1M</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
</tr>
<tr>
<td></td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>MES/IMI</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
</tr>
<tr>
<td></td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
<td>pH 6.5</td>
</tr>
<tr>
<td>F</td>
<td>8% PEG4K/16% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>9% PEG4K/18% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>10% PEG4K/20% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>11% PEG4K/22% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>12% PEG4K/24% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>8% PEG4K/16% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>9% PEG4K/18% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>10% PEG4K/20% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>11% PEG4K/22% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>12% PEG4K/24% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</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>
</tr>
<tr>
<td>F</td>
<td>8% PEG4K/16% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>9% PEG4K/18% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>10% PEG4K/20% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>11% PEG4K/22% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>12% PEG4K/24% GLY, 0.02M ALC, 0.1M MES/IMI pH 6.9</td>
<td>8% PEG4K/16% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>9% PEG4K/18% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>10% PEG4K/20% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>11% PEG4K/22% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
<td>12% PEG4K/24% GLY, 0.02M AA, 0.1M MES/IMI pH 6.9</td>
</tr>
</tbody>
</table>

**Table 4.1:** Fine screening conditions based on Morpheus D2 and H3. PEG = polyethylene glycol, EG = ethylene glycol, GLY = glycerol, AA = amino acids, ALC = alcohols, MES = 2-(N-morpholino)ethanesulfonic acid and IMI = Imidazole
4.4 Structure of ICL2

The ICL2 crystals belong to the $P2_1$ space group, with four monomers packed into the asymmetric unit forming a tetramer, which is the likely biological unit of this protein. ICL2 forms an elongated (200 Å) structure (Figure 4.2), with clear and well-defined electron density map for the entire molecule (residues 1-766), except for residues 1-10 and 593-605. There was, however, no apparent electron density for succinate, which was co-crystallised with the protein. Each monomer is composed of 28 $\alpha$-helices and 18 $\beta$-strands, arranged as two distinct N-terminal (residues 1-590) and C-terminal (residues 600-766) domains (Figure 4.3). The presence of the C-terminal domain is a striking feature of ICL2, as mycobacterial ICL1 comprises of only a single domain\textsuperscript{26}, which is comparable to the N-terminal domain of ICL2.

![Illustration of ICL2 structure]

Figure 4.2: The structure of ICL2 tetramer solved at 1.8 Å resolution. The picture shows distinct N- and C-terminal domains. Each colour represents a monomeric subunit. An N-terminal insertion (based on the sequence and structure alignments of ICL2 to ICL1) is present in each of the monomers, an example of which is shown in the brown square box.
The N-terminal domain is composed of 23 α-helices (α1-α23) and 11 β-strands (β1-β11). Superposition of ICL2 onto ICL1 gives an average root mean square difference (rmsd) of 1.24 Å over 390 Cα atoms, with 38% sequence identity (Figure 4.4). The alignments (structure and sequence) indicate a large insertion in ICL2 between residues 278 and 427 (Figure 4.4 and Table 4.2), which adds a helical structure (helices α10-α16) that plays a role in the interaction to the C-terminal domain (Figure 4.2).

The ICL2 C-terminal domain is comprised of five α-helices (α24-α28) and seven β-strands (β12-β18). Whereas the C-terminal domain shows no sequence homology to any known proteins, a structural search using PDBeFold suggests that this domain is structurally related to members of the Gcn5-related N-acetyl transferase (GNAT) superfamily, despite showing only 5-15% sequence homology to these proteins. In fact, the members of the GNAT superfamily possess low degree of sequence homology (3-23%). Nevertheless, the N-terminus of the members retains a conserved fold consisting of six or seven β-strands and four α-helices (Figure 4.5).27-30 This fold comprises of a first β-strand followed by two α-helices (α1 and α2), followed by three antiparallel β-strands (β2, β3, β4) followed by a central helix (α3), a fifth β-strand (β5), a fourth α-helix (α4) and a final β-strand (β6). The loop linking β4 to α3 in this conserved fold comprises a pyrophosphate binding site which shows a consensus sequence (R/Q-X-X-G-X-A/G) among the members of the GNAT superfamily. Another essential feature in the structure is the presence of a β-bulge at β4 forming V-shape for binding of acetyl-CoA.

A close analysis of the ICL2 structure indicates that most of the structural features of the GNAT superfamily (Figure 4.6) are conserved in the C-terminal domain. It consists β1 (β12 in ICL2), β2 (β13), β3 (β14), β4 (β15), followed by α3 (α25 in ICL2), β5 (β16), α4 (α26) and β6 (β17). Whereas the β-bulge (β15 in ICL2) seems to be conserved in the structure, the signature pyrophosphate binding sequence (R/Q-X-X-G-X-A/G) is not present. Furthermore, two α-helices (between β12 and β13 in ICL2) are missing compared with a typical member of GNAT
superfamily. The variation in the number of α-helices in members of GNAT superfamily has been reported previously. For example, a histone N-acetyl transferase from Saccharomyces cerevisiae (yHat1) misses α2 and an aminoglycoside-6'-N-acetyl transferase from Serratia marcescens [AAC(6')] possess additional α helix in between β1 and β2.

Figure 4.3: a) ICL2 monomer visualising N- and C-terminal domains. The loop connecting N- and C-terminal domains (residues 593-605) does not possess clear and well-defined electron density map; b) Topology of ICL2 monomer visualising 28 α-helices and 18 β-strands. The N-terminal domain composed of 23 α-helices and 11 β-strands and the C-terminal domain
composed of 5 α-helices and 7 β-strands. Green represents α-helices and yellow represents β-strands.

| ICL2 | ARAAPETCTEVHTPFGQFQGKDVATKRYEL-SSHRAGLULBYARQVEQ |
| ICL1 | ...........................................SAEAGQKDAQSFPRKDYTVYASKDVEDVAV |

| ICL2 | 50 RCTPTVFVHLABEANAGFEELRRELCAARKTTTFIRPTPGAVSMKQ |
| ICL1 | 39 QCSSVEHEAEAMRELQAEVLKHD---EWNNAASLAEINNAYQQEAE |

| ICL2 | 100 BBAYLDRATSKRSSEDPGFLASPLSLQDDAVVYRALLTQSNR |
| ICL1 | 85 BKAIALOSEQAVDDNLISHTPDQSSFLANSVQVYRVNVALQADQI |

| ICL2 | 150 QHYLRLQMSKQQRPAAYPDFPFAADALDHGQPHYRNLFAIFEV |
| ICL1 | 135 --AIEEDPSVENLAPILVGDAGSGLNYEYLGAA |

| ICL2 | 200 VYEEVRILDRGTPJKCGGSGKVLVECQICTRNARPEQDQMLMNVI |
| ICL1 | 176 VASAPEKALASEKCGBPSTGKVLETQQHITTSARAEADVFTV |

| ICL2 | 250 LARTDAEAAALDRAGERDQFPFLLGATKLDPYKSCFLAMVRFFYELG |
| ICL1 | 226 LARTDAEAAALDRAGERDQFPFLLG-- |

| ICL2 | 300 VKELNHLLYALGDOSEYAAAGGWRERQGFGLVSADAVNAWREDGQSIDG |
| ICL1 | 253 -- |

| ICL2 | 350 IFDQVESRFVAWEDDAGLMYGEAVADVLEFGQSEEGFIMAPEEWRAP |
| ICL1 | 253 -- |

| ICL2 | 400 AARASLHAARAKAHELGDPPWDELAATGCGOIGPGPYAIAKSLA |
| ICL1 | 253 --TEEGDRTYKNIICPDAARKA |

| ICL2 | 450 APFADLWMTTARLDRARQPEAVKAEPDQMLAYNSPSFNQGITT |
| ICL1 | 276 APFADLWMTTARLDRARQPEAVKAEPDQMLAYNSPSFNQK-H |

| ICL2 | 500 TEATAKREPEQLSKQGIGVFTYEGHQIDGVAAEFADLRQDGLLA |
| ICL1 | 325 LATAKREPEQLSKQGIGVFTYEGHQIDGVAAEFADLRQDGLLA |

| ICL2 | 550 KLRKMRMLVES-FRT---FELVGFRSDDALASSGRTATTKAMGKST |
| ICL1 | 375 KLRKMRMLVES-FRT---FELVGFRSQQSEMQLS |

| ICL2 | 597 OQHQLVQTEVPRLLEELWLLAMWSSGQHLKDKLRVQLRPQAGSEVLGL |
| ICL1 | 403 -- |

| ICL2 | 647 HGESDDKLANVFQFDQDRGRTILLVRDQNTFGAHELQKRLMTHLW |
| ICL1 | 403 -- |

| ICL2 | 697 VHRFKAQAVHYVTPTDDNLQYTSKMKSHGFTEVNGQEVGETIAYV |
| ICL1 | 403 -- |

| ICL2 | 747 AELIDDRVALRLK-ITEE--- |
| ICL1 | 405 IAETTDENSSTTALTGSTEIQ |

**Table 4.2:** Sequence alignment of ICL2 with ICL1. Highlights on black represents sequence identity, gray represents sequence similarity and yellow represents insertion in ICL2.
Figure 4.4: Superposition of the N-terminal domain of ICL2 onto ICL1. Green represents the N-terminal domain of ICL2 and magenta represents ICL1. Helices (α10-α16) represent insertion in ICL2 between residues 278 and 427.
Figure 4.5: Conserved structure of Gcn5-related N-acetyl transferase (GNAT) superfamily composed of 4 $\alpha$-helices, 7 $\beta$-strands and a pyrophosphate binding loop between $\beta4$ and $\alpha3$. Green colour represents $\alpha$-helices, yellow represents $\beta$-strands.
Figure 4.6: Structure of C-terminal domain of ICL2 composed of 5 α-helices, 7 β-strands and a pyrophosphate binding loop. Green represents α-helices, yellow represents β-strands.
4.5 Kinetic studies of ICL2

4.5.1 Kinetic of ICL2 in the presence of acetyl-CoA using the substrate DL-isocitrate

Given the relevance and importance of acetyl-CoA in the mycobacterial glyoxylate and methylcitrate cycles and the structural similarity of the ICL2 C-terminal domain to the GNAT superfamily, the role of acetyl-CoA on the catalytic activity of ICL2 was investigated. Using the NMR method described in Chapter 3, the activity of ICL2 was monitored with DL-isocitrate as a substrate, in the absence and presence of acetyl-CoA. A significant increase in ICL2 activity was observed when acetyl-CoA was added to the reaction mixture (Figure 4.7).

![Figure 4.7: Catalytic activity of ICL2 in the presence of acetyl-CoA. The sample contained 500 nM ICL2, 1 mM DL-isocitrate, 200 μM acetyl-CoA, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). Error bars are difficult to see because of the little variation between data.](image-url)
Under the experimental conditions (0.2 μM ICL2), only a small amount of acetyl-CoA (~10 μM) was found to be required to stimulate full ICL2 activity (Figure 4.8). We then measured the catalytic rate constant ($k_{\text{cat}}$) and Michaelis constant ($K_M$) of ICL2 in the absence and presence of acetyl-CoA (Figure 4.9a and 4.9b, Table 4.3). The $k_{\text{cat}}$ of ICL2 in the absence and presence of acetyl-CoA were $0.44 \pm 0.03 \text{ s}^{-1}$ and $4.1 \pm 0.3 \text{ s}^{-1}$, respectively. This shows that the catalytic activity of ICL2 was ~10 times higher in the presence of acetyl-CoA. The $k_{\text{cat}}$ value of ICL2 in the presence of acetyl-CoA ($4.1 \pm 0.03 \text{ s}^{-1}$) is similar to the value of ICL1 ($4.3 \pm 0.1 \text{ s}^{-1}$, see Chapter 3), suggesting the catalytic activity of ICL2 in the presence of acetyl-CoA is similar to ICL1. The $K_M$ values of ICL2 in the absence and presence of acetyl-CoA were $810 \pm 35 \mu\text{M}$ and $320 \pm 35 \mu\text{M}$, respectively. The $K_M$ value of ICL2 in the presence of acetyl-CoA ($320 \pm 35 \mu\text{M}$) is similar to ICL1 ($290 \pm 10 \mu\text{M}$, see Chapter 3). The results revealed that ICL2 binds DL-isocitrate more tightly in the presence of acetyl-CoA. Overall, ICL2 appears to have higher catalytic activity and tighter binding to its substrate in the presence of acetyl-CoA, suggesting a potential functional role in the *M. tuberculosis* glyoxylate cycle.
**Figure 4.8:** The measurement of the amount of acetyl-CoA that is required for full activation of ICL2 with dl-isocitrate as substrate. The sample contained 200 nM ICL2, 1 mM dl-isocitrate, varying concentrations of acetyl-CoA, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). The line is added to aid visualisation.
Figure 4.9: Hanes-Wolf plot of ICL2 (a) in the absence and (b) presence of acetyl-CoA. Sample contained 200 nM ICL2, 25 µM acetyl-CoA, varying concentrations of DL-isocitrate, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).
<table>
<thead>
<tr>
<th></th>
<th>ICL2</th>
<th>ICL1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>in the absence of acetyl-CoA</td>
<td>in the presence of acetyl-CoA</td>
</tr>
<tr>
<td>$K_M$</td>
<td>$810 \pm 35 \mu M$</td>
<td>$320 \pm 35$</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>$0.44 \pm 0.03 \text{ s}^{-1}$</td>
<td>$4.1 \pm 0.3 \text{ s}^{-1}$</td>
</tr>
</tbody>
</table>

Table 4.3: Kinetic parameters of ICL2 (in the absence and presence of acetyl-CoA) and ICL1 with DL-isocitrate as substrate. Measurements were made using samples contained ICL2 (200 nM)/ICL1 (190 nM), varying concentration of DL-isocitrate, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).

4.5.2 Kinetics parameters of ICL2 in the presence of propionyl-CoA, succinyl-CoA and CoA using the substrate DL-isocitrate

The higher activity of ICL2 with acetyl CoA encouraged us to investigate its activity in the presence of propionyl-CoA, succinyl-CoA and CoA. At 200 $\mu$M concentration, propionyl-CoA was found to stimulate the catalytic activity of ICL2 to a similar extent to that of acetyl-CoA. Succinyl-CoA and CoA were also found to increase the activity of ICL2, albeit to a lesser extent (Figure 4.10). This is interesting because a recent study by Bi et al. showed that the activity of ICLs may be regulated by posttranslational acetylation.\textsuperscript{31} GNATs are acetyltransferases and it is possible that the C-terminal domain auto-catalyses the acetylation of ICL2 to stimulate its catalytic activity. However, the activation of ICL2 by CoA, which lacks the acetyl group, demonstrates that the C-terminal domain is likely an acetyl-CoA binding domain rather than as an acetyltransferase.

To ensure ICL2 activation is due to CoA binding rather than acetate or propionate, the rate of DL-isocitrate turnover to succinate and glyoxylate was investigated in the presence of acetate and propionate (Figure 4.11). The results showed the turnover rate is not affected by acetate and
propionate, indicating that the activation takes place upon binding of acetyl CoA and propionyl-CoA.

The amount of propionyl-CoA required for complete activation of ICL2 was subsequently measured (Figure 4.12), indicating that ~40 μM of propionyl-CoA is required to fully stimulate ICL2 activity (with dl-isocitrate as substrate). Propionyl-CoA is one of the intermediates of the methylcitrate cycle, which is a key detoxification pathway of *M. tuberculosis* when odd-chain fatty acids are used as a carbon source. Therefore, the activation of ICL2 by propionyl-CoA is potentially significant as it may provide a role for ICL2 in the methylcitrate cycle of *M. tuberculosis* (see next section).
Figure 4.10: Catalytic activity of ICL2 in the presence of CoA analogues using dl-isocitrate as substrate. The graph shows the activation of ICL2 in the presence of CoA analogues (acetyl-CoA, propionyl-CoA, succinyl-CoA and CoA). Acetyl-CoA and propionyl-CoA showed the highest activation, whereas succinyl-CoA and CoA showed a half activity compared to acetyl-CoA and propionyl-CoA. The sample contained 500 nM ICL2, 1 mM dl-isocitrate, 200 μM propionyl-CoA/succinyl-CoA/CoA, 5 mM MgCl2 and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H2O and 10% D2O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).
Figure 4.11: Catalytic activity of ICL2 using DL-isocitrate as substrate in the presence of acetate and propionate. The graph shows no change in the activity of ICL2 in the presence of acetate and propionate. Measurements were made using samples contained 200 µM ICL2, 1 mM DL-isocitrate, 5 mM MgCl2, 1mM acetate/propionate and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H2O and 10% D2O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).
Figure 4.12: The measurement of the amount of propionyl-CoA required for full activation of ICL2 during DL-isocitrate turnover. Sample contained 200 nM ICL2, 1 mM DL-isocitrate, varying concentrations of propionyl-CoA, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). Error bars are difficult to see because of little variation between data. The line is added to aid visualisation.

4.5.3 Kinetics of ICL2 using (2R,3S)-2-methylisocitrate as substrate

The methylisocitrate lyase activity of ICL2 was investigated using (2R,3S)-2-methylisocitrate, an intermediate of the methylcitrate cycle, as the substrate. In the absence of acetyl-CoA or propionyl-CoA, ICL2 at 4 µM concentration showed very little or no turnover of (2R,3S)-2-methylisocitrate to succinate and pyruvate (Figure 4.13a). However, upon addition of acetyl-CoA and propionyl-CoA, fourfold less ICL2 (1 µM) showed a significant increase in the
reaction rate, as revealed by an increase in the concentrations of succinate and pyruvate (Figure 4.13b). The amount of acetyl-CoA and propionyl-CoA required for full activation of ICL2 with (2R,3S)-2-methylisocitrate as substrate were ~25 µM (Figure 4.14) and ~200 µM (Figure 4.15) respectively.

**Figure 4.13:** $^1$H NMR spectroscopy to monitor ICL2-catalysed turnover of (2R,3S)-2-methylisocitrate into succinate and pyruvate (a) in the absence and (b) presence of acetyl-CoA. (a) The ample contained 4 µM ICL2, 1 mM (2R,3S)-2-methylisocitrate, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. (b) The sample contained 1 µM ICL2, 1 mM (2R,3S)-2-methylisocitrate, 25 µM acetyl-CoA, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O.
Figure 4.14: The measurement of amount of acetyl-CoA required for full activation of ICL2 for turnover of (2R,3S)-2-methylisocitrate. Samples contained 1 µM ICL2, 1 mM (2R,3S)-2-methylisocitrate, varying concentrations of acetyl-CoA, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). Error bars are difficult to see because of little variation between data. The line is added to aid visualisation.
Figure 4.15: The measurement of amount of propionyl-CoA required for full activation of ICL2 for turnover of (2R,3S)-2-methylisocitrate. Samples contained 1 µM ICL2, 1 mM (2R,3S)-2-methylisocitrate, varying concentrations of propionyl-CoA, 5 mM MgCl₂ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H₂O and 10% D₂O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). Error bars are difficult to see because of little variation between data. The line is added to aid visualisation.

The $K_M$ and $k_{cat}$ values of ICL2 were then measured using (2R,3S)-2-methylisocitrate as substrate in the presence of acetyl-CoA (Figure 4.16 and Table 4.4). The $K_M$ of ICL2 is 915 ± 65 µM, which is three times higher compared when to DL-isocitrate was used as substrate (320 ± 35 µM). The $k_{cat}$ of ICL2 using (2R,3S)-2-methylisocitrate as substrate is 0.89 ± 0.042 s⁻¹.
suggesting the turnover rate of \((2R,3S)\)-2-methylisocitrate is much slower than \(DL\)-isocitrate \((k_{cat} = 4.1 \pm 0.3 \text{ s}^{-1})\). This result is not surprising as the additional methyl group may affect the binding affinity and geometry of the substrate to the enzyme active site. A similar trend was also observed for ICL1 (Chapter 3). Similar to ICL1 (see Chapter 3), inhibition of ICL2 at high \((2R,3S)\)-2-methylisocitrate concentration (above 1000 \(\mu\)M) was observed. (Figure 4.17). This is not surprising as the structure and active site of the N-terminal domain of ICL2 and ICL1 are very similar.

**Figure 4.16**: Hanes-Woolf plot of ICL2 using \((2R,3S)\)-2-methylisocitrate as a substrate. Samples contained 1 \(\mu\)M ICL2, 25 \(\mu\)M acetyl-CoA, varying concentrations of \((2R,3S)\)-2-methylisocitrate, 5 mM MgCl\(_2\) and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H\(_2\)O and 10% D\(_2\)O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). Error bars are difficult to see because of little variation between data.
<table>
<thead>
<tr>
<th></th>
<th><strong>DL-isocitrate</strong></th>
<th></th>
<th><strong>(2R,3S)-2-methylisocitrate</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_M$ (μM)</td>
<td>290 ± 10 μM</td>
<td>560 ± 28</td>
<td>320 ± 35 μM</td>
<td>915 ± 65</td>
</tr>
<tr>
<td>$k_{cat}$ (s$^{-1}$)</td>
<td>4.3 ± 0.1 s$^{-1}$</td>
<td>0.5 ± 0.01</td>
<td>4.1 ± 0.3 s$^{-1}$</td>
<td>0.89 ± 0.042</td>
</tr>
</tbody>
</table>

**Table 4.4:** Kinetic parameters of ICL2 (in the presence of acetyl-CoA) and ICL1 with DL-isocitrate and (2R,3S)-2-methylisocitrate as substrate. Measurements were made using samples containing ICL2 (1 μM)/ICL1 (2 μM), varying concentrations of (2R,3S)-2-methylisocitrate, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and 10% D$_2$O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement).

**Figure 4.17:** ICL2 kinetics using the substrate (2R,3S)-2-methylisocitrate in the presence of acetyl-CoA showing enzyme inhibition at higher substrate concentration (above 1000 μM). Measurements were made using samples containing 1 μM ICL2, varying concentrations of (2R,3S)-2-methylisocitrate, 5 mM MgCl$_2$ and 50 mM Tris/Tris-D11 (pH 7.5) in 90% H$_2$O and
10% D$_2$O. The errors shown are the standard deviation from three independent measurements (each sample prepared separately for the measurement). Error bars are difficult to see because of little variation between data. The line is added to aid visualisation.

4.6 Thermal shift assay of ICL2 in the presence of acetyl-CoA

The aforementioned kinetic experiments showing activation of ICL2 in the presence of acetyl-CoA were also supported by a thermal shift assay (Figure 4.18). The melting temperature ($T_m$) of ICL2 in the presence of MgCl$_2$ was 50.9 °C. Incubation of acetyl-CoA with ICL2 in the presence of MgCl$_2$ increased the $T_m$ by 1.9 °C (52.8 °C). This demonstrates acetyl-CoA binds to ICL2 to form a stable complex. The increase in $T_m$ as a result of a ligand-protein interaction has been reported previously; for example, the binding of tamoxifen to the estrogen receptor appeared to increase $T_m$ as a result of a protein-ligand stable complex. The increase in the $T_m$ of the protein upon binding of the ligand was also discussed in Chapter 3 of this thesis.
Figure 4.18: Protein melt curve for ICL2. Samples contained 20 μM ICL2, 100 μM MgCl₂ (where required) and 100 μM acetyl-CoA (where required) in 50 mM HEPES pH 7.5. Temperature was increased from 25 to 95 °C at 1 °C increment every 60 seconds. The melting temperature of ICL2 in the absence and presence of MgCl₂ was 49.7 °C and 50.9 °C. The melting temperatures of ICL1 in the presence of acetyl-CoA was 52.8 °C.

4.7 Crystallisation of ICL2 in the presence of acetyl-CoA

Considering the interesting kinetic results for ICL2 in the presence of acetyl-CoA and analogues, protein crystallography was subsequently attempted to solve the structure of ICL2 in the presence of acetyl-CoA. Using optimal conditions described previously (fine screening based on Morpheus D2 and H3) (see Table 4.1), crystals of acetyl-CoA bound ICL2 were obtained under two conditions (9 % PEG 4K/18 % GLY, 0.02 M AAs, 0.1 M MES/IMI pH 6.1 and pH 6.5). The structure was solved by molecular replacement using the acetyl-CoA free ICL2 as a search model, in which the N-terminal domain was first used to solve the structure.
followed by using this partially solved structure as a fixed model to search for the C-terminal domain. Model building using Phenix and Coot resulted in a complete model that was refined to 2.67 Å resolution, with crystallographic R<sub>value</sub> of 22.3 % and R<sub>free</sub> of 24.39 %. Ramachandran plots produced by MolProbity<sup>25</sup> showed that 97.45% of the residues are within the favoured regions with 0.28% outliers.

The acetyl-CoA-bound ICL2 structure also showed a tetrameric arrangement (Figure 4.19), with a clear electron density map for the entire molecule, except residues 1-10, 211-220 and 593-605. The structure revealed one molecule of acetyl-CoA bound to each of the C-terminal domains, inducing unprecedented conformational changes in ICL2. Upon binding of acetyl-CoA, the C-terminal domain from one monomer moves an average of 77 Å towards the centre of ICL2 and rotates ~176 degrees to form a dimer with the C-terminal domain of the opposing monomer (Figure 4.20). The distinctive conformational changes have also been confirmed in solution using small-angle X-ray scattering (SAXS) experiments, with acetyl- and propionyl-CoA (performed by Dr Ghader Bashiri).
Figure 4.19: Structure of the acetyl-CoA bound ICL2 tetramer solved at 2.67 Å. Each colour represents a monomeric subunit. Acetyl-CoA (coloured by element in PyMOL - cyan carbon backbone) and a water molecule (blue white coloured spheres) that involves in the interaction of pyrophosphate moiety to AA residues of ICL2 C-terminal domain.
Figure 4.20: Structures of acetyl-CoA free and bound ICL2 visualising the conformation change upon binding of acetyl-CoA. a) acetyl-CoA free ICL2; b) acetyl-CoA bound ICL2.

Binding of acetyl-CoA induces the conformational change to form a closed confirmation.

(Acetyl-CoA is coloured by element in PyMOL - cyan carbon backbone)

Acetyl-CoA binds in the cleft created by α25, α26 and β15 (Figure 4.21a and 4.22). The adenosine moiety makes only one hydrogen bond interaction with the side chain of Asp713, with the P3 phosphate group interacting with the side chains of Arg687, Lys764 and His724 residues (Figure 4.22). The pantothenate tail makes hydrogen bond interactions with Thr678, Thr711 and Gln676 residues. Similarly to members of the GNAT superfamily (Figure 4.21b), a β-bulge (on β15) participates in creating a hydrophobic pocket in which the acetyl moiety is tucked in between β15 and β16-α26 loop. The longer propionyl and succinyl moieties of
propionyl-CoA and succinyl-CoA could also fit in the hydrophobic cage formed by Val673, Gln676, His693, Tyr707 and Asn714 residues. The binding of the pyrophosphate moiety of acetyl-CoA is interesting because the R/Q-X-X-G-X-A/G sequence that is conserved amongst all GNATs (present in the loop connecting β4 and a3) for pyrophosphate binding is not conserved in ICL2. In members of GNAT superfamily, acetyl-CoA interacts via hydrogen bonding between the backbone of the conserved motifs and the oxygen atoms of the pyrophosphate group. In ICL2, in the absence of the conserved R/Q-X-X-G-X-A/G motif, six residues are present in the pyrophosphate binding loop of ICL2 (the loop connecting β15 and a25) are Gly680, Ala681, Glu682, Leu683, Arg684, Glu685 and Lys686. The pyrophosphate moiety of acetyl-CoA makes hydrogen bonds to Leu683, Glu685 and Lys686 present in the loop. The pyrophosphate moiety also interacts with Thr678 of β15 and Leu688 and Thr690 of a25. The interaction of pyrophosphate with the side chain hydroxyl group of a residue in a3 has been observed in some members of GNAT, but interactions with residues in β4 have not been reported. The pyrophosphate makes four direct hydrogen bond interactions with Thr678, Gln685, Arg687 and Thr690 residues. Interestingly, the structure also shows that pyrophosphate interacts with Leu683, Lys686, Leu688 residues via either a water molecule or a Mg²⁺ ion (Figure 4.22). It was tested by molecular dynamics experiment (performed by Dr Jóhannes Reynisson), which revealed that a water molecule required for binding of pyrophosphate to Leu683, Lys686 and Leu688 residues. The interactions of the pyrophosphate with residues in the loop via water molecule have also been reported previously.
**Figure 4.21**: a) Acetyl-CoA bound C-terminal domain of ICL2. Acetyl-CoA binds in the cleft created by $\alpha_{25}$, $\alpha_{26}$ and $\beta_{15}$. An acetyl-CoA and a water molecule are shown as a cyan carbon backbone stick and a blue white sphere, respectively. Amino acid residues interact to acetyl-CoA are shown as wheat coloured sticks and labelled in black; b) Acetyl-CoA bound NatD of *Schizosaccharomyces pombe*, PDB code 4U9V. An acetyl-CoA (a cyan carbon backbone stick) binds to the cleft created by $\alpha_{3}$, $\alpha_{4}$ and $\beta_{4}$. Amino acid residues interact to acetyl-CoA are shown as wheat coloured sticks and labelled in black.
Figure 4.22 Picture exported from LIGPLOT\textsuperscript{35} visualising hydrogen bond interactions of acetyl-CoA to AA residues in the C-terminal domain. Adenosine moiety interacts with Asp713, P3 phosphate group interacts with Arg 687, His724 and Lys764, pantothenate interacts with Gln676, Thr678 and Thr711, and pyrophosphate interacts directly with Thr678, Gln685, Arg687 and Thr690 and via water with Leu683, Lys686 and Leu688.
The overall conformational changes of ICL2 upon acetyl-CoA binding are extensive. Specifically, binding of acetyl-CoA induces dissociation of the C-terminal domain dimer observed in the acetyl-CoA free ICL2 structure and a new dimer is formed with the C-terminal domain of the opposing monomer. The C-terminal domain of acetyl-CoA bound ICL2 dimerises via interactions (electrostatic and hydrogen bonding) between the loops connecting β14/β15 and β12/β13 of opposing monomer (Figure 4.23a). Specifically, the positively charged Arg665 interacts with negatively charged Glu640 at the dimer interface. Three β-strands (β16 and β17 and β18) of the acetyl-CoA free ICL2 are disordered, suggesting a higher dynamic state of the dimer. Interestingly, the C-terminal domain of acetyl-CoA bound structure is dimerised involving the same three β-strands (β16 and β17 and β18) (Figure 4.23b). These three β-strands are well-ordered on top of each other in the less dynamic state, which led to the stable dimer. Overall, the binding of acetyl-CoA induces the C-terminal domain to adopt a well-ordered and less dynamic state, which could favour to dissociate the existing dimer and form a new more stable dimer with the opposing monomer. The decrease in the flexibility of β-strands upon binding of a ligand to a protein leading to form a stable dimer has been reported previously.36
Figure 4.23: Dimerisation of the C-terminal domain of ICL2. a) In the absence of acetyl-CoA, three $\beta$-strands ($\beta_{16}$, $\beta_{17}$ and $\beta_{18}$) are disordered suggesting to higher dynamic state. b) In the presence of acetyl-CoA, three $\beta$-strands ($\beta_{16}$, $\beta_{17}$ and $\beta_{18}$) are well ordered and also involved in dimerisation.
4.8 Active site of ICL2

Although the ICL2 structures solved in this study did not contain any active site metals or ligands, sequence alignment between ICL1 and ICL2 suggested that Cys215 is the catalytic residue in the latter. The metal binding residues are also conserved between ICL1 and ICL2 [Asp108, Asp153 and Glu155 in ICL1, Figure 4.24,\textsuperscript{26} and Asp123, Asp177 and Asp179 (acidic amino acid Glu155 in ICL1) in ICL2, Figure 4.25 (see also sequence alignment, Table 4.2)]. It is therefore reasonable to assume that metal and substrate binding at the active sites of ICL1 and ICL2 should be similar.

