Three enzymes - One substrate
Regulation of carbon flux through a “non-canonical” metabolic branchpoint

Audrey Crousilles
Department of Biochemistry
University of Cambridge
Queens’ College
September 2018

This dissertation is submitted for the degree of
Doctor of Philosophy
Deposit & Copying of Hardbound Dissertation Declaration

Please note that you will also need to bind a copy of this Declaration into your final, hardbound print copy of your dissertation - this has to be the very first page of the hardbound dissertation. Do not include this form in the electronic version of your dissertation.

<table>
<thead>
<tr>
<th>1</th>
<th>Surname (Family Name)</th>
<th>Forenames(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Croustilles</td>
<td>Audrey Laure</td>
<td>Miss</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>Title of Dissertation as approved by the Degree Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Three enzymes - One substrate</td>
</tr>
<tr>
<td></td>
<td>Regulation of carbon flux through a &quot;non-canonical&quot; metabolic branchpoint</td>
</tr>
</tbody>
</table>

In accordance with the University Regulations in Statutes and Ordinances for the PhD, MSc and MLitt Degrees, I agree to deposit one or more copies of my dissertation and summary with the Secretary of the Board of Graduate Studies in a form or forms approved by the Board. The Secretary of the Board of Graduate Studies shall deposit a hardbound print copy of my dissertation and summary in the University Library under the following terms and conditions:

1. **Dissertation Author Declaration**

I am the author of this dissertation and hereby give the University the right to make my dissertation available in print form as described in 2. below.

My dissertation is my original work and a product of my own research endeavours and includes nothing which is the outcome of work done in collaboration with others except as declared in the Preface and specified in the text. I hereby assert my moral right to be identified as the author of the dissertation.

The deposit and dissemination of my dissertation by the University does not constitute a breach of any other agreement, publishing or otherwise, including any confidentiality or publication restriction provisions in sponsorship or collaboration agreements governing my research or work at the University or elsewhere.

2. **Access to Dissertation**

I confirm that I have selected an appropriate access and permission usage level for my dissertation, which is detailed on the Access Confirmation Form that I submitted with the electronic version of my dissertation. I understand that one print copy of my dissertation will be deposited in the University Library for archival and preservation purposes, and that, unless upon my application an embargo or restricted access has been granted for a specified period of time by my Supervisor or Degree Committee respectively prior to this deposit, the dissertation will be made available by the University Library for consultation by readers in accordance with University Library Regulations and copies of my dissertation may be provided to readers in accordance with applicable legislation, or on other access and permission usage terms detailed on my Access Confirmation Form and/or as may be marked on the dissertation.

3. **Agreement to terms**

By including this Declaration in my Hardbound Dissertation I confirm that any statements on this Declaration are correct and that I agree to the terms for access to my dissertation.
Declaration

This dissertation is the result of my own work carried out under the supervision of Dr. Martin Welch in the Department of Biochemistry and includes nothing which is the outcome of work done in collaboration except as specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This dissertation does not exceed the word limit specified by the Biology Degree Committee.
Abstract

Three enzymes – One substrate
Regulation of carbon flux through a “non-canonical” metabolic branchpoint

Audrey Crousilles

*Pseudomonas aeruginosa* is a common opportunistic pathogen. Recent work indicates that in many infection scenarios, *P. aeruginosa* exhibits an exquisite predilection for metabolizing fatty acids to yield acetyl-CoA. In most higher organisms, acetyl-CoA cannot be used for biomass production because the two carbon atoms which enter the TCA cycle are lost as CO₂. However, many bacteria are able to bypass these oxidative decarboxylation steps, allowing them to conserve carbon for gluconeogenesis. They perform this by using the “glyoxylate shunt”. Here, isocitrate is cleaved by isocitrate lyase (ICL) to yield succinate and glyoxylate (which, in a subsequent reaction, is combined with a further acetyl-CoA unit to yield the gluconeogenic precursor, malate). However, ICL has to compete with the TCA cycle enzyme, isocitrate dehydrogenase (ICD), for the available isocitrate, and it is the outcome of this “metabolic tussle” which dictates the flux of carbon through the glyoxylate shunt. In *E. coli*, ICD is inactivated by AceK-dependent phosphorylation, allowing flux through the glyoxylate shunt. However, *P. aeruginosa* is “wired up” differently because it employs not one, but two highly-expressed isocitrate dehydrogenases (ICD and IDH). For this PhD project, I focused on these three enzymes (ICD, IDH and ICL). I cloned, overexpressed and purified them at high yield to perform a thorough investigation of their kinetics, regulation and more interestingly crystal structures. I found that only one of these (the *E. coli*-like ICD) is regulated by AceK-mediated phosphorylation. The other, IDH, is allosterically regulated, as is the isocitrate lyase. These findings demonstrate that in *P. aeruginosa* the rerouting of the carbon flux through the glyoxylate shunt is delicately regulated via allostery mainly. The conditions in which the cells grow and access to either poor or rich carbon sources heavily influence the partitioning of the central metabolism. In *P. aeruginosa*, the TCA cycle remains more active (than in *E. coli* for example) even during growth on poor nutrient and this is probably an important aspect to manage oxidative stress accompanying growth. Finally, I have solved the x-ray crystal structures of ICD, IDH and ICL. These are entirely novel structures that have not been defined previously and are new entries to the Protein Data Bank. The structure solving work highlighted very interesting peculiarities to these enzymes when compared with other bacterial pathogens. This emphasizes the growing idea that *Pseudomonas aeruginosa* is a unique bacterium that cannot be modelled by the well-studied *Escherichia coli*. All this work crystallizes the knowledge to build up a picture of how flux is likely to be regulated at this “non-canonical” metabolic branchpoint and features new interesting directions for downstream applications such as drug-design.
Acknowledgements

Firstly and foremost, I would like to thank Dr. Martin Welch for this wonderful opportunity to work in his group and for his supervision and guidance throughout my time here in Cambridge. I am very appreciative of all the encouragement, training and support. Thank you to Pr. George Salmond for his advice and input throughout my time in the department as my PhD adviser. Thanks also to the Cambridge Trust and Cambridge Philosophical Society for funding this PhD over the first three years and then the fourth year. An additional thank you to the Microbiology Society and Queens’ College for their supportive travel grants.

Thank you to all members of the Welch and Salmond groups for creating such an amazing working environment and for making every day so enjoyable. I have met so many wonderful people over the years, some continued their career elsewhere, but they remain in my heart and spirit. Special thanks go to Dr. Dimitri Chirgadze for his thorough training on the X-ray facility and structure solving. This work would have not been possible without his concours. Thank you also to Prof. Ben Luisi for helping me out at a time of slight despair while solving IDH. Thanks also to Dr. Paul Brear for his helpful expertise and advice in structure solving and in participating to a better understanding. A special thank you to Dr. Katherine Stott for teaching me how to perform sedimentation velocity analytical ultracentrifugation.

Finally, I also need to thank my relatives for their continuous support during this time in the UK. Another thanks to all my friends (old and new) and my boyfriend who influenced me in the greatest way to carry on research whatever the odds.
# Table of contents

List of figures .............................................................................................................................i
List of tables ............................................................................................................................... iii
List of appendixes ......................................................................................................................... iv
Abbreviations ............................................................................................................................... v

## 1. Introduction ..........................................................................................................................1

1.1. *Pseudomonas aeruginosa* .................................................................................................1
   1.1.1. Generalities ............................................................................................................... 1
   1.1.2. Infections scenario ................................................................................................. 2
       1.1.2.1. Acute infections .......................................................................................... 2
       1.1.2.2. Chronic infections ...................................................................................... 3
       1.1.2.3. Cystic fibrosis (CF) .................................................................................... 4
   1.1.3. Antibiotic resistance ................................................................................................ 5
       1.1.3.1. Intrinsic resistance ...................................................................................... 5
       1.1.3.2. Acquired resistance ..................................................................................... 6
   1.1.4. Virulence factors .................................................................................................... 7

1.2. A metabolic choice: 2C or not 2C? ....................................................................................8
   1.2.1. TCA cycle .............................................................................................................. 8
       1.2.1.1. Context and generalities ........................................................................... 8
       1.2.1.2. Regulation ................................................................................................... 9
   1.2.2. The glyoxylate shunt ........................................................................................... 10
       1.2.2.1. Generalities .............................................................................................. 10
       1.2.2.2. Branchpoint effect .................................................................................... 12
   1.2.3. Metabolism and virulence are linked ...................................................................... 13

1.3. The enzymology at the branchpoint ................................................................................... 14
   1.3.1. The isocitrate dehydrogenases ............................................................................ 14
   1.3.2. The isocitrate dehydrogenase kinase/phosphatase ............................................ 15
   1.3.3. The isocitrate lyase ......................................................................................... 16

1.4. Aims of this PhD project .................................................................................................. 17

## 2. Materials and methods .......................................................................................................18

2.1. General microbiological procedures ..................................................................................18
   2.1.1. Strains ............................................................................................................... 18
   2.1.2. Media components and supplements .................................................................. 18
   2.1.3. Growth conditions ............................................................................................ 19
   2.1.4. Growing and harvesting bacterial samples ....................................................... 20

2.2. DNA manipulation .............................................................................................................20
   2.2.1. DNA extraction, purification, sequencing and storage ....................................... 20
   2.2.2. Polymerase Chain Reaction (PCR) ..................................................................... 20
   2.2.3. Agarose gel electrophoresis ............................................................................. 22

2.3. Strains preparation ............................................................................................................22
   2.3.1. Plasmid construction ......................................................................................... 22
2.11. Crystallisation and X-ray experiments ...................................................... 34
  2.11.1. Approach to crystallisation ................................................................. 34
    2.11.1.1. Crystallisation screening and optimization .................................................. 34
    2.11.1.2. Crystallisation conditions for ICD ............................................................. 35
    2.11.1.3. Crystallisation conditions for IDH ......................................................... 35
    2.11.1.4. Crystallisation conditions for ICL .......................................................... 35
  2.11.2. Crystal “fishing” and storage of the samples ........................................ 35
  2.11.3. Sending samples and Synchrotron “shooting” ........................................ 35
  2.11.4. Crystal structure determination, model building and refinement ............... 36
5.3. Purification of ICL protein