\textbf{Figure 4.24:} a) ICL1 monomer showing the position of Mg\textsuperscript{2+} (blue sphere) and the three amino acid residues (Asp108, Asp153 and Glu155) it interacts with in the active site; b) Close up of the interaction between Mg\textsuperscript{2+} and Asp108, Asp153 and Glu155 within the ICL1 active site.
Figure 4.25: A picture visualising the conserved Mg$^{2+}$ binding site in ICL2.

ICL1 undergoes a conformational change from the open to the closed conformation upon binding of the substrate (dl-isocitrate or (2R,3S)-2-methylisocitrate). The C-terminal end of the adjacent subunit then moves towards the active site loop forming a lid over the active site loop, keeping the protein in a closed conformation. Electrostatic interactions between positively charged lysine residues (Lys189 and Lys190 of ICL1) at the active site loop and the negatively charged glutamate residues (Glu423 and Glu424 of ICL1) at the C-terminal lid are important to keep enzyme in its catalytically active conformation.$^{37}$ ICL2 also undergoes a conformational change upon acetyl-CoA binding, albeit via a different mechanism to ICL1. ICL2 lacks the negatively charged lid from the adjacent subunit, likely a reason for the weak
catalytic activity of ICL2 in the absence of acetyl-CoA. Upon binding of acetyl-CoA, rearrangement of the C-terminal domains occurs, resulting in closed tetrameric conformation. The loop connecting β13/β14 (residues 635-641) of the C-terminal is sandwiched between helices α20 and α21 of the N-terminal domain. However, the loop is ~32 Å away from the catalytic Cys215, which indicates this conformation change is not directly involved in the increase of the catalytic activity of ICL2.

Upon close observation of two adjacent monomers in both acetyl-CoA free and bound ICL2 tetramers, it is possible to provide a plausible framework for the allosteric modulation of the enzyme. The active site loop (loop connecting β4/β5) is well defined in the acetyl-CoA free structure. The loop connecting N- and C- terminal domains (residues 586-601) of the adjacent monomer is facing away from catalytic loop providing space for the active site loop. This results in an open conformation that allows the binding of Mg²⁺ and the substrate (Figure 4.26). This also explains the weak catalytic activity of ICL2 in the absence of acetyl-CoA. Upon binding of acetyl-CoA, the orientation of the C-terminal domain changes, i.e. the C-terminal domain rotates towards the N-terminal domain. The loop connecting N- and C-terminal domains from the adjacent monomer thus move closer to the space that was occupied by the active site loop in the acetyl-CoA free structure, thus pushing the active site loop closer to the active site, i.e. the close conformation in a similar manner to ICL1 (Figure 4.27). This is aided by charge-charge repulsion between the positively charged residues (Lys213 and Lys214) on the active site loop and the positively charged loop that connects the N- and C-terminal domains (pI =10). Interestingly, instead of a well-defined active site loop covering the entrance of the active site, the binding of acetyl-CoA resulted in a patchy and unresolved density for the active site loop in three chains of the tetramer. This is likely due to increase in flexibility of the active site loop. The presence of Mg²⁺ and the substrates may have stabilised the active site loop. However, the absence of Mg²⁺ and the substrates in the structure leaves the active site loop
flexible. Further experiments, including molecular dynamics calculations, may help provide further insight into the activation mechanism.

**Figure 4.26**: Two monomers of the acetyl-CoA free ICL2 showing the active site loop, metal binding site and the loop connecting N- and C- terminal domains of adjacent monomer. The picture shows the loop connecting N- and C- terminal domain of adjacent monomer facing away from the active site loop.
Figure 4.27: Two monomers of acetyl-CoA bound ICL2 showing the change in orientation of C-terminal domain upon binding of acetyl-CoA. The picture shows the loop connecting the N- and C-terminal domains of the adjacent monomer moves closer to the active site loop, resulting the patchy and unresolved active site loop. The patchy and unresolved active site loop is highlighted in red circle and the loop connecting N- and C-terminal domain is highlighted in black circle.
4.9 Conclusion

The work that is described in this chapter summarises efforts in the structural and biochemical analysis of ICL2. The structure of ICL2 was solved at 1.8 Å resolution using X-ray crystallography. ICL2 was found to be an elongated tetramer with two-distinct N- and C-terminal domains. Sequence and structure alignments demonstrated the N-terminal domain is similar to ICL1 and a structural similarity search using PDBeFold indicated the C-terminal domain is structurally similar to members of GNAT acetyltransferase superfamily. Biochemical studies revealed that the binding of acetyl-CoA (or analogues) increased the catalytic activity of ICL2. Structural analyses showed that acetyl-CoA binds to the C-terminal domain of ICL2, which induces a significant conformational change. The conformational change locks ICL2 in a closed tetrameric conformation that stimulates its catalytic activity. The conformational change was also confirmed in solution state using SAXS experiments. Overall, the experiments provided valuable insight into the long-standing dilemma on the role of ICL2 in the metabolism of *M. tuberculosis.*
Chapter 5: Itaconic acid as a green and sustainable substrate for pharmaceuticals

5.1 Introduction

5.1.1 Green chemistry

Green chemistry is the development of new technologies that decrease or exclude the use and production of harmful substances.\textsuperscript{1,2} It is based on twelve principles developed by Anastas and Warner.\textsuperscript{1} As can be seen in Figure 5.1, the key principles focus on the practice of sustainable, safe procedures that minimise waste.\textsuperscript{3}

![Twelve principles of green chemistry](image)

**Figure 5.1:** Twelve principles of green chemistry.\textsuperscript{1}

\textsuperscript{1} Part of the work from this chapter was published as described in Bhusal, R. P.; Sperry, J., Flexible synthesis of diverse \textit{N}-heterocycles from substrates attainable from biomass. *Green Chem.* 2016, \textit{18}, 2453-2459.
5.1.2 Green pharmaceutical chemistry

Pharmaceutical companies are ranked first among chemical manufacturers in the production of waste including by-products, contaminated solvents, depleted reagents and air pollutants.\(^4\) Implementing green chemistry practices into the pharmaceutical industry will ensure drug discovery and production becomes cost effective and environmental friendly. Some leading pharmaceutical companies have already started to incorporate the principles of green chemistry into manufacturing processes. For example, Pfizer developed a green synthetic pathway to the antidepressant drug sertraline by incorporating benign solvents and preventing the formation of harmful by-products (Scheme 5.1).\(^5\) In the first commercial synthesis, tetralone **20** was condensed with methylamine in the presence of titanium tetrachloride in a mixture of solvents (toluene, hexanes and THF) to give compound **21**, which was hydrogenated using palladium on carbon (Pd/C) in THF to give compound **22** as a mixture of *cis* and *trans* diastereomers in 6:1 ratio. The desired *cis*-diastereomer could be isolated from the mixture by resolution with D-mandelic acid. The process generates hazardous solid by-products, titanium dioxide and methylamine hydrochloride and uses environmentally hostile solvents like toluene, hexanes and THF. The improved synthesis substituted harmful solvents with ethanol and avoided use of titanium tetrachloride in conversion of **20** to **22**, eliminating formation of titanium dioxide and monomethylamine hydrochloride. Moreover, the stereoselectivity and yield of the hydrogenation was improved by using Pd/CaCO\(_3\) in ethanol leading to 20:1 diastereoselectivity in >94% yield. Overall, the new synthesis eliminates 440 tons of TiO\(_2\)-MeNH\(_2\).HCl and 40 tons of *trans* isomer waste per year.
The synthesis of the HIV integrase inhibitor, raltegravir, was redesigned according to green chemistry principles by Merck, leading to a 33% increase in overall yield with 65% organic and aqueous solvent waste reduction (Schemes 5.2 and 5.3). In 1st generation synthesis (Scheme 5.2), compound 24 was N-methylated with methyl iodide and Mg(OMe)_2 in DMSO and methanol, followed by addition of 4-fluorobenzylamine (4-FBA) in ethanol to afford compound 25 in 63% yield. The regioselectivity of N-Me to O-Me is 78:22. The Cbz group in
Compound 25 was hydrogenated using Pd/C and methanesulfonic acid in methanol, which upon neutralization gave compound 26 in 99% yield. The amine 26 was coupled to acyl chloride 27 in presence of N-methylmorpholine (4-NMM) in THF to give a mixture of raltegravir 28 and ester 29, which upon treatment with aqueous methylamine and neutralisation gave pure raltegravir 28 in 88% yield.

Scheme 5.2: First generation synthesis of raltegravir

In the second generation synthesis (Scheme 5.3), compound 24 was first treated with 4-FBA and trimethylamine (TEA) in presence of methanol to give an amide, which was further reacted with Mg(OH)₂, Me₃S(O)I in N-methyl-2-pyrrolidone (NMP) to afford 25 in 91% yield. The N-methylation process did not produce any O-methylated by-products. In this step, the yield was increased from 63 to 91% and also solvent waste was reduced. Compound 25 was treated
with pivaloyl chloride to form a pivalate ester, which upon hydrolysis of the Cbz group using Pd/C in methanol and glycolic acid to afford compound 30, which was crystallized using triethylamine. The coupling of amine 30 to acyl chloride 27 in presence of 4-NMM in acetonitrile gave compound 31. The new process is more productive, uses 3 times less solvent and half the amount of compound 27 in comparison to 1st generation synthesis. Finally, basic hydrolysis of pivaloyl group using potassium hydroxide followed by crystallization using aqueous acetic acid gave raltegravir 28 at 97 % yield.

Scheme 5.3: Second generation synthesis of raltegravir
5.1.3 Biomass for synthesis of fine chemicals

The chemical industry is almost entirely reliant on fossil resources.\textsuperscript{7} As these finite reserves dwindle, a switch to biorenewable resources is required to ensure a supply of these chemicals at levels which we enjoy today. Biomass from plants and animals is potential source of platform chemicals for the chemical industry.\textsuperscript{8} In contrast to hydrocarbons in an oil refinery, biorefineries comprise oxygen rich heterogeneous solids, so different processing approaches are required to obtain core building blocks.\textsuperscript{9} The first challenge is to develop efficient methods for large scale processing of biomass, likely through technological advances in thermochemical, biochemical, mechanical and chemical processes.\textsuperscript{7} The use of these building blocks for fine chemical synthesis may not be an economical option at present, but it does present the only realistic approach for the future of chemical manufacturing.\textsuperscript{9} Some applications of the use of biomass-derived platforms in fine chemical synthesis are presented henceforth.

5.1.3.1 Synthesis of Ranitidine and \(\delta\)-aminolevulinic acid from cellulose-derived 5-(chloromethyl)furfural

Mascal and co-worker reported the synthesis of ranitidine, a drug for treatment of gastric and duodenal ulcers (Scheme 5.4)\textsuperscript{10} from the cellulose derived platform chemical 5-(chloromethyl)furfural 32. Upon treatment of 32 with commercially available \(N\)-acetylcysteamine, the thiol 33 was formed. Reductive amination gave tertiary amine 34 that upon hydrolysis gave diamine 35. Finally, condensation of 35 with 36 gave ranitidine 37.
Scheme 5.4: Synthesis of ranitidine from 5-(chloromethyl)furfural

The same building block was also used in the synthesis of the natural herbicide δ-aminolevulinic acid (Scheme 5.5).\textsuperscript{11} This natural product is also a prodrug (a precursor of porphyrin) used in photodynamic therapy for the treatment of actinic keratosis.\textsuperscript{12} 5-(Chloromethyl)furfural 32 was treated with sodium azide to give azidofuran 38. Upon reaction of 38 with singlet oxygen, a mixture of lactone 39 and β-carbonyl acrylic acid 40 were produced, which upon catalytic hydrogenation gave δ-aminolevulinic acid 41.
Scheme 5.5: Synthesis of δ-aminolevulinic acid from 5-(chloromethyl)furfural

5.1.3.2 Synthesis of enantiopure cyclopropyl esters from (−)-levoglucosenone

(−)-Levoglucosenone (LGO), which is obtained by acid hydrolysis of lignocellulose, is a chiral synthon with plethora of applications in chemical synthesis. The chiral centres present in LGO makes it a valuable building block for the sustainable synthesis of fine chemicals in enantiopure form. Recent advances in the large scale production of LGO make it a realistic candidate as sustainable platform chemical. Greatrex and co-workers reported the synthesis of enantiopure cyclopropyl esters starting from LGO (Scheme 5.6A). The cyclopropyl esters are potential precursors to the central nervous system (CNS) active cyclopropylamino acid derivatives such as the GABAc receptor antagonist (1R, 2R)-trans-2-aminomethycyclopropanecarboxylic acid [(−)-TAMP] and milnacipran (Scheme 5.6B). LGO was hydrogenated to give compound 43, which upon Baeyer-Villiger oxidation gave lactone 44. The chiral lactone 44 was mesylated to give (S)-45, which upon reaction with sodium ethoxide or (1-hydroxycyclohexyl)lithium gave (S)-46 and (S)-47 respectively. Finally, lithium bis(trimethylsilyl)amide (LHMDS) promoted intramolecular cyclization of (S)-46 and (S)-47 gave enantiopure cyclopropyl ester (1S, 2S)-48 and (1S, 2S)-49 respectively. In a similar fashion, enantiomer (1R, 2R)-49 was also synthesized. Treatment of (S)-46 with
trifluoroacetic acid followed by mesylation gave \((R)-47\), which upon treatment with LHMDS gave \((1R, 2R)-49\).

**Scheme 5.6:** [A] Synthesis of enantiopure cyclopropyl esters from \((-\)–levoglucosenone [B] CNS active cyclopropyl amino acid derivatives
5.1.3.3 Synthesis of muscarine-like compounds from \((1R, 5S)\)-anhydroisosaccharino-\(\delta\)-lactone (LAC)

The chiral compound \((1R, 5S)\)-anhydroisosaccharino-\(\delta\)-lactone (LAC), attainable on a gram scale (overall 6 % yield) by pyrolysis of cellulose, is a promising candidate for exploratory studies as a platform chemical.\(^{16}\) Mancini and co-workers reported the synthesis of potential bioactive muscarine-like compounds from LAC (Scheme 5.7A).\(^{17}\) Methanolysis of LAC \(50\) gave a 3-hydroxytetrahydrofuran which was converted to tosylate \(51\). Conversion to azido-tetrahydrofuran \(52\) followed by ester hydrolysis, hydrogenation and filtration through cationic resin gave the amino acid \(53\). The amino acid \(53\) was treated with excess of methyl iodide to give muscarine-like compound \(54\). Treatment of tosylate \(51\) with various secondary amines gave the muscarine derivatives \(55, 56\) and \(57\). All the derivatives \((54 - 57)\) displayed good binding affinity towards all five muscarinic receptors (M1 - M5) with selectivity towards M1, M2 and M3, that was comparable to muscarine (Scheme 5.7B).
Scheme 5.7: [A] Synthesis of muscarine-like compounds from (1R, 5S)-anhydroisosaccharino-\(\delta\)-lactone (LAC) [B] Muscarine

5.1.4 Importance of nitrogen-heterocycles

Nitrogen-heterocycles (cyclic molecules where one or more carbon atoms are replaced with a nitrogen) are present in many classes of natural products and a large majority of commercially important drugs, agrochemicals, functional materials, dyes, nutraceuticals, fragrances and speciality polymers.\(^\text{18-20}\) For example, pyrrole is an important nitrogen heterocycle present in various pharmaceuticals, including atorvastatin, obatoclax, and aloracetam (Figure 5.2A).\(^\text{21}\) Pyridine is the most common aromatic N-heterocycle in FDA approved drugs, including pharmaceuticals like loratadine, lansoprazole, and crizotinib (Figure 5.2B).\(^\text{18}\) Indole is ubiquitous in pharmaceuticals such as ondansetron, delavirdine and tadalafil (Figure 5.2C).\(^\text{19}\)
5.1.5 Nitrogen heterocycles from biomass

5.1.5.1 Conversion of cellulose to N-heterocycles

Because of the prevalence of nitrogen heterocycles, methods for their sustainable production from biomass are of great importance. Cellulose is a biopolymer present in plants that represents $1.5 \times 10^{12}$ tons of total biomass produced in a year.\textsuperscript{22} Xu and co-workers reported the production of pyrroles, indoles and pyridines upon thermo-catalysis and ammonisation (TCC-A) of cellulose using a zeolite (HZSM-5) catalyst (Scheme 5.8).\textsuperscript{23} The formation of
pyrroles was promoted by \textit{in situ} TCC-A at 500 °C and a short interaction time. Along with pyrroles, a small amount of indoles and pyrazines were also formed using this method. \textit{Ex situ} TCC-A at 600 °C using the catalyst with a high Si/Al ratio led to the production of pyridines as the major products, also producing a small amount of indoles, pyrroles and anilines. The production of indoles as the major component was promoted by low Si/Al ratio in the catalyst and a long residence time, which also produced a small amount of pyridines, pyrroles and anilines. An issue with these methods is formation of N-containing biochar, non-nitrogen aromatics and gases, along with the cumbersome purification of the products.

![Scheme 5.8: From biomass to N-heterocycles (pyridines, pyrroles and indoles).](image)

**Scheme 5.8:** From biomass to \textit{N}-heterocycles (pyridines, pyrroles and indoles).\textsuperscript{23}

### 5.1.5.2 Conversion of biomass-derived building blocks to \textit{N}-heterocycles

Xu and co-workers also reported the production of \textit{N}-heterocycles from bio-derived furan, furfural and glycerol.\textsuperscript{24-26} Furan and furfural are cellulose-derived platform chemicals\textsuperscript{27,28} and glycerol, a component of all natural fats and oils in the form of fatty acid esters, is produced in large amount as a by-product of biodiesel processing (1 kg of glycerol produced per 9 kg of biodiesel).\textsuperscript{29} Upon TCC-A of furan in the presence of a zeolite (HZSM-5) catalyst, indoles
were formed as the major product, along with a small amount of pyrroles and a trace amount of pyridines and anilines (Scheme 5.9A).\textsuperscript{24} The TCC-A of furfural in the presence of a zeolite catalyst gave indoles as the major product, along with a small amount of pyridines and a trace amount of pyrroles and anilines (Scheme 5.9B).\textsuperscript{25} Upon TCC-A of glycerol in presence of the same zeolite catalyst, pyridines were selectively produced (Scheme 5.9C).\textsuperscript{26} An issue with all three processes is the generation of solid carbon residues, gases and non-nitrogen aromatic compounds in the reactor. However, the yields were higher compared to when cellulose was used as the substrate.

\textbf{Scheme 5.9:} Thermo-catalytic production of $N$-heterocycles from biomass-derived platform chemicals.\textsuperscript{24-26}

5.1.5.3 Synthesis of $N$-heterocycles from biomass derived building blocks

Afonso and co-worker reported the synthesis of a functionalized $N$-alkyl-pyridinium salt by ring expansion of cellulose-derived 5-hydroxymethylfurfural (HMF).\textsuperscript{30} The treatment of HMF 58 with n-butylamine 59 in the presence of formic acid gave the n-butyl pyridinium salt 60 in good yield (Scheme 5.10A). Pyridinium scaffolds are present in natural products and
biologically important compounds such as desmosine, a biomarker for the diagnosis of chronic obstructive pulmonary disease (COPD) (Scheme 5.10B).

Scheme 5.10: [A] synthesis of the N-butyl pyridinium salt from HMF [B] Desmosine

Garcia and co-workers reported the synthesis of pyrrolidones and quinolines from the biorenewable feedstock levulinic acid (Scheme 5.11).\textsuperscript{31} Reductive amination of levulinic acid 61 with amines (aliphatic or benzyl or cyclohexyl) in the presence of a ruthenium catalyst, gave a series of pyrrolidones 62 along with a small amount of compounds 63 and/or 64 and/or 65 (Scheme 5.11A). The condensation of levulinic acid 61 and 2-ethynylaniline 66 gave the quinoline 67 (Scheme 5.11B). Furthermore, the condensation of levulinic acid 61 with 2-ethynylaniline derivatives 68 in the presence of p-toluenesulfonic acid gave quinolines 69 (Scheme 5.11C).
Scheme 5.11: [A] Synthesis of pyrrolidones from levulinic acid [B] and [C] synthesis of quinolines from levulinic acid

To date, several other efforts have been employed for the synthesis of pyrrolidones from levulinic acid in the presence of different heterogeneous catalysts including an iron-nickel alloy,\textsuperscript{32} a platinum\textsuperscript{33} and a ruthenium-based catalyst.\textsuperscript{34, 35}
5.1.6 Itaconic acid (IA)

Itaconic acid (IA) 1 (Figure 5.3) is a fully sustainable, non-toxic, five-carbon chemical building block that features on the U.S. DOE National Renewable Energy Laboratory's Top 12 list of renewable chemicals attainable from biomass.\(^{36}\)

![Chemical structure of itaconic acid](image)

**Figure 5.3:** Itaconic acid

It was first reported in 1837 by Baup as a product of the thermal decomposition of citric acid.\(^{37}\) In 1932, Kinoshita reported the production of IA by fungal fermentation of carbohydrates using *Aspergillus itaconicus*.\(^{37, 38}\) At present, *Aspergillus terreus* NRRL1960 is the best microorganism to produce IA, delivering 91 g L\(^{-1}\) IA from glucose.\(^{39, 40}\)

In 2012, Weastra compiled a report addressing the market potential of IA.\(^{41}\) The market price of IA is US$1800-2000/MT with the global IA market estimated at 41,400 MT worth US$ 74.5 million in 2011 and is expected to reach 407,790 MT worth US$ 567.4 million in 2020. The production costs have halved from US$ 4.0 kg\(^{-1}\) in 2001 to US$ 2.0 kg\(^{-1}\) in 2013.\(^{42}\) In order to substitute petrochemical based platform chemicals, it is estimated the production costs need to be decreased to around US$ 1.5 kg\(^{-1}\).\(^{43}\) Over 25% of the production costs are dictated by the feedstock used in the fermentation.\(^{44}\) Molasses, a by-product of sugarcanes and sugar beets processing to produce sucrose, is a rich source of glucose\(^{45}\) and has been frequently used for the production of IA, but it is not economical because of the sensitivity of *A. terreus* to impurities present and issues with downstream purification process.\(^{46}\) An attractive alternative is xylose, attainable by acid hydrolysis of lignocellulose present in the bagasse of sugarcanes,\(^{47}\) hardwoods\(^{48}\) and agricultural residues.\(^{49}\) At present, the production yield of IA from xylose is
38 g L\(^{-1}\), which is low compared to glucose (91 g L\(^{-1}\)).\(^{50, 51}\) Saha and co-workers reported that the pentose sugar, arabinose delivers 34 g L\(^{-1}\) IA upon fermentation by *Aspergillus terreus* NRRL1961.\(^ {51}\)

At present, the applications of IA includes the production of plastics, rubbers, resins, detergents and textiles.\(^ {41}\) IA has the potential to replace acrylic acid (used in the production of superabsorbent polymers), acetone cyanohydrin (methyl methacrylate), maleic anhydride (unsaturated polyester resin) and sodium tripolyphosphate (phosphate-free detergents).\(^ {41}\)

### 5.1.7 Pyrrole

At present, pyrrole 71 is produced industrially from furan 70 and ammonia in the presence of a silica or alumina catalyst (Scheme 5.12A).\(^ {52}\) Furan is mass produced by a palladium catalysed decarbonylation of furfural 72 or partial oxidation of 1,3-butadiene 73 (Scheme 5.12B).\(^ {53}\) Ammonia is mass-produced by the Haber process, whereby nitrogen and hydrogen are reacted in presence of a catalyst at high temperature (Scheme 5.12C).\(^ {54}\) The Haber process accounts for 1.2 % global energy demand and represents 0.93 % of global greenhouse gas emissions.\(^ {55}\)

Furthermore, it is not sustainable given that hydrogen is obtained from natural gas.

**Scheme 5.12:** [A] Synthesis of pyrrole from furan [B] Synthesis of furan [C] Haber process
Pyrrole is a realistic sustainable platform chemical because both furan and ammonia are available from biorenewable sources. The precursor to furan, furfural, is commercially produced by the dehydration of pentose sugars like xylose and arabinose, which are attainable by the hydrolysis of hemicellulose in biomass including corncob and bagasse of sugarcane.⁵³ There are also ongoing efforts regarding the development of more efficient procedures for furfural production from pentose sugars. Recently, Li and co-workers reported the synthesis of furfural by catalytic decarbonylation of xylose using γ-valerolactone as solvent (Scheme 5.13A).⁵⁶ Ammonia is also a biorenewable substrate as it can be obtained from protein waste. For example, Liao and co-workers reported the production of ammonia along with biofuels by treating protein wastes such as peptone, trypticase peptone, and casein with the protease producing Bacillus subtilis (Scheme 5.13B).⁵⁷ The protease hydrolysed polypeptides to amino acids, which upon deamination gave ammonia and the biofuel.


There are also ongoing efforts to develop efficient biomass gasification procedures for the production of ammonia. Gilbert and co-workers reported the production of ammonia from
wood chip gasification. Wood chips and nitrogen gas were processed at 1200-1600 °C and 30-40 atm in O₂-enriched air (55 wt % O₂) to produce 1 kg of ammonia per 2.71 kg of biomass (35 % moisture content), reducing greenhouse gas emission by 65 % when compared to the Haber process. Andersson and co-workers reported a techno-economic analysis of ammonia production between a stand-alone plant and a pulp mill integrated gasification process. The price of ammonia production is estimated to be £581-£882 per ton from the stand-alone plant and £509-£774 per ton from the integrated system, values that correlate well with current price of fossil-derived ammonia (£514 per ton).

5.2 Objective

The objective of the work described in this chapter was to synthesise several N-heterocycles from two simple biorenewable substrates, itaconic acid and pyrrole. As shown in Scheme 5.14, the overall plan hinges on the conjugate addition of pyrrole 71 to dimethyl itaconate 74 to access the mono- and di-addition products 75 and 76. With this initial stage successful, the presence of the electron-rich heteroaromatic ring and electrophilic esters will be used to affect distinct, selective intramolecular C-acylation and N-acylation reactions that would result in a diverse array of N-heterocycles.
Scheme 5.14: Proposed assembly of N-heterocycles

5.3 Result and Discussion

5.3.1 Synthesis of dimethyl itaconate 74

The synthesis of dimethyl itaconate 74 was achieved by heating a solution of itaconic acid 1 in methanol in the presence of a small amount of sulfuric acid (Scheme 5.15).

Scheme 5.15: Esterification of itaconic acid to dimethyl itaconate

In keeping with the theme of this project, we used methanol (wood alcohol) in the esterification as it is attainable from the pyrolysis of biomass including tree trunks (yield = 18 wt %), oil
from bark wastes (10 %), corn stalks (7 %), sugarcane bagasses (6 %), roots and sawdusts (4 %). Furthermore, methanol is attainable by the pyrolysis of hazelnut shells (8 %), softwoods (1 %) and hardwoods (1.7 %).

5.3.2 Conjugate addition of pyrrole 71 to dimethyl itaconate 74 to give mono-addition product 75 and di-addition product 76

With dimethyl itaconate 74 in hand, a study investigating its conjugate addition reaction with pyrrole 71 was initiated. Lewis acids were investigated to find optimum conditions for the conjugate addition reaction. Pyrrole 71 and dimethyl itaconate 74 were subjected to BF$_3$-OEt$_2$ in THF and CH$_2$Cl$_2$, but reaction was not observed. The Lewis acids InCl$_3$, TiCl$_4$, Cu(OTf)$_2$, and ZnCl$_2$, all resulted in no reaction. When AlCl$_3$ was used (in CH$_2$Cl$_2$, toluene and DCE), trace amounts of the mono- and di-addition products 75 and 76 were formed. It has been reported that AlCl$_3$ mediated Friedel-Crafts reactions in nitromethane are accelerated by in situ formation of an AlCl$_3$-MeNO$_2$ complex. Pleasingly, a solution of AlCl$_3$ in a 1:1 mixture of CH$_2$Cl$_2$ and MeNO$_2$ gave both the mono-addition 75 and di-addition products 76 (Scheme 5.16). When three equivalents of pyrrole were used, the mono-addition product 75 was formed as the major product, along with a small amount of the di-addition product 76, in good yield (Scheme 5.16A). Attempts to eliminate the formation of the di-addition product by dilution, decreasing the reaction temperature and the duration of the reaction met with failure. The di-addition product 76 was exclusively formed when two and half equivalents of dimethyl itaconate was employed (Scheme 5.16B).
The general mechanism of the conjugate addition reaction proceeds by initial activation of the \( \alpha,\beta \)-unsaturated carbonyl oxygen of dimethyl itaconate 74 by the Lewis acid AlCl\(_3\) (Scheme 5.17). Subsequently, a \( \pi \)-bond electrons of pyrrole attacks electron deficient \( \beta \)-carbon to give 77. The intermediate 77 undergoes rearrangement leaving AlCl\(_3\) to give the mono-addition product 75.

Scheme 5.16: Conjugate addition of pyrrole to dimethyl itaconate

We first investigated the viability of the mono-addition product 75 to undergo selective C- and N-migration.
5.3.3 Acid catalysed intramolecular C-acylation and aromatisation

In an effort to promote intramolecular C-acylation of mono-addition product 75, we investigated various Bronsted and Lewis acids. The mono-addition product 75 was treated with p-toluene sulfonic acid (PTSA) in toluene, but no reaction was observed. The Lewis acid AlCl₃ in toluene and ZnCl₂ in CH₂Cl₂ were employed, but again no reaction was observed. We became aware that polyphosphoric acid (PPA) had been used to promote intramolecular C-acylation reactions.⁶⁵ Pyrrole 75 was subjected to PPA in toluene at 100 °C, and a trace amount of the tetrahydroindolone 78 was formed, as determined by ¹H NMR. Pleasingly, the treatment of pyrrole 75 with PPA under solvent-free conditions gave the tetrahydroindolone 78 in decent yield (Scheme 5.18A). Importantly, this reaction was selective for C-acylation and no other products were formed. The tetrahydroindolone scaffold can be found in fine chemicals such as the antipsychotic medication molindone⁶⁶ and tetrahydroindolone arylpiperazine derivatives,⁶⁷ possible candidates for the treatment of the post-traumatic stress disorder (Scheme 5.18B).

Scheme 5.18: [A] Acid catalysed C-acylation [B] Molindone and Tetrahydroindole arylpiperizine
The mechanism for the intramolecular C-acylation is shown in Scheme 5.19. Protonation of the carbonyl oxygen of the pyrrole 75 initiates attack by the C-3 position the adjacent pyrrole ring. Aromatisation followed by loss of methanol give the tetrahydroindolone 78. In general, the acylation at C-2 position is favoured over C-3. In this instance, the C-3 acylation of pyrrole dominates due to the presence of the substituent at C-2.

Scheme 5.19: Mechanism for the intramolecular C-acylation

With tetrahydroindole 78 in hand, dehydrogenation was attempted in an effort to access the aromatic indole. Palladium on carbon mediated dehydrogenation of 78 in diphenyl ether and mesitylene at reflux gave methyl-4-hydroxyindole-6-carboxylate 79 (Scheme 5.20A). Given the enormous importance of the indole heterocycle in pharmaceuticals (Scheme 5.20B), sustainable synthetic methodologies that provide indoles are of great importance. Furthermore, there are only limited strategies in the literature for the synthesis of 4,6-disubstituted indoles.
Scheme 5.20: [A] Aromatisation of tetrahydroindolone to indole [B] Pharmaceuticals containing indole ring

5.3.4 Base mediated N-acylation and aromatisation

Given that pyrrole 75 underwent selective C-acylation to the dihydroindolone, we subsequently switched to basic conditions in an effort to effect selective N-acylation. Upon treating pyrrole 75 with potassium carbonate in N,N-dimethylformamide, the dihydroindolizinone 80 was selectively formed (Scheme 5.21A). The dihydroindolizinone motif is present in the natural products polygonatine and kinganone (Scheme 5.21B). This methodology could also find utility in the synthesis of related alkaloids such as swainsonine and exochomine (Scheme 5.21C)

The mechanism for the intramolecular $N$-acylation is shown in Scheme 5.22. Deprotonation of pyrrole N-H by hydroxide followed by nucleophilic attack onto the ester gives the dihydroindolizinone product with concomitant loss of methanol.

Scheme 5.22: Mechanism for the intramolecular $N$-acylation.
In an effort to access a complementary heterocyclic ring system, the dehydrogenation of dihydroindolizinone 80 to the corresponding indolizine 81 was attempted (Table 5.1). Dihydroindolizinone 80 was heated with Pd/C in various solvents including Ph-O-Ph (entry 1), Ph-O-Ph/mesitylene (entry 2), 1,2-dichlorobenzene (entry 3), AcOH (entry 4), xylene (entry 5), diethylene glycol methyl ether (entry 6), but no reaction was observed in all instances. Dihydroindolizinone 80 was also heated with Pd(TFA)₂ in the presence of DMAP and PTSA in DMSO₇₈ (entry 7), but again no reaction was observed. Dihydroindolizinone 80 was heated with Cu(OAc)₂ in DMF at 130 °C (entry 8) and 160 °C₇⁹ (entry 9), but not reaction was observed at 130 °C, with degradation occurring at 160 °C. Dihydroindolizinone 80 was heated with DDQ at reflux in numerous solvents including xylene (entry 10), 1,2-dichlorobenzene (entry 11), THF (entry 12), MeCN (entry 13) and dioxane (entry 14), but again no reaction was observed in all cases. Dihydroindolizinone 80 was also heated at 110 °C with DDQ in presence of AcOH in toluene (entry 15), but in this instance degradation was observed. I₂ in DMSO (entry 16 and 17), IBX in DMSO (entry 18) and chloranil in toluene at reflux (entry 19), all gave negative results. Dihydroindolizinone 80 was also refluxed with MnO₂ in toluene (entry 20), CuBr₂/HBr in MeCN (entry 21), Ag₂O in CH₂Cl₂ (entry 22), in all cases no reaction was observed. When dihydroindolizinone 80 was heated with Pd/C in n-octanol at 195 °C (entry 23), a new spot was observed by TLC and subsequently identified as 2-pyridone 82, as a result of isomerisation (Scheme 5.23A). The synthesis of 2-pyridones from biorenewable substrate possesses great potential as it is a heterocycle motif found in many bioactive natural products and synthetic drugs.₈₀-₈³ It is present in camptothecin family chemotherapeutic agents like topotecan,₈₄ alkaloid mappicine₈⁵ and preclinical anti-tumor compound 7-(2-fluoro-4-iodophenyl)amino)-N-(2-hydroxyethoxy)-6-methyl-5-oxo-1,2,3,5-tetrahydroindolizine-8-carboxamide₈₆ (Scheme 5.23B).
Table 5.1: Attempted aromatization of dihydroindolizinone 80

<table>
<thead>
<tr>
<th>Entry</th>
<th>Oxidising agents</th>
<th>Solvents</th>
<th>Temp/ °C</th>
<th>Time</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pd/C</td>
<td>Ph-O-Ph</td>
<td>250</td>
<td>72 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>2</td>
<td>Pd/C</td>
<td>Ph-O-Ph and Mesitylene</td>
<td>250</td>
<td>72 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>3</td>
<td>Pd/C</td>
<td>1,2-Dichloro Benzene</td>
<td>reflux</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>4</td>
<td>Pd/C</td>
<td>AcOH</td>
<td>110</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>5</td>
<td>Pd/C</td>
<td>xylene</td>
<td>reflux</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>6</td>
<td>Pd/C</td>
<td>Diethylene glycol methyl ether</td>
<td>reflux</td>
<td>24 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>7</td>
<td>Pd(TFA)$_2$, 2-DMAP, PTSA, O$_2$</td>
<td>DMSO</td>
<td>80</td>
<td>24 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>8</td>
<td>Cu(OAc)$_2$, O$_2$</td>
<td>DMF</td>
<td>80-130</td>
<td>24 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>9</td>
<td>Cu(OAc)$_2$</td>
<td>DMF : DMSO (9:1)</td>
<td>160</td>
<td>24 h</td>
<td>Degradation</td>
</tr>
<tr>
<td>10</td>
<td>DDQ</td>
<td>Xylene</td>
<td>reflux</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>11</td>
<td>DDQ</td>
<td>1,2-dichloro benzene</td>
<td>reflux</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>12</td>
<td>DDQ</td>
<td>THF</td>
<td>reflux</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>13</td>
<td>DDQ</td>
<td>MeCN</td>
<td>reflux</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>14</td>
<td>DDQ</td>
<td>Dioxane</td>
<td>reflux</td>
<td>48 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>15</td>
<td>DDQ/AcOH</td>
<td>Toluene</td>
<td>110</td>
<td>8 h</td>
<td>Degradation</td>
</tr>
<tr>
<td>16</td>
<td>I$_2$</td>
<td>DMSO</td>
<td>110</td>
<td>3 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>17</td>
<td>I$_2$</td>
<td>DMSO</td>
<td>110</td>
<td>12 h</td>
<td>Degradation</td>
</tr>
<tr>
<td>18</td>
<td>IBX</td>
<td>DMSO</td>
<td>100</td>
<td>15 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>19</td>
<td>Chloranil</td>
<td>Toluene</td>
<td>reflux</td>
<td>12 h</td>
<td>Degradation</td>
</tr>
<tr>
<td>20</td>
<td>MnO$_2$</td>
<td>Toluene</td>
<td>reflux</td>
<td>24 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>21</td>
<td>CuBr$_3$/ HBr</td>
<td>MeCN</td>
<td>reflux</td>
<td>24 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>22</td>
<td>Ag$_2$O</td>
<td>CH$_2$Cl$_2$</td>
<td>reflux</td>
<td>24 h</td>
<td>No reaction</td>
</tr>
<tr>
<td>23</td>
<td>Pd/C</td>
<td>n-octanol</td>
<td>195</td>
<td>24 h</td>
<td>No reaction</td>
</tr>
</tbody>
</table>
Scheme 5.23: [A] Isomerisation by Pd/C [B] camptothecin, mappicine and 7-((2-fluoro-4-iodophenyl)amino)-N-(2-hydroxyethoxy)-6-methyl-5-oxo-1,2,3,5-tetrahydroindolizine-8-carboxamide

The isomerisation of dihydroindolizinone 80 to 2-pyridone 82 appears to be solvent dependent. The isomerisation occurs in the presence of the protic solvent n-octanol (Table 5.1, entry 23), but did not proceed in the apolar solvents diphenyl ether and mesitylene (Table 5.1, entry 1 and 2). This result warrants discussion.

Pd/C is commercially manufactured directly from PdCl₂, whereby the PdCl₂ is adsorbed on an activated charcoal surface and reduced in the presence of concentrated HCl. Thus, commercially available Pd/C retains trace amounts of PdCl₂ and HCl. Therefore, our understanding is that isomerisation is correlated to HCl, which the literature suggests is liberated in the presence of protic solvents.
In 2003, Hirota and co-worker documented a solvent effect on the cleavage of a silyl protecting group under hydrogenation conditions in the presence of Pd/C.\textsuperscript{88} Tert-butyl(cinnamyloxy)dimethylsilane \textbf{83} was subjected to hydrogenation in various solvents leading to a mixture of the dehydrogenated product \textbf{84} and the dehydrogenated and deprotected product \textbf{85} depending on the solvent used (Table 5.2). Upon hydrogenation of \textbf{83} in methanol, 3-phenylpropan-1-ol \textbf{85} was exclusively observed, a result of hydrogenation and desilylation (entry 1). When the hydrogenation was performed in ethanol, the same product \textbf{85} was observed, along with the reduced product \textbf{84} that had retained the silyl group (entry 2). Upon switching to aprotic solvents, no significant amounts of desilylation occurred (entries 3, 4 and 5).

\textbf{Table 5.2:} Solvent effects in the hydrogenation of tert-butyl(cinnamyloxy)dimethylsilane \textbf{83}

<table>
<thead>
<tr>
<th>Entry</th>
<th>Solvent</th>
<th>\textbf{84}</th>
<th>\textbf{85}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MeOH</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>EtOH</td>
<td>34</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>THF</td>
<td>98</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Toluene</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>EtOAc</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

The authors postulated the silyl ether cleavage was promoted by HCl liberated from Pd/C in protic solvents (entries 1 and 2). The same group followed up the work and validated this hypothesis in a subsequent publication (Table 5.3).\textsuperscript{89} Pd/C was stirred in methanol under H\textsubscript{2} (entry 1) and air (entry 2) followed by filtration. TES ether \textbf{86} was stirred with the filtrate and in both cases silyl ether cleavage was observed, resulting in n-decanol \textbf{87} as the major product. However, the cleavage was completely suppressed by the addition of the acid-scavenging,
basic resin Amberlite®IRA-45 when pre-stirring the methanol and Pd/C (entries 3 and 4). This finding justified that the desilylation was mediated by liberated acid rather than Pd/C itself.

**Table 5.3:** Cleavage of TES ether in the Pd/C-methanol filtrate

<table>
<thead>
<tr>
<th>Entry</th>
<th>Conditions</th>
<th>Yield %</th>
<th>86</th>
<th>87</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H₂</td>
<td>37</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Air</td>
<td>41</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>H₂ + Amberlite®IRA-45</td>
<td>100</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Air + Amberlite®IRA-45</td>
<td>100</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Overall, the aforementioned findings validate the release of HCl from Pd/C in a polar solvent with or without H₂. Therefore, we conclude the isomerisation of dihydroindolizinone 80 to 2-pyridone 82 in n-octanol was mediated by acid liberated from the Pd/C during the reaction.

The selective formation of indole 79 and dihyroindolizinone 80 from the mono-addition product 75 provides efficient access to these heterocyclic motifs from biorenewable substrates (Scheme 5.24A). For comparison purposes, it is informative to compare our work to that of Vedejs and co-workers, who examined the cyclisation of half ester 90 obtained from the Stobbe condensation of pyrrole-2-carboxaldehyde 88 with dimethyl succinate 89 (Scheme 5.24B). Upon subjecting the half ester 90 to triethylamine in acetic anhydride, indolizine 91 was exclusively formed. However, upon heating the half ester 90 in acetic anhydride, a 1:1 mixture of indoles 92 and 93 was formed. Moreover, heating the half ester 90 with AcOH in toluene gave an 8:1 mixture of indole 92 and indolizine 91. Not only did Vedej’s route fail to provide...
selective access to both the indole and indolizine, the products are all isolated as their acetate(s), which would likely need removal at some point during subsequent use.

Scheme 5.24: [A] Selective routes to indole and indolizinone [B] Vedej’s cyclisation of half ester 90

Having established the viability of C-acylation and N-acylation of mono-addition product 75, the methodology was subsequently applied to the di-addition product 76.
5.3.5 Acid mediated C-acylation of di-addition product 76 and aromatisation

In an effort to promote C-acylation, PPA catalysed cyclisation of the di-addition product 76 was investigated (Scheme 5.25A). The di-addition product 76 was heated in PPA, affording indole 94 in low yield (condition A). The use of toluene as a solvent in PPA mediated reactions is well documented in the literature. Therefore, the di-addition product 76 was subjected to PPA in toluene, which lead to much improved yield of 94 (condition B). Surprisingly, none of tetrahydrocarbazole dione 95 was ever formed. The reason for the diacylation product not forming is likely due to the electron-withdrawing carbonyl group at C-4 decreasing the electron density of the pyrrole ring by resonance effect, resisting the intramolecular C-acylation reaction to from tetrahydrocarbazole dione 95 (Scheme 5.25B).

Scheme 5.25: [A] C-acylation of the di-addition product 76; [B] deactivation of pyrrole ring by a resonance effect.
The previous dehydrogenation method using Pd/C in phenyl ether and mesitylene to successfully obtain indole 79 from tetrahydroindolone 78 was subsequently attempted to the more substituted tetrahydroindolone 94. Upon subjecting 94 to Pd/C in phenyl ether and mesitylene, the 2,4,6-trisubstituted indole 96 was afforded in moderate yield (Scheme 5.26).

**Scheme 5.26:** Aromatisation of the substituted tetrahydroindolone 94

### 5.3.6 Acid mediated C-acylation of indole 96 and aromatisation

With the indole 96 in hand, its cyclisation to the corresponding carbazolone was investigated. The indole 96 was subjected to PPA in toluene, forming the dihydrocarbazolone 97 in moderate yield (Scheme 5.27A). The dihydrocarbazolone 97 is a heterocyclic motif present in pharmaceuticals including the anti-emetic drug ondansetron. One can envisage that the carbazolone could be readily reduced to the corresponding tetrahydroindolone, a motif found in numerous pharmaceuticals, such as frovatriptan, which is used for treatment of migraine headache (Scheme 5.27B).
Scheme 5.27: [A] C-acylation of the indole 96 [B] Ondansetron and Frovatriptan

With the dihydrocarbazolone 97 in hand, it was subjected to Pd/C mediated dehydrogenation using conditions employed previously. A mixture of carbazole 98 and O-methylated carbazole 99 was formed in a 2.5:1 ratio (Scheme 5.28A). The carbazole ring system is present in a plethora of natural products including murrayanine and hyellazole and pharmaceuticals such as carvedilol, a drug used in treatment of heart failure and hypertension, and carprofen, a non-steroidal anti-inflammatory drug (NSAID) (Scheme 5.28B). Carbazoles are also very important in materials science. For example, the functional polymers poly(γ-2-(N-carbazolyl)ethyl-L-glutamate) and poly(N-epoxypropyl-carbazole) are used as charge transporting materials in photocopiers (Scheme 5.28C).
The formation of O-methylated carbazole 99 is an interesting result that requires discussion. It is possible the residual water and HCl present in the Pd/C hydrolyses the ester(s) in any of compounds 97, 98 and 99 liberating methanol in the process (Scheme 5.29A). The O-methylated carbazole 99 could result from two pathways (Scheme 5.29B). Methanol could attack carbazolone 97 leading to the hemiacetal 103. Upon protonation, either methanol or water could serve as the leaving group. As methanol and water are comparable in their leaving group ability,98 either of the two carbazoles 104 and 105 could result. Secondly, the liberated methanol could directly alkylate the carbazole 104. This is possible as methanol is known to serve as an electrophile at high temperatures in the presence of acidic catalysts.99
Scheme 5.29: [A] Hydrolysis of esters liberating methanol; [B] Mechanism of O-methylated carbazole formation
5.3.7 Based mediated N-acylation and isomerization of di-addition product 76

After the successful C-acylation methodology enabled access to the indole 96 and carbazoles 98 and 99 from the di-addition product 76, our attention turned towards the base mediated N-acylation of the same substrate using the conditions employed previously in this thesis (Scheme 5.30). The di-addition product 76 was subjected to potassium carbonate in DMF at 110 ºC, giving the substituted dihydroindolizinone 106 as a result of selective N-acylation. With the substituted dihydroindolizinone 106 in hand, its isomerisation to the 2-pyridone 107 was investigated using conditions described previously in this thesis (Scheme 5.30). Upon subjecting the substituted dihydroindolizinone 106 to the Pd/C in n-octanol at elevated temperature, the substituted 2-pyridone 107 was formed.

Scheme 5.30: Base mediated N-acylation of the di-addition product 76 followed by isomerisation of the dihydroindolizinone 106

5.3.8 Acid mediated C-acylation of dihydroindolizinone 106

With the substituted dihydroindolizinone 106 in hand, we also attempted the PPA mediated C-acylation method (Scheme 5.31A). Upon subjecting 106 to PPA, C-acylation occurred to give the tricyclic N-heterocycle 108. We hypothesised 108 would undergo dehydrogenation to give
the aromatic pyrido[1,2-a]indole 109. Upon subjecting 108 to the dehydrogenation conditions successfully employed in this thesis, no reaction was observed. Several other dehydrogenation conditions were trialled (refer to Table 5.1), all without success. The reason for the unsuccessful dehydrogenation is likely due to the electron withdrawing acyl group at N-1 making the heterocycle electron deficient and thus resistant to dehydrogenation. Given our failure to access the pyrido[1,2-a]indole 109 from the compound 108, we changed the synthetic strategy and postulated it could be obtained by the base mediated N-acylation of the previously synthesised indole 96 (Scheme 5.31A). Indeed, the indole 96 underwent selective N-acylation upon treatment with potassium carbonate, affording the desired pyrido[1,2-a]indole 109. This heteroaromatic scaffold is present in alkaloids including strychnine and eburnamonine,100 and the potential drug fabesetron,101 which is undergoing clinical trials for the treatment of irritable bowel syndrome (Figure 5.31B).
5.4 Conclusion

The chapter describes our efforts towards synthesising several \( N \)-heterocycles from two substrates attainable from biomass, dimethyl itaconate 74 and pyrrole 71.

Figure 5.4: Dimethyl itaconate and pyrrole

The study was initiated with the conjugate addition of pyrrole 71 to dimethyl itaconate 74 (Scheme 5.32). When excess pyrrole was employed, the mono-addition product 75 was formed as a major product, along with a small amount of the di-addition product 76. The di-addition product 76 was exclusively formed when excess dimethyl itaconate was used. The reactions were found to work best using aluminium trichloride as the Lewis acid to promote the reaction.

Scheme 5.32: Conjugate addition of pyrrole 71 to dimethyl itaconate 74

The cyclisation of the mono-addition product 75 were first investigated (Scheme 5.33). In the presence of PPA, 75 underwent completely selective \( C \)-acylation to give the dihydroindolone 78, which underwent the facile dehydrogenation giving the indole 79 (Scheme 5.33). Drawing
Inspiration from successful C-acylation, we attempted a base mediated N-acylation on the mono-addition product 75, affording the dihydroindolizinone 80. When attempting the dehydrogenation of 80 to give the indolizine 81, the isomerised product 2-pyridone 82 was observed. This isomerisation is likely promoted by liberation of HCl from the Pd/C in the presence of protic solvent n-octanol at elevated temperatures.

Scheme 5.33: Selective C- and N-acylations of the pyrrole 75

With successful C- and N-acylations achieved on the mono-addition product 75, we applied this methodology to the di-addition product 76. C-Acylation in the presence of PPA in toluene led to the dihydroindolone 94, which underwent Pd/C mediated dehydrogenation giving the indole 96 (Scheme 5.34). Here, the use of toluene as a solvent increased yield of the reaction. The C-acylation methodology was subsequently applied to the indole 96. When attempting PPA mediated C-acylation to 96, the dihydrocarbazolone 97 was formed, which underwent Pd/C mediated dehydrogenation to give a mixture of the carbazole 98 and 99 (Scheme 5.34). The formation of the O-methylated carbazole is due to residual methanol generated by ester hydrolysis caused by residual H2O and HCl on the Pd/C during the reaction.
Scheme 5.34: Preparation of the dihydroindolone, indole, carbazolone and carbazole (94-99)

With successful C-acylation of the di-addition product 76, the base mediated N-acylation approach was applied to 76 (Scheme 5.35). In the presence of base, 76 underwent selective N-acylation giving the substituted dihydroindolizinone 106, which isomerised in the presence of Pd/C in n-octanol to the substituted 2-pyridone 107. The PPA mediated C-acylation of dihydroindolizinone 106 gave the tricyclic dione 108, which unfortunately could not be converted to the pyrido[1,2-a]indole 109 by dehydrogenation. However, the pyrido[1,2-a]indole 109 could be readily obtained by the base mediated N-acylation of indole 96.
Scheme 5.35: Preparation of dihydroindolizinone, 2-pyridone, tricyclic dione and pyrido[1,2-a]indole (106-109)

A suite of diverse N-heterocycles including dihydroindolone, indole, dihydroindolizinone, 2-pyridone, pyrido[1,2-a]indole dihydrocarbazolone and carbazole have all been prepared from dimethyl itaconate and pyrrole, two compounds attainable from biomass.
Chapter 6: Materials and methods

6.1 Protein production and purification

6.1.1 Materials

All chemicals were obtained from Sigma-Aldrich, Environmental Control Products (ECP), Bio-Rad or Thermo Fisher Scientific unless otherwise stated. Restriction enzymes and buffers were obtained from New England Biolabs (NEB). Synthetic genes and primers were obtained from Integrated DNA Technologies. The pNIC28-Bsa4 vector was a gift from Opher Gileadi (Addgene plasmid #26103). If necessary, pipette tips (Eppendorf or HTL) and microtubes (Eppendorf or VWR) were sterilised using a benchtop autoclave (3850 EL Autoclave, Tuttnauer) at 121 ºC for 20 min.

6.1.2 Preparation of stock and working solutions of primers and synthetic genes

Synthetic gene fragments were dissolved in 1 × Tris-EDTA buffer (TE) (Table 6.1) to a stock solution of 10 ng/μL concentration. Primers were dissolved in 1 × TE to make a stock solution of 100 μM concentration. The stock solutions were diluted ten times with ultrapure water to make working solutions of 10 μM concentration.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per 100 mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 M Tris-HCl buffer (pH 8.0)</td>
<td>1 mL</td>
</tr>
<tr>
<td>0.5 M EDTA</td>
<td>200 μL</td>
</tr>
<tr>
<td>Ultrapure water</td>
<td>799 μL</td>
</tr>
</tbody>
</table>

Table 6.1: Preparation of 1 × TE buffer

6.1.3 Polymerase chain reaction (PCR)

DNA of interest was amplified by PCR from an appropriate plasmid DNA template using PrimeSTAR HS DNA polymerase (Takara Bio Inc.) in a Mastercycler pro S thermocycler (Eppendorf) following manufacturer’s specifications. The components of a PCR reaction mixture are described in Table 6.2. A negative control in the absence of a primer and a template
was run to detect DNA contamination. PCR products were purified when necessary using GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific) to remove primers and enzymes or using GeneJET Gel extraction kit (Thermo Fisher Scientific) to isolate a specific size DNA after gel electrophoresis. Purified PCR products were either used immediately or stored at -20 °C for posterior use.

<table>
<thead>
<tr>
<th>Components</th>
<th>Amounts per sample (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR enhancer</td>
<td>10</td>
</tr>
<tr>
<td>2 × buffer</td>
<td>25</td>
</tr>
<tr>
<td>dNTPs</td>
<td>4</td>
</tr>
<tr>
<td>Forward primer</td>
<td>2</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>2</td>
</tr>
<tr>
<td>Template</td>
<td>1</td>
</tr>
<tr>
<td>Polymerase</td>
<td>0.5</td>
</tr>
<tr>
<td>Ultrapure water</td>
<td>5.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
</tr>
</tbody>
</table>

Table 6.2: A PCR reaction mixture

### 6.1.4 Electrophoresis

#### 6.1.4.1 Agarose gel electrophoresis

Agarose gel electrophoresis was performed to estimate DNA fragments size. 1 % w/v agarose gel was prepared in 1 × Tris acetate-EDTA (TAE) buffer, which was prepared from 50 × TAE stock buffer (Table 6.3).

<table>
<thead>
<tr>
<th>Components</th>
<th>Amounts per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris base</td>
<td>242 g</td>
</tr>
<tr>
<td>Glacial acetic acid</td>
<td>57.1 mL</td>
</tr>
<tr>
<td>500 mM EDTA (pH 8.0)</td>
<td>100 mL</td>
</tr>
</tbody>
</table>

Table 6.3: Components for 50 × TAE buffer

The setting of the gel was initiated by adding 1% agarose gel in 1 × TAE buffer onto a gel slab (Bio-Rad). A comb (Bio-Rad) was immediately inserted to form wells, followed by the addition
of ethidium bromide (5-6 µL). Samples were prepared by mixing 5 µL of reaction mixture and 2 µL of DNA loading buffer. 7 µL of the prestained DNA ladder (Bio-Rad) was loaded onto the first well of the gel for molecular weight visualisation followed by loading 7 µL of samples to other wells. Electrophoresis was run by a Bio-Rad PowerPac Power Supply and Mini-Sub cell GT DNA electrophoresis system using 1 × TAE buffer at a constant voltage of 90 V for 30 minutes at room temperature (rt). The DNA gel was visualised using Molecular imager Gel Doc XR+ with image lab software (Bio-Rad).

### 6.1.4.2 Denaturing Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed to estimate protein’s molecular weight. The composition of the resolving gel and the stacking gel employed for SDS-PAGE are described in Table 6.4 and Table 6.5, respectively. 10% Ammonium persulphate (APS) and 10% sodium dodecyl sulphate (SDS) solutions were freshly made for each gel preparation. The setting of the gel was initiated by adding APS and tetramethylethylenediamine (TEMED). The resolving gel was firstly solidified in a gel cast (Bio-Rad), and the stacking gel was then cast upon the resolving gel, and a comb (Bio-Rad) was immediately inserted to form 10 wells on the stacking gel.

<table>
<thead>
<tr>
<th>Composition</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrapure water</td>
<td>3.3 mL</td>
</tr>
<tr>
<td>Acrylamide/Bis-acrylamide (30%/0.8% w/v)</td>
<td>4 mL</td>
</tr>
<tr>
<td>1.5 M Tris (pH 8.8)</td>
<td>2.5 mL</td>
</tr>
<tr>
<td>10% (w/v) SDS</td>
<td>100 µL</td>
</tr>
<tr>
<td>10%(w/v) APS</td>
<td>100 µL</td>
</tr>
<tr>
<td>TEMED</td>
<td>5 µL</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>10 mL</strong></td>
</tr>
</tbody>
</table>

**Table 6.4**: Components for two SDS-PAGE resolving gels with 12% acrylamide
<table>
<thead>
<tr>
<th>Composition</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrapure water</td>
<td>1.485 mL</td>
</tr>
<tr>
<td>Acrylamide/Bis-acrylamide (30%/0.8% w/v)</td>
<td>335 µL</td>
</tr>
<tr>
<td>0.5 M Tris (pH 6.8)</td>
<td>625 mL</td>
</tr>
<tr>
<td>10% (w/v) SDS</td>
<td>25 µL</td>
</tr>
<tr>
<td>10%(w/v) APS</td>
<td>25 µL</td>
</tr>
<tr>
<td>TEMED</td>
<td>5 µL</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>2.5 mL</strong></td>
</tr>
</tbody>
</table>

**Table 6.5:** Components for two SDS-PAGE stacking gels with 4% acrylamide

10 µL of an analyte was mixed with 10 µL of 2 × sample loading buffer (Table 6.6) and then heated for at least 3 minutes at 95 °C using the dry bath (Corning LSE). 8 µL of the Kaleidoscope prestained protein marker (Bio-Rad) was loaded onto the first well of the gel for molecular weight visualisation and 15 µL of the mixture of the analyte and the loading buffer was loaded onto another well. Electrophoresis was run on a Mini-PROTEAN Tetra Cell System (Bio-Rad) using the running buffer (Table 6.6) at a constant voltage of 160 V for 1 hour at room temperature. The gel was removed from the cast and stained for 30 seconds microwave heat followed by 30 minutes room temperature by a Coomassie brilliant blue staining solution (Table 6.7). It was then destained overnight using the destain solution (Table 6.7)
### Solutions

<table>
<thead>
<tr>
<th>2 × Sample loading buffer</th>
<th>Composition</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 M Tris-HCl (pH 6.8)</td>
<td>1.25 mL</td>
<td></td>
</tr>
<tr>
<td>10% SDS</td>
<td>4 mL</td>
<td></td>
</tr>
<tr>
<td>Glycerol</td>
<td>2.5 mL</td>
<td></td>
</tr>
<tr>
<td>β-mercaptoethanol</td>
<td>1 mL</td>
<td></td>
</tr>
<tr>
<td>1% Bromophenol blue</td>
<td>0.5 mL</td>
<td></td>
</tr>
<tr>
<td>Ultrapure water</td>
<td>0.75 mL</td>
<td></td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>10 mL</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1 × Running Buffer</th>
<th>Composition</th>
<th>Amount per Litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>1.44 g</td>
<td></td>
</tr>
<tr>
<td>Tris Base</td>
<td>3.03 g</td>
<td></td>
</tr>
<tr>
<td>SDS</td>
<td>1 g</td>
<td></td>
</tr>
<tr>
<td>Ultrapure water</td>
<td>1 L</td>
<td></td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>1 L</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.6: Components for 2X sample loading buffer and 1X running buffer for SDS-PAGE

<table>
<thead>
<tr>
<th>Solution</th>
<th>Composition</th>
<th>Amount per Litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coomassie stain</td>
<td>Coomassie blue</td>
<td>1 g</td>
</tr>
<tr>
<td></td>
<td>Methanol</td>
<td>400 mL</td>
</tr>
<tr>
<td></td>
<td>Acetic acid</td>
<td>100 mL</td>
</tr>
<tr>
<td></td>
<td>DI water</td>
<td>500 mL</td>
</tr>
</tbody>
</table>

| Destain solution    | Methanol    | 400 mL          |
|                     | Acetic acid | 100 mL          |
|                     | DI water    | 500 mL          |

Table 6.7: Components for Coomassie stain solution and destain solution

### 6.1.5 Protein concentration determination

Concentration of proteins were determined using NanoDrop® 2000c spectrophotometer (Thermo Scientific). The Protein A<sub>280</sub> and Other protein (E & MW) options were used. The buffer solution was used as a blank. Protein molecular weight and extinction coefficient were calculated using the ProtParam tool. Typically, 2 μL of the analyte was used per measurement. All measurements were done in triplicate.
6.1.6 Ligation independent cloning (LIC)

6.1.6.1 Preparation of vector

First of all, the pNIC28-Bsa4 vector was digested with BsaI restriction enzyme in a 100 μL reaction volume (Table 6.8) by incubating for 3 hrs at 50 ºC. The cut vector was cleaned up using the GeneJET PCR purification kit (Thermo Scientific) and eluted with 25 μL elution buffer.

<table>
<thead>
<tr>
<th>Component</th>
<th>Reaction Volume/μL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrapure Water</td>
<td>57.6</td>
</tr>
<tr>
<td>NEB Buffer 3</td>
<td>10</td>
</tr>
<tr>
<td>BSA 10mg/mL</td>
<td>1</td>
</tr>
<tr>
<td>Vector (0.176 μg/μL)</td>
<td>28.4 (5 μg)</td>
</tr>
<tr>
<td>BsaI-HF (30 unit)</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 6.8: Reaction mix for BsaI cut of pNIC28-Bsa4

The purified BsaI cut vector was further digested with T4 DNA polymerase in the presence of dGTP (Table 6.9). The mixture was incubated for 30 mins at 22 ºC and then inactivate T4 DNA polymerase at 75 ºC for 20 min.