5.3.1. Preliminary results
5.3.2. Gel filtration analysis
5.3.3. AUC data

5.4. Gene expression of aceA

5.5. Kinetic analyses

5.5.1. Michaelis-Menten data
5.5.2. Screening for regulators
5.5.3. Inhibition kinetic study
5.5.4. Thermal shift assay

5.6. Crystal structure of ICL

5.6.1. Structure description
5.6.2. Structural comparison and contact analysis
5.6.3. Active site analysis of ICL_P

5.7. Discussion

6. Conclusion

References

Appendixes
List of figures

Figure 1: Prevalence of the bacterial population in cystic fibrosis lungs per age cohort ............................................ 4
Figure 2: TCA cycle schematic ................................................................. 8
Figure 3: Glyoxylate shunt schematic ....................................................... 10
Figure 4: Branchpoint model in E. coli .................................................... 12
Figure 5: Schematic representation of the monomeric and dimeric isocitrate dehydrogenases .............................. 14
Figure 6: Genetic context of the icd gene in P. aeruginosa PAO1 and other Pseudomonas species ............................. 39
Figure 7: ClustalOmega alignment of ICD amino acid sequence ................................................................. 40
Figure 8: Coomassie stained SDS-PAGE analysis after purification of ICDpa .................................................... 41
Figure 9: Gel filtration of ICD ................................................................ 42
Figure 10: SV-AUC data of ICD protein ................................................... 43
Figure 11: Gene expression results of Picd::lacZ fusion .............................................. 44
Figure 12: Kinetics results of ICD ................................................................ 46
Figure 13: Kinetic impact of potential regulators of ICDpa ..................................... 47
Figure 14: Study of the effect of ATP on ICD activity ........................................ 49
Figure 15: Lineweaver-Burk plots of ATP inhibition ........................................ 50
Figure 16: Effect of other nucleotides on ICD activity ...................................... 51
Figure 17: Effect of AceK on ICD ................................................................ 52
Figure 18: ICD crystal structure ................................................................. 53
Figure 19: Analysis of the interface of ICD .................................................. 54
Figure 20: Superposition of ICDpa and ICDec apo monomers ....................... 55
Figure 21: Active site of ICDpa superposed with ICDec .................................. 57
Figure 22: Proposed catalytic mechanism of isocitrate dehydrogenase .......... 59
Figure 23: Genetic context of the idh gene in P. aeruginosa PAO1 and other Pseudomonas species ............................ 63
Figure 24: Alignment of idh promoter and transcriptional silencer ................. 64
Figure 25: ClustalOmega alignment of IDH amino acid sequence .................... 65
Figure 26: Coomassie-stained SDS-PAGE analysis after purification of IDHpa .................................................... 66
Figure 27: Gel filtration of IDH ................................................................ 67
Figure 28: SV-AUC data of IDH protein ..................................................... 68
Figure 29: Gene expression results of Pidh::lacZ fusion .................................. 70
Figure 30: Kinetic analysis of IDH activity ............................................... 71
Figure 31: AceK activity on IDH compared with ICD ..................................... 72
Figure 32: Screening of potential regulators on IDH ...................................... 73
Figure 33: Activators of IDH ................................................................ 74
List of tables

Table 1: Bacterial strains
Table 2: Growth media
Table 3: 1X MOPS medium
Table 4: 10X MOPS medium
Table 5: 100X micronutrient stock
Table 6: PCR primers
Table 7: Typical PCR program for Taq/Phusion polymerase
Table 8: Plasmids used in this study
Table 9: Solutions for β-galactosidase assay
Table 10: Buffers for His\textsubscript{6}-tagged proteins expression
Table 11: Buffers for His\textsubscript{6}-MBP-tagged proteins expression
Table 12: Buffers used for preparation and running gels
Table 13: Buffers for His\textsubscript{6}-tagged proteins expression
Table 14: Extinction coefficient and predicted weight of all proteins
Table 15: Regulators showing an impaired response in the direct assay
Table 16: Data collection and refinement statistics
Table 17: Kinetics parameters of ICD
Table 18: Kinetics parameters of IDH
Table 19: Kinetics parameters of ICL with inhibitors (non-competitive)
Table 20: Residues comparison of ICL active site
List of appendixes

Appendix 1: Michaelis-Menten plots ........................................................................................................133
Appendix 2: SDS-PAGE analysis of His$_6$-tagged proteins .................................................................134
Appendix 3: Full gel filtration results of II(D ..............................................................135
Appendix 4: ESPript representation of *P. aeruginosa* ICD .................................................................136
Appendix 5: Extracted information from the validation report of ICD PDB deposition .......................137
Appendix 6: Full gel filtration results of IDH ......................................................................................138
Appendix 7: ESPript representation of *P. aeruginosa* IDH.................................................................139
Appendix 8: Extracted information from the validation report of IDH PDB deposition .......................140
Appendix 9: Full gel filtration results of ICL .....................................................................................141
Appendix 10: ESPript representation of ICL ......................................................................................142
Appendix 11: Full results from COCOMAPS analysis on ICL .............................................................143
Appendix 12: Extracted information from the validation report of ICL PDB deposition .....................144
Abbreviations

AceK: Isocitrate dehydrogenase kinase/phosphatase
Acetyl-CoA: Acetyl-coenzyme A
AIDS: Acquired immunodeficiency syndrome
ADP: Adenosine diphosphate
AMP: Adenosine monophosphate
ATP: Adenosine triphosphate
ASA: Accessible surface area
BLAST: Basic local alignment search tool
Bp: Base pair
CCR: Carbon catabolite repression
CF: Cystic fibrosis
CFTR: Cystic fibrosis transmembrane conductance regulator
COPD: Chronic obstructive pulmonary disease
DNA: Deoxynucleic acid
dNTPs: Deoxyribonucleotide triphosphates
EDP: Entner-Doudoroff pathway
EMP: Embden-Meyerhof-Parnas
ICD: Isocitrate dehydrogenase
G-C: Guanine-Cytosine
GDP: Guanosine diphosphate
GTP: Guanosine triphosphate
ICL: Isocitrate lyase
IDH: Isocitrate dehydrogenase
IPTG: Isopropyl β-D-1-thiogalactopyranoside
KDPG: 3-deoxy-2-keto-6-phosphogluconate
LB: Luria-Bertani (broth or agar)
LDH: Lactate dehydrogenase
LPS: Lipopolysaccharide
Mbp: Million base pairs
MS: Malate synthase
4-MUG: 4-Methylumbelliferyl-β-D-galactopyranoside
MWCO: Molecular weight cut-off
NAD(H): Nicotinamide adenine dinucleotide
NADP(H): Nicotinamide adenine dinucleotide phosphate
Ni-NTA: Nickel-nitrilotriacetic acid
OD: Optical density
ORF: Open reading frame
PCR: Polymerase chain reaction
PDB: Protein databank
PEP: Phosphoenolpyruvate
RFU: Relative fluorescence unit
RGP: Region of genome plasticity
ROS: Reactive oxygen species
rpm: rotations per minute
SDS: Sodium dodecyl sulphate
SDS-PAGE: SDS polyacrylamide gel electrophoresis
SV-AUC: Sedimentation velocity analytical ultra-centrifugation
TCA cycle: Trichloroacetic acid cycle
TGB: TCA cycle Glyoxylate shunt Branchpoint
TRIS: Tris(hydroxymethyl)aminomethane
UTI: Urinary tract infection
1. Introduction

1.1. Pseudomonas aeruginosa

1.1.1. Generalities

*Pseudomonas aeruginosa* is a Gram-negative, rod-shaped and mono-flagellated bacterium of the Pseudomonadaceae family. Pseudomonads are renowned for their metabolic versatility and their widespread distribution. *P. aeruginosa* is defined as a ubiquitous microorganism; it can grow in soil and coastal marine habitats but also on plant and animal tissues. It is capable of forming biofilms as a trait to grow on wet surfaces including rocks and soil. *Pseudomonas* was initially an environmental organism, but it evolved to colonise and establish infections in more complex hosts including humans, animals and plants.

In humans, *P. aeruginosa* is an opportunistic pathogen of immunocompromised or immunodeficient individuals. It accounts for around 10% of all nosocomial infections in hospitals in the European Union, and in February 2017, made it to the top-three priority pathogens list for new antibiotics R&D after reports of carbapenem-resistant strains. So, *P. aeruginosa* is a significant source of bacteraemia in burns victims, urinary-tract infections in catheterized patients and hospital-acquired pneumonia treated with respirators. The ubiquitous trend stems from an array of abilities to facilitate adaptation and survival in diverse ecological niches. *P. aeruginosa* has broad capabilities to transport, metabolize and grow on organic substances. It is well equipped to export compounds (antibiotics included) with a large number of pumps and efflux systems. Furthermore, *P. aeruginosa* has the ability to form adhesive communities known as biofilms and possesses an impressive arsenal of virulence factors. More specifically, the metabolic versatility is provided by genes encoding not only the enzymes participating in metabolic pathways, but also a surprisingly large number of transcriptional regulators. With more than 500 regulatory genes, *P. aeruginosa* is particularly complex and demonstrates fine regulation of multiple pathways. This PhD project will highlight one of them and its consequences for carbon management in the cell.

Another interesting aspect of *Pseudomonas aeruginosa* is its genome. It has one of the largest and most complex genomes among bacteria. Within the species, the size varies between 5.5 and 7 Mbp and the G-C content fluctuates around 65-67%. The genome itself is made up of a single circular chromosome and a variable number of plasmids. The divergence in genome size is explained by its structure, there is the core genome and the accessory genome. The latter increments the variable length. The conserved core component of the genome is largely collinear among *P. aeruginosa* strains. Ongoing genome projects indicate that the core genome consists of somewhat more than 4,000 genes. It is highly conserved among divergent strains within the species, and overall shows very low interclonal sequence diversity of 0.5 – 0.7 %. Very few loci in the core genome are subject to diversifying selection. As an example of the integrity of this genome, a study performed on a subset
of characterized clinical strains identified 5,021 genes that are conserved across all five genomes, with at least 70% sequence identity. This set of genes was therefore indicative of the core genome. Among these, more than 90% of them shared at least 98% sequence identity. Furthermore, a similar study analysed twenty clinical isolates and concluded that 89 to 97% of the PAO1 open reading frames (ORFs) were detected revealing a conserved pattern of genome mosaicism. However, the variable genomic regions appeared as 38 islands dispersed in the chromosome. These sites of dispersion of additional genetic information remain complex and very variable, they participate in the genome mosaicism of *P. aeruginosa* and explain its intrinsic capability to adapt so easily and perfectly to its environment.

The accessory genome on the other hand consists of fragments of DNA from a few hundreds to 200 kbp long. The minimum size of an accessory element was defined to be at least four contiguous ORFs. These so-called regions of genome plasticity (RGP) add about a further 10,000 or so genes to the whole genome. The actual number of accessory elements diverges in length and number which leads to the greatest variability between clones of the same species, but overall there are between 38 and 53 islands that have been identified between all strains. As an example, the laboratory strain PAO1 exhibits very little accessory element with only one RGP while another laboratory strain PA14 exhibits seven of them. The usual definition of an RGP refers to a horizontally acquired genetic element present in the chromosome of individual strains. Very often phages, transposons or insertion sequences make up these RGP and indicate how mobile these DNA elements are; they have been acquired then kept by the host strain. Many elements were eventually fixed by secondary mutation or deletion, but a few of them remain mobile and are passed on to a new host. The accessory genome is central to *P. aeruginosa* biology. The horizontal transfer of these elements implements the genome evolution of the species and more importantly it also confers specific phenotypes that are advantageous in selective conditions. Overall, the accessory genome encodes genes involved in the pathogenicity with lipopolysaccharide (LPS) O antigen, pyoverdine, pili and even flagellum. These are great assets for the persistence of the species by encoding virulence factors and resistance to antibiotics.

### 1.1.2. Infections scenario

#### 1.1.2.1. Acute infections

*P. aeruginosa* is an opportunistic pathogen, infections occur in hospitalised patients, immunocompromised hosts and patients with cystic fibrosis. Patients with a greater risk of infection include individuals with acquired immunodeficiency syndrome (AIDS) or cancer. Perhaps, one of the most evident infections following an immunocompromising event is in burns victims. Extensive breaches in the skin barrier open a door to opportunistic pathogens; and the environmental occurrence of *P. aeruginosa* makes it a prime candidate. Treating these infections is difficult and the mortality rate among the patients is likely to reach up to 40-50%. It is then urgent to prevent the spread of the pathogen from the environment or between patients. This example illustrates how most acute
infections are nosocomial in nature. Out of all hospital-acquired infections caused by Gram-negative bacilli, *P. aeruginosa* ranks as the second cause associated with nosocomial infection. However, among the anatomic sites of *P. aeruginosa* infection, the lung is associated with the highest mortality rate. Respiratory tract infections by *P. aeruginosa* remain the most frequent, they include hospital-acquired and ventilator-associated pneumonia. The evolution of such infections is characterized by haemorrhagic and necrotizing lung pathology with vasculitis (inflammation of the blood vessels) and microabcesses. The case of ulcerative keratitis of the cornea exemplifies the versatility of the microorganism. Without necessarily a point of entry into the body, these infections became common in users of extended-wear soft contact lenses. The intrinsic change of the tears composition as a consequence of use of the lenses promotes the adhesion and colonisation of the cornea by *P. aeruginosa*. The transenveloppe apparatus of the microorganism (flagellum and pili) play an important role for the adhesion to the host tissue. Following that, *P. aeruginosa* is then capable of producing several exotoxin proteins and proteases that are important virulence factors in keratitis. These products include exotoxin A, phospholipase C, elastase (LasB), alkaline protease and the LasA protease (endopeptidase); and eventually degrade the corneal tissue. Bloodstream infections by *P. aeruginosa* persist in patients admitted in intensive care units because of the multiple point of entries into the body. Surgical site infections fall into this category. However, the prognosis of *P. aeruginosa* bacteraemia remains poor with a mortality exceeding 38%. Finally, urinary tract infections (UTIs) remain the most common bacterial infections an individual will encounter. In this case, *P. aeruginosa* ranks as the third species to cause hospital-acquired catheter-associated UTIs.

### 1.1.2.2. Chronic infections

The chronic, long-term infections associated with *P. aeruginosa* generally occur in the lungs of patients who exhibit subjacent respiratory conditions. The most common of these conditions is cystic fibrosis that we will investigate in the next section. However, there are many cases of chronic infections in patients with bronchiectasis or chronic obstructive pulmonary disease (COPD). Bronchiectasis is defined by the presence of permanent and abnormal dilation of the bronchi. The inflammatory process commences in the small airways, releases mediators such as proteases which damage the large airways. With progression of the disease, the follicles enlarge in size and cause airflow obstruction to the small airways. COPD is characterised by poorly reversible airflow obstruction and an abnormal inflammatory response in the lungs. The airflow is limited due to an abnormal inflammatory response of the lungs to noxious chemicals (particles, gases, cigarette smoking). These three diseases show a similar pattern of accumulation of mucus in the lungs due to poor mucosal clearance and this results in the formation of a favourable environment for the establishment of *Pseudomonas*. Additionally, chronic infections with *P. aeruginosa* are associated with high-density bacterial assemblages, such as biofilms that favour the emergence of variants mostly arising through homologous recombination, recombinatorial DNA repair, and DNA mismatch repair deficiency.
1. Introduction

1.1.2.3. Cystic fibrosis (CF)

Cystic fibrosis is an autosomal recessive disease caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR). This gene is located on the human chromosome 7 and encodes a cyclic adenosine monophosphate-regulated chloride channel found at the apical surface of epithelial cells. So far, more than 500 mutations of the CFTR gene have been identified which showcases the complexity of the disease and its causes. In the lungs, inactive or inefficient functioning CFTR results in impaired chloride transport and enhanced sodium absorption across airway epithelial cells. This leads to a net increase in water absorption. Alternatively, the volume of the liquid that sits on the outer surface of the airway epithelium is reduced (due to hyper absorption by the cells) and the mucus eventually becomes thicker. Because of this thick mucus, ciliary clearance is impaired which provides a favourable environment for bacteria to proliferate. There is no longer a threat that would expulse them out of the lungs\textsuperscript{48–54}. The widespread presence of CFTR throughout the body helps to explain the actual pleiotropic effect of this genetic condition. Other organs such as the liver, pancreas, intestines and bones will eventually develop signs of complication\textsuperscript{55,56}. In terms of figures, in the European Union, CF affects on average 1 in 3,000 births\textsuperscript{57}. It is the most common genetic disease within the Caucasian population and it is the leading cause to premature respiratory failure. There are approximately 60,000 individuals currently living with CF in North America and Europe, 40\% of whom are adults\textsuperscript{58}.

Although the survival of CF patients has significantly improved in the past decades, CF remains a life-limiting condition. The median survival age ranges from 23 to 40 years old in Europe depending on the country studied. One aspect of the condition is its impact on the nutrition and growth rate of the individuals. Due to pancreatic malfunction, CF leads to malnutrition which would eventually impact the overall health of the patient. Compared to the early recognized cases of CF in the forties, medical care and antibiotic treatment improved to extend the survival as we know it nowadays\textsuperscript{59–62}. However, the consequence is now that morbidity and mortality is caused by recurrent and chronic bacterial infections. Colonization of the lungs by a polymicrobial population leads to complicated situations with advanced respiratory failure and infections\textsuperscript{63–65}.

The CF airways accumulate thick mucus which provides an environment highly susceptible to colonization by opportunistic pathogens, including bacteria, fungi and viruses. Bacterial colonization is probably promoted by the unique composition of the sputum. It is particularly rich in mucin, lipids,

\textbf{Figure 1: Prevalence of the bacterial population in cystic fibrosis lungs per age cohort. For the first 12 years of the patient life, the bacterial population in the CF airways is diverse with \textit{S. aureus} dominance. Then, it appears clearly that by the age of 18, \textit{P. aeruginosa} population becomes more prominent (Figure adapted from Paranjape et Mogayzel, 2014\textsuperscript{71}).}
proteins, amino acids, ions and DNA released by the host’s neutrophils. The life of a CF individual will inevitably be marked by the colonization of the lungs by successive bacterial species. In the early stages, as seen in figure 1 (adapted from Paranjape et Mogayzel71), Staphylococcus aureus and Haemophilus influenzae dominate. Then, by the age of 18, Pseudomonas aeruginosa takes over and typically becomes predominant. Other species remain steady, such as S. aureus, Burkholderia cepacia and Stenotrophomonas maltophilia. The complexity of the bacterial population encourages contact and interaction between bacterial cells themselves and between bacterial and the host’s cells via cell-to-cell signalling molecules, via notably quorum sensing in the former or TLR5-activated cascade by flagellin in the latter73-74.

1.1.3. Antibiotic resistance

Pseudomonas aeruginosa is intrinsically resistant to a variety of antimicrobials and consequently has joined the category of “superbugs” due to its capacity to develop antibiotic resistance. The bacterium constitutively demonstrates decreased susceptibility to most antibiotics including aminoglycosides, fluoroquinolones and β-lactams. This is because of low outer membrane permeability which prevents the penetration of these molecules into the cells due to inefficient porins as an uptake route for antibiotics. This then allows secondary adaptive mechanisms to work more efficiently, including increased efflux and enzymatic antibiotic modifications (e.g. β-lactamase). Like immunity in humans, antibiotic resistance is either intrinsic or acquired75-77.

1.1.3.1. Intrinsic resistance

The outer membrane permeability plays a major role in P. aeruginosa resistance. It constitutes a semi-permeable barrier for the uptake of small molecules. However, the uptake of hydrophilic molecules such as β-lactams is limited to the use of water-filled channels of porin proteins which prevent the movement of such molecules into the cell. This is a characteristic of all Gram-negative bacteria but this is particularly true for P. aeruginosa as it has up to a 100-fold lower outer membrane permeability compared with E. coli. The major porin in P. aeruginosa is OprF and accounts for up to 65% of nonspecific permeation through the outer membrane. Smaller channel proteins, OprD and OprB, mediate the movement of other antibiotic-like molecules. Despite the high number of copies of OprF, it remains an inefficient antibiotic uptake route because of either heterogeneous formation (only a very small proportion of porins form large channels) or inefficient architecture (the porins have been reported to be extremely narrow). That means that the remaining 35% of nonspecific permeation is guided by the other channels, which are intrinsically less efficient as porins. Specific channels are responsible for the uptake of carbapenem β-lactams, such as OprD. Loss of OprD is frequently associated with resistance to imipenem81,82. And more recently, OprD impaired strains of P. aeruginosa have been detected. A point mutation leading to an amino acid substitution at codon 170 gives the advantage of escaping the last line of antibiotic treatment83-84.
1. Introduction

*P. aeruginosa* is geared up with efflux pumps that will eventually extrude any molecule that has been taken up via the porins. The systems are composed of three protein components: an energy-dependent pump in the cytoplasmic membrane (Mex component), an outer membrane porin (Opr component) and a linker protein, also a Mex component, which couples the two previous together. These pumps belong to the resistance-nodulation-cell division (RND) superfamily of transporters. To date 12 out of 17 sequenced RND have been characterized in *P. aeruginosa*. Four of them have been studied in depth and these are MexAB-OprM, MexXY-OprM, MexCD-OprJ and MexEF-OprN. MexAB-OprM is responsible for extrusion of β-lactams, quinolones and a range of disinfectants. MexXY-OprM extrudes aminoglycosides while MexEF-OprN extrudes carbapenems and quinolones. Finally, MexCD-OprJ accommodates a variety of macrolides, tetracycline and some β-lactams.