<table>
<thead>
<tr>
<th>Component</th>
<th>Reaction Volume/μL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrapure water</td>
<td>10.75</td>
</tr>
<tr>
<td>BsaI-digested vector</td>
<td>25</td>
</tr>
<tr>
<td>10 × T4 polymerase buffer</td>
<td>5</td>
</tr>
<tr>
<td>dGTP 25 mM</td>
<td>5</td>
</tr>
<tr>
<td>BSA 10mg/mL</td>
<td>0.5</td>
</tr>
<tr>
<td>100 mM DTT</td>
<td>2.5</td>
</tr>
<tr>
<td>T4 DNA polymerase</td>
<td>1.25</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>50</strong></td>
</tr>
</tbody>
</table>

Table 6.9: Reaction mix for T4 DNA polymerase digestion of pNIC28-Bsa4
6.1.6.2 Preparation of insert

A synthetic gene fragment was digested by T4 DNA polymerase in 10 µL reaction mixture (Table 6.10). The reaction mixture was incubated for 30 mins at 22 ºC and then T4 DNA polymerase was inactivated by incubating for 20 mins at 75 ºC.

<table>
<thead>
<tr>
<th>Component</th>
<th>Reaction Volume/ µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA 10 ng/µL</td>
<td>5</td>
</tr>
<tr>
<td>Ultrapure water</td>
<td>2.15</td>
</tr>
<tr>
<td>NEB buffer 2</td>
<td>1</td>
</tr>
<tr>
<td>dCTP 25 mM</td>
<td>1</td>
</tr>
<tr>
<td>100 mM DTT</td>
<td>0.5</td>
</tr>
<tr>
<td>BSA 10mg/mL</td>
<td>0.1</td>
</tr>
<tr>
<td>T4 DNA polymerase</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td><strong>10</strong></td>
</tr>
</tbody>
</table>

Table 6.10: Reaction mix for T4 DNA polymerase digesting of insert

6.1.6.3 Annealing

1 µL of the T4 DNA polymerase digested insert and 2 µL of the BsaI/T4 DNA polymerase digested vector were mixed in a sterilised 1.5 mL microcentrifuge tube (Eppendorf) and incubated for 30 minutes at rt.

6.1.7 Cloning in the presence of ligase

6.1.7.1 Preparation of insert and vector

A PCR product and pYUB28b vector plasmid were double digested by NdeI and HindIII restriction enzymes in a reaction mixture of 60 µL (Table 6.11 and 6.12). The reaction mixtures were centrifuged for 3 seconds and incubated for 3 hrs at 37 ºC and then agarose gel electrophoresis was run for the confirmation of digestion. To ensure plasmid was digested with both the restriction enzymes, controls with single digestion were also employed. The DNAs were purified using GeneJET Genomic DNA Purification Kit (Thermo Scientific).
### Table 6.11: Reaction mix for NdeI and HindIII digestion of PCR product

<table>
<thead>
<tr>
<th>Components</th>
<th>Amounts (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR product (73 ng/µL)</td>
<td>36</td>
</tr>
<tr>
<td>Buffer 2.1</td>
<td>6</td>
</tr>
<tr>
<td>NdeI</td>
<td>3</td>
</tr>
<tr>
<td>HindIII</td>
<td>3</td>
</tr>
<tr>
<td>Ultrapure water</td>
<td>12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>60</strong></td>
</tr>
</tbody>
</table>

### Table 6.12: Reaction mix for NdeI and HindIII digestion of pYUB28b vector

<table>
<thead>
<tr>
<th>Components</th>
<th>Amounts (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pYUB28b vector</td>
<td>5</td>
</tr>
<tr>
<td>Buffer 2.1</td>
<td>6</td>
</tr>
<tr>
<td>NdeI</td>
<td>3</td>
</tr>
<tr>
<td>HindIII</td>
<td>3</td>
</tr>
<tr>
<td>Ultrapure water</td>
<td>12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>60</strong></td>
</tr>
</tbody>
</table>

### 6.1.7.2 Ligation

Ligation was performed at 18 °C for 18 hours using one to three vector to insert ratio. The components of ligation mixture are described in Table 6.13.

### Table 6.13: Ligation reaction mix

<table>
<thead>
<tr>
<th>Components</th>
<th>Amounts (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pYUB28b plasmid</td>
<td>8.4 (40 ng)</td>
</tr>
<tr>
<td>Insert</td>
<td>1 (56 ng)</td>
</tr>
<tr>
<td>Ligase buffer</td>
<td>1</td>
</tr>
<tr>
<td>Ligase</td>
<td>1</td>
</tr>
</tbody>
</table>

### 6.1.8 Transformation

#### 6.1.8.1 Heat shock general procedure

A frozen aliquot of competent cells was removed from -80 °C and allowed to thaw on ice. The cells (40 µL) were then mixed with 1-3 µL plasmid and incubated for 20 minutes on ice. The mixture was then heat shocked for 45 seconds at 42 °C in a water bath (Thermo Fisher Scientific), and was immediately transferred into ice and kept it for 2 minutes. 200 µL of super
optimal broth with catabolite repression (SOC) medium (Table 6.14) or LB medium (Table 6.15) was then added to the mixture, followed by incubation in a water bath (TechNe) at 37 ºC for 60 minutes. Finally, 100 µL of the transformation mixture was plated on an appropriate LB agar plate with appropriate antibiotic(s) and incubated overnight at 37°C.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>20 g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>0.58 g</td>
</tr>
<tr>
<td>KCl</td>
<td>0.186 g</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>2.03 g</td>
</tr>
<tr>
<td>Glucose</td>
<td>3.6 g</td>
</tr>
</tbody>
</table>

*Table 6.14: Recipes of SOC medium*

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>10 g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>5 g</td>
</tr>
</tbody>
</table>

*Table 6.15: Recipes of LB medium*

### 6.1.8.2 Electroporation general procedure

A frozen aliquot of electro-competent cells was removed from -80 ºC and allowed to thaw on ice. To 40 µL cells, 2 µL of a plasmid was added and incubated for 2 minutes on ice. The mixture was transferred to a pre-chilled 0.1 cm sterile electroporation cuvette (Bio-Rad Laboratories, Hercules, CA, USA). The TOP10 *E.coli* cells were electroporated at 1.8 kV, 25 µF, 200 Ω using Micropulser (BioRad). 1 mL of LB medium was added and then transferred to 1.5 mL microcentrifuge tube, followed by incubation in a water bath (TechNe) at 37 ºC for 60 minutes. 100 µL of transformation mixture was plated on an appropriate LB agar plate with appropriate antibiotic(s) and incubated overnight at 37°C.
6.1.8.3 Transformation of recombinant pNIC28-Bsa4 vector to E. coli XL10-Gold cells

Heat shock general procedure was followed using an LB agar plate containing 5 % sucrose and 50 µg/mL kanamycin.

6.1.8.4 Transformation of recombinant pYUB28b vector to E. coli Top10 cells

Electroporation general procedure was followed using a low salt LB-agar plate with 50 µg/mL hygromycin B.

6.1.9 Plasmid preparation and DNA sequencing

The transformed E.coli XL10-GOLD cells were grown overnight at 37 ºC in 100 mL 2YT medium containing 50 µg/mL kanamycin and the transformed E. Coli Top10 cells were grown in 10 mL LB medium containing 50 µg/mL hygromycin. GeneJET plasmid miniprep kit (Thermo Fisher Scientific) was used to extract and purify plasmids from the transformed competent cells. The recombinant plasmid was confirmed by DNA sequencing (DNA Sequencing Centre, The University of Auckland).

6.1.10 Small scale optimisation and large scale protein production of ICL1, ICL2a and ICL2b using BL21 (DE3)

6.1.10.1 Transformation of recombinant pNIC28-Bsa4 vector to BL21 (DE3) cells

Heat shock general procedure was followed using an LB agar plate containing 50 µg/mL kanamycin.

6.1.10.2 Starter culture

A colony of a transformed cell was picked and inoculated into 100 mL 2YT medium (Table 6.16) containing 50 µg/mL kanamycin in 500 mL Erlenmeyer culture flask (Schott). It was
then incubated overnight at 37 °C in an incubator shaker (Thermo Scientific) with a shaking rate 180 rpm.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>16 g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>10 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>5 g</td>
</tr>
</tbody>
</table>

**Table 6.16: 2YT medium composition**

6.1.10.3 Small scale growth optimisation

Protein expression was attempted in three isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations (0.2 mM, 0.5 mM and 1 mM) and induced at three temperatures (18 °C, 28 °C and 37 °C). 200 μL of freshly prepared starter culture was then inoculated into nine flasks containing 100 mL 2YT medium in the presence of 50 μg/mL kanamycin and incubated at 37 °C in an incubator shaker (Thermo Scientific) till the cell density reached an optical density (OD$_{600}$) of 0.6-0.8. The OD$_{600}$ was measured using Nanodrop 2000c spectrophotometer (Thermo Scientific). Once the OD$_{600}$ was attained, an appropriate amount (0.2 mL or 0.5 mL or 1 mL) of 100 mM IPTG was added to a set of 3 flasks making final concentrations, 0.2 mM, 0.5 mM and 1 mM and were incubated overnight at three different temperatures, 18 °C, 28 °C and 37 °C. The cell pellets were harvested at 4°C by centrifugation using Mulifuge 3S-R centrifuge (Thermo Electron Corporation) at 4,000 rpm and resuspended into 5 times (w/v) of HEPES buffer pH 7.8 (50 mM HEPES, 500 mM NaCl, 5 mM imidazole). The mixture was sonicated on ice using a Misonix S-4000 sonicator at an amplitude of 60 for 15 seconds with a 45-seconds interval for three runs. The cell lysates were collected by centrifugation and amount of protein expressed was quantified running SDS-PAGE.
6.1.10.4 Large scale growth of ICL1 and ICL2b

Large scale growth was conducted using 2000 mL Erlenmeyer culture flasks (Schott) each containing 500 mL 2YT medium in the presence of 50 µg/mL kanamycin. Freshly prepared starter culture (1 mL) was added to each flask, which was then incubated at 37 ºC in an incubator shaker (Thermo Scientific) with shaking rate 180 rpm until cell density reached an OD$_{600}$ between 0.6 and 0.8. Once the OD$_{600}$ was reached, appropriate amount IPTG (for ICL1, 0.2 mM final concentration and for ICL2b, 0.5 mM) was added to each flask and incubated in a shaking incubator for 16 hrs at 18 ºC and 180 rpm. The cells were harvested by centrifugation using SORVALL LYNX400 centrifuge (Thermo Scientific) for 30 minutes at 4 ºC and 15000 rpm. The cells were stored in a resealable plastic bag at -80 ºC in Innova U101 ultra-low temperature Freezer (New Brunswick).

6.1.10.5 Cell lysis

20 g cell pellets were resuspended in 100 mL HEPES buffer pH 7.8 (50 mM HEPES, 500 mM NaCl and 5 mM imidazole) containing 1% (v/v) Halt Protease Inhibitor Cocktail (EDTA-Free) by stirring on a hot plate stirrer using ice. The cells were lysed by sonication (60 % amplitude, four cycles of 20 seconds with 40 seconds rest in between) on ice using a Misonix S-4000 sonicator. Then, insoluble debris was removed by centrifugation for 20 mins at 4ºC and 15000 rpm. The supernatant was then filtered using a 0.45 µM filter (Millipore).
6.1.11 Small scale optimisation and large scale protein production of ICL2 and ICL2a using BL21 LOBSTR and *M. smegmatis*

6.1.11.1 Transformation of recombinant pYUB28b vector to plasmid pGro7 co-transformed BL21 LOBSTR

Heat shock general procedure was followed using an LB agar plate containing 50 μg/mL hygromycin B and 50 μg/mL chloramphenicol.

6.1.11.2 Transformation of recombinant pYUB28b vector to *M. smegmatis* MC<sup>2</sup>4517

Slightly modified electroporation general procedure was followed. 260 μL of 10 % glycerol was added to the mixture of cells and plasmids and electroporated at 2500 V, 25 μF, 100 Ω using 2 mm cuvette by Gene Pulser Xcell (BioRad). LB medium (1 mL) was added followed by incubation for 3 hours at 37 ºC and plated on LB agar plate containing 50 μg/mL hygromycin B and 50 μg/mL kanamycin and incubated for 3 days at 37 ºC.

6.1.11.3 Starter culture

6.1.11.3.1 BL21 LOBSTR

A colony of a transformed cell was picked and inoculated into 10 mL LB medium in a 50 mL falcon tube (Thermo Fisher Scientific) containing appropriate antibiotics (50 μg/mL hygromycin and 50 μg/mL chloramphenicol). It was then incubated overnight at 37 ºC in an incubator shaker (Thermo Fisher Scientific) with a shaking rate 180 rpm.

6.1.11.3.2 *M. smegmatis*

A colony of a transformed cell was picked and inoculated into 10 mL LB medium in a 50 mL falcon tube containing appropriate antibiotics (50 μg/mL hygromycin B and 50 μg/mL
kanamycin). It was then incubated for 3 days at 37 °C in an incubator shaker (Thermo Scientific) with a shaking rate 180 rpm.

### 6.1.11.4 Small scale growth optimisation

#### 6.1.11.4.1 BL21 LOBSTR

Protein expression was attempted in three conditions (leaky, 0.5 mM IPTG and autoinduction) and induced at two temperatures (18 °C and 37 °C). 20 µL of freshly prepared starter culture was inoculated into 4 flasks containing 10 mL terrific broth medium (Table 6.17) and 2 flasks containing 10 mL auto-inducing terrific broth medium (Table 6.18) in the presence of appropriate antibiotics (50 µg/mL hygromycin and 50 µg/mL chloramphenicol) and incubated at 37 °C for 4 hours. Arabinose (final concentration 0.1 %) was added and continued incubation at 37 °C till cell density reached an optical density (OD_{600}) of 0.6-0.8. The OD_{600} was measured using NANODROP 2000c spectrophotometer (Thermo Scientific). Once the OD_{600} reached, an appropriate amount (0.5 mL of 100 mM IPTG) was added to two flasks making 0.5 mM final concentration. Cultures were divided into two sets and were incubated at 18 °C and 37 °C for overnight. The cell pellets were harvested at 4°C by centrifugation using Mulifuge 3S-R centrifuge (Thermo Electron Corporation) at 4,000 rpm and resuspended into 1 mL of HEPES buffer pH 7.5 (20 mM HEPES pH 7.5, 300 mM NaCl, 5 % glycerol, 20 mM imidazole and 1mM β-mercaptoethanol). The resuspended cells were transferred to lysis tube containing zirconia/silica. Cells were burst using Fastprep FP120 (SAVANT) for 45 seconds at 60 speed. The cell lysates were collected by centrifugation for 10 minutes at 4 °C and 15000 rpm and then transferred into a 1.5 microcentrifuge tube containing cobalt beads. It was incubated for an hour, followed by centrifugation for a minute at 4 °C and 5000 rpm. The supernatant was discarded and beads were washed with HEPES buffer pH 7.5. Finally, beads were resuspended.
in 100 µL HEPES buffer pH 7.5 containing 250 mM of imidazole and centrifuged to collect the supernatant. The amount of soluble protein expressed was quantified running SDS-PAGE.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>12 g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>24 g</td>
</tr>
<tr>
<td>Glycerol</td>
<td>4 mL</td>
</tr>
<tr>
<td>K$_2$HPO$_4$ (0.72 M)</td>
<td>50 mL</td>
</tr>
<tr>
<td>KH$_2$PO$_4$ (0.17 M)</td>
<td>50 mL</td>
</tr>
</tbody>
</table>

Table 6.17: Terrific broth medium composition

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone</td>
<td>12 g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>24 g</td>
</tr>
<tr>
<td>Glycerol</td>
<td>8 mL</td>
</tr>
<tr>
<td>Lactose</td>
<td>5 g</td>
</tr>
<tr>
<td>Glucose</td>
<td>0.15 g</td>
</tr>
</tbody>
</table>

Table 6.18: Auto-induction terrific broth medium composition

6.11.4.2 M. smegmatis

Protein expression was attempted at 37 °C using the complex autoinduction medium ZYM 5052 (Table 6.19). 300 µL freshly prepared starter culture was inoculated into the medium containing appropriate antibiotics (50 µg/mL hygromycin B and 50 µg/mL kanamycin) and 0.05 % Tween 80 and incubated at 37 °C for 3 days, followed by the procedure described in 6.11.4.1.
<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZY</td>
<td>1 % N-Z-amine AS 957 mL</td>
</tr>
<tr>
<td></td>
<td>0.5 % Yeast extract</td>
</tr>
<tr>
<td>MgSO₄ (1 M)</td>
<td>2 mL</td>
</tr>
<tr>
<td>50 × M</td>
<td>25 mM Na₂HPO₄ 20 mL</td>
</tr>
<tr>
<td></td>
<td>25 mM KH₂PO₄</td>
</tr>
<tr>
<td></td>
<td>5 mM Na₂SO₄</td>
</tr>
<tr>
<td></td>
<td>50 mM NH₄Cl</td>
</tr>
<tr>
<td>50 × 5052</td>
<td>55 mM Glycerol 20 mL</td>
</tr>
<tr>
<td></td>
<td>2.8 mM Glucose</td>
</tr>
<tr>
<td></td>
<td>5.6 mM α-D-Lactose</td>
</tr>
<tr>
<td>1000 × metals</td>
<td>50mM FeCl₃ 1 mL</td>
</tr>
<tr>
<td></td>
<td>20mM CaCl₂</td>
</tr>
<tr>
<td></td>
<td>10mM MnCl₂</td>
</tr>
<tr>
<td></td>
<td>10mM ZnSO₄</td>
</tr>
<tr>
<td></td>
<td>2mM CoCl₂</td>
</tr>
<tr>
<td></td>
<td>2mM CuCl₂</td>
</tr>
<tr>
<td></td>
<td>2mM NiCl₂</td>
</tr>
<tr>
<td></td>
<td>2mM Na₂MoO₄</td>
</tr>
<tr>
<td></td>
<td>2mM Na₂SeO₃</td>
</tr>
<tr>
<td></td>
<td>2mM H₃BO₃</td>
</tr>
</tbody>
</table>

Table 6.19: ZYM 5052 medium composition (1 L)

6.1.11.5 Large scale growth of ICL2

Large scale growth was conducted using 2000 mL Erlenmeyer culture flasks (Schott) each containing 500 mL terrific broth medium containing 50 µg/mL hygromycin and 50 µg/mL chloramphenicol. 1 mL of freshly prepared starter culture was added to each flask, which was then incubated at 37 °C in an incubator shaker with shaking rate 180 rpm until cell density reached an OD₆₀₀ between 0.4-0.6. Once the OD₆₀₀ was reached, arabinose (final concentration 0.1 %) was added to each flask and incubated in shaking incubator for 16 hrs at 37 °C and 180
rpm. The cells were harvested by centrifugation using Sorvall Lynx400 centrifuge (Thermo Scientific) for 30 minutes at 4 °C and 15000 rpm. The cells were stored in 50 mL falcon tubes at -20 °C room.

6.1.11.6 Cell lysis

10 mL cell pellets in the 50 mL falcon tubes were resuspended in 20 mL HEPES buffer pH 7.5 (20 mM HEPES, 150 mM NaCl, 20 mM imidazole and 1 mM β-mercaptoethanol) containing 1% (v/v) Halt Protease Inhibitor Cocktail (EDTA-Free) by vortexing. The cells were lysed by cell bursting system (Microfluidics). Then, insoluble debris was removed by centrifugation for 20 mins at 4°C and 15000 rpm. The supernatant was then filtered using a 0.45 µM filter (Millipore).

6.1.12 Chromatography

All columns were operated with AKTA start (ICL1 and ICL2b) or AKTA primer (ICL2) liquid chromatography system (GE Healthcare) and monitored by UV detector at 280 nm. The overall procedures of protein purification were performed at 4 °C.

6.1.12.1 ICL1 and ICL2b

6.1.12.1.1 Affinity Chromatography

The components of different buffers used for affinity chromatography are described in Table 6.20-6.24. All the buffers were filtered through 0.2 µM filter (Millipore) and degassed under vacuum for 15 mins before use. 5 mL HisTrap HP column with immobilised metal chromatography (IMC) sepharose matrix (GE Healthcare) charged with NiCl₂ was washed with six column volumes of filtered and degassed ultrapure water and equilibrated with six column volumes of binding buffer (Table 6.20). About 100 mL clear cell lysate was loaded onto the
column with flow rate 1 mL/min, followed by a wash with six column volumes of wash buffer (Table 6.21). Protein was eluted in 3 mL fraction with the help of elution buffer (Table 6.22). All fractions with UV absorbance were analysed by SDS-PAGE to determine which fractions contained desired purified protein. All fractions with protein were mixed together and concentrated to 1 mL using Amicon® Ultra-15 centrifugal filter devices (Merck Millipore) with a 30 kDa MWCO (molecular weight cut-off) and centrifugation at 4 °C and 3000 rpm using bench-top centrifuge (Multifuge 3S-R, Thermo Fisher Scientific). The column was finally stripped with strip buffer (Table 6.23) followed by a wash with binding buffer and Mili-Q water (5 column volume each) and recharged with charge buffer (Table 6.24) and stored in 20 % ethanol.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imidazole</td>
<td>0.34 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>29.2 g</td>
</tr>
<tr>
<td>HEPES</td>
<td>11.9 g</td>
</tr>
</tbody>
</table>

Table 6.20: Binding buffer contained 5 mM imidazole, 500 mM NaCl and 50 mM HEPES buffered at pH 7.8.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imidazole</td>
<td>2.0 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>29.2 g</td>
</tr>
<tr>
<td>HEPES</td>
<td>11.9 g</td>
</tr>
</tbody>
</table>

Table 6.21: Wash buffer contained 30 mM imidazole, 500 mM NaCl and 50 mM HEPES buffered at pH 7.8.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imidazole</td>
<td>34.0 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>29.2 g</td>
</tr>
<tr>
<td>HEPES</td>
<td>11.9 g</td>
</tr>
</tbody>
</table>

Table 6.22: Elution buffer contained 500 mM imidazole, 500 mM NaCl and 50 mM HEPES buffered at pH 7.8.
<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDTA</td>
<td>37.2 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>29.2 g</td>
</tr>
<tr>
<td>HEPES</td>
<td>11.9 g</td>
</tr>
</tbody>
</table>

**Table 6.23:** Strip buffer contained 100 mM EDTA, 500 mM NaCl and 50 mM HEPES buffered at pH 7.8.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>NiSO₄·7H₂O</td>
<td>14.0 g</td>
</tr>
</tbody>
</table>

**Table 6.24:** Charge buffer contained 50 mM NiSO₄ (no pH adjustment).

### 6.1.12.1.2 Size exclusion chromatography (SEC)

The components of the buffer used are described in Table 6.25. The buffer was filtered through 0.2 µM filter (Millipore) and degassed under vacuum for 15 mins before use. SEC was conducted using HiPrep 16/60 sephacryl S-100 HR (GE Healthcare) at a flow rate of 0.5 mL/min. The column was washed with two column volume of ultrapure water followed by two column volume of SEC buffer. The protein (1 mL) was loaded onto the column with a 1 mL injection loop. The protein was eluted with SEC buffer in 1 mL fraction. All fractions with UV absorbance were analysed by SDS-PAGE to determine which fractions contained purified protein. The fractions with proteins were mixed and concentrated to 400 µL. 200 µL was aliquoted 20 µL each to 1.5 mL microcentrifuge tube and stored in -80 °C after rapid freezing in liquid nitrogen. Remaining 200 µL was buffer exchanged into NMR buffer (50 mM tris-D11, pH 7.5) by spin concentration using Amicon Ultra-15 centrifugal filter devices (Merck Millipore) with a 30 kDa MWCO at 4 °C and 3000 rpm for at least 5 rounds of dilution. Finally, it was aliquoted and stored at -80 °C until use.
<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris HCl</td>
<td>6.05 g</td>
</tr>
<tr>
<td>NaCl</td>
<td>8.76 g</td>
</tr>
</tbody>
</table>

Table 6.25: SEC buffer used for purification of ICL1 and ICl2b containing 50 mM Tris and 150 mM NaCl buffered at pH 7.5.

6.1.12.2 ICL2

6.1.12.2.1 Affinity Chromatography

All the buffers were filtered through 0.2 μM filter (Millipore) and degassed under vacuum for 15 mins before use. All the process were conducted maintaining flow rate 1 mL/min. 5 mL HisTrap HP column with immobilised metal chromatography (IMC) sepharose matrix (GE Healthcare) charged with NiCl₂ was washed with six column volumes of filtered and degassed Mili-Q water and equilibrated with six column volumes of slightly modified binding buffer (20 mM HEPES, 150 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol and buffered at pH 7.5). 100 mL clear cell lysate was loaded onto the column using the peristaltic pump (GE Healthcare). The column was attached to AKTA followed by a wash with aforementioned modified binding buffer till UV flattened out. Protein was eluted using slightly modified elution buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 500 mM imidazole and 1 mM β-mercaptoethanol buffered at pH 7.5) in 3 mL fraction using gradient elution technique setting 100 % elution buffer in 150 minutes. All fractions with UV absorbance were analysed by SDS-PAGE to determine which fractions contained desired purified protein. All fractions with protein were mixed and concentrated to 2 mL by centrifugation at 4 °C and 3000 rpm using bench-top centrifuge (Multifuge 3S-R, Thermo electron) using Amicon Ultra-15 centrifugal filter devices (Merck Millipore) with a 100 kDa MWCO.
6.1.12.2.2 Size exclusion chromatography (SEC)

The components of the buffer used are described in Table 6.26. The buffer was filtered through 0.2 μM filter (Millipore) and degassed under vacuum for 15 mins before use. SEC was conducted using HiPrep 16/60 sephacryl S-100 HR (GE Healthcare) at a flow rate of 0.5 mL/min. The column was washed with two column volume of ultrapure water followed by two column volume of SEC buffer. The protein (0.5 mL) was loaded onto the column with a 0.5 mL injection loop. The protein was eluted with SEC buffer (Table 6.26) in 1 mL fraction. All fractions with UV absorbance were analysed by SDS-PAGE to determine which fractions contained purified protein. The process of injection and elution were repeated for 4 times. The main fractions from each run mixed and used for crystallisation. The others fractions containing proteins were mixed and concentrated 500 μL using Amicon® Ultra-15 centrifugal filter devices (Merck Millipore) with a 100 kDa MWCO and stored in -80 °C after rapid freezing in liquid nitrogen.

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount per litre</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 M HEPES pH 7.5</td>
<td>20 mL</td>
</tr>
<tr>
<td>5 M NaCl</td>
<td>30 mL</td>
</tr>
<tr>
<td>β-Mercaptoethanol</td>
<td>34.965 μL</td>
</tr>
</tbody>
</table>

Table 6.26: SEC buffer for ICL2 purification containing 20 mM HEPES and 150 mM NaCl and 1 mM β-mercaptethanol buffered at pH 7.5.

6.1.13 Digestion of polyhistidine tag by tobacco etch virus (TEV) protease

The plasmid of TEV protease was a gift from A/Prof. Christopher J. Squire laboratory (School of Biological Sciences, The University of Auckland). TEV protease was produced and purified according to procedures described by Tropea et al.2 25 mg of ICL2b was mixed with 0.5 mg of TEV protease making 1 to 50 of TEV protease to protein, which was incubated at 4 °C for 18 hrs. The mixture was then applied through affinity chromatography to remove the cleaved polyhistidine tag and TEV protease. The protein was collected in the flow through.
6.2 NMR Methods

6.2.1 Materials

D$_2$O and DMSO-D$_6$ were obtained from Cambridge Isotope Laboratories, USA. Tris-D$_{11}$ was obtained from Euriso-top or Cortecnet. All other chemicals obtained from Sigma-Aldrich otherwise stated.

6.2.2 Buffer for NMR experiment

50 mM Tris-D$_{11}$ buffer was made up in Ultrapure water and buffered at pH 7.5 using concentrated HCl. pH adjustments were carried out using pH meter (INESA). Electrodes were calibrated either to pH 7.0 or 4.0 or 9.0 before use. All buffers were filter sterilised through 0.22 µM filters (Millipore) and were stored in 50 mL falcon tubes at 4 ºC.

6.2.3 Sample preparation

Stock solutions of magnesium chloride, DL-isocitrate and (2S,3R)-2-methylisocitrate were prepared in Tris-D$_{11}$ and all other compounds were prepared in DMSO-D$_6$. The stock solutions stored at -20 ºC before use. All samples (500 µL) for NMR were prepared just before the experiment in 1.5 mL Eppendorf tubes, mixed by vortexing and transferred to standard 5 mm NMR tubes (Wilmad) followed by centrifugation.

6.2.3.1 ICL1

For kinetic parameters, all samples contained appropriate amount of ICL1 (for DL-isocitrate, 190 nM; for (2S,3R)-2-methylisocitrate, 2 µM), 5 mM MgCl$_2$, variable concentration of DL-isocitrate (50 µM-1 mM)/(2S,3R)-2-methylisocitrate (250 µM-2 mM) and buffered with 50 mM Tris-D$_{11}$ (pH 7.5) in 90% H$_2$O and 10% D$_2$O. For single concentration inhibition study, all samples contained 190 nM ICL1, 1 mM DL-isocitrate, 5 mM MgCl$_2$, 100 µM inhibitors and
buffered with 50 mM Tris-D_{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O. For IC\textsubscript{50} measurement, all samples contained 190 nM ICL1, 1 mM DL-isocitrate, 5 mM MgCl\textsubscript{2}, variable concentration inhibitors (0.1 µM-300µM where possible) and buffered with 50 mM Tris-D\textsubscript{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O.

**6.2.3.2 ICL2**

For kinetic parameters in the absence of acetyl-CoA, all samples contained 2 µM ICL2, 5 mM MgCl\textsubscript{2}, variable concentration of DL-isocitrate (250 µM-2 mM) and buffered with 50 mM Tris-D\textsubscript{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O. For single concentration modulation study, all samples contained 500 nM ICL2, 1 mM DL-isocitrate, 5 mM MgCl\textsubscript{2}, and 200 µM acetyl-CoA/propionyl-CoA/succinyl-CoA/CoA and buffered with 50 mM Tris-D\textsubscript{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O. For the measurement of amount of acetyl-CoA and propionyl-CoA needed for full activation, all samples contained 200 nM ICL2, 1 mM DL-isocitrate/1mM (2S,3R)-2-methylisocitrate, 5 mM MgCl\textsubscript{2}, variable concentration of acetyl-CoA (0.5 µM-50 µM)/propionyl-CoA (10 µM-400µM) and buffered with 50 mM Tris-D\textsubscript{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O. For kinetic parameters in the presence of acetyl-CoA using DL-isocitrate as substrate, all sample contained 200 nM ICL2, variable concentration of DL-isocitrate (100 µM-1 mM), 5 mM MgCl\textsubscript{2}, 25 µM acetyl-CoA and buffered with 50 mM Tris-D\textsubscript{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O. For kinetic parameters in the presence of acetyl-CoA using (2S,3R)-2-methylisocitrate as substrate, all sample contained 1 µM ICL2, variable concentration of (2S,3R)-2-methylisocitrate (250 µM-2 mM), 5 mM MgCl\textsubscript{2}, 25 µM acetyl-CoA and buffered with 50 mM Tris-D\textsubscript{11} (pH 7.5) in 90% H\textsubscript{2}O and 10% D\textsubscript{2}O.
6.2.4 Sample acquisition and processing

NMR experiments were conducted at a $^1$H frequency of 500 MHz using a Bruker Avance III HD spectrometer equipped with a BBFO probe. The instrument was operated by TopSpin 3.1 software. Experiments were conducted at 300 K. The pulse tip-angle calibration using the single-pulse nutation method (Bruker pulsecal routine) was undertaken for each sample. Water suppression was achieved using the excitation sculpting method. Unless otherwise stated, the number of transients was 16, and the relaxation delay was 2 seconds. A sample without protein was prepared and run as a control followed by addition of protein to the sample. The lag time between the addition of an enzyme and the end of the first experiment was usually 4 minutes. All measurements were performed in triplicate.