*P. aeruginosa* typically carries two endogenous β-lactamases: cephalosporinase, AmpC and an oxacillinase, PoxB. AmpC is well known and is commonly linked to β-lactam resistance in clinical isolates. AmpC is not unique to *P. aeruginosa*, it is common to a large number of Gram-negative bacteria. It is inducible by a number of β-lactam antibiotics and thus contributes to the intrinsic resistance of *P. aeruginosa*. The right amount of antibiotic will trigger the up-regulation of the enzyme. Resistance to β-lactams is further improved by mutational derepression of *ampC* gene providing a quicker response in case of selective pressure from the environment. More β-lactamases have been acquired as the antibiotic treatment became harsher, as an example this includes the extended-spectrum β-lactamase (ESBL) enzymes capable of hydrolysing a broader range of β-lactams.

1.1.3.2. Acquired resistance

There are two types of acquired resistance that implement the intrinsic resistance of *P. aeruginosa* to become even less susceptible to antibiotics. These are horizontal transfer of genetic elements and mutational resistance. DNA elements such as plasmids, transposons, integrons or resistance islands can be acquired by conjugation, transformation or transduction. This promotes antibiotic resistance acquisition and multi-drug resistance as some plasmids can potentially carry multiple antibiotic cassettes. Such horizontal transfer has been mainly reported in cases of aminoglycoside and β-lactam resistance. Another mechanism for acquired resistance is mutational resistance. The normal mutation rate (10^6 to 10^9 per nucleotide per generation for individual antibiotics) can significantly increase under selective conditions. For example, the frequency of selection of ciprofloxacin-resistant mutants of *P. aeruginosa* increased 100-fold when the organism was grown as a biofilm in comparison with free-living cells. The hypermutators represent this extreme adaptation of *P. aeruginosa* that can acquire resistance to several different antibiotics. These strains contain mutations in genes involved in the repair of damaged DNA and they are frequently found in CF airways.
1.1.4. Virulence factors

The opportunistic profile of *P. aeruginosa* is recognisable for having an arsenal of virulence factors that help to facilitate successful infection and colonisation across a wide range of environments. The established versatility of its genome justifies how the bacterium is able to exploit any niche it colonizes. *P. aeruginosa* utilizes the secretion of compounds, toxins and proteins as virulence factors. This can occur via simple transport or using a complete machinery dedicated to the transport of extracellular virulence factors. Alternatively, the bacteria are capable of adapting their type of growth, from free cells to biofilms, introducing an alternative pathogenesis. The production and regulation of these virulence factors is under the control of quorum sensing (QS)\(^93\)–\(^95\). The production of extracellular compounds includes exotoxin A and exoenzyme S. Exotoxin A leads to the inhibition of protein biosynthesis and local tissue damage\(^96\),\(^97\). Exoenzyme S causes direct tissue destruction through its action on the actin cytoskeleton of the host’s cells\(^98\). The secretion of phospholipase C and rhamnolipid promotes a synergetic break down of the host’s lipids forming the surfactant and eventually aid the scavenging of nutrients for the bacteria’s good\(^99\). This is particularly clear in the CF airways in which the surfactant is cleaved, transported and metabolised by *P. aeruginosa*. This provides a sustainable environment for the colonisation of the bacteria and the worsening of the patient’s condition. The package of secreted proteases includes LasB, LasA, and alkaline protease. The ability of *P. aeruginosa* to destroy the protein elastin promotes loss of the epithelial barrier of the infected tissues and opens a wider door to invasive infections\(^100\).

*P. aeruginosa* possesses four dedicated secretion systems that are involved in the secretion of virulence factors, they are Type I, II, III and VI. There are two Type I secretion systems (T1SS), four Type II (T2SS), only one Type III system (T3SS) and three Type VI secretion systems (T6SS) that have been identified and located on the chromosome\(^101\),\(^102\). The Apr T1SS is known to secrete AprA, the alkaline protease along with AprX of unknown function\(^103\),\(^104\). The second, Has T1SS secretes haem scavenging proteins\(^105\). The two T2SS are the Xcp and Hxc systems. The Xcp system secretes a number of toxins and proteases that will eventually disrupt the host’s cells’ integrity (exotoxin A, lipases and elastases)\(^106\),\(^107\). The Hxc system on the other hand is dedicated to the secretion of the alkaline phosphatase LapA\(^108\). The T3SS in *P. aeruginosa* is of a more complex nature in terms of architecture with its needle like apparatus capable of injecting the products directly into the host’s cells. It secretes a number of enzymes known as ExoS, ExoT, ExoU and ExoY. These have a potent cytotoxic effect on the eukaryotic cells. Finally, the genome of *P. aeruginosa* contains 3 loci encoding T6SS components, called H1, H2 and H3. Also, much is now known about these effectors of this system. To date, they have been characterized to be either cell wall-degrading effectors (Tse1, Tae, VgrG these are amidase, glycosidase and muramidase respectively) or cell membrane-targeting (Tle, VasX are phospholipase or pore forming respectively)\(^109\),\(^110\). The regulation of the T3SS and T6SS is under the control of the two-component system RetS/LadS as well as cyclic-di-GMP (c-di-GMP). These two systems thus coordinate the two secretion systems to further establish *Pseudomonas* pathogenesis\(^111\).
1. Introduction

The motility of *Pseudomonas aeruginosa* can be an important factor in allowing first establishment then colonization of new niches. The microorganism can swim, swarm or twitch. Swimming is aided by the rotation of a single polar flagellum. It has been demonstrated that it also triggers phagocytosis by neutrophils, the first line of defence in human infections\(^\text{112}\). Swarming motility defines a multicellular phenomenon involving the coordinated and rapid movement of a bacterial population across semi-solid surfaces. This is usually linked to increased antibiotic resistance\(^\text{113}\). The twitching motility involves the extension and retraction of type IV pili and therefore plays a vital role in bacterial attachment and initial colonisation, particularly on mucosal cell surfaces\(^\text{114}\). These factors highlight the complexity of *Pseudomonas aeruginosa* pathogenesis, and it is particularly well equipped to take over a niche and lead to acute infections, and depending on the scenario, also a chronic infection.

1.2. A metabolic choice: 2C or not 2C?

1.2.1. TCA cycle

1.2.1.1. Context and generalities

Historically, the tricarboxylic acid cycle was identified in 1937 by Hans Krebs in animal tissues. Until about the late forties little was known about the cycle in microorganisms. The first steps of understanding of this cycle in unicellular organisms were in yeast although work slowly drifted to bacteria in the early fifties with the research conducted by Krebs, Kornberg and Monod among others\(^\text{115,116}\). The tricarboxylic acid (TCA) cycle was characterized in *Pseudomonas* in 1951\(^\text{117}\).

![TCA cycle schematic](image)

The TCA cycle is part of the wider metabolic network, and especially glucose oxidation. The complete assimilation of a molecule of glucose eventually produces 38 molecules of ATP. In bacteria, the complete oxidation of glucose involves three fundamental pathways. The first is the glycolytic or Embden-Meyerhof-Parnas (EMP) pathway, the second is the Krebs cycle (also called the citric acid cycle or TCA), and the third is the series of membrane-bound electron transport complexes coupled to oxidative phosphorylation. Respiration is the metabolic process by which pyruvate or acetyl-CoA is completely
oxidised to carbon dioxide\(^\text{118}\). This means that the TCA cycle plays a major role in the catabolism of organic fuel molecules, including glucose and other sugars, fatty acids, and some amino acids. In case of access to glucose in the environment, the glycolytic or EMP pathway first cleaves it into two molecules of pyruvate. However, there are two other glucose-catabolizing pathways found in bacteria: the oxidative pentose phosphate pathway and the Entner-Doudoroff pathway (EDP), which is almost exclusively found in obligate aerobic bacteria. Most *Pseudomonas* species (*aeruginosa* included) do not perform glycolysis via the EMP pathway because of the absence of key enzymes. As a result, sugars are metabolized in *P. aeruginosa* primarily by the EDP which yields two molecules of pyruvate out of one molecule of glucose\(^\text{119-121}\).

Before any of these molecules can enter the TCA cycle they first have to be broken down into acetyl-CoA by pyruvate dehydrogenase. Once fed into the cycle, acetyl-CoA is eventually converted into carbon dioxide and in doing so, yields reducing equivalents for electron transport. So, the TCA cycle consists of eight steps catalysed by eight different enzymes.

The cycle is initiated (1) when acetyl-CoA reacts with the compound oxaloacetate to form citrate and to release coenzyme A. Then, in a succession of reactions, (2) citrate is rearranged to form isocitrate; (3) isocitrate loses a molecule of carbon dioxide and then undergoes oxidation to form \(\alpha\)-ketoglutarate; (4) \(\alpha\)-ketoglutarate loses a molecule of carbon dioxide again and is oxidized to form succinyl-CoA; (5) succinyl-CoA is enzymatically converted to succinate; (6) succinate is oxidized to fumarate; (7) fumarate is hydrated to produce malate; and, to end the cycle, (8) malate is oxidized to oxaloacetate. Each complete turn of the cycle results in the regeneration of oxaloacetate and the formation of two molecules of carbon dioxide (see figure 2). One advantage of the TCA cycle is the production of energy. In step 5, one molecule of ATP is regenerated directly after dephosphorylation of GTP by a nucleoside diphosphate kinase. The production of molecules of \(\text{NAD}^+\), \(\text{NADP}^+\) and \(\text{FAD}\) indirectly lead to the formation of ATP during the final step of respiration which includes electron transport and oxidative phosphorylation\(^\text{122}\). The TCA cycle also has an important anabolic role in the cell. It works as the central metabolic “hub” of the cell for the biosynthesis of key cellular intermediates for anabolic reactions. For example, many amino acids are synthesized starting with the transamination of \(\alpha\)-ketoglutarate and oxaloacetate. Porphyrins and heme are synthetized from succinyl-CoA while citrate initiates the synthesis of fatty acids and sterols\(^\text{123}\).

**1.2.1.2. Regulation**

The expression of the TCA cycle enzymes is regulated at the transcriptional level. In *Escherichia coli*, three regulatory systems interact to balance the levels of these enzymes. Firstly, the substrate uptake relies on an intracellular sensor (usually a transcription factor) that will trigger a positive feedback loop. The accumulation of an intermediate signals nutrient availability, the information is transferred via a transcription factor to amplify the flux. A typical example of this system is the *Lac* operon and the output of the presence or absence of lactose as an inducer\(^\text{124,125}\). Secondly, the carbon catabolite repression (CCR) integrates several regulatory circuits which can
prioritize one substrate over another by sensing the presence of preferred carbon sources\textsuperscript{126–128}. It consequently reduces the uptake of alternative carbon sources. This system centres around the transcription factor Crp in \textit{E. coli}, or alternatively Crc in \textit{P. aeruginosa}. Crc (Catabolite Repression Control)\textsuperscript{129,130} is a post-transcriptional global regulator of carbon metabolism. Crc binds an A-rich region located close to the ribosomal binding site of the target mRNAs and inhibits translation \textit{initiation}\textsuperscript{131}. The mechanism of action is similar to that of RsmA, which means that the actual regulation of Crc itself is dependent on a small RNA (CrcZ) that binds the protein preventing it from binding to its target\textsuperscript{132}. Finally, the CbrAB two-component system was discovered in \textit{P. aeruginosa} as important regulatory elements for the expression of several catabolic pathways. \textit{cbrAB} mutants are unable to utilize a variety of organic compounds as sole carbon source\textsuperscript{133,134}. Such a two-component system consists of a sensor protein kinase (or transmitter) paired with a cognate response regulator. The environmental signals (carbon source available) trigger autophosphorylation of the transmitter that will subsequently transfer its phosphate to the acceptor domain of the regulator resulting in its \textit{activation}\textsuperscript{135,136}.

\subsection*{1.2.2. The glyoxylate shunt}

\subsection*{1.2.2.1. Generalities}

The glyoxylate shunt is present in a wide range of plants, yeasts and bacteria. It allows the net conversion of acetyl-CoA to malate and is thus defined as an anaplerotic or replenishing pathway. Acetyl-CoA plays a pivotal role in metabolism for being the end product of multiple catabolic routes (glycolysis, fatty acid β-oxidation, amino acid degradation). Moreover, in case of access to poor nutrient source such as acetate during infection scenarios, acetyl-CoA is the end product of the conversion catalysed by acetyl-CoA synthetase. This will feed into the first steps of the TCA cycle as seen previously. However, synthesis of sugar compounds from acetyl-CoA molecules alone is impossible, unless the organism develops a pathway termed the glyoxylate cycle, which is usually regarded as a bypass or shunt of the more generally occurring TCA cycle\textsuperscript{137}. Two enzymes are specific to this cycle,
isoctirate lyase (ICL) and malate synthase (MS). So, the glyoxylate cycle consists of six of the eight reactions of the TCA cycle but by-passes the two steps generating carbon dioxide (see figure 3). Thus, instead of oxidative decarboxylation in the TCA cycle, isoctirate is enzymatically cleaved to yield glyoxylate and succinate in a reaction catalysed by isoctirate lyase. This reaction is then followed by a condensation of glyoxylate and acetyl-CoA leading to the formation of malate by malate synthase\textsuperscript{138,139}. By saving two atoms of carbon, the glyoxylate shunt allows microorganisms to produce precursors necessary for gluconeogenesis and other biosynthetic processes while growing on limited metabolic inputs with a C\textsubscript{2} backbone such as acetate.

Interestingly, the glyoxylate shunt was first discovered by Kornberg and Krebs in 1957 in \textit{Pseudomonas KB1}\textsuperscript{140}. The understanding of the implications of producing metabolic intermediates from C\textsubscript{2} sources opened the door to the study of bacteria capable of surviving on poor nutrient sources\textsuperscript{141}. Glucose, as the preferred carbon source, is not systematically present in the environment and that is particularly the case during infections\textsuperscript{69,142}. Despite the fact that the shunt was first discovered in a pseudomonal species, most of the knowledge about it is based on \textit{E. coli}. In this bacterial species, the two enzymes (ICL and MS) that form the bypass are induced when the bacteria grow on substrates such as acetate and ethanol as sole carbon source. On the other hand, they are completely repressed when carbohydrates or succinate are present in the growth medium\textsuperscript{143,144}. In \textit{E. coli}, the structural genes of isoctirate lyase and malate synthase (\textit{aceA} and \textit{aceB}, respectively) constitute an operon and they are expressed on a polycistronic mRNA\textsuperscript{145}. A third enzyme accompanies the glyoxylate operon, this is isoctirate dehydrogenase kinase/phosphatase (AceK). It plays a major role in the regulation or partitioning of carbon flux between the TCA cycle and the glyoxylate shunt and this will be discussed in the next section. In terms of transcriptional regulation, in \textit{E. coli}, the \textit{aceBAK} operon is expressed from a single promoter after acetate induction and it is regulated by a repressor protein encoded by \textit{iclR}\textsuperscript{146,147}. IclR binds to a site which overlaps the -35 region of the \textit{aceBAK} promoter\textsuperscript{148}. Release of this repression upon adaptation to growth on acetate, fatty acids, or limiting glucose is presumably responsible for induction of the operon expression. The expression of the operon seems to respond to the general metabolic state of the cell, not necessarily to acetate only since growth on fatty acids triggers the expression without free acetate production\textsuperscript{148}. Another regulator, encoded by \textit{fadR}, participates in the negative transcriptional control of the operon along the regulation of fatty acid metabolism. FadR was initially identified for repressing the genes encoding the enzymes for fatty acid degradation\textsuperscript{149}. However, it is now known that it also activates the expression of \textit{iclR} by binding to its promoter region\textsuperscript{150}.

In \textit{Pseudomonas aeruginosa} however, the genes encoding the glyoxylate shunt enzymes do not follow that same organisation; \textit{aceA}, \textit{aceK} and \textit{glcB}, encoding ICL, AceK and MS respectively, are not in an operon and are separated on the genome. This highlights already the differences and the unique trait of \textit{P. aeruginosa} compared with the well-studied bacterial model. In terms of genetic regulation of the bypass, a study attempted to investigate the effect of IclR on \textit{aceA}\textsuperscript{151}. The disruption
of the ORF predicted to encode IclR did not affect aceA expression in PAO1. The same study, however, found that RpoN negatively regulates aceA expression in PAO1.

1.2.2.2. Branchpoint effect

The branchpoint effect is the partitioning between the TCA cycle and the glyoxylate shunt; here, isocitrate plays a pivotal role. It is the substrate for isocitrate lyase on one hand, and isocitrate dehydrogenase on the other. Thus, there is a competition between these two enzymes for the substrate. In E. coli and S. enterica serovar Typhimurium, isocitrate dehydrogenase is regulated by phosphorylation. The function of this reversible phosphorylation is to regulate the branchpoint between the glyoxylate bypass and the TCA cycle during steady-state growth on acetate or fatty acids. Phosphorylation of isocitrate dehydrogenase diverts some of the flux from the TCA cycle to the glyoxylate shunt. The immediate effect of phosphorylation is to render isocitrate dehydrogenase inactive, forcing isocitrate through the glyoxylate shunt. So, inhibition of isocitrate dehydrogenase decreases flux through the TCA cycle and thus forces isocitrate through the bypass\textsuperscript{152,153}. The resulting increase in the level of isocitrate further increases the velocity of isocitrate lyase, the first enzyme of the shunt. Partitioning of the flux can therefore be achieved through control of isocitrate dehydrogenase activity\textsuperscript{153}. The single increase of isocitrate levels which occurs during growth on acetate does not trigger isocitrate lyase activity\textsuperscript{154}. It has been demonstrated that mutant strains which are deficient in isocitrate dehydrogenase kinase activity failed to grow on acetate, suggesting the importance of the branchpoint regulation at that stage\textsuperscript{155}. It is known that during growth on acetate, about 75\% of the dehydrogenase enzyme is converted to its inactive phosphorylated form\textsuperscript{156}.

The phosphorylation cycle also controls the shunt during transition between carbon sources. For example, addition of a preferred carbon source such as glucose to a culture growing on acetate renders the bypass unnecessary. Under these conditions, the cells eventually shut down the bypass by dephosphorylating isocitrate dehydrogenase. Inhibition of the bypass occurs because the reactivation of isocitrate dehydrogenase draws isocitrate back to the TCA cycle. And because the Michaelis-Menten constant $K_m$ of isocitrate dehydrogenase ($K_m = 8$ $\mu$M) for isocitrate is significantly higher than that of isocitrate lyase ($K_m = 600$ $\mu$M), isocitrate is naturally redirected through the TCA cycle\textsuperscript{157}. This is also called the branchpoint effect. Because of the striking difference in the affinity of the two enzymes for the substrate, there is a

![Figure 4: Branchpoint model in E. coli. The affinity of ICD ($K_m = 8$ $\mu$M) for isocitrate outcompetes ICL ($K_m = 600$ $\mu$M) and isocitrate goes through the TCA cycle. Flux through the glyoxylate shunt is redirected when ICD is phosphorylated, thus inactive, by AceK.](image)
need to rewire the metabolism to push isocitrate through the bypass when necessary. The regulation of the glyoxylate bypass appears to be exquisitely sensitive to the metabolic state of the cell. Access to poor vs. rich carbon source and transition between the two if any, triggers a sensitive control at the branchpoint before partitioning of the carbon flux. For example, during growth on acetate, it appears that isocitrate and 3-phosphoglycerate (second last intermediate in the glycolysis before the production of PEP) participate in the control of the phosphorylation cycle. These intermediates probably act as indicators of the metabolic state of the cell and thus of the need to redirect isocitrate through the glyoxylate shunt. In case of depletion of these metabolites, phosphorylation of isocitrate dehydrogenase would increase, forcing isocitrate through the bypass. However, in case of transition to a preferred carbon source such as glucose, the levels of these two metabolites become normal but they do not prompt dephosphorylation of ICD. Dephosphorylation is more likely to be promoted by pyruvate as the levels of this metabolite rise significantly upon addition of glucose. The effect of metabolite levels, the sensitivity of the phosphorylation cycle and the branchpoint effect are combined to produce a system in which very subtle changes in metabolic signals have the potential for producing profound changes in the flux between the TCA cycle and the glyoxylate shunt. This knowledge of the branchpoint has been gathered after extensive study on E. coli drawing the known model as seen in figure 4.