6.2.5 Data treatment

All data were processed using TopSpin 3.5pl7. Initial rates were calculated by linear fitting using Excel 2013 (Microsoft) for data points up to 20% turnover (typically 4 minutes). Kinetic parameters were obtained using the Hanes plot. For single concentration inhibition assay, concentration vs time was plotted using Excel 2013. IC$_{50}$ values were obtained by SigmaPlot 13 and fitted with the sigmoid, 3 parameter model.

6.3 Thermal shift assay

Thermal shift assay was carried out using a MyiQ real time PCR instrument (Bio-Rad) or a QuantStudio 3 (applied Biosystems) by increasing temperature from 25 to 95 °C at 1 °C increment every 60 seconds. A sample was prepared in a mass mixture of 80 µL in 1.5 mL Eppendorf tube each containing 20 µM enzyme, 1 mM compound and 1 mM MgCl$_2$ in 50 mM Tris-HCl pH 7.5 and 8 µL of SYPRO Orange dye (Sigma Aldrich) which was prepared to 200X dilution in 50 mM Tris-HCl (pH 7.5) of stock (5000X concentrate). The sample was incubated
on ice for an hour and was aliquoted to 3 x 25 µL in a 96 well plate (BioRad). A standard was prepared by excluding compounds. Protein unfolding was monitored by measuring the fluorescence of the dye. For determination of protein melting temperature values, a melting curve for each data set was analysed by SigmaPlot 13 (USA) and fitted with the Sigmoid, 3 parameter model.

6.4 X-ray crystallography

6.4.1 Acetyl-CoA free ICL2

6.4.1.1 Sample preparation

A sample mixture containing 8 mg/mL ICL2 in buffer (20 mM HEPES, 150 mM NaCl and 1 mM β-mercaptoethanol), 1 mM MgCl2 and 1 mM succinate was prepared followed by incubation on ice for at least 10 minutes.

6.4.1.2 Initial Screening

Initial crystallisation screening was carried out using the sitting-drop vapour diffusion method in 96-well plate format (Intelliplate™, Art Robbins Instrument) using an Oryx4 robot (Douglas Instruments) dispensing system at 18 °C. Three commercial screens were employed for initial screening, JCSG-plus5, Morpheus6 and PACT premier7. 50 µL of each crystallization screen solution was manually dispensed in the reservoirs of the crystallisation plate. The plate was then transferred to the Oryx robotic system, which was set up to dispense drops by consecutive dispensing of protein solution and crystallisation reservoir solutions. 50 µL of a protein sample was prepared and used for the screening. Protein and reservoir solution were mixed in one to one ratio (0.15 µL protein plus 0.15 µL precipitant).
6.4.1.3 Fine Screening

A fine screening was designed based on the initial conditions that gave crystals with slightly different parameters such as protein precipitant and salt concentrations. The fine screening was carried out in 96 well CrystalClear P strips (HAMPTON RESEARCH) using sitting drop method. 80 μL of the crystallisation reservoir solutions were manually transferred into wells. Protein and screen were mixed to a final volume of 2 μL in one to one ratio of protein to crystallisation reservoir solution and sealed with crystal clear tape.

6.4.1.4 Additive screening

Morpheus additive screening\(^8\) was used to further optimise the best condition obtained from the fine screening experiments. The additive screening was typically carried out in 96 well CrystalClear P strips (HAMPTON RESEARCH) using the sitting drop method. Crystallisation reservoir solutions were prepared by manually transferring 70 μL of mother liquor (the best condition from fine screening) and 10 μL of additive screen into wells. Protein and crystallisation reservoir solutions were mixed to a final volume of 2 μL in one to one ratio of protein to crystallisation reservoir solution and sealed with crystal clear tape.

6.4.1.5 X-ray Data collection and processing

The X-ray diffraction data sets were collected at the Australian Synchrotron, Melbourne, Australia. This facility is equipped with two beamlines, MX1 (with a bending magnet) and MX2 (with an undulator). The data were collected with both beamlines MX1 and MX2 using the Blu-Ice software.\(^9\) All initial data sets were indexed and processed using XDS\(^10\), and scaled with AIMLESS\(^11\) from the CCP4 program suite.\(^12\) The structure was solved by molecular replacement using Phaser\(^13\) with the isocitrate lyase of Aspergillus nidulans (PDB code:1DQU) as a search model.\(^14\) The initial structure was refined using REFMACS\(^15\) and cycles of
automatic (phenix.autobuild\textsuperscript{16}) and manual model building (COOT\textsuperscript{17}) resulted in a protein model that contained the structure for N-terminal of the ICL2 protein. A new data set from a crystal grown in the presence of sodium tungstate was subsequently collected that was then used to solve the structure of the ICL2 using the afore-mentioned N-terminal structure of the ICL2 protein. Repeated building cycles with phenix.autobuild\textsuperscript{16} and COOT\textsuperscript{17} resulted in a protein model that contained both the N- and C-terminal domains. The final structure was then refined using REFMAC5. The PDB\_redo program\textsuperscript{18} was used in the final stages of refinement. Water molecules were identified by their spherical electron density and appropriate hydrogen bond geometry with the surrounding structure Preliminary coordinates were visualised and figures produced using Pymol.\textsuperscript{19}

6.4.2 Acetyl-CoA bound structure

6.4.2.1 Sample preparation

A sample mixture containing 6 mg/mL ICL2 in buffer (20 mM HEPES, 150 mM NaCl and 1 mM β-mercaptoethanol), 2 mM MgCl\textsubscript{2}, 1 mM succinate and 1 mM acetyl-CoA was prepared followed by incubation on ice for at least 10 minutes.

6.4.2.2 Screening

Screening using the fine screens (described in 6.4.1.3 ) was carried out in 96 well CrystalClear P strips (HAMPTON RESEARCH) using sitting drop method. 80 μl of the crystallisation reservoir solutions were manually transferred into wells. Protein and screen were mixed to a final volume of 2 μl in one to one ratio of protein to crystallisation reservoir solution and sealed with crystal clear tape.

198
6.4.2.3 X-ray Data collection and processing

The data collection was done as described in 6.4.1.5. The structure was solved by molecular replacement using MOLREP\textsuperscript{20} with the apo-ICL2 structure as a search model. The N-terminal domain was separately used to solve the structure, followed by using the resulting model to search for the C-terminal domain. The structure was completed by cycles of automated (phenix.autobuild\textsuperscript{16}) and manual (COOT\textsuperscript{17}) building and refinement using REFMAC5\textsuperscript{15}. Water molecules were identified by their spherical electron density and appropriate hydrogen bond geometry with the surrounding structure Preliminary coordinates were visualised and figures produced using Pymol\textsuperscript{19}. 
6.5 Synthesis of N-heterocycles

6.5.1 General

Commercially available reagents were used throughout without purification unless otherwise stated. Anhydrous solvents were used as supplied. Dichloromethane and toluene were dried using LC Technology Solutions Inc. SP-1 solvent purification system under an atmosphere of dry nitrogen. All reactions were routinely carried out in oven-dried glassware under a nitrogen atmosphere unless otherwise stated. Analytical thin layer chromatography was performed using 0.2 mm silica plates and compounds were visualized under 365 nm ultraviolet irradiation followed by staining with either alkaline permanganate or ethanolic vanillin solution. Infrared spectra were obtained using a Perkin Elmer spectrum One Fourier Transform Infrared spectrometer as thin films between sodium chloride plates. Absorption maxima are expressed in wavenumbers (cm\(^{-1}\)). Melting points were recorded on an Electrothermal melting point apparatus and are uncorrected. NMR spectra were recorded as indicated on an NMR spectrometer operating at 500, 400 and 300 MHz for \(^1\)H nuclei and 125, 100 and 75 MHz for \(^{13}\)C nuclei. Chemical shifts are reported in parts per million (ppm) relative to the tetramethylsilane peak recorded as δ 0.00 ppm in CDCl\(_3\)/TMS solvent, or the residual acetone (δ 2.05 ppm), chloroform (δ 7.26 ppm), DMSO (δ 2.50 ppm) or methanol (δ 3.31 ppm) peaks. The \(^{13}\)C NMR values were referenced to the residual acetone (δ 29.9 ppm) chloroform (δ 77.1 ppm), DMSO (δ 39.5 ppm) or methanol (δ 49.0 ppm) peaks. \(^{13}\)C NMR values are reported as chemical shift δ and assignment. \(^1\)H NMR shift values are reported as chemical shift δ, relative integral, multiplicity (s, singlet; d, doublet; t, triplet; q, quartet; quin, quintet; m, multiplet), coupling constant (J in Hz) and assignment. Assignments are made with the aid of DEPT 90, DEPT 135, COSY, NOESY and HSQC experiments. High resolution mass spectra were obtained by electrospray ionization in positive ion mode at a nominal accelerating voltage of 70 eV on a microTOF mass spectrometer.
6.5.2 General Procedure A; C-acylation

To polyphosphoric acid (PPA) at 100 °C was added a solution of substrate in toluene and the mixture was stirred at 100 °C for the time stated. The reaction mixture was quenched with ice cold water (10 mL), the aqueous suspension extracted with ethyl acetate and the organic layer was washed with water, brine, dried (Na₂SO₄), filtered and concentrated in vacuo. The crude residue was purified by flash chromatography on silica gel with eluent stated to give the desired product.

6.5.3 General Procedure B; N-acylation

A solution of substrate and potassium carbonate in DMF (4 mL) was stirred at 110 °C for the time stated. The solvent was concentrated in vacuo and water was added to the residue. The resulting aqueous suspension was extracted with ethyl acetate and the organic layer was washed with water, brine, dried (Na₂SO₄), filtered and concentrated in vacuo. The crude residue was purified by flash chromatography on silica gel with eluent stated to give the desired product.

6.5.4 General Procedure C; Dehydrogenation

A suspension of substrate and Pd/C (10 %) in phenylether-mesitylene or n-octanol was heated at the temperature and time stated. The reaction mixture was filtered through Celite® and the residual catalyst washed with ethyl acetate. The filtrate was concentrated in vacuo and the residue purified by flash chromatography on silica gel with eluent stated to give the desired product.
6.5.5 Synthesis

Dimethyl 2-((1H-pyrrol-2-yl)methyl)succinate 75

![Chemical Structure](image)

To a solution of dimethyl 2-methylene succinate (500 mg, 3.16 mmol) in dichloromethane (5 mL) and nitromethane (5 mL) was added pyrrole (0.62 mL, 9.48 mmol) and aluminium trichloride (443 mg, 9.48 mmol) and the solution was stirred at room temperature for 45 min. The reaction mixture was then quenched with water (20 mL). The aqueous layer was extracted with dichloromethane and the organic layer washed with brine, dried (Na$_2$SO$_4$), filtered and concentrated in vacuo. The crude residue was purified by flash chromatography on silica gel using hexanes-ethyl acetate (3:1) as eluent to afford the title compound (427 mg, 1.90 mmol, 60%) as light yellow liquid; $\nu_{\text{max}}$ (neat)/cm$^{-1}$ 3383, 2953, 1722, 1570, 1509, 1436, 1409, 1357, 1269, 1194, 1165, 1121, 1095, 1027, 1004, 885, 837, 789, 717; $\delta_H$ (400 MHz, CDCl$_3$) 8.36 (1 H, br s, NH), 6.69-6.67 (1 H, m, ArH), 6.09 (1 H, q, $J$ 2.9 Hz, ArH), 5.93-5.92 (1 H, m, ArH), 3.71 (3 H, s, Me), 3.67 (3 H, s, Me), 3.10 (1 H, quint, $J$ 6.6, CH), 3.00 (1 H, dd, $J$ 14.8 and 6.4, CH of CH$_2$), 2.90 (1 H, dd, $J$ 14.8 and 6.4, CH of CH$_2$), 2.68 (1 H, dd, $J$ 16.4 and 7.2, CH of CH$_2$), 2.50 (1 H, dd, $J$ 16.5 and 6.4, CH of CH$_2$); $\delta_C$ (100 MHz, CDCl$_3$) 175.0 (C), 172.4 (C), 128.0 (C), 117.2 (CH), 108.3 (CH), 107.3 (CH), 52.2 (Me), 51.9 (Me), 41.9 (CH), 35.0 (CH$_2$), 29.0 (CH$_2$); HRMS (ESI, [M + Na]$^+$) found 248.0893. [C$_{11}$H$_{15}$NO$_4$ + Na]$^+$ requires 248.0898.
Tetramethyl 2,2’-((1H-pyrrole-2,5-diyl)bis(methylene))disuccinate 76

To a solution of pyrrole (0.2 mL, 2.88 mmol) in dichloromethane (5 mL) and nitromethane (5 mL) was added dimethyl 2-methylene succinate (1.40 g, 7.2 mmol) and aluminium trichloride (960 mg, 7.2 mmol) and the solution was stirred at room temperature for 45 min. The mixture was then quenched with water (20 mL). The resulting solution was extracted with dichloromethane and the organic layer washed with brine, dried (Na$_2$SO$_4$), filtered and concentrated in vacuo. The crude residue was purified by flash chromatography on silica gel using hexanes-ethyl acetate (2:1) as eluent to afford the title compound (1.08 g, 2.82 mmol, 98%) as brown liquid; $\nu_{\text{max}}$ (neat)/cm$^{-1}$ 3383, 1727, 1436, 1355, 1263, 1193, 1163, 1094, 1038, 998, 890, 839, 773; $\delta_H$ (400 MHz, CDCl$_3$) 8.33 (1 H, br s, NH), 5.76 (2 H, d, $J$ 2.6, 2 x ArH), 3.70 (6 H, s, 2 x Me), 3.67 (6 H, s, 2 x Me), 3.06 (2 H, quin, $J$ 6.7, 2 x CH), 2.93 (2 H, dd, $J$ 14.8 and 6.4, 2 x CH of CH$_2$), 2.83 (2 H, dd, $J$ 14.8 and 6.7, 2 x CH of CH$_2$), 2.66 (2 H, dd, $J$ 16.5 and 7.6, 2 x CH of CH$_2$), 2.46 (2H, dd, $J$ 16.5 and 6.2, 2 x CH of CH2); $\delta_C$ (100 MHz, CDCl$_3$) 175.04 (C), 175.00 (C), 172.5 (2 x C), 127.60 (C), 127.58 (C), 107.43 (CH), 107.42 (CH), 52.2 (2 x Me), 51.9 (2 x Me), 42.07 (CH), 42.02 (CH), 35.10 (CH$_2$), 35.06 (CH$_2$), 29.4 (CH$_2$), 29.3 (CH$_2$); HRMS (ESI, [M + Na]$^+$) found 406.1472. [C$_{18}$H$_{25}$NO$_8$ + Na]$^+$ requires 406.1478.
Methyl 4-oxo-4,5,6,7-tetrahydro-1H-indole-6-carboxylate 78

General procedure A was performed using PPA (2 g) and 75 (100 mg, 0.45 mmol) in toluene (1 mL) for 3 h. Workup and column chromatography eluting with hexanes-ethyl acetate (1:1) gave the title compound (27 mg, 0.14 mmol, 31%) as a colourless solid; M.p. 127-131 °C; $\nu_{\text{max}}$ (neat)/cm$^{-1}$ 3289, 3130, 2952, 2922, 2851, 1725, 1632, 1569, 1505, 1474, 1435, 1377, 1350, 1293, 1271, 1244, 1234, 1193, 1152, 1116, 1077, 1003, 913, 902, 881, 866, 799, 764, 720, 692; $\delta_H$ (400 MHz, CDCl$_3$) 8.80 (1 H, br s, NH), 6.70 (1 H, t, $J$ 2.6, ArH), 6.54 (1 H, t, $J$ 2.8, ArH), 3.72 (3 H, s, Me), 3.29-3.22 (1 H, m, CH), 3.13-3.10 (2 H, m, CH$_2$), 2.82-2.69 (2 H, m, CH$_2$); $\delta_C$ (100 MHz, CDCl$_3$) 191.6 (C), 173.6 (C), 141.1 (C), 120.3 (C), 119.5 (CH), 106.2 (CH), 52.4 (Me), 41.3 (CH), 40.0 (CH$_2$), 25.5 (CH$_2$); HRMS (ESI, [M + Na]$^+$) found 216.0631. [C$_{10}$H$_{11}$NO$_3$ + Na]$^+$ requires 216.0637.

Methyl 4-hydroxy-1H-indole-6-carboxylate 79

General procedure C was performed using 78 (100 mg, 0.52 mmol) and Pd/C (100 mg) in phenylether (3.6 mL) and mesitylene (0.4 mL) for 240 °C for 36 h. Workup and column chromatography eluting with hexanes then hexanes-ethyl acetate (1:1) gave the title compound (52 mg, 0.27 mmol, 54%) as a brown solid; $\delta_H$ (400 MHz, CD$_2$OD) 7.68 (1 H, d, $J$ 1.0, ArH), 7.30 (1 H, d, $J$ 3.3, ArH), 7.06 (1 H, d, $J$ 1.4, ArH), 6.58 (1 H, dd, $J$ 0.8, 3.3, ArH),
3.88 (3 H, s, Me); δ_C (100 MHz, CD_3OD) 170.2 (C), 151.1 (C), 138.6 (C), 127.4 (C), 124.9 (C), 123.5 (CH), 107.4 (CH), 104.5 (CH), 100.0 (CH), 52.3 (Me); spectroscopic data consistent with the literature.\(^{21}\)

Methyl 5-oxo-5,6,7,8-tetrahydroindolizine-7-carboxylate 80

![Methyl 5-oxo-5,6,7,8-tetrahydroindolizine-7-carboxylate 80](image)

General procedure B was performed using 75 (375 mg, 1.7 mmol) and potassium carbonate (470 mg, 3.4 mmol) in DMF (4 mL) for 12 h. Workup and column chromatography eluting with hexanes-ethyl acetate (4:1) gave the title compound (198 mg, 1.02 mmol, 60%) as a fawn solid; M.p. 83-85 °C; v_max (neat)/cm\(^{-1}\) 1716, 1575, 1449, 1429, 1379, 1354, 1302, 1229, 1214, 1199, 1151, 1100, 1068, 1031, 1022, 1005, 917, 870, 825, 802, 750, 621, 594; δ_H (400 MHz, CDCl\(_3\)) 7.35 (1 H, dd, J 3.6 and 1.4, ArH), 6.22 (1 H, t, J 3.3, ArH), 6.03-6.02 (1 H, m, ArH), 3.73 (3 H, s, Me), 3.23-3.17 (1 H, m, CH of CH\(_2\)), 3.16-3.10 (1 H, m, CH), 3.03-2.97 (1 H, m, CH of CH\(_2\)), 2.94-2.85 (2 H, m, CH\(_2\)); δ_C (100 MHz, CDCl\(_3\)) 172.8 (C), 166.4 (C), 130.1 (C), 116.6 (CH), 113.1 (CH), 110.0 (CH), 52.5 (Me), 39.2 (CH), 35.3 (CH\(_2\)), 25.9 (CH\(_2\)); HRMS (ESI, [M + Na]\(^+\)) found 216.0631. [C\(_{10}\)H\(_{11}\)NO\(_3\) + Na]\(^+\) requires 216.0637.

Methyl 5-oxo-1,2,3,5-tetrahydroindolizine-7-carboxylate 82

![Methyl 5-oxo-1,2,3,5-tetrahydroindolizine-7-carboxylate 82](image)

General procedure C was performed using 80 (40 mg, 0.21 mmol) and Pd/C (40 mg) in n-octanol (2 mL) at 195 °C for 12 h. Workup and column chromatography eluting with ethyl
acetate-methanol (9.5:0.5) gave the title compound (30 mg, 0.155 mmol, 75%) as a colourless solid; M.p. 75-80 ºC; ν_{max} (neat)/cm^{-1} 1714, 1577, 1383, 1360, 1231, 1200, 1155, 1034, 1008, 827, 754; δ_{H} (400 MHz, CDCl_{3}) 7.03 (1 H, d, J 1.2, ArH), 6.61 (1 H, d, J 1.3, ArH), 4.16 (2 H, t, J 7.3, CH_{2}), 3.89 (3 H, s, Me), 3.12 (2 H, t, J 7.8, CH_{2}); ν_{max} (neat)/cm^{-1} 13714, 1577, 1383, 1360, 1306, 1231, 1200, 1155, 1034, 1008, 827, 754; δ_{H} (400 MHz, CDCl_{3}) 7.03 (1 H, d, J 1.2, ArH), 6.61 (1 H, d, J 1.3, ArH), 4.16 (2 H, t, J 7.3, CH_{2}), 3.89 (3 H, s, Me), 3.12 (2 H, t, J 7.8, CH_{2}); 3.22 (2 H, quin, J 7.5, CH_{2}); δ_{C} (100 MHz, CDCl_{3}) 165.6 (C), 161.9 (C), 151.1 (C), 141.2 (C), 119.2 (CH), 99.8 (CH), 52.7 (Me), 48.9 (CH_{2}), 31.8 (CH_{2}), 21.3 (CH_{2}); HRMS (ESI, [M + Na]^+) found 216.0634. [C_{10}H_{11}NO_{3} + Na]^+ requires 216.0637.

Dimethyl 2-((6-(methoxycarbonyl)-4-oxo-4,5,6,7-tetrahydro-1H-indol-2-yl)methyl)succinate

General procedure A was performed using PPA (2 g) and 76 (321 mg, 0.84 mmol) in toluene (2 mL) for 45 min. Workup and column chromatography eluting with hexanes-ethyl acetate (1:1.5) gave the title compound (216 mg, 0.65 mmol, 73%) as a brown solid; M.p. 74-77 ºC; ν_{max} (neat)/cm^{-1} 3249, 2955, 1729, 1631, 1484, 1436, 1365, 1239, 1168, 1192, 1052, 1003, 915, 891, 813, 729; δ_{H} (400 MHz, CDCl_{3}) 8.93 (2 H, br s, 2 x NH), 6.234 (1 H, s, ArH), 6.228 (1 H, s, ArH), 3.72 (12 H, s, 4 x Me), 3.670 (3 H, s, Me), 3.26-3.18 (2 H, m, 2 x CH), 2.96-2.84 (4 H, m, 2 x CH_{2}), 2.77-2.64 (6 H, m, 2 x CH of CH_{2} + 2 x CH_{2}), 2.56-2.50 (2 H, m, 2 x CH of CH_{2}); δ_{C} (100 MHz, CDCl_{3}) 191.2 (2 x C), 175.12 (C), 175.05 (C), 173.6 (2 x C), 172.3 (C), 172.2 (C), 141.2 (2 x C), 130.3 (2 x C), 120.3 (2 x C), 104.72 (CH), 104.69 (CH), 52.5 (2 x Me), 52.4 (2 x Me), 52.1 (2 x Me), 41.41 (CH), 41.38 (CH), 41.2 (2 x CH), 39.8 (2 x CH_{2}), 35.1 (2 x CH_{2}), 28.7 (2 x CH_{2}), 25.4 (2 x CH_{2}); HRMS (ESI, [M + Na]^+) found 374.1210. [C_{17}H_{21}NO_{7} + Na]^+ requires 374.1216.
Dimethyl 2-((4-hydroxy-6-(methoxycarbonyl)-1H-indol-2-yl)methyl)succinate 96

![Chemical Structure of 96]

General procedure C was performed using 94 (156 mg, 0.44 mmol) and Pd/C (237 mg) in phenylether (2.7 mL) and mesitylene (0.3 mL) for 240 °C for 18 h. Workup and column chromatography eluting with hexanes-ethyl acetate (1:1) gave the title compound (100 mg, 0.29 mmol, 66%) as a yellow oil; ν max (neat)/cm⁻¹ 3351, 2953, 1710, 1686, 1591, 1546, 1435, 1368, 1279, 1222, 1166, 1092, 1070, 1001, 867, 795, 769; δH (400 MHz, CDCl₃) 8.80 (1 H, brs, NH), 7.71 (1 H, d, J 1.0, ArH), 7.20 (1 H, d, J 1.3, ArH), 6.38 (1 H, d, J 1.1, ArH), 5.43 (1 H, br s, OH), 3.91 (3 H, s, Me), 3.73 (3 H, s, Me), 3.67 (3 H, s, Me), 3.25-3.16 (2 H, m, CH₂ of CH₂), 3.13-3.07 (1 H, m, CH of CH₂), 2.76-2.71 (1 H, m, CH of CH₂), 2.58-2.53 (1 H, m, CH of CH₂); δC (100 MHz, CDCl₃) 174.9 (C), 172.3 (C), 168.2 (C), 148.1 (C), 137.8 (C), 137.2 (C), 124.3 (C), 122.0 (C), 106.9 (CH), 105.1 (CH), 98.9 (CH), 52.6 (Me), 52.2 (2 x Me), 41.5 (CH), 35.2 (CH₂), 29.5 (CH₂); HRMS (ESI, [M + Na]⁺) found 372.1054. [C₁₇H₁₉NO₇ + Na]⁺ requires 372.0962.

Dimethyl 5-hydroxy-4-oxo-2,3,4,9-tetrahydro-1H-carbazole-2,7-dicarboxylate 97

![Chemical Structure of 97]

General procedure A was performed using PPA (2 g) and 96 (40 mg, 0.12 mmol) in toluene (1 mL) for 45 min. Workup and column chromatography eluting with hexanes-ethyl acetate
(1:1.5) gave the title compound (20 mg, 0.06 mmol, 55%) as a brown solid; M.p. 255-259 °C; \( \nu_{\text{max}} \) (neat)/cm\(^{-1} \) 3287, 2953, 1706, 1589, 1476, 1436, 1404, 1344, 1315, 1226, 1199, 1147, 1086, 1062, 994, 896, 875, 788, 768, 690; \( \delta_{\text{H}} \) (400 MHz, DMSO-d\(_6\)) 12.49 (1 H, br s, NH), 10.76 (1 H, s, OH), 7.52 (1 H, d, \( J \) 1.4, ArH) 7.03 (1 H, d, \( J \) 1.4, ArH), 3.84 (3 H, s, Me), 3.65 (3 H, s, Me), 3.56-3.49 (1 H, m, CH), 3.35-3.22 (2 H, m, CH\(_2\)), 2.84-2.83 (2 H, m, CH\(_2\)); \( \delta_{\text{C}} \) (100 MHz, DMSO-d\(_6\)) 193.9 (C), 172.9 (C), 166.5 (C), 153.3 (C), 150.2 (C), 137.3 (C), 126.7 (C), 117.5 (C), 112.0 (C), 106.6 (CH), 105.2 (CH), 52.1 (Me), 52.0 (Me), 39.9 (CH), 37.9 (CH\(_2\)), 24.8 (CH\(_2\)); HRMS (ESI, [M + Na]\(^+\)) found 340.0792. \([\text{C}_{16}\text{H}_{15}\text{NO}_6 + \text{Na}]^+\) requires 340.0797.

Dimethyl 4,5-dihydroxycarbazole-2,7-dicarboxylate 98 and dimethyl 4-hydroxy-5-methoxycarbazole-2,7-dicarboxylate 99

![Chemical Structure](image)

General procedure C was performed using 97 (100 mg, 0.31 mmol) and Pd/C (170 mg) in phenylether (2.7 mL) and mesitylene (0.3 mL) for 240 °C for 18 h. Workup (filtration through silica gel instead of Celite\(^{\text{®}}\)) and column chromatography eluting with hexanes-ethyl acetate (1:1) gave the title compounds:

14 (25 mg, 0.08 mmol, 25%) as a brown solid; M.p. 325-330 °C; \( \nu_{\text{max}} \) (neat)/cm\(^{-1} \) 3313, 1692, 1590, 1518, 1456, 1436, 1427, 1360, 1334, 1313, 1226, 1190, 1089, 1074, 1010, 992, 912, 861, 848, 796, 759, 722; \( \delta_{\text{H}} \) (400 MHz, DMSO-d\(_6\)) 11.82 (1 H, s, NH), 11.19 (2 H, s, 2 x OH), 7.64 (2 H, d, \( J \) 1.37, 2 x ArH), 7.18 (2 H, d, \( J \) 1.34, 2 x ArH), 3.88 (6 H, s, 2 x Me); \( \delta_{\text{C}} \) (100 MHz, DMSO-d\(_6\)) 166.5 (2 x C), 150.3 (2 x C), 141.2 (2 x C), 128.7 (2 x C), 113.3 (2 x C), 208
104.6 (4 x CH), 52.2 (2 x Me); HRMS (ESI, [M + Na]+) found 338.0635. [C_{16}H_{13}NO_{6} + Na]^+ requires 338.0641.

15 (11 mg, 0.03 mmol, 10%) brown solid; M.p. 245-250 °C; ν_{max}(neat)/cm^{-1} 3302, 1695, 1637, 1571, 1518, 1498, 1432, 1332, 1315, 1256, 1211, 1116, 1090, 1083, 1046, 1012, 964, 862, 790, 759, 729; δ_{H}(400 MHz, DMSO-d_6) 11.98 (1 H, s, NH), 9.31 (1 H, s, OH), 7.82 (1 H, d, J 1.0, ArH), 7.64 (1 H, d, J 1.2, ArH), 7.33 (1 H, s, ArH), 7.12 (1 H, d, J 1.0, ArH), 4.19 (3 H, s, Me), 3.91 (3 H, s, Me), 3.88 (3 H, s, Me); δ_{C}(100 MHz, DMSO-d_6) 166.5 (C), 166.3 (C), 152.1 (C), 151.1 (C), 141.4 (C), 140.8 (C), 129.1 (C), 128.4 (C), 113.4 (C), 112.6 (C), 107.2 (CH), 105.0 (CH), 104.2 (CH), 100.7 (CH), 56.8 (Me), 52.3 (Me), 52.2 (Me); HRMS (ESI, [M + Na]+) found 352.0792. [C_{17}H_{15}NO_{6} + Na]^+ requires 352.0797.

Dimethyl 2-((7-(methoxycarbonyl)-5-oxo-5,6,7,8-tetrahydroindolizin-3-yl)methyl)succinate 106

![Structure of 106]

General procedure B was performed using 76 (420 mg, 1.09 mmol) and potassium carbonate (757 mg, 5.47 mmol) in DMF (5 mL) for 12 h. Workup and column chromatography eluting dichloromethane-ethyl acetate (15:1) gave the title compound (222 mg, 0.63 mmol, 58%) as a yellow oil; ν_{max}(neat)/cm^{-1} 1720, 1436, 1388, 1350, 1305, 1252, 1195, 1167, 1075, 1049, 1003, 920, 889, 835, 789; δ_{H}(400 MHz, CDCl_{3}) 5.93 (1 H, s, ArH), 5.92 (1 H, s, ArH), 5.88-5.87 (2 H, m, 2 x ArH), 3.72 (6 H, s, 2 x Me), 3.65 (3 H, s, Me), 3.64 (3 H, s, Me), 3.62 (3 H, s, Me), 3.61 (3 H, s, Me), 3.34-2.84 (16 H, m, 4 x CH + 6 x CH_{2}), 2.73-2.65 (2 H, m, 2 x CH of CH_{2}), 2.53-2.47 (2 H, m, 2 x CH of CH_{2}); δ_{C}(100 MHz, CDCl_{3}) 175.14 (C), 175.11 (C), 172.8 (2 x...
C), 172.3 (2 x C), 168.1 (C), 168.0 (C), 131.0 (4 x C), 113.93 (CH), 113.88 (CH), 108.6 (2 x CH), 52.5 (2 x Me), 52.0 (2 x Me), 51.8 (2 x Me), 40.8 (2 x CH), 38.9 (2 x CH), 36.7 (2 x CH), 35.7 (CH2), 35.6 (CH2), 31.6 (CH2), 31.5 (CH2), 26.6 (2 x CH2); HRMS (ESI, [M + Na]+) found 374.1210. [C17H21NO7 + Na]+ requires 374.1216.