Studies on other prokaryotes, such as Mycobacterium tuberculosis and Salmonella enterica, have confirmed the importance of the glyoxylate shunt. The bypass becomes functional when the bacterium has access to C2 substrate as the main carbon source. A study conducted on a ΔaceA mutant strain S. enterica serovar Typhimurium demonstrated that the bacteria were enabled to grow on acetate as sole carbon source. Alternatively, a similar approach in M. tuberculosis concluded that a Δicl1 Δicl2 double mutation eliminated growth on fatty acids. This establishes the importance of the glyoxylate shunt and more particularly the first enzyme of the bypass in scenarios of growth on limited carbon sources including acetate and fatty acids. Considering the abundance of these pathogens in human infections, the question of infection attenuation and link between metabolism and virulence ensues.

1.2.3. Metabolism and virulence are linked

Work performed in M. tuberculosis and P. aeruginosa demonstrated that isocitrate lyase activity influences virulence. The main studies performed on M. tuberculosis looked at the severity of mice’s lungs infection, primarily infected with M. tuberculosis wild-type. The infective strain was either a single mutant of isocitrate lyase (Δicl1 or Δicl2) or a double mutant (Δicl1 Δicl2). Mycobacterium tuberculosis is equipped with two isocitrate lyases. The loss of either one of the isocitrate lyase had no dramatic effect on the growth of M. tuberculosis during the first two weeks of infection. Modest reduction of the bacterial loads in infected lungs was seen during the time course of the infection between the second and sixteenth week. The double mutant however did not show any lung pathology and the mice remained healthy until the experiment was terminated. By contrast,
mice infected with the wild-type strain showed all signs of acute infection including splenomegaly along lung inflammatory lesions and lung enlargement\textsuperscript{161, 162}.

A similar study was conducted in \textit{Pseudomonas aeruginosa} with a \textit{ΔaceA} mutant of PAO1 in a rat chronic lung infection model. Disruption of \textit{aceA} in PAO1 led to a four-fold reduction in histopathology in rat lungs. This is a clear indication that isocitrate lyase is indeed required for optimal virulence\textsuperscript{164}. More studies investigated the glyoxylate shunt as a novel drug-target. One study performed on \textit{P. aeruginosa} investigated new antimicrobial therapeutics that would target the glyoxylate shunt only. By cutting down the possibility for the bacteria to use the bypass, they would no longer survive on acetate and/or fatty acids and eventually be cleared of the infection site. A screening of nearly 150,000 molecules identified compounds that were able to inhibit the growth of \textit{P. aeruginosa} in minimal media supplemented with acetate. Eight compounds eventually showed significant inhibition of both isocitrate lyase and malate synthase opening a door to further characterization of drugs targeting this pathway\textsuperscript{165}.

1.3. The enzymology at the branchpoint

1.3.1. The isocitrate dehydrogenases

In the TCA cycle, isocitrate dehydrogenase catalyzes the decarboxylation of isocitrate into \textit{α}-keto glutarate. This enzyme has been extensively studied in \textit{E. coli} and isocitrate dehydrogenase activity has been reported in multiple microorganisms: \textit{Corynebacterium glutamicum}\textsuperscript{166}, \textit{Mycobacterium tuberculosis}\textsuperscript{167}, \textit{Vibrio parahaemolyticus}\textsuperscript{168}, \textit{Rhodobacter vanniieli}\textsuperscript{169}, \textit{Azotobacter vinelandii}\textsuperscript{170}, \textit{Colwellia psychrerythraea}\textsuperscript{171} and another species of \textit{Pseudomonas}\textsuperscript{172}. Some of these species exhibit only one form of the enzyme (ICD), whereas others seem to possess a different isozyme (IDH). Some species contain both ICD and IDH such as \textit{P. aeruginosa}, \textit{Colwellia maris}, \textit{Vibrio} sp. and \textit{M. tuberculosis}.

The ICD enzyme is dimeric, consisting of two subunits with an average molecular mass of 40–45 kDa each. In contrast, all IDH enzymes characterized to date are monomeric, with a molecular mass of 80–100 kDa.
In both *C. psychrerythraea* and *C. maris*, optimum temperature for activity and thermostability of ICD indicate that it is a suitable enzyme in mesophilic organisms. IDH seems extremely labile above 25°C, with an optimum temperature for activity of 20°C, revealing that it is cold-adapted in *C. maris*\(^{71}\). However, both isozymes present in *Pseudomonas psychrophila* have been studied and are adapted for a mesophilic organism\(^{172}\). Furthermore, the two encoding genes in *Colwellia* (previously described as *Vibrio* sp. strain ABE-1) are located in tandem, with the IDH gene followed by the ICD gene, although the expression of those two genes is independently regulated\(^{171,173}\). Interestingly, in *P. psychrophila*, the two genes are divergently transcribed\(^{172}\), and the *P. aeruginosa* *icd* and *idh* genes are similarly organized (PA2623 and PA2624 respectively). Nevertheless, the bacterial monomeric and dimeric enzymes catalyse the same reaction, despite having little homology in amino acid sequence\(^{174,175}\).

ICD and IDH belong to a large, ubiquitous, and very ancient protein family whose members play central roles in energy metabolism\(^{176}\), amino acid biosynthesis\(^{177,178}\) and vitamin production\(^{179-181}\). Both are NADP\(^{+}\)-dependent enzymes which catalyse the oxidative decarboxylation of D-isocitrate to form α-ketoglutarate, CO\(_2\) and NADPH. In *C. glutamicum*, the K\(_{m}\) values allow to conclude that the monomeric IDH displays an about 3-fold lower K\(_{m}\) for both substrate and co-factor, and its overall catalytic efficiency at 21°C, calculated is 10-fold lower than the dimeric enzyme\(^{182}\).

The crystal structures of ICD from *E. coli*\(^{183}\) and *M. tuberculosis*\(^{184}\) have been solved (PDB entries 4AJA and 4HCX respectively) as has that of IDH from *A. vinelandii*\(^{185,186}\) (PDB entry 1J1W). ICD consists of three domains: a large domain (domain A), a clasp-like domain (domain B) and a small domain (domain C)\(^{183}\). Although the amino acid sequences of both isozymes are different, the folding topology of IDH is related to that of ICD. Indeed, the monomeric IDH is created by fusing one subunit of ICD with the smaller domain of the second subunit, albeit with some differences. The structure essentially consists of all three domains from the first subunit of ICD and domains B'-C' from the second subunit (see figure 5). As stated above, phosphorylation of ICD inactivates the enzyme. This occurs on Ser113 of the *E. coli* enzyme\(^{183}\) but nothing has been concluded concerning the *M. tuberculosis* version as there is no AceK in this microorganism. Amino acid sequences demonstrate that Ser115 from *P. aeruginosa* aligns with Ser113 from *E. coli*. This is of particular interest for further investigation.

**1.3.2. The isocitrate dehydrogenase kinase/phosphatase**

In *E. coli*, isocitrate dehydrogenase (ICD) is regulated by a bifunctional protein, ICD kinase/phosphatase (also named AceK). Interestingly, it possesses the two opposing activities within one protein, and specifically recognizes only intact ICD\(^{158,187}\). The transfer of the phosphate from ATP to Ser113 causes complete inhibition of ICD activity by blocking isocitrate from binding through a combination of electrostatic and steric effects. Interestingly, NADP\(^{+}\) and NADPH bind to both the phosphorylated and dephosphorylated forms of ICD\(^{188}\). The phosphatase activity has an absolute requirement for ATP or ADP; and this protein has an intrinsic ATPase activity as well\(^{189}\).
From structural and computational data, isocitrate dehydrogenase kinase/phosphatase appears to be a monomer of about 65-68 kDa. AceK structure has been extensively studied in *E. coli* and a regulatory mechanism has been proposed based on those data. The protein structure contains two functional domains. The C-terminal domain resembles eukaryotic protein kinase and represents the kinase/phosphatase domain. The N-terminal or regulatory domain has allosteric binding pockets involved in the regulation of the function of the catalytic domain. ATP is found in a pocket between the two domains. AceK displays highly specific binding to ICD. However, considering that the phosphorylation site lies buried deep inside the ICD structure, Zheng et Jia have proposed that ICD forms a tight homodimer with the two active sites on opposite sides. In the open conformation, the phosphorylation loop undergoes a significant conformational change with the residue Ser113 moving and rotating. As a result, the serine changes from a buried, inwards-facing orientation to an exposed, outwards-facing position, rendering Ser113 accessible to ICD kinase/phosphatase. The ICD phosphorylation cycle is controlled by a variety of metabolites; AMP, pyruvate and 3-phosphoglycerate activate ICD phosphatase but inhibit ICD kinase. Other compounds such as NADPH inhibit both activities. Isocitrate seems to inhibit ICD kinase but its effect on ICD phosphatase remains controversial.

1.3.3. The isocitrate lyase

In *E. coli*, isocitrate lyase has a much lower affinity for isocitrate than ICD. Under growth conditions supplemented with glucose, ICL cannot compete with isocitrate dehydrogenase. However, flux regulation can be simply achieved by changing the activity of either one of the two enzymes, ICL or ICD. In *E. coli*, aceA gene expression is increased when the bacterium grows on acetate as a sole carbon source. In that case, AceK inactivates ICD by phosphorylation, and so pushes isocitrate through the glyoxylate shunt. However, when glucose is added to the medium, ICD is reversibly dephosphorylated resulting in the restoration of its activity. After the branchpoint, ICL represents the first enzyme that catalyses the cleavage of isocitrate into succinate and glyoxylate. The enzyme simply binds the substrate, there is no co-factor; however magnesium seems indispensable to coordinate the binding of the substrate and for optimal activity.