Dimethyl 2-((7-(methoxycarbonyl)-5-oxo-1,2,3,5-tetrahydroindolizin-3-yl)methyl)succinate

![Image of the compound](image.png)

General procedure C was performed using 106 (32 mg, 0.09 mmol) and Pd/C (32 mg) in n-octanol (2 mL) at 195 ºC for 12 h. Workup and column chromatography eluting with hexanes-ethyl acetate (1:1) gave the title compound (25 mg, 0.07 mmol, 78%) as a yellow oil; νmax (neat)/cm⁻¹ 2954, 1727, 1663, 1587, 1540, 1436, 1368, 1251, 1163, 1073, 988, 878, 839, 722; δH (400 MHz, CDCl3) 6.98 (1 H, s, ArH), 6.97 (1 H, s, ArH), 6.56 (2 H, s, ArH), 4.84-4.78 (1 H, m, CH), 4.77-4.72 (1 H, m, CH), 3.88 (6 H, s, 2 x Me), 3.74 (3 H, s, Me), 3.69 (3 H, s, Me), 3.68 (3 H, s, Me), 3.67 (3 H, s, Me), 3.25-3.13 (2 H, m, CH2), 3.05-2.94, (4 H, m, 2 x CH + CH2), 2.89-2.78 (2 H, m, CH2), 2.70-2.57 (2 H, m, CH2), 2.43-2.34 (1 H, m, CH of CH2), 2.32-2.22 (2 H, m, CH2), 2.11-2.02 (3 H, m, CH2 + CH of CH2), 1.99-1.92 (1 H, m, CH of CH2), 1.72-1.65 (1 H, m, CH of CH2); δC (100 MHz, CDCl3) 174.6 (C), 174.2 (C) 172.29 (C), 172.25 (C), 165.6 (2 x C), 161.8 (C) 161.7 (C), 150.7 (2 x C), 141.4 (C), 141.3 (C), 120.0 (CH), 119.9 (CH), 99.93 (CH), 99.91 (CH), 59.8 (CH), 59.2 (CH), 52.9 (2 x Me), 52.35 (Me), 52.33 (Me), 52.1 (Me), 52.0 (Me), 39.1 (CH), 38.9 (CH), 36.0 (CH2), 35.9 (CH2), 33.9 (CH2), 33.8 (CH2), 30.14 (CH2), 30.12 (CH2), 26.9 (CH2), 26.8 (CH2); HRMS (ESI, [M + Na]+) found 374.1210. [C17H21NO7 + Na]+ requires 374.1216.
Dimethyl 1,6-dioxo-1,2,3,4,6,7,8,9-octahydropyrido[1,2-\textit{a}]indole-3,8-dicarboxylate 108

![108]

General procedure A was performed using PPA (2 g) and 106 (170 mg, 0.48 mmol) in toluene (2 mL) for 3 h. Workup and column chromatography eluting with hexanes-ethyl acetate (1:1) gave the \textit{title compound} (82 mg, 0.26 mmol, 53%) as a colourless solid; M.p. 82-87 °C; $\nu_{\text{max}}$ (neat)/cm$^{-1}$ 1725, 1657, 1574, 1435, 1373, 1356, 1274, 1246, 1177, 1134, 1069, 1029, 1003, 916, 756; $\delta_H$ (400 MHz, CDCl$_3$) 6.34 (2 H, t, $J$ 1.3, 2 x ArH), 3.735 (3 H, s, Me), 3.730 (3 H, s, Me), 3.726 (3 H, s, Me), 3.721 (3 H, s, Me), 3.70-3.68 (1 H, m, CH of CH$_2$), 3.65-3.63 (1 H, m, CH of CH$_2$), 3.44-3.37 (2 H, m, CH$_2$), 3.28-3.10 (6 H, m, 4 x CH + 2 x CH$_2$), 3.08-3.01 (2 H, m, 2 x CH of CH$_2$), 2.98-2.96 (4 H, m, 2 x CH$_2$), 2.75 (2 H, s, CH$_2$), 2.73 (2 H, s, CH$_2$); $\delta_C$ (100 MHz, CDCl$_3$) 192.12 (C), 192.09 (C), 173.5 (C), 173.4 (C), 172.5 (C), 172.4 (C), 168.3 (2 x C), 141.8 (2 x C), 131.8 (C), 131.7 (C), 123.5 (2 x C), 105.4 (2 x CH), 52.7 (2 x Me), 52.4 (2 x Me), 40.9 (CH), 40.7 (CH), 39.74 (CH$_2$), 39.69 (CH$_2$), 38.4 (CH), 38.3 (CH), 36.54 (CH$_2$), 36.51 (CH$_2$), 27.4 (2 x CH$_2$), 26.3 (CH$_2$), 26.2 (CH$_2$); HRMS (ESI, [M + Na]$^+$) found 342.0948. [C$_{16}$H$_{17}$NO$_6$ + Na]$^+$ requires 342.0998.
Dimethyl 1-hydroxy-6-oxo-6,7,8,9-tetrahydropyrido[1,2-α]indole-3,8-dicarboxylate 109

General procedure B was performed using 96 (20 mg, 0.057 mmol) and potassium carbonate (24 mg, 0.17 mmol) in DMF (1 mL) for 2 h. Workup and column chromatography eluting with hexanes–ethyl acetate (1.5:1) gave the *title compound* (12 mg, 0.038 mmol, 67%) as a colourless solid; M.p. 225-230 °C; ν<sub>max</sub> (neat)/cm<sup>-1</sup> 1731 1695, 1578, 1432, 1326, 1313, 1233, 1198, 1056, 1024, 996, 886, 824, 771; δ<sub>H</sub> (400 MHz, DMSO-d<sub>6</sub>) 10.12 (1 H, br s, OH), 8.43 (1 H, d, J 1.1, ArH), 7.27 (1 H, d, J 1.1, ArH), 6.60 (1 H, s, ArH), 3.85 (3 H, s, Me), 3.62 (3 H, s, Me), 3.46-3.35 (2 H, m, CH + CH of CH<sub>2</sub>), 3.20-3.14 (1 H, m, CH of CH<sub>2</sub>), 3.01-3.00 (2 H, m, CH<sub>2</sub>); δ<sub>C</sub> (100 MHz, DMSO-d<sub>6</sub>) 172.8 (C), 168.0 (C), 166.7 (C), 149.5 (C), 137.9 (C), 135.1 (C), 126.0 (C), 122.7 (C), 109.3 (CH), 108.8 (CH), 102.2 (CH), 52.1 (Me), 52.0 (Me), 37.4 (CH), 35.5 (CH<sub>2</sub>), 25.7 (CH<sub>2</sub>); HRMS (ESI, [M + Na]<sup>+</sup>) found 340.0792. [C<sub>16</sub>H<sub>15</sub>NO<sub>6</sub> + Na]<sup>+</sup> requires 340.0797.
References

Chapter 1 references


Chapter 2 references


Chapter 3 references


30. Musser, S. M.; Stowell, M. H.; Lee, H. K.; Rumbley, J. N.; Chan, S. I., Uncompetitive substrate inhibition and noncompetitive inhibition by 5-n-undecyl-6-hydroxy-4, 7-dioxobenzothiazole (UHDBT) and 2-n-nonyl-4-hydroxyquinoline-N-oxide (NQNO) is


Chapter 4 references


21. PACT premier™


23. The Morpheus® Additive Screen


Chapter 5 references


241


Chapter 6 references


5. JCSG-plus™


7. PACT premier™

8. The Morpheus® Additive Screen


248
Targeting isocitrate lyase for the treatment of latent tuberculosis

Ram Prasad Bhusal¹, Ghader Bashiri², Brooke X.C. Kwai¹, Jonathan Sperry¹ and Ivanhoe K.H. Leung¹

¹School of Chemical Sciences, The University of Auckland, Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand
²School of Biological Sciences, The University of Auckland, Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand

Tuberculosis (TB) is an infectious disease caused by Mycobacterium tuberculosis that can remain dormant for many years before becoming active. One way to control and eliminate TB is the identification and treatment of latent TB, preventing infected individuals from developing active TB and thus eliminating the subsequent spread of the disease. Isocitrate lyase (ICL) is involved in the mycobacterial glyoxylate and methylisocitrate cycles. ICL is important for the growth and survival of M. tuberculosis during latent infection. ICL is not present in humans and is therefore a potential therapeutic target for the development of anti-TB agents. Here, we explore the evidence linking ICL to persistent survival of M. tuberculosis. The structure, mechanism and inhibition of the enzyme is also discussed.

Introduction

Tuberculosis (TB) is an infectious disease caused by Mycobacterium tuberculosis [1,2]. M. tuberculosis can live inside the human body for years without causing disease – resulting in a syndrome that is known as latent TB. TB has a latency period that is longer than any other infectious disease [3–5]. It is estimated that at least a quarter of the world’s population is infected by the bacteria [6], of which around 5–15% will develop active TB in their lifetime, a probability that increases dramatically if the infected individual becomes immunocompromised.

The World Health Organisation End TB Strategy aims to reduce the mortality rate by 90% and the incidence rate by 80% by 2030 [7]. Treatment of latent TB infection, especially for people from high-risk groups such as those who are infected by HIV, is a viable strategy to control the disease because M. tuberculosis can only spread from people who have developed active pulmonary TB [8–10]. Current medication regimens used to treat latent TB infection require high patient compliance, which typically involves regular (sometimes daily) intake of antimicrobial drugs for up to 9 months [11,12]. In addition, these drugs can induce severe hepatotoxicity and other unpleasant

Corresponding authors: Sperry, J. (j.sperry@auckland.ac.nz), Leung, Ivanhoe K.H. (i.leung@auckland.ac.nz)
side effects [13]. The development of more-effective and less toxic drugs to treat latent TB infection is therefore required if the goals set out by the WHO are to be met.

**ICL in *M. tuberculosis***

Isocitrate lyase (ICL) is an Mg$^{2+}$-dependent enzyme that catalyses the reversible lysis of a C–C bond of $\\beta$-isocitrate to form glyoxylate and succinate (Fig. 1a) [14]. ICL is present in bacteria (including mycobacteria), fungi and plants, but not in humans or animals [15]. In *M. tuberculosis* there are two known isoforms of ICL: ICL1 and ICL2, which are encoded by the genes *icl1* and *aceA* (also known as *icl2*), respectively [16,17]. Exceptions can be found in some mycobacterial species including *M. tuberculosis* H37Rv, one of the most studied laboratory strains, in which the gene *aceA* is split into two open reading frames: *aceAa* and *aceAb* [17]. It is not known whether these two genes (*aceAa* and *aceAb*) encode stable proteins and/or whether the proteins that these genes encode possess ICL activities. Sequence analyses indicate that, although ICL1 and ICL2 share 27% sequence identity, ICL1 is a prokaryotic-like ICL isoform whereas ICL2 is a eukaryotic-like ICL isoform [16,17]. To date, the most studied isoform of *M. tuberculosis* ICLs is ICL1, because recombinant ICL2 was found to be unstable in vitro [14]. ICL1 has a stronger affinity to the substrate $\\beta$-isocitrate than ICL2 (as reflected by the Michaelis constant $K_M$), and is more active than ICL2 in vitro [18].

ICL was first proposed to play a part in fatty acid metabolism of *M. tuberculosis*. Genetic studies showed that the expression level of *icl1* and *aceA* increased when mycobacteria were grown in media containing palmitate or acetate (i.e., metabolites of fatty acids) [14]. *M. tuberculosis* mutants lacking the *icl1* and *aceA* genes were

---

**FIGURE 1**

(a) Isocitrate lyase (ICL) catalyses the reversible conversion of isocitrate or methylisocitrate to succinate and glyoxylate or pyruvate. (b) Consensus mechanism for ICL catalysis. An unidentified residue acts as a general base to deprotonate the hydroxyl group of isocitrate. In ICL1 this residue was proposed to be Arg228, Tyr89 or His180. The catalytic cysteine (Cys191) serves as a general acid to aid the formation of the succinate co-product. The histidine residue (His193) interacts with the catalytic cysteine to aid the formation of the aci-carboxylate intermediate.

---

www.drugdiscoverytoday.com 1009
found unable to grow and survive in media with fatty acids as the sole carbon source and a carbon mixture that contained fatty acids. However, the growth of the same mutant was not affected in media with carbohydrates as the carbon source [19]. Interestingly, complementation of the same mutant with plasmids carrying icl1 or aceA genes restored growth on fatty acids [19]. These experiments suggest that ICL1 and ICL2 are essential for fatty acid metabolism of M. tuberculosis during the actively replicating state, although there might be some functional redundancy among these two enzymes in their fatty acid metabolism roles.

ICLs also have a functional role in the survival of M. tuberculosis during nonreplicating persistent or latent infection. The gene that encodes ICL was found to be upregulated when M. tuberculosis was grown under hypoxic conditions [20] and in growth-arrested bacilli in mice (to mimic persistent infection) [21]. ICL was also found to be essential for the survival of nonreplicating M. tuberculosis in nutrient-deprived (but oxygen-rich) media, as demonstrated by experiments using an icl1 deletion mutant of M. tuberculosis H37Rv, which also lacks the aceA gene [22]. These experiments showed that ICL activity is essential for the survival of M. tuberculosis in the nonreplicating dormant phase.

Unlike the functional redundancy of ICL1 and ICL2 in fatty acid metabolism during the actively replicating state of M. tuberculosis [19], the roles of ICL1 and ICL2 in the survival of M. tuberculosis in the nonreplicating dormant phase are not well-understood. An early study using an icl1 deletion mutant of M. tuberculosis was found unable to survive long-term within macrophages or cause virulence in mice, although the role of aceA was not addressed in that study [23]. A follow-up study by the same authors found that ICL1 and ICL2 were required for the survival of M. tuberculosis in macrophages and mice by using a mutant with icl1 and aceA deleted [19]. The results were further supported by studies supplementing wild-type M. tuberculosis with an inhibitor that blocks ICL1 and ICL2: 3-nitropropionate (see below), in which the bacteria showed a similar growth pattern in fatty acids to that of the icl1 and aceA double-deletion mutant [19]. In contrast to the earlier findings [23], single deletion of icl1 or aceA showed little effect on bacterial growth in macrophages and in mice. It is likely that, during the persistent phase of infection, M. tuberculosis exists in a mixture of different metabolic states and ICL1 and ICL2 are important for growth and survival. Further studies, for example using chemical probes that selectively inhibit ICL1 and ICL2, would be an informative strategy to elucidate the roles of ICL1 and ICL2 in latent M. tuberculosis at different metabolic states during persistent infection.

**ICL and the glyoxylate cycle**

One of the most well-known roles of ICL is its involvement as the first enzyme in the glyoxylate cycle, which was first identified from cell-free extracts of Pseudomonas aeruginosa in 1953 [24]. The glyoxylate cycle is an alternative pathway to the tricarboxylic acid (TCA) cycle (Fig. 2) [25,26]. The early steps of the glyoxylate cycle resemble those in the TCA cycle, in which acetyl-coenzyme A (CoA) is converted into α-isocitrate via citrate and cis-aconitate. The major point of difference between these two cycles is the role of α-isocitrate [27]. In the TCA cycle, α-isocitrate is first converted into 2-oxoglutarate, and subsequently into succinyl-CoA. In the glyoxylate cycle, the two decarboxylation steps in the TCA cycle are bypassed. ICL competes with isocitrinate dehydrogenase (an enzyme of the TCA cycle) for the substrate α-isocitrate [27], and catalyses the conversion of α-isocitrate into succinate and glyoxylate. The second enzyme in the glyoxylate cycle, malate synthase (encoded by the gene glcB), then catalyses the formation of malate from glyoxylate and acetyl-CoA. Other TCA cycle intermediates, including oxaloacetate, are replenished by anaplerotic reactions from succinate and malate, products of the glyoxylate cycle.

Unlike other pathogens that tend to consume one carbon source before adapting to the next source, M. tuberculosis can catabolise multiple carbon sources simultaneously [28]. These carbon sources include fatty acids, which are abundant for actively replicating and latent M. tuberculosis inside macrophages [29–33]. Fatty acids and lipids may even be the preferred carbon sources of M. tuberculosis during infection and dormancy [31–33]. When fatty acids are utilised as a carbon source, β-oxidation can lead to an increase in acetyl-CoA. The glyoxylate cycle thus enables assimilation of these extra acetyl-CoAs as carbon substrates into the intermediates of the bacteria’s central carbon metabolism. At the same time, carbon loss in the form of carbon dioxide is also prevented. Because M. tuberculosis utilises fatty acids as a carbon source during the actively replicating and nonreplicating latent phases [28], the role that ICL has in the glyoxylate cycle can be attributed to the functional essentiality of the enzyme to the growth and survival of the bacteria. The glyoxylate cycle can be particularly important for latent M. tuberculosis that lays dormant within macrophages. Because the conditions inside macrophages are relatively hypoxic and lacking in external nutrients [29], the glyoxylate cycle can enable the bacteria to utilise these carbons (which otherwise might be lost as carbon dioxide) for carbohydrate synthesis (gluconeogenesis). This upregulation of the glyoxylate cycle (regarding ICL and malate synthase) during persistent infection inside macrophages is well documented for many other bacteria [34]. However, in M. tuberculosis, although the expression of icl1 was induced, the level of glcB was not significantly altered in the active state when grown in media containing fatty acids as the carbon source [14]. The level of glcB was even found to be down-regulated in the dormant state [21]. This lack of coordination in the regulation of ICL and malate synthase in the glyoxylate cycle is unique to M. tuberculosis and has not been found in other bacteria. There are proposals that succinate and glyoxylate formed as a result of ICL catalysis could have other roles rather than their canonical roles in fatty acid metabolism and gluconeogenesis. There might exist a glyoxylate-to-glycine shunt, the proposed main function of which is to replenish the pool of NAD to support the minimal hypoxic metabolism in the bacteria’s latent state [35,36]. However, such a pathway has not been fully characterised. It has also been shown that, under hypoxic conditions, succinate, another product of ICL catalysis, might be the preferred respiratory electron donor to sustain succinate dehydrogenase activity and facilitate oxidative synthesis of ATP [37,38].

**ICL and the methylcitrate cycle**

During infection, the pools of carbon that M. tuberculosis can utilise include fatty acids from the host as well as the bacteria’s internal lipid reserves [39]. Although natural animal fatty acids are composed of an even number of carbons, bacteria including mycobacteria possess the ability to synthesise odd-chain fatty acids by the methylcitrate cycle (e.g., D-isocitrate) [40]. The methylcitrate cycle is initiated by the enzyme ICL1 in bacteria (whereas ICL2 is not involved in this pathway) [40]. D-isocitrate is first converted into 2-oxoglutarate and subsequently into succinyl-CoA. In the glyoxylate cycle, the two decarboxylation steps in the TCA cycle
However, \( \beta \)-oxidation of odd-chain fatty acids is potentially harmful to the bacteria, because the process generates propionyl CoA and propionate [41–43], both of which are toxic [44–46]. In addition to fatty acids, mycobacteria can also utilise cholesterol as a carbon source [47,48], the metabolism of which also yields propionyl CoA as a by-product [41–43].

The methylcitrate cycle is a well-studied pathway that bacteria employ to transform propionate and propionyl CoA into less toxic forms (Fig. 2) [44–46]. Three enzymes, including methylcitrate synthase, methylcitrate dehydratase and methylisocitrate lyase are involved in the bacterial methylcitrate cycle. However, bioinformatics studies infer that the genome of \( M. \) \( \text{tuberculosis} \) encodes only the homologues of methylcitrate synthase and methylcitrate dehydratase, but not methylisocitrate lyase, even though \( M. \) \( \text{tuberculosis} \) can survive in a concentrated propionate environment [29].
To identify the enzyme that is responsible for the lysis of methylisocitrate in *M. tuberculosis*, its growth in media containing propionate was studied. Interestingly, ICL was found to be essential for the growth of *M. tuberculosis* in propionate, because the *icl1* deletion mutant of *M. tuberculosis* cannot grow in propionate or cholesterol-containing media, despite the presence of saturated amounts of carbohydrates and fatty acids [19,46,49]. A *M. tuberculosis* mutant that lacks the *icl1* and *aceA* genes cannot grow in media containing propionate as the sole carbon source [19,44]. However, complementing the double-deletion mutant with plasmids encoding *icl1* and *aceA* could restore mycobacterial growth in propionate media [19]. These experiments showed that ICLs probably play a part in the glyoxylate and methylcitrate cycles in *M. tuberculosis*.

In addition, *in vitro* experiments with purified ICL1 and ICL2 proteins showed that both enzymes could use isocitrate and methylisocitrate as substrates (Fig. 1a), although isocitrate is the preferred substrate over methylisocitrate for ICL1 and ICL2 [18]. To understand the substrate selectivity of ICL, crystallisation of the C191S variant of ICL1, in which the catalytic cysteine residue in the active site was mutated to a serine residue to slow down the catalytic activity of the enzyme, was attempted in the presence of methylisocitrate [18]. Interestingly, in the resulting crystal structures the reaction products pyruvate and succinate were observed inside the enzyme active site instead of methylisocitrate, indicating the mutant was still active during crystallisation. Nevertheless, the results showed that ICL1 could accommodate the methyl group of methylisocitrate in the hydrophobic pocket of the enzyme, further supporting the proposal that ICL can have a dual role in the survival of *M. tuberculosis* in a high fatty acid content environment that is characteristic of active and persistent infections.

**ICL and its potential role in antibiotic tolerance of *M. tuberculosis***

Recently, ICL was linked to the development of antibiotic resistance in *M. tuberculosis*. By using metabolomics and gene expression analyses it was demonstrated that ICL was activated when *M. tuberculosis* was subjected to sublethal doses of three different anti-TB drugs: rifampicin, streptomycin or isoniazid [50]. *M. tuberculosis* with an *icl1* deletion showed a 100-fold increase in sensitivity to these antibiotics [50]. Furthermore, studies that used a *M. tuberculosis* mutant complemented with a K189E variant of ICL1, the mutation of which kept the enzyme in an open conformation and hence reduced its catalytic activity (see below), showed increased sensitivity towards rifampicin and streptomycin but not to isoniazid [51]. These experiments suggest that the catalytic activity of ICL could play a part in mycobacterial defence against (at least some) antibiotics, although the exact molecular mechanism is not clear. Interestingly, when the *icl1* deletion mutant was co-incubated with antioxidants including thiourea or Tempol, the heightenened susceptibility towards the antibiotics was reduced [50]. The results suggest that ICL could play a part in the metabolic defence against antibiotic-induced oxidative stress. This discovery is significant because long-term treatments and increasing antibiotic resistance are two of the biggest challenges preventing the eradication of TB. These recent studies increase the appeal of ICL as a target in anti-TB drug discovery programmes.

**Structure and catalytic mechanism of ICL**

The exact mechanism by which ICL converts isocitrate into glyoxylate and succinate is not fully understood, but a retro Claisen-type condensation pathway has been inferred (Fig. 1b) [52]. The first step involves deprotonation of the isocitrate hydroxyl group followed by fragmentation of the isocitrate to form glyoxylate and succinate [53,54]. Mutagenesis and bioinformatics studies showed that the highly conserved KKCCH sequence motif at the enzyme active site (residues 189–193 in ICL1, also found in ICLs from other organisms) is essential for the catalytic activity of the enzyme [53,55]. Three residues are proposed to be involved in the catalytic cycle, including the cysteine and histidine residues in the conserved KKCCH sequence motif, along with an unidentified residue that is close to the substrate at the active site [53,54]. The catalytic cysteine residue (Cys191 in ICL1) serves as a general acid to aid the formation of the succinate co-product, whereas the histidine residue (His193 in ICL1) interacts with the catalytic cysteine, decreasing its pKₐ and aiding the formation of the aci-carboxylate intermediate that forms upon C(2)–C(3) bond cleavage [53]. The unidentified residue presumably acts as a general base to deprotonate the C(2) hydroxyl group of the isocitrate substrate. In ICL1 this residue is proposed to be either an arginine (Arg228), tyrosine (Tyr89) or histidine (His180) [54].

Divalent metal ions play an important part in the catalytic activity of ICL1 and ICL2 [14]. The substrate α-isocitrate binds to ICL by chelating the active site metal ion via the C(1) carboxylic group and C(2) hydroxyl group. Mg²⁺ is required for optimal catalytic activity, whereas replacement of the Mg²⁺ by Mn²⁺ results in a 60% loss of catalytic activity. Other divalent cations are not effective substitutes for Mg²⁺, and negligible activities were detected in the absence of divalent cations.

To date, only the crystal structure of ICL1 has been solved (PDB IDs: 5DQL, 1N8I, 1F8I, 1F8M and 1F61) [52]. ICL2 was found to be unstable and no structural information is available [14]. The crystal structures of ICL1 in its apo and inhibitor-bound forms were first reported in 2000 [52]. ICL1 crystallised as either a dimer or tetramer, with each monomer subunit containing 14 α-helices and 14 β-strands. The core of ICL1 contains eight α-helices and eight β-strands forming an α/β-barrel-like structure. The highly conserved KKCCH sequence motif extends from β5 of the α/β-barrel forming a loop structure (residues 185–196) and connects to a small β-sheet domain that consists of five short β-strands (β6, β7, β9, β10, β11). At the C-terminal, α12–α14 project away from the barrel and interact with the nearby subunit (Fig. 3a).

When comparing the crystal structure of apo-ICL1 and the holo forms with glyoxylate, nitropropionate (a mimic of the product succinate) or bromopyruvate (an inhibitor of ICL) conformational changes are observed (Fig. 3b) [52]. In particular, big changes are apparent around the active site loop region that includes the catalytically essential sequence motif KKCCH (residues 185–196 of ICL1), and the C-terminal end (residues 411–428 of ICL1) of the adjacent subunit. In the open conformation, the catalytic cysteine (Cys191) is relatively far away from other active site residues, thus allowing the substrate (and other solvent molecules) to access the binding site. Upon binding of the substrate, the enzyme adopts a closed conformation and access to the active site is then blocked.
by the active site loop (residues 185–196). The C-terminal end of the adjacent subunit then moves to the space that was opened up as a result of the closure of the active site loop, forming a lid. Electrostatic interactions between positively charged lysine residues at the active site loop (Lys189 and Lys190) and the negatively charged glutamate residues at the C-terminal lid (Glu423 and Glu424) are important to keep the enzyme in its catalytically active conformation [56]. Disrupting these interactions can lead to a reduction in the enzyme’s catalytic activity [56].

Interestingly, His180, one of the residues that was proposed to act as a general base during the catalytic cycle of ICL1, is also an important residue for the enzyme’s catalytic activity and oligomerisation state [56]. Mutagenesis studies showed that ICL1 H180A is a monomer that is catalytically inactive. Molecular dynamics calculations showed that loss of interactions between His180 and Tyr89 (another residue that was proposed to be important for ICL1 activity) could cause changes in the orientation of α6 and the C-terminal helices (α12–α14), thus affecting the protein’s ability to interact with other monomer subunits to form an oligomer. The loss of such interactions also affects the stability of ICL1 in its catalytically active conformation, because the active site loop is no longer being held together by the lid of the adjacent subunit in the tetramer.

It has long been proposed that bacterial and plant ICLs could be switched on and off by reversible acetylation and succinylation, respectively [57–59], although the exact molecular and structural basis of such control is not known. Recently, two post-translation modifications were found on M. tuberculosis ICL, including lysine acetylation and lysine succinylation [51,60,61]. Interestingly, one of the succinylated lysine residues (Lys189) is located in the catalytically essential conserved sequence motif (KKCGH) [51]. Although it is not possible to recombinantly produce (at least in large quantity) succinylated ICL1 for structural and mechanistic studies, the effect of lysine succinylation can be mimicked by a lysine to glutamate mutation. Structural and kinetic studies using ICL K189E showed that replacing the positively charged lysine with the negatively charged glutamate (a similar effect to lysine succinylation) could keep the enzyme in the open form, thus rendering the enzyme catalytically inactive. The regulation of the TCA and glyoxylate cycles is important for the survival of M. tuberculosis because both cycles take α-ketoglutarate as an intermediate [27]. A recent study showed that glyoxylate (a product of the
glyoxylate cycle) can positively control the activity of *M. tuberculosis* isocitrate dehydrogenase (an enzyme in the TCA cycle) [62]. The presence of another potential feedback mechanism involving reversible succinylation of the active site lysine in ICL1 to activate and deactivate the glyoxylate cycle is not surprising given the tight control that is required to regulate these two closely related cycles. Further studies into the interplay and control mechanisms between the TCA and glyoxylate cycles could enable the development of new anti-TB drugs that can selectively activate and deactivate ICL and isocitrate dehydrogenase at different stages of infection.

**Inhibitors of ICL**

Given the central role ICL has in the glyoxylate and methylisocitrate cycles, ICL is a current inhibition target for antimicrobial applications including (but not limited to) latent TB. However, despite considerable efforts by academia and industry, no compounds have progressed through to the clinical trial stage. There are three major challenges in targeting ICLs: (i) the polar nature of the ICL binding pocket; (ii) the small size of the natural substrates; and (iii) the need to target ICL1 and ICL2.

The polar nature of the ICL binding pocket favours small, polar molecules. It is therefore not surprising that early inhibitor discovery work that involved plant, fungal and bacterial ICLs has resulted in compounds that are analogues of the substrate isocitrate, or succinate and glyoxylate. These included itaconic acid [63], 3-nitropropionate [64] and 3-bromopropionate. Itaconic acid and 3-nitropropionate are noncovalent inhibitors of ICL [63,64] but 3-bromopropionate was found to covalently bind to ICL via the catalytic cysteine residue at the active site [65]. Although these compounds are relatively potent inhibitors against *M. tuberculosis* ICL, they are nonselective and possess high toxicity, rendering them unsuitable as potential drug candidates [66–70]. For example, whereas 3-nitropropionate can be a potent inhibitor of ICL1 and ICL2, it also inhibits (human) succinate dehydrogenase [71], an important enzyme in the TCA cycle. Interestingly, recent studies have suggested that animals, including mice and humans, could produce itaconic acid as a natural antibiotic in macrophages during inflammation [72,73]. These findings raise the possibility of using itaconic acid (or analogues) as a starting point for ICL inhibitor discovery programmes, although the design of structural analogues that can balance the polar nature of the molecule and permeability of the hydrophobic TB cell wall will be a challenge.

A potential method to bypass the polar active site is to target the oligomerisation interface of ICL1, because monomeric ICL1 was shown to be catalytically inactive. Liu *et al.* used molecular docking to help design peptide-based ICL1 inhibitors [74]. Although the best reported peptidyl inhibitor was only moderately potent in *vitro* (IC₅₀ = 126 μM), it demonstrates that the oligomerisation interface could be a potential inhibition target. Future peptide-based inhibitor work would focus on improving the potency, *in vivo* stability and cell permeability of the peptides.

Recent efforts to discover new ICL inhibitors have focused on high-throughput screening (HTS), with at least five studies published in the past decade [75–79]. The relatively small size of the ICL1 binding pocket is challenging for HTS because the number of compounds that can fit the binding pocket is limited, which also restricts the scope of any SAR studies. The two most notable HTS studies involved GlaxoSmithKline (GSK) and the Global Alliance for TB Drug Development (GATB), who screened 900 000 compounds, and the Tuberculosis Antimicrobial Acquisition and Coordinating Facility (TAACF), who screened just under 101 000 compounds. The TAACF screen was moderately successful [77]. For instance, 4-(2-pyridyl)-2-thiazolamine was identified as a potentially new scaffold for TB inhibitors. In addition, heterocyclic benzopyranones were also identified as ICL inhibitors. Because there are few previous reports of compounds containing these scaffolds as anti-TB drug candidates they constitute interesting starting points for future medicinal chemistry programmes.

In addition to these industrial efforts, several academic research groups have worked on the development of ICL1 inhibitors. Vinová and co-workers have focused on synthesising derivatives of salicylanilides and benzanilide [80–84]. The group of Sirram synthesised derivatives of 3-nitropropionamides, 5-nitro-2-furoic acid hydrazones, 5-nitro-2,6-dioxohexahydro-4-pyrimidinecarboxamides, phthalazin-4-ylacetamides and isatyl semicarbazone [85–90]. However, the lack of complete SAR data of these derivatives has, at least to date, limited the potential of these compounds to synthetic medicinal chemistry projects. A comprehensive list of ICL inhibitors that have been developed to date can be found in two reviews [91,92].

**Concluding remarks**

ICL is an attractive inhibition target for the treatment of latent TB because it is vital for bacterial survival and, in addition, humans do not possess this enzyme. Existing ICL inhibitors are mimics of the native substrate (or products) that keeps the enzyme in the closed confirmation, as observed in the inhibitor-bound structures of ICL1. The inherent toxicity of these inhibitors is probably caused by their binding to other human enzymes, for example those that take isocitrate or succinate as substrates. Designing inhibitors that are selective for ICLs presents a significant challenge, the answer to which might lie in the mechanism that regulates the *M. tuberculosis* glyoxylate and TCA cycles through post-translational modification of the active site lysine residue, which could enable the design of compounds that selectively target the catalytically inactive open form of the enzyme. Although ICL1 and ICL2 are both required for the survival of *M. tuberculosis* during persistent infection, further research into understanding the roles of the two isoforms at different stages of infection could enable the development of more-efficient treatments. One way to approach this is through the development of isoform-specific ICL1 and ICL2 inhibitors as chemical probes, a tactic that is generally milder than genetic manipulations such as gene knockouts. However, an obstacle is the lack of structural information on ICL2, because recombinant ICL2 was found to be unstable [14]. The design of alternative ICL2 constructs could enhance the stability of the recombinant enzyme for structural and mechanistic studies.

Overall, it is evident that ICL plays an important part in the survival of *M. tuberculosis* during persistent infection, including the metabolism of odd- and even-chain fatty acids and cholesterol as carbon sources to produce energy. ICL is also linked to the development of antibiotic resistance of the bacteria, although the mechanism of this process is not fully understood. A detailed understanding of the structure and catalytic mechanism of ICL1 has inspired the development of some ICL1 inhibitors, but further
Drug Discovery Today  Volume 22, Number 7  July 2017

complementary studies are required (especially those that focus on
ICL2) before ICL inhibitors can be seriously considered as potential
therapeutics against latent TB.

REVIEWS

Acknowledgement
We thank the University of Auckland for a Doctoral Scholarship (R.
P.B) and the Maurice and Phyllis Paykel Trust for funding.

J. Immunol. 185, 15–22
372, 2127–2135
5 Esmail, H. et al. (2014) The ongoing challenge of latent tuberculosis. Phil. Trans. R.
Soc. B 369, 20130437
6 Houben, R.M.G.J. and Dodd, P.J. (2016) The global burden of latent tuberculosis
8 Getahun, H. et al. (2015) Management of latent Mycobacterium tuberculosis infection:
9 Sulis, G. et al. (2016) Recent developments in the diagnosis and management of
tuberculosis. NPJ Prim. Care Respir. Med 26, 16078
10 Lönnroth, K. et al. (2010) Tuberculosis control and elimination 2010–50: cure, care,
and social development. Lancet 375, 1814–1829
11 Norton, B.L. and Holland, D.P. (2012) Current management options for latent
tuberculosis: a review. Infect. Drug Resist 5, 163–173
latent TB infection. Infect. Drug Resist 3, 63–72
13 Saukkonen, J.J. et al. (2006) An official ATS statement: hepatotoxicity of
14 Höner zu Bentrup, K. et al. (1999) Characterization of activity and expression of
181, 7161–7167
15 Kondrashov, F.A. et al. (2006) Evolution of glyoxylate cycle enzymes in Metazoa:
evidence of multiple horizontal transfer events and pseudogene formation. Biol.
Direct 1, 31
the complete genome sequence. Nature 393, 537–544
tuberculosis clinical and laboratory strains. J. Bacteriol. 184, 5479–5490
18 Gould, T.A. et al. (2006) Dual role of isocitrate lyase 1 in the glyoxylate and
methylcitrate cycles in Mycobacterium tuberculosis. Mol. Microbiol. 61, 940–947
lyases 1 and 2 are jointly required for in vivo growth and virulence. Nat. Med. 11,
638–644
upregulated during anaerobic persistence by fluorescence and kanamycin resistance
selection. Tuberculosis 88, 518–525
growth arrest. Mol. Microbiol. 78, 1199–1215
tuberculosis requires respiration, ATP synthase and isocitrate lyase for maintenance
of ATP homeostasis and viability. Microbiology 156, 81–87
23 McKinney, J.D. et al. (2000) Persistence of Mycobacterium tuberculosis in
macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature
406, 735–738
24 Campbell, J.J.R. et al. (1953) A deviation from the conventional tricarboxylic acid
cycle in Pseudomonas aeruginosa. Biochim. Biophys. Acta 11, 594
acetate assimilation. Mol. Microbiol. 61, 274–276
26 Kornberg, H.L. and Krebs, H.A. (1957) Synthesis of cell constituents from C2-units
by a modified tricarboxylic acid cycle. Nature 179, 988–991
point of two metabolic cycles. The tricarboxylic acid cycle and the glyoxylate shunt.
J. Biol. Chem. 259, 9646–9654
compartmentalized co-catabolism of carbon substrates. Chem. Biol. 17, 1122–1131
29 Schnappinger, D. et al. (2003) Transcriptional adaptation of Mycobacterium
tuberculosis within macrophages: insights into the phagosomal environment. J. Exp.
Med. 198, 693–704
Microbe 8, 68–76

bacteria. Cell Microbiol. 8, 10–22
33 Forrellad, M.A. et al. (2014) Role of the Mce1 transporter in the lipid homeostasis of
Mycobacterium tuberculosis. Tuberculosis 94, 170–177
34 Michael, L. et al. (2002) Life and death in a macrophage: role of the glyoxylate cycle
in virulence. Eukaryot. Cell 1, 657–662
35 Wayne, L.G. and Lin, K.-Y. (1982) Glyoxylate metabolism and adaptation of
Mycobacterium tuberculosis to survive under anaerobic conditions. Infect. Immun. 37,
1042–1049
Mycobacterium tuberculosis. Annu. Rev. Microbiol. 55, 139–163
110, 6554–6559
38 Pecsi, I. et al. (2014) Essentiality of succinate dehydrogenase in Mycobacterium
smegmatis and its role in the generation of the membrane potential under hypoxia.
mBio 5, e01093–14
39 Lee, W. et al. (2013) Intracellular Mycobacterium tuberculosis exploits host-derived
fatty acids to limit metabolic stress. J. Biol. Chem. 288, 6788–6800

40 Rezanka,
T. and Sigler, K. (2009) Odd-numbered very-long-chain fatty acids from
the microbial, animal and plant kingdoms. Prog. Lipid Res. 48, 206–238
41 Griffin, J.E. et al. (2012) Cholesterol catabolism by Mycobacterium tuberculosis
42 Yang, X. et al. (2009) Cholesterol metabolism increases the metabolic pool of
propionate in Mycobacterium tuberculosis. Biochemistry 48, 3819–3821
43 Thomas, S.T. et al. (2011) Pathway profiling in Mycobacterium tuberculosis:
elucidation of cholesterol-derived catabolite and enzymes that catalyze its
metabolism. J. Biol. Chem. 286, 43668–43678
44 Muñoz-Elı́as, E.J. et al. (2006) Role of the methylcitrate cycle in Mycobacterium
tuberculosis metabolism, intracellular growth, and virulence. Mol. Microbiol. 60,
1109–1122
methylmalonyl pathway in Mycobacterium tuberculosis. Implications for propionate
metabolism during growth on fatty acids. J. Bacteriol. 190, 3886–3895
46 Upton, A.M. and McKinney, J.D. (2007) Role of the methylcitrate cycle in
propionate metabolism and detoxification in Mycobacterium smegmatis. Microbiology
153, 3973–3982
47 Pandey, A.K. and Sassetti, C.M. (2008) Mycobacterial persistence requires the
48 Soto-Ramirez, M.D. et al. (2017) Cholesterol plays a larger role during Mycobacterium
tuberculosis in vitro dormancy and reactivation than previously suspected.
Tuberculosis 103, 1–9
49 Eoh, H. and Rhee, K.Y. (2014) Methylcitrate cycle defines the bactericidal
essentiality of isocitrate lyase for survival of Mycobacterium tuberculosis on fatty
50 Nandakuma, M. et al. (2014) Isocitrate lyase mediates broad antibiotic tolerance in
Mycobacterium tuberculosis. Nat. Commun. 5, 4306
35, 1030–1041
52 Sharma, V. et al. (2000) Structure of isocitrate lyase, a persistence factor of
53 Moynihan, M.M. and Murkin, A.S. (2014) Cysteine is the general base that serves in
54 Jongkon, N. et al. (2015) Probing the catalytic mechanism involved in the isocitrate
lyase superfamily: hybrid quantum mechanical/molecular mechanical calculations
role in catalysis by isocitrate lyase from Escherichia coli. Curr. Microbiol. 35,
267–269
56 Shukla, H. et al. (2015) Insight into the structural flexibility and function of
Mycobacterium tuberculosis isocitrate lyase. Biochimie 110, 73–80

www.drugdiscoverytoday.com

1015

Reviews  KEYNOTE REVIEW

References


1016 www.drugdiscoverytoday.com


68 Greene, J.G. et al. (1998) 3-Nitropropionic acid exacerbates N-methyl-D-aspartate toxicity in striatal culture by multiple mechanisms. *Neuroscience* 84, 503–510


79 Liu, Y. et al. (2016) Identification of a novel inhibitor of isocitrate lyase as a potent antitubercular agent against both active and non-replicating *Mycobacterium tuberculosis*. *Tuberculosis* 97, 38–46


Development of NMR and thermal shift assays for the evaluation of *Mycobacterium tuberculosis* isocitrate lyase inhibitors†

Ram Prasad Bhusal, ‡a Krunal Patel, ‡a Brooke X. C. Kwai, ‡a Anne Swartjes, ‡a,‡ab Ghader Bashiri, ‡cd Jóhannes Reynisson, ‡a Jonathon Sperry ‡a,* and Ivanhoe K. H. Leung ‡a,*

The enzymes isocitrate lyase (ICL) isoforms 1 and 2 are essential for *Mycobacterium tuberculosis* survival within macrophages during latent tuberculosis (TB). As such, ICLs are attractive therapeutic targets for the treatment of tuberculosis. However, there are few biophysical assays that are available for accurate kinetic and inhibition studies of ICL in vitro. Herein we report the development of a combined NMR spectroscopy and thermal shift assay to study ICL inhibitors for both screening and inhibition constant (IC50) measurement. Operating this new assay in tandem with virtual high-throughput screening has led to the discovery of several new ICL1 inhibitors.

Introduction

Tuberculosis (TB) is a high burden infectious disease that is caused by *Mycobacterium tuberculosis*.1–2 In 2015, there were over 10 million new TB cases and around 1.8 million TB-related deaths.3 TB has a long latency period; once a human is infected with *M. tuberculosis*, the bacteria may stay inactive within macrophages for many years leading to a syndrome that is known as latent TB.4–6 The environment inside macrophages is relatively hypoxic and lacking in external nutrients. In order to survive under these conditions, *M. tuberculosis* is able to simultaneously catabolise different carbon sources, including fatty acids and cholesterol that are available in relative abundance inside macrophages.7–9 The enzymes isocitrate lyase (ICL) isoforms 1 and 2 play essential roles in this metabolic adaptation.10 ICLs are key enzymes in both the *M. tuberculosis* glyoxylate and methylcitrate cycles. In the glyoxylate cycle, ICLs catalyse the conversion of the tricarboxylic acid (TCA) cycle intermediate isocitrate into glyoxylate and succinate (Fig. 1a), thus bypassing the two decarboxylation steps in the TCA cycle and preserving these carbons for gluconeogenesis.11,12 In the methylcitrate cycle, ICLs catalyse the conversion of methylisocitrate, an intermediate of the propionate degradation pathway, to pyruvate and succinate. Propionate, which is toxic to the bacteria, is generated by β-oxidation of odd chain fatty acids and cholesterol that *M. tuberculosis* may utilise as carbon sources.13

Given the pivotal roles ICLs play in the survival of *M. tuberculosis* inside macrophages, ICLs are attractive inhibition targets for the treatment of latent TB.14–16 We recently initiated a research programme aimed at identifying new inhibitors of ICLs, but it was readily apparent that few accurate biophysical assays are available to study ICL kinetics and inhibition *in vitro*. Most ICL assays rely on ultraviolet/visible (UV/vis) spectrophotometry to determine the amount of glyoxylate that is formed as a result of ICL-catalysed reactions. For example, one method uses lactate dehydrogenase (LDH) to catalyse the reduction of glyoxylate to glycolate, during which NADH (a cosubstrate of LDH) is oxidised to NAD+. However, these assays have several drawbacks. Auto-oxidation of NADH to NAD⁺ may affect the accuracy of the LDH-coupled assay.20 In addition, this method is not suitable for measuring the methylisocitrate lyase activity of ICLs because...
LDH cannot take pyruvate as substrate. For the phenylhydrazine-coupled assay, the accuracy of the assay may be compromised by the rate of the glyoxylate-phenylhydrazone complex formation, which is pH dependent and gets slower above pH 7. Phenylhydrazine is also unstable at pH 7 and above, with the breakdown products may lead to a slow increase in the UV absorption, thus affecting the accuracy of the measurements.

1H nuclear magnetic resonance (NMR) spectroscopy is an established technique for the study of enzyme kinetics that has been used to characterise different enzyme systems including (but not limited to) carbohydrate-processing enzymes, enzymes related to antibiotic resistance and oxygenases. 1H NMR spectroscopy enables the direct monitoring of reaction kinetics in real time and accurate, quantitative information can be obtained by following changes in the peak area of the resonances associated with the substrate and/or reaction product(s). In contrast, thermal shift assay is a simple and high-throughput method that can be used to study protein–ligand binding interactions by measuring the melting temperature of a protein by the use of a fluorescence dye that is sensitive to changes in hydrophobic environment. When a protein unfolds, it exposes its hydrophobic core. This enables the dye to bind to the exposed hydrophobic regions, which lead to fluorescence. Ligand binding may stabilise or destabilise the protein towards

Fig. 1 (a) Isocitrate lyase catalyses the conversion of isocitrate to glyoxylate and succinate; (b) 1H NMR spectroscopy to monitor ICL1-catalysed turnover of isocitrate into succinate; (c) corresponding plot of the isocitrate turnover data. The curve was added to aid visualisation. Sample contained 190 nM ICL1, 1 mM α-isocitrate, 5 mM MgCl₂ and 50 mM Tris/Tris-D₁₁ (pH 7.5) in 90% H₂O and 10% D₂O. The hashtag (#) indicates Tris/Tris-D₁₁ peak. The errors shown are the standard deviation from three separate measurements.
Results and discussion

ICL1 enzyme kinetics by ¹H NMR

We first tested the use of ¹H NMR spectroscopy to monitor the ICL1-catalysed turnover of isocitrate to succinate and glyoxylate. DL-Isocitrate, which is available commercially, was used as the substrate. MgCl₂ was added to the reaction mixture as it was previously shown to be important for ICL1 activity.¹⁷ ¹H spectra were recorded at ∼1.3 minute intervals. Upon addition of the enzyme, the peaks corresponding to isocitrate dropped in intensity, which was accompanied by the appearance of a new singlet peak at 2.3 ppm, corresponding to succinate (Fig. 1b). Integration of the isocitrate and succinate peaks showed that the reaction appeared to slow down when ∼50% of the isocitrate was consumed (Fig. 1c). As the isocitrate was a racemic mixture, this result infers that ICL1 has a preference for one enantiomer, which is in agreement with a previous study that showed D-isocitrate is the preferred substrate of the enzyme.³³

Divalent metals play important role in the activity of ICL1. Previous studies showed that Mg²⁺ (and to a lesser extent, Mn²⁺) are required for optimal ICL1 activity.¹⁷ ³⁴ In order to confirm the concentration of divalent magnesium that is required for optimal activity of the enzyme, the reaction was run using different concentrations of MgCl₂. Under our reaction conditions, 500 μM of MgCl₂ was required for the optimal activity (Fig. S1†). At least 500 μM of MgCl₂ was used in all subsequent kinetic and inhibition assays.

The kinetic parameters for ICL1 with DL-isocitrate were then evaluated by ¹H NMR. The Michaelis constant (Kₘ) was found to be 290 ± 10 μM and the catalytic constant (k₉cat) was determined to be 4.3 ± 0.1 s⁻¹ (Fig. S2†). These values were similar to those obtained by Gould et al. using the aforementioned LDH assay, which were 190 μM and 5.24 s⁻¹ respectively (Table 1).¹² The slight discrepancy between the two measurements is likely due to differences in the reaction conditions. Overall, this validated the accuracy of our ¹H NMR assay to study ICL1 kinetics.

<table>
<thead>
<tr>
<th></th>
<th>Kₘ [μM]</th>
<th>k₉cat [s⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>This study</td>
<td>290 ± 10</td>
<td>4.3 ± 0.1</td>
</tr>
<tr>
<td>Gould et al.¹²</td>
<td>190</td>
<td>5.24</td>
</tr>
</tbody>
</table>

We then tested the use of ¹H NMR spectroscopy to monitor the ICL1-catalysed turnover of methylisocitrate to pyruvate and succinate (Fig. S3a†). (2S,3R)-2-Methylisocitrate was used as substrate. In the presence of ICL1 and Mg²⁺, two new singlet signals at ∼2.3 ppm, one corresponded to succinate and the other corresponded to pyruvate, were found to increase in intensity over time (Fig. S3b†). This was coupled with a drop in intensity of the methylisocitrate signals. We then repeated the experiments at different methylisocitrate concentrations. Interestingly, substrate inhibition was observed when methylisocitrate concentration exceeded 1 mM (Fig. S4†). Substrate inhibition was not observed when isocitrate was used as substrate. Further investigations are required to fully understand the biological significance of these observations. Overall, our results showed that ¹H NMR spectroscopy is a versatile and informative tool to study ICL1 kinetics that allows the use of different substrates and enables kinetic information to readily be measured and quantified.

Inhibition studies of ICL1 by ¹H NMR

We then applied our new NMR-based assay to study ICL1 inhibition. Initially, we chose four known ICL inhibitors, including three so-called first generation inhibitors itaconic acid,³⁵ 3-nitropropionate³⁶ and 3-bromopyruvate.³⁷ Itaconic acid and 3-nitropropionate are noncovalent inhibitors of ICL1 whereas 3-bromopyruvate inhibits ICL1 in a covalent manner. Methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate, an inhibitor that was discovered last year by Liu et al. using high-throughput screening, was also evaluated (Table 2).³⁸

Single concentration inhibition experiments were first conducted (Fig. S5†). In agreement with previous studies,¹⁷ our ¹H NMR assay showed that 3-nitropropionate was the most potent inhibitor amongst 3-nitropropionate, 3-bromopyruvate and itaconic acid. Under our assay condition, methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate was the weakest of the four tested. We then repeated our measurements at different inhibitor concentrations in order to obtain quantitative inhibition information (IC₅₀; Table 2 and Fig. S6–S9†). The IC₅₀ values of 3-nitropropionate, 3-bromopyruvate, itaconic acid and methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate were found to be 14.7 ± 1.8 μM, 17.5 ± 1.0 μM, 29.4 ± 4.1 μM and 250 ± 7 μM respectively. The reported IC₅₀ value for methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate was 30.9 μM.³⁸ The slight discrepancy in our measured and reported
IC$_{50}$ values for methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate is likely due to the different reaction conditions and assays used in the two studies. Overall, our results show that $^1$H NMR is a useful tool to study ICL1 inhibition in vitro, enabling a rapid evaluation of inhibitor strength as well as providing quantitative information such as IC$_{50}$.

**Thermal shift assay to study ICL1-inhibitor interactions**

Although $^1$H NMR spectroscopy was found to be a useful method to study ICL1 inhibition, it is relatively low throughput and labour intensive. A high throughput assay is needed to facilitate the efficient screening and development of new ICL inhibitors. Thermal shift assays are a widely-used method to study protein–ligand interactions. The principle of a thermal shift assay is based on the premise that ligand binding can stabilise or destabilise protein to thermal denaturing, and therefore lead to a shift in the protein's melting temperature.

First, the melting temperature of ICL1 was measured. As MgCl$_2$ is important for the activity of the enzyme, a saturating concentration of 1 mM was used. The melting temperature of ICL1 in the presence of MgCl$_2$ was found to be 43.0 °C. Next, the melting temperatures of ICL1 in the presence of a saturating concentration (1 mM) of the aforementioned inhibitors and MgCl$_2$ were measured. Addition of 3-bromopyruvate or itaconic acid were found to stabilise ICL1, with positive shifts to melting temperatures of 52.5 °C and 53.3 °C respectively. Interestingly, 3-nitropropionate and methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate were found to destabilise the protein, with negative thermal shifts to 40.9 °C and 37.6 °C respectively (Fig. S10†).

A negative thermal shift upon ligand binding has been previously observed for other protein systems. A positive thermal shift may be observed if the ligand induces the protein to adapt a more stable 'closed' conformation, whilst negative thermal shift may be observed if the ligand keeps the protein in a less stable 'open' conformation. Previous structural studies by Sharma et al. showed that ICL1 may undergo a two-step conformation change upon substrate binding (Fig. S11†). Indeed, a crystal structure of ICL1 in the presence of both 3-nitropropionate and glyoxylate was found to adapt a 'closed' conformation (PDB id: 1F8I). 3-Nitropropionate is a structural analogue of succinate. We reasoned that the binding of 3-nitropropionate on its own may keep ICL1 in the open conformation in order to allow

<table>
<thead>
<tr>
<th>Inhibitor Structure</th>
<th>IC$_{50}$/μM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Nitropropionate</td>
<td>14.7 ± 1.8</td>
</tr>
<tr>
<td>3-Bromopyruvate</td>
<td>17.5 ± 1.0</td>
</tr>
<tr>
<td>Itaconic acid</td>
<td>29.4 ± 4.1</td>
</tr>
<tr>
<td>Methyl-4-(4-methoxyphenyl)-4-oxobut-2-enoate</td>
<td>250 ± 7</td>
</tr>
<tr>
<td>Compound 29</td>
<td>&gt;100</td>
</tr>
<tr>
<td>Compound 38</td>
<td>&gt;100</td>
</tr>
</tbody>
</table>
glyoxylate to bind. However, in the presence of both 3-nitropropionate and glyoxylate, the protein can then undergo conformational change to the ‘closed’ conformation, as suggested by the crystal structure, to catalyse the reverse reaction. This proposal is consistent with the mechanism suggested by Sharma et al. It should also be noted that binding of 3-nitropropionate or methyl 4-(4-methoxyphenyl)-4-oxobut-2-enoate may induce a change in the oligomeric state of ICL1, which exists as a tetramer in solution. However, based on evidence from X-ray crystallography and molecular docking, both compounds are not known to bind to the oligomerisation interface between the ICL1 monomers, or interfere with the residues that were previously identified as important for the protein’s oligomerisation.

Application of the combined $^1$H NMR and thermal shift assays with virtual high-throughput screening

Virtual high-throughput screening is a cost-effective and efficient strategy to identify chemical structures that are potentially important for binding to a target protein. Obviously, the hits identified by virtual high-throughput screening needed to be verified experimentally. In order to test the applicability of the $^1$H NMR and thermal shift assays and to identify new inhibitors of ICL1, a virtual screen was conducted. Using the crystal structure of ICL1 (PDB ID: 1F8I, resolution 2.25 Å), a screen was performed with the InterBioScreen Ltd natural product collection. 9050 compounds were screened and four different scoring functions, including GoldScore (GS), ChemScore (CS), Piecewise Linear Potential (ChemPLP) and Astex Statistical Potential (ASP), were used. Ten docking runs were allowed for each compound with virtual screening setting (30%). Based on the scores, ligands with predicted low GS (<45), CS (<20), ChemPLP (<45), ASP (<20) as well as those with no hydrogen bonding (HB = 0) were eliminated, which resulted in 840 compounds. These compounds were screened again with high search efficiency (100%) and fifty docking runs. Candidates with low GS (<38), CS (<17), ChemPLP (<38), ASP (<17) as well as those with predicted limited hydrogen bonding (HB < 1) were eliminated, resulting in 205 candidates. For both rounds of screening the cut-off values of the scores were determined based on the scores of the known inhibitors itaconic acid, 3-nitropropionate and 3-bromopyruvate. Furthermore, only compounds with predicted hydrogen bonding were taken forward since hydrogen bond is important not only for the affinity but also the specificity of the ligand binding. These candidates were then visually inspected for consensus of the best predicted configuration of the ligands between the scoring functions. Ligands that showed plausible configurations, i.e., not strained, lipophilic moieties not pointing into the water environment resulting in an entropy penalty, were taken forward. Furthermore, compounds that did not contain undesirable moieties that are linked to general cell toxicity (e.g. thiourea and aliphatic ketones) and chemical reactivity (e.g. Michael acceptors and imines), were chosen. This screening methodology has been successfully applied previously to find active ligands for various bio-molecular systems. In total, 41 compounds were selected for experimental testing (Fig. S12 and Table S1†).

We then applied the $^1$H NMR and thermal shift assays to verify the hits obtained from the virtual screen. First, we tested the compounds using the thermal shift assay. Out of the 41 compounds, 19 induced a shift of more than 0.5 °C (positive or negative) in the melting temperature of ICL1 (Fig. S13†). This was followed with $^1$H NMR-based single concentration inhibition experiment to test the 19 hits. The result showed that two molecules significantly inhibited ICL1 (compounds 29 and 38, Fig. S14†). The IC$_{50}$ of the molecules were both >100 μM (Fig. S15 and S16† and Table 2). Molecular modelling suggests compounds 29 and 38 both occupy the substrate and Mg$^{2+}$ binding sites. The reason for the relatively low IC$_{50}$ values is due to the removal of the Mg$^{2+}$ ion from the binding pocket upon inhibitor binding. Mg$^{2+}$ sits within a cavity that is predicted to be occupied by aliphatic moieties of the inhibitors. Thus, the inhibitors need to displace the magnesium ion to bind efficiently, which would require a considerable energy expenditure due to the saturation concentration (5 mM) of the ion in the experimental setup. The main stream molecular descriptors (molecular weight, log $P$, hydrogen bond donors/acceptors, polar surface area and rotatable bonds see Table S2†) for compounds 29 and 38 were calculated and they conform to drug-like chemical space except log $P$ and numbers of hydrogen bond donors, which are in lead-like chemical space (for the definition of chemical space see ref. 57). Furthermore, the molecular weight for the hits is in the low to mid 300s, making them excellent starting points for chemical modification and further development. Finally, nine close structural derivatives were purchased to generate a structural activity (SAR) profile (Fig. S17†), but none showed any activity. In general, docking to the binding site showed that these compounds are too bulky to fit into it. Overall, our results show that combining $^1$H NMR and thermal shift assays is an effective strategy for screening potential ICL1 inhibitors.

Conclusions

ICL isoforms 1 and 2 are important enzymes for the survival of $M. tuberculosis$ in macrophages, enabling the bacteria to utilise fatty acids and cholesterol as carbon sources. ICLs are attractive inhibition targets for the treatment of latent TB. By using ICL1 as a model system, we have demonstrated the general applicability of a combined $^1$H NMR and thermal shift assays to screen for and evaluate ICL inhibitors. Both methods presented herein are relatively simple to carry out. In contrast to current fluorescence-based assays that rely on enzyme or chemically coupled reactions, the NMR assay described herein enables a direct observation of substrate consumption and product formation and is therefore less prone to errors. One minor drawback of the NMR
procedure reported by Gileadi et al. pNIC28-Bsa4 vector were prepared and cloned using the protocol reported by Gileadi et al. with XL10-Gold. The recombinant plasmid was confirmed by DNA sequencing (DNA Sequencing Centre, The University of Auckland). The correct plasmid were then used to transform BL21 (DE3) competent cells for protein expression.

Experimental section

Materials

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich/Merck, Thermo Fisher Scientific, Environmental Control Products (ECP), AK Scientific, Global Science – a VWR/Bio-Strategy Company and Bio-Rad. Tris-D11 and D2O were from Cambridge Isotope Laboratories or Cortecnet. Restriction enzymes Bam-HF and T4 DNA polymerase were obtained from New England Biolabs. Competent cells XL10-Gold and BL21 (DE3) were obtained from Agilent. The Bio-Rad Precision Plus Protein Kaleidoscope Prestained Protein Standards were used for sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE). Methyl 4-(4-methoxy-phenyl)-4-oxobut-2-enoate was obtained from Enamine. Compounds from virtual screening were obtained from InterBioScreen.

Cloning of ICL1

Synthetic gene (gBlocks) encoding M. tuberculosis ICL1 (Table S3) were obtained from Integrated DNA Technologies. The pNIC28-Bsa4 vector was a gift from Opher Gileadi (Addgene plasmid #26103).28 In order to clone onto the pNIC28-Bsa4 vector, the tattctcaatccatg sequence was added to the 5’ end and taaagttaagggata was added to the 3’ end of the DNA sequence encoding M. tuberculosis ICL1 (Table S3) when designing the synthetic gene. The synthetic gene and the pNIC28-Bsa4 vector were prepared and cloned using the procedure reported by Gileadi et al. with XL10-Gold.29 The recombinant plasmid was confirmed by DNA sequencing (DNA Sequencing Centre, The University of Auckland). The correct plasmid were then used to transform BL21 (DE3) competent cells for protein expression.

Production and purification of ICL1

The recombinant plasmid was first transformed in E. coli BL21(DE3). Starter culture was incubated overnight at 37 °C with shaking in 2YT media. The starter culture was then diluted with fresh 2YT media, which was then incubated at 37 °C with shaking until OD600 of 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG; 200 μM final concentration) was then added and further incubated at 18 °C with shaking for a further 16 to 20 hours. Cells were harvested by centrifugation. Cell pellets were obtained by centrifugation and resuspended in 50 mM HEPES buffer pH 7.8 with 5 mM imidazole and 500 mM NaCl. The cells were lysed on ice by sonication (4 × 20 seconds) burst at 60% amplitude with 40 seconds rest in between. The protein was purified by 5 mL His-trap column eluting with Tris-HCl buffer pH 7.8 with 500 mM imidazole and followed by gel filtration using 50 mM Tris-HCl buffer pH 7.5.