The protein structure of ICL from two species of prokaryote has been resolved: *E. coli* (PDB entry 1IGW) and *M. tuberculosis* (PDB entry 1F61) and one species of eukaryote, *Aspergillus nidulans* (PDB entry 1DQU). Those studies have established that the quaternary structure of ICL is a tetramer. However, analysis of the amino acid sequences reveals that the subunit molecular mass varies between 48 kDa for the prokaryotic enzymes and 67 kDa for the eukaryotic version. The *P. aeruginosa* enzyme has an intermediate mass of 59 kDa and 97 extra residues compared with the *E. coli* enzyme. As noted by Britton et al., in the eukaryotic enzyme, the extra 104 residues form an additional head domain, so this could also be the case in the *P. aeruginosa* ICL. In *P. aeruginosa*, there is a single copy of ICL-encoding gene (PA2634, *aceA*) and the protein sequence shares just 30% similarity with the species presented above aka *E. coli*.
1. Introduction

*Isocitrate lyase* became of major interest in *Mycobacterium tuberculosis* as a potential drug target; and is well-studied in *E. coli*. The link between virulence and metabolism established in mice models with lungs infection attracted research in targeting the glyoxylate shunt to “starve” the mycobacteria\(^{198-200}\). In *Pseudomonas aeruginosa*, isocitrate lyase mediates survival on poor nutrient sources (as described above) and is linked with T3SS expression. Indeed, a study conducted on PAO1 under oxygen-limited conditions, concluded that in a \(\Delta aceA\) mutant the expression of PcrV (T3SS core component), PopN (T3SS translocation regulator), ExoS (T3SS effector protein) and ExsD (T3SS regulator) was greatly reduced\(^{201}\). However, very little is known about the enzymology and structure of ICL in *Pseudomonas aeruginosa*. This work attempts to investigate for the first time this glyoxylate bypass that was first discovered in the *Pseudomonas* genus.

1.4. Aims of this PhD project

The TCA-glyoxylate branchpoint has been well-characterized in *E. coli*, the architecture remains simple. There are only two competitors for *isocitrate*, ICD in the TCA cycle and ICL in the glyoxylate shunt. Upon growth in poor nutrient sources, *isocitrate* is redirected towards ICL after ICD has been phosphorylated thus inactivated. ICL and ICD compete for the same substrate, but the phosphorylation-mediated inhibition of ICD rewires the carbon flux through the glyoxylate shunt. So, there are two branches, TCA cycle and glyoxylate shunt, represented by two competitors ICD and ICL; and depending on the growth conditions, one branch is “on” while the other is “off”. However, in *P. aeruginosa*, that architecture is different. Three enzymes have to compete for the same substrate, ICD, IDH and ICL. The rerouting of carbon flux between the TCA cycle and the glyoxylate shunt is certainly different and more importantly flux control must be exerted differently. Furthermore, despite the fact that the glyoxylate shunt has first been discovered in a *Pseudomonas* KB1, very little is known about the multiple enzymes found at the branchpoint. The aim of this project was to determine the crystal structure, the function and the regulation of the TCA-glyoxylate branchpoint enzymes from *P. aeruginosa* in considerable detail. To investigate this further, I pursued the following objectives:

- To resolve the X-ray crystal structure of ICD, IDH and ICL. The first three have been successful and have been deposited in the PDB.
- To study the oligomeric status of the enzymes and capture the differences with other known enzymes in bacterial species.
- To characterize the kinetic properties of the enzymes.
- To examine the impact of potential regulatory metabolites on the activity of each enzyme.
2. Materials and methods

2.1. General microbiological procedures

2.1.1. Strains

The strains used in this study are listed in table 1. For long-term storage, cells were kept in 25% glycerol solution at -80°C. For short-term storage, *E. coli* colonies on agar plates were stored at 4°C, while *P. aeruginosa* colonies were kept on the bench at room temperature.

<table>
<thead>
<tr>
<th>Table 1: Bacterial strains</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genotype/Phenotype</strong></td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>DH5α</td>
</tr>
<tr>
<td>[DH5α]:His:icd</td>
</tr>
<tr>
<td>[DH5α]:His:idh</td>
</tr>
<tr>
<td>[DH5α]:His:aceA</td>
</tr>
<tr>
<td>[DH5α]:His:aceK</td>
</tr>
<tr>
<td>[DH5α]:His-MBP:icd</td>
</tr>
<tr>
<td>[DH5α]:His-MBP:idh</td>
</tr>
<tr>
<td>[DH5α]:His-MBP:aceA</td>
</tr>
<tr>
<td>[DH5α]:His-MBP:aceK</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em></td>
</tr>
<tr>
<td>PAO1</td>
</tr>
<tr>
<td>PAO1::Picd</td>
</tr>
<tr>
<td>PAO1::Pidh</td>
</tr>
<tr>
<td>PAO1::Pacea</td>
</tr>
<tr>
<td>PAO1::Pacek</td>
</tr>
</tbody>
</table>

2.1.2. Media components and supplements

For standard bacterial growth, Luria Bertani broth (LB) was used. The LB broth preparation was sterilised by autoclaving for 20 min at 121°C. For growth study, MOPS minimal medium was used. The non-sterilisable elements of the MOPS medium were sterilised using 0.22 μm filter.

For growth curves, MOPS minimal medium supplemented with carbon sources was used. See tables 2-5 for further details of media components and supplements.

<table>
<thead>
<tr>
<th>Table 2: Growth media</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Growth media</strong></td>
</tr>
<tr>
<td>Luria broth (LB)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
2. Materials and methods

<table>
<thead>
<tr>
<th>Component</th>
<th>10X MOPS medium 1 M</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon source</td>
<td>per Litre</td>
<td></td>
</tr>
<tr>
<td>Acetate 1 M</td>
<td>40 mL</td>
<td>40 mM</td>
</tr>
<tr>
<td>Glucose 1 M</td>
<td>20 mL</td>
<td>20 mM</td>
</tr>
<tr>
<td>K₂SO₄ 27.5 mM</td>
<td>10.5 mL</td>
<td>0.29 mM</td>
</tr>
<tr>
<td>KH₂PO₄ 172.8 mM</td>
<td>7.6 mL</td>
<td>1.32 mM</td>
</tr>
</tbody>
</table>

Table 4: 10X MOPS medium

<table>
<thead>
<tr>
<th>Component</th>
<th>per Litre</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOPS pH 7.5 1M</td>
<td>400 mL</td>
<td>400 mM</td>
</tr>
<tr>
<td>Tricine pH 7.5 1M</td>
<td>40 mL</td>
<td>40 mM</td>
</tr>
<tr>
<td>FeSO₄ 18.4 mM</td>
<td>9.7 mL</td>
<td>0.18 mM</td>
</tr>
<tr>
<td>NH₄Cl 1.9 M</td>
<td>50 mL</td>
<td>95 mM</td>
</tr>
<tr>
<td>CaCl₂ 53 mM</td>
<td>100 mL</td>
<td>5.3 μM</td>
</tr>
<tr>
<td>MgCl₂ 6 H₂O 512 mM</td>
<td>10 mL</td>
<td>5.1 μM</td>
</tr>
<tr>
<td>NaCl 5 M</td>
<td>100 mL</td>
<td>500 mM</td>
</tr>
<tr>
<td>Micronutrient stock 100 X</td>
<td>10 mL</td>
<td>1X</td>
</tr>
</tbody>
</table>

Table 5: 100X micronutrient stock

<table>
<thead>
<tr>
<th>Component</th>
<th>mg/100ml</th>
<th>Stock concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ammonium molybdate tetrahydrate</td>
<td>0.3</td>
<td>3 μM</td>
</tr>
<tr>
<td>Boric acid</td>
<td>2.4</td>
<td>400 μM</td>
</tr>
<tr>
<td>Cobalt chloride</td>
<td>0.7</td>
<td>30 μM</td>
</tr>
<tr>
<td>Cupric sulfate</td>
<td>0.3</td>
<td>10 μM</td>
</tr>
<tr>
<td>Manganese sulfate</td>
<td>1.6</td>
<td>80 μM</td>
</tr>
<tr>
<td>Zinc sulfate</td>
<td>0.3</td>
<td>10 μM</td>
</tr>
</tbody>
</table>

2.1.3. Growth conditions

All growth experiments were conducted at 37°C. For growth on solid media, 1.5% w/v agar was added to the appropriate liquid medium. The plates were usually incubated for 18-24 h before any further experiment. For small-scale liquid culture or overnight culture, a single colony was inoculated in 10 mL LB into a 30 mL screw-cap plastic tube. The *E. coli* or PAO1 strain was incubated at 37°C on a rotary drum at 120 rpm for 18-24 h.
2. Materials and methods

2.1.4. Growing and harvesting bacterial samples

For medium-scale liquid culture (growth curve), a small aliquot of overnight culture was used to inoculate 50 mL of liquid media in a baffled 250 mL Erlenmeyer flask to adjust to a starting OD$_{600}$ = 0.05. Antibiotics and other supplements, if required, were added at the start. The flasks were shaken at 250 rpm and 37°C in a water bath for the course of the experiments. Growth was monitored by optical density (OD) measurement at 600 nm wavelength using an Eppendorf BioSpectrometer.

For large-scale liquid culture (protein overexpression), $9.6 \times 10^8$ cells/mL of overnight *E. coli* culture was used to inoculate 2 to 6 L of LB media supplemented with carbenicillin in baffled 2 L Erlenmeyer flasks. The flasks were incubated in an orbital shaker at 250 rpm and 37°C. After reaching OD$_{600}$ ~ 0.5/0.6, 1 mM (His$_6$-tagged protein) or 0.3 mM (His$_6$-MBP-tagged protein) of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the bacterial culture to induce protein expression.

Harvesting of bacterial cells was carried out by centrifugation of liquid culture using a Beckman Coulter Avanti J26-XP1 centrifuge (7,455 g, 20 min, 4°C). For smaller-scale samples (i.e. 10 or 50 ml), sedimentation was achieved in 30 ml screw-cap plastic tubes or 50 ml Falcon tubes in an Eppendorf 5810R centrifuge (3,214 g, 20 min, 4°C).

2.2. DNA manipulation

2.2.1. DNA extraction, purification, sequencing and storage

Genomic DNA was extracted from bacterial cells using a GeneJET genomic DNA purification kit (Thermo Fisher) following the manufacturer’s guidelines. The concentration of DNA samples was measured using a Nanodrop (ND-1000, Thermo Scientific) to measure absorbance at A$_{260}$/A$_{280}$. PCR products were purified using a GeneJET PCR purification kit (Thermo Fisher), and when necessary, migrated DNA fragments were purified after agarose gel electrophoresis with a GeneJET gel extraction kit. All DNA samples were stored at -20°C for further analysis.

DNA sequencing was performed by the GATC Biotech company after preparing the samples according to their guidelines. The data were analysed using either ApE or SnapGene software.