Thermal shift assay

Thermal shift assay was carried out using a BioRad MyiQ real time PCR instrument. The assay was carried out with 2 μM ICL1, 1 mM compounds and 1 mM MgCl2 in 50 mM Tris-HCl pH 7.5. Protein unfolding was monitored by measuring the fluorescence of the SYPRO Orange dye. The dye stock (5000× concentrate) was first diluted in 50 mM Tris-HCl (pH 7.5) to a 200× concentrate before diluting by 5 times into the sample. Temperature was increased from 25 to 95 °C at 1 °C increment every 60 seconds. All measurements were performed in triplicate. For determination of protein melting temperature values, melting curve for each data set was analysed by SigmaPlot 13 (USA) and fitted with the Sigmoid, 3 parameter model.

NMR experiments

NMR experiments were conducted at a 1H frequency of 500 MHz using a Bruker Avance III HD spectrometer equipped with a BBFO probe. Experiments were conducted at 300 K. Standard 5 mm NMR tubes (Wilmad) using a sample volume of 500 μL were used in all experiments. The pulse tip-angle calibration using the single-pulse notation method (Bruker pulsecal routine) was undertaken for each sample.60 All measurements were performed in triplicate.

Time course experiments were monitored by standard Bruker 1H experiments with water suppression by excitation sculpting. Unless otherwise stated, the number of transients was 16, and the relaxation delay was 2 seconds. The lag time between addition of enzyme and the end of the first experiment was usually 4 minutes. Initial rates were calculated by linear fitting using Excel 2013 (Microsoft) for data points up to 20% turnover. Kinetic parameters were obtained using the Hanes plot. Linear fitting was done using Excel 2013 (Microsoft). All NMR samples contained 190 nM ICL1, 1 mM DL-isocitrate and 5 mM MgCl2 buffered with 50 mM Tris-D3 (pH 7.5) in 90% H2O and 10% D2O. For kinetic parameter measurements, the isocitrate concentrations ranged from 50 μM to 2160 μM.
to 750 μM. For single concentration inhibition assay, 100 μM inhibitors was used. For IC<sub>50</sub> measurements, varying concentrations of inhibitors were used. IC<sub>50</sub> values were obtained by SigmaPlot 13 and fitted with the Sigmoid, 3 parameter model.

### Virtual high throughpout screening

The compounds were docked to the crystal structure of ICL I (PDB ID: 1F8I),<sup>39</sup> which was obtained from the Protein Data Bank (PDB).<sup>61,62</sup> The Seigress version FJ 2.6 program (Scigress: Version FJ 2.6; Fujitsu Limited, 2008-2016) was used to prepare the crystal structure for docking, i.e., hydrogen atoms added, the co-crystallised succinic acid and glyoxylic acid were removed from protein, the magnesium ion as well as crystallographic water molecules. The Seigress software suite was also used to transfer the structures from 2D to 3D followed by structural optimisation using the MM2 force field.<sup>63</sup> The centre of the binding was defined on the co-crystallised ligand with coordinates (x = 5.931, y = 56.950, z = 83.843) with 10 Å radius. For the initial screen 30% search efficiency was used (virtual screen) with ten runs per compound. For the second phase (re-dock) 100% efficiency was used in conjunction with fifty docking runs. The GoldScore,<sup>45</sup> ChemScore,<sup>46,47</sup> ChemPLF<sup>48</sup> and ASP<sup>49</sup> scoring functions were implemented to validate the predicted binding modes and relative energies of the ligands using GOLD v5.2 software suite. The InterBioScreen Ltd natural product collection was used for the screening.<sup>44</sup> The robustness of the protocol was tested by re-docking the co-crystallised ligand (succinic acid) with these results: RMSD (root-mean-square deviation) GS < 1.750 Å, CS < 0.929 Å, PLP < 0.747 Å and ASP < 1.882 Å, verifying the validity of the procedure. The QikProp v3.21 (QikProp v3.2, Schrödinger, New York, 3.2, 2009) software package was used to calculate the molecular descriptors of the compounds. The reliability of the prediction power of QikProp is established for the molecular descriptors used in this study.<sup>64</sup>

### Synthesis of (2S,3R)-2-methylisocitrate

(2S,3R)-2-Methylisocitrate was prepared according to the procedure reported by Darley et al.<sup>65</sup>

### Conflicts of interest

The authors declare no competing interests.

### Acknowledgements

We thank the University of Auckland for a Doctoral Scholarship (R. P. B) and the Maurice and Phyllis Paykel Trust for funding. G. B. is supported by a Sir Charles Hercus Fellowship awarded through the Health Research Council of New Zealand. We thank Professor Bernard Golding (Newcastle University, UK) for his advice when preparing (2S,3R)-2-methylisocitrate. We thank Dr M. Schmitz for maintenance of the NMR facility and Ms K. Boxen for the DNA sequencing service.

### References


36 J. V. Schloss and W. W. Cleland, Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analogue, Biochemistry, 1982, 21, 4420–4427.


44 InterBioScreen Ltd., 121019 Moscow, P.O. Box 218, Russia, http://www.ibscreen.com (accessed August 16, 2017).


Flexible synthesis of diverse N-heterocycles from substrates attainable from biomass†

Ram Prasad Bhusal and Jonathan Sperry*

A suite of diverse N-heterocycles including dihydroindolone, indole, dihydroindolizinone, 2-pyridone, pyrrole, 1,2-dihydroindole, dihydrocarbazolone and carbazole have all been prepared from dimethyl itaconate and pyrrole, two compounds attainable from biomass.

Introduction

Nitrogen-heterocycles (N-heterocycles) are present in many society-reliant chemicals¹ including pharmaceuticals, agrochemicals, dyes, nutraceuticals, fragrances and specialty polymers. At present, these compounds are mainly produced by synthetic chemistry using building-blocks derived from finite resources. Flexible synthetic routes to N-heterocycles that employ substrates attainable from biomass will ensure the supply of this important class of compounds will continue to meet demand as fossil-fuel levels decline.

Itaconic acid (1, Fig. 1) is produced by the fungal fermentation of xylose (extracted from hardwood and agricultural residues) or glucose (typically in the form of molasses from sugar cane or beets).²,³ Itaconic acid is classed as a fully sustainable chemical building block that features on the U.S. DOE National Renewable Energy Laboratory’s Top 12 list of renewable chemicals attainable from biomass.⁴ An ongoing interest in the synthetic utility of the bio-based molecule ¹ led us to examine its potential in the sustainable synthesis of important N-heterocycles, a strategy that requires a nitrogenous building block also attainable from biomass. Pyrrole was chosen as it is currently mass-produced from furan and ammonia,⁵ both of which are available from biorenewable sources.⁶–⁹

The overall plan hinges on the conjugate addition of pyrrole to dimethyl itaconate (2)¹⁰ selectively providing the mono- and di-addition products 3 and 4 (Scheme 1). With this initial stage successful, the presence of the electron-rich heteroaromatic ring and electrophilic esters in 3 and 4 will be used to effect distinct, selective intramolecular cyclization reactions that would result in a series of N-heterocycles derived from two simple substrates attainable from biomass (Scheme 1).

Results and discussion

A study investigating the conjugate addition of pyrrole to dimethyl itaconate (2) was initiated (Scheme 2). Lewis acids performed better than Bronsted acids in this reaction and specifically, aluminium trichloride was optimal. When an excess of pyrrole was used, the mono-addition product 3 was formed in good yield, along with a small amount of the di-addition product 4 (Scheme 2A). The di-addition product 4 could exclusively formed when an excess of dimethyl itaconate was employed (Scheme 2B).¹¹

The intramolecular cyclization reactions of pyrrole 3 were investigated. In the presence of polyphosphoric acid (PPA), pyrrole 3 underwent intramolecular C-acylation¹² to form the tetrahydroindolone 5, a heterocyclic motif present in the antipsychotic medication molindole.¹³ Dehydrogenation of 5 delivered methyl-4-hydroxyindole-6-carboxylate 6 in what constitutes an overall type 9 indole synthesis.¹⁴ Given the enormous importance of the indole heterocycle,¹⁵–¹⁷ sustainable synthetic methodologies that provide indoles are of great importance. The mode of cyclization could be switched upon treating pyrrole 3 with base, effecting selective N-acylation to give the dihydroindolizine 7, a motif present in the alkaloids polygonatine and kiganone.¹⁸ In an interesting result, attempts to aromatize 7 to the indolizine 8 only effected isomerization¹⁹ to the 2-pyridone 9, a heterocycle found in many bioactive natural products and synthetic drugs.²⁰ The synthesis outlined in Scheme 3 provides a complementary alternative to the synthesis of 4-hydroxyindole-6-carboxylates and indolizines from the half ester obtained from Stobbe condensation of pyrrole-2-carboxaldehyde with a dialkyl succinate.²¹–²⁴
The methodology outlined in Scheme 3 was subsequently applied to the double conjugate addition product 4 (Scheme 4). Upon exposing 4 to PPA, a single C-acylation occurred to give the tetrahydroindolone 10. None of tetrahydrocarbazole dione 11 was ever formed despite extensive optimization efforts. However, dehydrogenation of 10 gave indole 12 which did undergo C-acylation to give the dihydrocarbazolone 13, a heterocyclic motif present in the anti-emetic drug ondansetron.25 Dehydrogenation of 13 gave carbazoles 14 and 15.26 Carbazole is a tricyclic heterocycle present in a plethora of natural products27 and functional polymers.28 As the synthesis of carbazoles is often accomplished by the construction of the central ring using one of the many known C–N bond forming processes, the synthesis of 14/15 from pyrrole follows a relatively uncommon strategy for the construction of this heterocycle.29

Treating 4 with base led to exclusive N-acylation and formation of dihydroindolizinone 16, which underwent the same Pd/C-mediated isomerization described in Scheme 3 to give the 2-pyridone 17. Upon subjecting 16 to PPA, C-acylation occurred to give the tricycle 18 which resisted attempts at dehydrogenation to the desired tricycle 19. However, base-mediated cyclization of indole 12 did provide the desired pyrido[1,2-a]indolone 19, a heterocyclic motif present in many alkaloids including strychnine.30
Conclusions

In conclusion, a series of chemically diverse N-heterocycles have been rapidly assembled from two substrates available from biomass, demonstrating that biorenewable building blocks can provide a sustainable alternative for the synthesis of valuable small molecules currently acquired from finite resources. By only using compounds attainable from biomass in synthetic programmes, interesting chemical transformations can be discovered that lead to the production of new ‘sustainable chemical space’. Current focus is geared towards adopting environmentally friendly purification procedures and eliminating toxic solvents during the exploration of new sustainable chemical space. Results from our efforts in this area will be reported in due course.

Experimental

General

All reactions were carried out in oven-dried glassware under a nitrogen atmosphere unless otherwise stated. Analytical thin layer chromatography was performed using 0.2 mm silica plates and compounds were visualized under 365 nm ultraviolet irradiation followed by staining with either alkaline permanganate or ethanolic vanillin solution. Infrared spectra were obtained as thin films between sodium chloride plates. Absorption maxima are expressed in wavenumbers (cm\(^{-1}\)). Melting points were recorded on a melting point apparatus and are uncorrected. NMR spectra were recorded as indicated on an NMR spectrometer operating at 500, 400 and 300 MHz for \(^1\)H nuclei and 125, 100 and 75 MHz for \(^{13}\)C nuclei. Chemical shifts are reported in parts per million (ppm) relative to the tetramethylsilane peak recorded as \(\delta\) 0.00 ppm in CDCl\(_3\)/TMS solvent, or the residual acetone (\(\delta\) 2.05 ppm), chloroform (\(\delta\) 7.26 ppm), DMSO (\(\delta\) 2.50 ppm) or methanol (\(\delta\) 3.31 ppm) peaks. The \(^{13}\)C NMR values were referenced to the residual acetone (\(\delta\) 29.9 ppm) chloroform (\(\delta\) 77.1 ppm), DMSO (\(\delta\) 39.5 ppm) or methanol (\(\delta\) 49.0 ppm) peaks. \(^1\)H NMR values are reported as chemical shift \(\delta\) and assignment. \(^{13}\)C NMR values are reported as chemical shift \(\delta\) and assignment. \(^1\)H NMR shift values are reported as chemical shift \(\delta\), relative integral, multiplicity (s, singlet; d, doublet; t, triplet; q, quartet; quin, quintet; m, multiplet), coupling constant (\(J\) in Hz) and assignment. Assignments are made with the aid of DEPT 90, DEPT 135, COSY, NOESY and HSQC experiments. High resolution mass spectra were obtained by electrospray ionization in positive ion mode at a nominal accelerating voltage of 70 eV on a microTOF mass spectrometer.

General procedure A; C-acylation

To polyphosphoric acid (PPA) at 100 °C was added a solution of substrate in toluene and the mixture was stirred at 100 °C for the time stated. The reaction mixture was quenched with ice cold water (10 mL), the aqueous suspension extracted with ethyl acetate and the organic layer was washed with water, brine, dried (Na\(_2\)SO\(_4\)), filtered and concentrated in vacuo. The crude residue was purified by flash chromatography on silica gel with eluent stated to give the desired product.

General procedure B; N-acylation

A solution of substrate and potassium carbonate in DMF (4 mL) was stirred at 110 °C for the time stated. The solvent...
was concentrated in vacuo and water was added to the residue. The resulting aqueous suspension was extracted with ethyl acetate and the organic layer was washed with water, brine, dried (Na₂SO₄), filtered and concentrated in vacuo. The crude residue was purified by flash chromatography on silica gel with eluent stated to give the desired product.

**General procedure C; dehydrogenation**

A suspension of substrate and Pd/C (10%) in phenylether-mesitylene or n-octanol was heated at the temperature and time stated. The reaction mixture was filtered through Celite® and the residual catalyst washed with ethyl acetate. The filtrate was concentrated in vacuo and the residue purified by flash chromatography on silica gel with eluent stated to give the desired product.

**Dimethyl 2-((pyrrole-2-yl)methyl)succinate (3)**

To a solution of dimethyl 2-methylenesuccinate (500 mg, 3.16 mmol) in dichloromethane (5 mL) and nitromethane (5 mL) was added pyrrole (0.62 mL, 9.48 mmol) and aluminium trichloride (1.08 g, 7.2 mmol) and the solution was stirred at room temperature for 45 min. The reaction mixture was then quenched with water (20 mL). The aqueous layer was extracted with dichloromethane and the organic layer washed with brine, dried (Na₂SO₄), filtered and concentrated in vacuo. The crude residue was purified by flash chromatography on silica gel using hexanes–ethyl acetate (3:1) as eluent to afford the title compound (427 mg, 1.90 mmol, 60%) as a light yellow liquid; \( \nu_{\text{max}} \) (neat)/cm\(^{-1} \) 3383, 2953, 1722, 1570, 1436, 1409, 1357, 1269, 1194, 1165, 1121, 1095, 1027, 1004, 885, 837, 789, 717; \( \delta_{\text{H}} \) (400 MHz, CDCl₃) 8.37 (1 H, br s, NH), 6.68 (1 H, td, J 2.7 and 1.5, ArH), 6.09 (1 H, dt, J 5.8 and 2.7, ArH), 5.93–5.92 (1 H, m, ArH), 3.71 (3 H, s, Me), 3.67 (3 H, s, Me), 3.10 (1 H, q, J 6.4, CH), 3.00 (1 H, dd, J 14.8 and 6.4, CH of CH₂), 2.90 (1 H, dd, J 14.8 and 6.4, CH of CH₂), 2.68 (1 H, dd, J 16.4 and 7.2, CH of CH₂), 2.50 (1 H, dd, J 16.4 and 6.4, CH of CH₂); \( \delta_{\text{C}} \) (100 MHz, CDCl₃) 175.0 (C), 172.3 (C), 128.0 (C), 117.2 (CH), 108.3 (CH), 107.3 (CH), 52.2 (Me), 51.9 (Me), 42.1 (CH), 29.0 (CH₂); HRMS (ESI, \([\text{M} + \text{Na}]^+\) found 406.1472. \([\text{C}_{18}\text{H}_{25}\text{NO}_8 + \text{Na}]^+ \) requires 406.1478.

**Methyl 4-oxo-4,5,6,7-tetrahydroindole-6-carboxylate (5)**

General procedure A was performed using PPA (2 g) and 3 (100 mg, 0.45 mmol) in toluene (1 mL) for 3 h. Workup and column chromatography eluting with hexanes–ethyl acetate (1:1) gave the title compound (27 mg, 0.14 mmol, 31%) as a colourless solid; m.p. 127–131 °C; \( \nu_{\text{max}} \) (neat)/cm\(^{-1} \) 3299, 3130, 2952, 2922, 2851, 1725, 1632, 1569, 1505, 1474, 1435, 1377, 1350, 1293, 1271, 1244, 1234, 1193, 1152, 1116, 1077, 1003, 913, 902, 881, 866, 799, 764, 720, 692; \( \delta_{\text{H}} \) (400 MHz, CDCl₃) 8.80 (1 H, br s, NH), 6.70 (1 H, t, J 2.7, ArH), 6.54 (1 H, t, J 2.7, ArH), 3.72 (3 H, s, Me), 3.29–3.32 (1 H, m, CH), 3.13–3.10 (2 H, m, CH₂), 2.82–2.69 (2 H, m, CH₂); \( \delta_{\text{C}} \) (100 MHz, CDCl₃) 191.6 (C), 173.6 (C), 141.1 (C), 120.3 (C), 119.5 (CH), 106.2 (CH), 52.4 (Me), 41.3 (CH), 40.0 (CH₂), 25.5 (CH₂); HRMS (ESI, \([\text{M} + \text{Na}]^+\) found 216.0631. \([\text{C}_{16}\text{H}_{16}\text{NO}_5 + \text{Na}]^+ \) requires 216.0637.

**Methyl 4-hydroxyindole-6-carboxylate (6)**

General procedure C was performed using 5 (100 mg, 0.52 mmol) and Pd/C (100 mg) in phenylether (3.6 mL) and mesitylene (0.4 mL) for 240 °C for 36 h. Workup and column chromatography eluting with hexanes then hexanes–ethyl acetate (1:1) gave the title compound (52 mg, 0.27 mmol, 54%) as a brown solid; \( \delta_{\text{H}} \) (400 MHz, CD₃OD) 7.68 (1 H, t, J 0.9, ArH), 7.30 (1 H, d, J 3.5, ArH), 7.06 (1 H, d, J 1.4, ArH), 6.58 (1 H, d, J 3.5 and 0.9, ArH), 3.88 (3 H, s, Me); \( \delta_{\text{C}} \) (100 MHz, CD₃OD) 170.2 (C), 151.1 (C), 138.6 (C), 127.4 (C), 124.9 (C), 123.5 (CH), 107.4 (CH), 104.5 (CH), 100.0 (CH), 52.3 (Me); spectroscopic data consistent with the literature.²¹

**Methyl 5-oxo-5,6,7,8-tetrahydroindolizine-7-carboxylate (7)**

General procedure B was performed using 3 (375 mg, 1.7 mmol) and potassium carbonate (470 mg, 3.4 mmol) in DMF (4 mL) for 12 h. Workup and column chromatography eluting with hexanes–ethyl acetate (4:1) gave the title compound (198 mg, 1.02 mmol, 60%) as a fawn solid; m.p. 83–85 °C; \( \nu_{\text{max}} \) (neat)/cm\(^{-1} \) 1716, 1575, 1449, 1429, 1379, 1354, 1302, 1229, 1214, 1199, 1151, 1100, 1068, 1031, 1022, 1005, 917, 870, 825, 802, 750, 621, 594; \( \delta_{\text{H}} \) (400 MHz, CDCl₃) 7.35 (1 H, dd, J 3.4 and 1.4, ArH), 6.22 (1 H, t, J 3.4, ArH), 6.03–6.02 (1 H, m, ArH), 7.33 (3 H, s, Me), 3.23–3.17 (1 H, m, CH of CH₂), 3.16–3.10 (1 H, m, CH), 3.03–2.97 (1 H, m, CH of CH₂), 2.94–2.85 (2 H, m, CH₂); \( \delta_{\text{C}} \) (100 MHz, CDCl₃) 172.8 (C), 166.4 (C), 130.1 (C), 116.6 (CH), 113.1 (CH), 110.0 (CH), 52.5 (Me),
Methyl 5-oxo-1,2,3,5-tetrahydroindolizine-7-carboxylate (9)

General procedure C was performed using 7 (40 mg, 0.21 mmol) and Pd/C (40 mg) in n-octanol (2 mL) at 195 °C for 12 h. Workup and column chromatography eluting with ethyl acetate–methanol (9:5; 0.5 g) of the title compound (30 mg, 0.155 mmol, 75%) as a colourless solid; m.p. 75 °C.

Dimethyl 5-hydroxy-4-oxo-2,3,4,9-tetrahydrocarbazole-2,7-dicarboxylate (13)

General procedure A was performed using PPA (2 g) and 12 (40 mg, 0.12 mmol) in toluene (1 mL) for 45 min. Workup and column chromatography eluting with hexanes–ethyl acetate (1:1.5) gave the title compound (20 mg, 0.06 mmol, 55%) as a brown solid; m.p. 255–259 °C; $\nu_{\text{max}}$ (neat)/cm$^{-1}$ 3287, 2953, 1706, 1589, 1476, 1436, 1404, 1344, 1315, 1226, 1199, 1147, 1086, 1062, 994, 876, 785, 788, 768, 690; $\delta_{\text{H}}$ (400 MHz, DMSO-d$_6$) 12.49 (1 H, br s, NH), 10.76 (1 H, s, OH), 7.52 (1 H, d, J 1.4, ArH) 7.03 (1 H, d, J 1.4, ArH), 3.84 (3 H, s, Me), 3.65 (3 H, s, Me), 3.56–3.49 (1 H, m, CH), 3.35–3.22 (2 H, m, CH$_3$), 2.84–2.83 (2 H, m, CH$_2$); $\delta_{\text{C}}$ (100 MHz, DMSO-d$_6$) 193.9 (C), 172.9 (C), 166.5 (C), 153.3 (C), 150.2 (C), 137.3 (C), 126.7 (C), 117.5 (C), 112.0 (C), 106.6 (CH), 105.2 (CH), 52.1 (Me), 52.0 (Me), 39.9 (CH), 37.9 (CH$_3$), 24.8 (CH$_2$); HRMS [EI, [M + Na]$^+$] found 340.0792. $[C_{16}H_{15}NO_6 + Na]^+$ requires 340.0797.

Dimethyl 4,5-dihydroxy carbazole-2,7-dicarboxylate (14) and dimethyl 4-hydroxy-5-methoxy carbazole-2,7-dicarboxylate (15)

General procedure C was performed using 13 (100 mg, 0.31 mmol) and Pd/C (170 mg) in phenylether (2.7 mL) and mesitylene (0.3 mL) for 240 °C for 18 h. Workup (filtration through silica gel instead of Celite®) and column chromatography eluting with hexanes–ethyl acetate (1:1) gave the title compounds:

14 (25 mg, 0.08 mmol, 25%) as a brown solid; m.p. 325–330 °C; $\nu_{\text{max}}$ (neat)/cm$^{-1}$ 3313, 1692, 1590, 1518, 1456, 1436, 1427, 1360, 1334, 1313, 1226, 1190, 1089, 1074, 1010, 992, 912, 861, 484, 796, 759, 722; $\delta_{\text{H}}$ (400 MHz, DMSO-d$_6$) 11.82 (1 H, s, NH), 11.19 (2 H, s, 2 × OH), 7.64 (2 H, d, J 1.3, 2 × ArH), 7.18 (2 H, d, J 1.3, 2 × ArH), 3.88 (6 H, s, 2 × Me); $\delta_{\text{C}}$ (100 MHz, DMSO-d$_6$) 166.5 (2 × C), 150.3 (2 × C), 141.2 (2 × C), 128.7 (2 × C), 113.3 (2 × C), 104.6 (4 × CH); 52.2 (2 × Me); HRMS [EI, [M + Na]$^+$] found 338.0635. $[C_{16}H_{14}NO_6 + Na]^+$ requires 338.0641.

15 (11 mg, 0.03 mmol, 10%) brown solid; m.p. 245–250 °C; $\nu_{\text{max}}$ (neat)/cm$^{-1}$ 3302, 1695, 1637, 1571, 1518, 1498, 1432, 1332, 1315, 1256, 1211, 1116, 1090, 1083, 1046, 1012, 964, 862, 790, 759, 729; $\delta_{\text{H}}$ (400 MHz, DMSO-d$_6$) 11.98 (1 H, s, NH), 9.31 (1 H, s, OH), 7.82 (1 H, d, J 1.0, ArH), 7.64 (1 H, d, J 1.2, ArH), 7.33 (1 H, s, ArH), 7.12 (1 H, d, J 1.0, ArH), 4.19 (3 H, s, Me), 3.91 (3 H, s, Me), 3.88 (3 H, s, Me); $\delta_{\text{C}}$ (100 MHz, DMSO-d$_6$) 166.5 (C), 166.3 (C), 152.1 (C), 151.1 (C), 141.4 (C), 140.8 (C), 129.1 (C), 128.4 (C), 113.4 (C), 112.6 (C), 107.2 (C), 105.0 (CH), 104.2 (CH), 100.7 (CH), 56.8 (Me), 52.3 (Me), 52.2 (Me); HRMS [EI, [M + Na]$^+$] found 352.0792. $[C_{17}H_{16}NO_7 + Na]^+$ requires 352.0797.

Green Chemistry 2016, 18, 2453–2459 | 2457
Dimethyl 2-((7-(methoxycarbonyl)-5-oxo-5,6,7,8-tetrahydroindolizin-3-yl)ethyl)succinate (16)

General procedure B was performed using 4 (420 mg, 1.09 mmol) and potassium carbonate (757 mg, 5.47 mmol) in DMF (5 mL) for 12 h. Workup and column chromatography eluting dichloromethane–ethyl acetate (15:1) gave the title compound (222 mg, 0.63 mmol, 58%, 1:1 mixture of diastereomers) as a yellow oil; δH (400 MHz, CDCl3) 9.59 (1 H, s, ArH), 6.92 (1 H, s, ArH), 6.56 (2 H, s, ArH), 4.84–4.78 (1 H, m, CH), 4.77–4.72 (1 H, m, CH), 3.88 (6 H, s, 2 × Me), 3.74 (3 H, s, Me), 3.69 (3 H, s, Me), 3.68 (3 H, s, Me), 3.67 (3 H, s, Me), 3.25–3.13 (2 H, m, CH2), 3.05–2.94, (4 H, m, 2 × CH + CH2), 2.89–2.78 (2 H, m, CH2), 2.70–2.57 (2 H, m, CH2), 2.43–2.34 (1 H, m, CH of CH2), 2.32–2.22 (2 H, m, CH2), 2.11–2.02 (3 H, m, CH2 + CH of CH2), 1.99–1.92 (1 H, m, CH of CH2), 1.72–1.65 (1 H, m, CH of CH2); δC (100 MHz, CDCl3) 174.6 (C), 174.2 (C) 172.9 (C), 172.5 (C), 165.6 (2 × C), 165.2 (C), 161.9 (C), 150.7 (2 × C), 141.4 (C), 141.3 (C), 120.0 (CH), 119.9 (CH), 99.93 (CH), 99.91 (CH), 95.8 (CH), 59.2 (CH), 52.9 (2 × Me), 52.4 (Me), 52.3 (Me), 52.1 (Me), 52.0 (Me), 39.1 (CH), 38.9 (CH), 36.0 (CH), 35.9 (CH), 33.9 (CH), 32.0 (CH), 30.14 (CH2), 30.12 (CH2), 26.9 (CH2), 26.8 (CH2); HRMS (ESI, [M + Na]+) found 374.1210. [C16H21NO7 + Na]+ requires 374.1216.

Dimethyl 2-((7-(methoxycarbonyl)-5-oxo-2,3,5-tetrahydroindolizin-3-yl)ethyl)succinate (17)

General procedure C was performed using 16 (32 mg, 0.09 mmol) and Pd/C (32 mg) in n-octanol (2 mL) at 195 °C for 12 h. Workup and column chromatography eluting hexanes–ethyl acetate (1:1) gave the title compound (25 mg, 0.038 mmol, 67%) as a colourless solid; δH (400 MHz, CDCl3) 6.98 (1 H, s, ArH), 6.60 (1 H, s, ArH), 6.56 (2 H, s, ArH), 4.84–4.78 (1 H, m, CH), 4.77–4.72 (1 H, m, CH), 3.88 (6 H, s, 2 × Me), 3.74 (3 H, s, Me), 3.69 (3 H, s, Me), 3.68 (3 H, s, Me), 3.67 (3 H, s, Me), 3.25–3.13 (2 H, m, CH2), 3.05–2.94, (4 H, m, 2 × CH + CH2), 2.89–2.78 (2 H, m, CH2), 2.70–2.57 (2 H, m, CH2), 2.43–2.34 (1 H, m, CH of CH2), 2.32–2.22 (2 H, m, CH2), 2.11–2.02 (3 H, m, CH2 + CH of CH2), 1.99–1.92 (1 H, m, CH of CH2), 1.72–1.65 (1 H, m, CH of CH2); δC (100 MHz, CDCl3) 174.6 (C), 174.2 (C) 172.9 (C), 172.5 (C), 165.6 (2 × C), 165.2 (C), 161.9 (C), 150.7 (2 × C), 141.4 (C), 141.3 (C), 120.0 (CH), 119.9 (CH), 99.93 (CH), 99.91 (CH), 95.8 (CH), 59.2 (CH), 52.9 (2 × Me), 52.4 (Me), 52.3 (Me), 52.1 (Me), 52.0 (Me), 39.1 (CH), 38.9 (CH), 36.0 (CH), 35.9 (CH), 33.9 (CH), 32.0 (CH), 30.14 (CH2), 30.12 (CH2), 26.9 (CH2), 26.8 (CH2); HRMS (ESI, [M + Na]+) found 374.1210. [C16H21NO7 + Na]+ requires 374.1216.

Acknowledgements

We thank the University of Auckland for financial support (FRDF grant 3704091 and Doctoral Scholarship to R. P. B.).

References

10 Dimethyl itaconate is obtained by the acid-catalysed die
14 D. F. Taber and P. K. Tirunahari, *The Chemistry of Indoles*
Green Chem. This journal is © The Royal Society of Chemistry 2016
19 The isomerization does not occur in the absence of Pd/C.
12 For a similar cyclization using a carboxylic acid, see: (a) J. Andersson and J. Lundgren, *Appl.
Prod.*, 2014, 64, 581.
9 For a recent publication on the direct synthesis of N-heterocycles (including pyrroles) from biomass, see: L. J. Xu, Q. Yao, J. Deng, Z. Han, Y. Zhang, Y. Fu, G. H. Huber and Q. Guo, *ACS Sustainable Chem. Eng.*, 2015, 3, 2890.
11 Despite screening several different catalysts and solvents, AlCl₃ in CH₂Cl₂-MeNO₂ was the only combination that gave acceptable yields.
26 The formation of O-methylated carbazole 15 requires comment. It is possible water contained in the Pd/C hydrolyses the ester(s) during the reaction. At the high temperatures, the liberated methanol then attacks the carbonyl in compound 13 or directly methylates 14, both of which would result in compound 15.