2.2.2. Polymerase Chain Reaction (PCR)

PCR reaction mix components were usually combined in final concentrations specified by each manufacturer’s guidelines. In general, the reaction mix (50 µL) contained 1 µL of template DNA, 1 µL of each forward and reverse primers (stock 100 µM, Sigma-Aldrich), 1 µL of dNTP mix (stock 10 µM, Bioline), 5 µL of dimethyl sulfoxide (DMSO, Sigma-Aldrich), 10 µL of Phusion buffer HF and 0.5 µL of Phusion polymerase enzyme (either NEB or produced in the laboratory).

DMSO acts as a polar aprotic solvent; it inhibits secondary structure formation in DNA and decreases the melting temperature of complementary DNA. DMSO was usually added at a final concentration of 5-10% v/v in all PCRs due to the overall GC-rich content of the DNA templates.
and primers. The template DNA used was either PAO1 genomic DNA, plasmid DNA, or DNA directly from a bacterial colony in “colony PCR”. Colony PCR followed the same program as classic Phusion PCR; it simply required an additional step of 5 min at 95°C in order to lyse the cells and release the DNA.

PCR primer design was done manually, based on the available sequence of the four genes (icd, idh, aceA and aceK), and using NEB Tm calculator to anticipate the melting temperature. Table 6 below lists the PCR primers and their characteristics. The restriction enzymes sequences (bold) were incorporated within the primers for cloning purposes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Oligo name</th>
<th>Sequence (5’-3’)</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>icd</td>
<td>ICD-fw</td>
<td>CCCCCCCGATCCATGGGATACCACCAAAAAAGATCCA</td>
<td>Cloning for protein overexpression from pMAL-c2x</td>
</tr>
<tr>
<td></td>
<td>ICD-rv</td>
<td>CCCCCCCGATCTCTACATCTTGGCAATCATCG</td>
<td></td>
</tr>
<tr>
<td>idh</td>
<td>IDH-rv</td>
<td>CCCCCCCGATCTGTTTCTGCTTCGATCGGCTGTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICL-fw</td>
<td>CCCCCCCGATCCATGTCGGCATATCAAGAAGCA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICL-rv</td>
<td>CCCCCCCGATTTCAGTGGAACGTGGTCATGG</td>
<td></td>
</tr>
<tr>
<td>aceA</td>
<td>AceK-fw</td>
<td>CCCCCCCGATCCATGGTCCAGAGGCGGCGGCAGCC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AceK-rv</td>
<td>CCCCCCCGATCTTCACTCCGAGCAACTGCTCGG</td>
<td></td>
</tr>
<tr>
<td>aceK</td>
<td>Picd-fw</td>
<td>GATCGAATTCTTGTCGTCGCTCAGTCCCG</td>
<td>Cloning for gene expression study with LacZ fusion</td>
</tr>
<tr>
<td></td>
<td>Picd-rv</td>
<td>GATCGGATCCCCGCGGCTGCTCGTAAAAACCC</td>
<td></td>
</tr>
<tr>
<td>Pidh</td>
<td>Pidh-fw</td>
<td>CCCCCCCGATCTCATGACCGGCGCTGATCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pidh-rv</td>
<td>CCCCCCCGATCCCTACGGTAACTCTTTTTT</td>
<td></td>
</tr>
<tr>
<td>Pacea</td>
<td>Pacea-fw</td>
<td>CCCCCCCGATTTCTCAGGACCAGACAGGCCAGCATC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pacea-rv</td>
<td>CCCCCCCGATCCGGCGGAGGCTATCCAGGT</td>
<td></td>
</tr>
<tr>
<td>Pacek</td>
<td>Pacek-fw</td>
<td>CCCCCCCGATCCGCTGGCGGCGCTTCTTGTGGTT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pacek-rv</td>
<td>CCCCCCCGATCCGCTGGCGGCGCTTCTTGTGGTT</td>
<td></td>
</tr>
</tbody>
</table>

PCR cycle conditions varied depending on the length of the fragment being amplified, the melting temperature (Tm) of the primers and the type of template DNA. A Veriti 96-well thermocycler was used for all reactions. Table 7 below shows the program used for most reactions.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Denaturation</td>
<td>95°C, 10min</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Denaturation</td>
<td>95°C, 1 min</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Annealing</td>
<td>62°C, 30 sec</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elongation</td>
<td>72°C, 1 min</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Elongation</td>
<td>72°C, 8 min</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Hold</td>
<td>4°C, indefinitely</td>
<td></td>
</tr>
</tbody>
</table>
2.2.3. Agarose gel electrophoresis

Linear DNA was separated by electrophoresis through agarose gels. DNA samples were mixed with BioLine loading dye and loaded onto a 1% agarose gel (1% w/v agarose dissolved in 1X Tris-Acetate-EDTA (TAE) buffer). Before casting, ethidium bromide was added to the molten agarose at a final concentration of 0.4 µg/mL.

Migration was performed using a Mini-Sub® Cell GT Cell (Bio-Rad) at 80 V for about one hour in 1X TAE buffer. Then, the DNA fragments stained with ethidium bromide were visualised using a UV transilluminator at 365 nm. HyperLadder I (200 – 10,000 bp, BioLine) was used as a size indicator.

2.3. Strains preparation

2.3.1. Plasmid construction

The plasmids used in this study are summarized in the Table 8 below. The fidelity of the cloned PCR-amplified products was confirmed by sequencing. All the His6-tagged protein encoding genes (icd, idh, aceA and aceK) were the results of the work of a PartII student in the laboratory. He cloned the four subsequent genes into pQE80. The same icd, idh, aceA and aceK genes were later cloned into a modified version of pMAL-c2x. Vector pMAL-c2x encodes an MBP-tag, the modified version of it produced in the laboratory encodes a His6-tag located at the N-terminus of the MBP-tag. The modification work of the pMAL-c2x vector, renamed His-pMAL-c2x, was performed by Dr. Martin Welch who added the corresponding genetic element into the original vector using PCR amplification. After modification of the vector, icd, idh, aceA and aceK genes were cloned into His-pMAL-c2x. The purpose of this new construct was to allow the production of a native protein, with no tag attached to it as His-pMAL-c2x carries a Factor Xa cleavage site.

Finally, the promoter regions of all genes were cloned into pLP170 vector (non-commercial vector, first developed by the Iglewski group) to study the gene expression in different carbon sources (glucose and acetate supplemented minimal medium).

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pQE80</td>
<td>Bacterial expression, CbR</td>
<td>Qiagen</td>
</tr>
<tr>
<td>His-pMAL-c2x</td>
<td>Bacterial expression, CbR</td>
<td>This study (originally from NEB)</td>
</tr>
<tr>
<td>pLP170</td>
<td>lacZ transcriptional fusion, CbR</td>
<td>Preston et al., 1997</td>
</tr>
</tbody>
</table>

2.3.2. Cloning

His6-MBP-tagged constructs were cloned into His-pMAL-c2x vector. After PCR amplification of genomic DNA, the icd, idh, aceA and aceK amplicons and the plasmid were individually purified.
and digested by BamHI and HindIII for 2h at 37°C. Both insert and vector were then incubated with T4 DNA ligase (NEB) for 1h on ice then 1h at room temperature. Promoter regions of all four genes were cloned into the pLP170 vector after amplification and digestion with EcoRI and BamHI for 2h at 37°C.

2.3.3. Competent cells

E. coli DH5α electrocompeotent cells were used to incorporate the four His-pMAL-c2x constructs (icd, idh, aceA and aceK) into the cells. After overnight growth and subculture, the cells were pelleted, washed three times with ice-cold water and suspended in 10% v/v ice-cold glycerol for electroporation.

Similarly, electrocompeotent cells of P. aeruginosa PAO1, provided by Dr. Stephen Dolan, were used to introduce into the cells the four pLP170 constructs carrying the promoter regions of either icd, idh, aceA and aceK. Prior to transformation, the cells were grown overnight and subcultured into 10 mL of LB. After growth, the cells were pelleted, washed three times with PBS and resuspended in 10% v/v glycerol before electroporation.

2.3.4. Transformation by electroporation

Introduction of plasmid DNA into E. coli DH5α or P. aeruginosa PAO1 was performed by electroporation. A volume of 100 µL of competent cells was electroporated at 2.5 kV with 2 µL of ligated product. Four E. coli DH5α samples were transformed with the four His-pMAL-c2x constructs carrying either icd, idh, aceA or aceK genes. Alternatively, four P. aeruginosa PAO1 samples were transformed with the four pLP170 constructs, Picd, Pidh, PaceA or PaceK respectively. Immediately after electroporation, 900 µL of LB were added to the cells, which were then incubated at 20°C for 1h on a rotating drum. Finally, 100 µL were plated on carbenicillin-supplemented LBA for overnight growth and selection.

2.4. Gene expression analysis

Four gene promoter lacZ fusions were generated using the promoter-less lacZ plasmid pLP170, as described in Table 6. The transcriptional activity of the promoters that were fused to a lacZ gene can be measured by β-galactosidase activity.

2.4.1. Growth curves

The four P. aeruginosa wild-type PAO1 strains were transformed as described in section 2.3.2. with pLP170 constructs carrying the promoter of each gene upstream of lacZ. Ahead of the experiment, overnight cultures of each strain were prepared to reach stationary phase in order to inoculate 50 mL flasks starting with a normalised OD<sub>600</sub> of 0.05.

The cells were grown in MOPS medium supplemented with either acetate or glucose as single carbon source. Every hour (up to 10 h), 1 mL of culture was harvested for growth analysis. The
growth rate was monitored by optical density of each sample was measured at 600 nm using an Eppendorf BioSpectrometer. Finally, 100 µL of each sample was saved in a 96-well plate to carry on the β-galactosidase analysis. All experiments were performed in triplicate.

2.4.2. β-Galactosidase assay

The β-galactosidase assay (MUG) measures the cleavage by β-galactosidase of 4-methylumbelliferyl-β-D-galactopyranoside (4-MUG) into fluorescent 4-methylumbelliferone. The harvested cells previously put aside were diluted by a factor of 1/100, following which 10 µL were then distributed into another 96-well plate. To that, 100 µL of lysozyme and 4-MUG mixture (see Table 9) was added immediately before measuring the production of 4-methylumbelliferone.

<table>
<thead>
<tr>
<th>Components</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-Mu-β-D-galactoside</td>
<td>1X</td>
</tr>
<tr>
<td>Lysozyme in PBS buffer</td>
<td>18.1 mg/mL</td>
</tr>
</tbody>
</table>

Monitoring of the release of 4-methylumbelliferone (4-MU) by β-galactosidase was done using a Gemini XPS microplate reader operated by the SoftMax Pro software. The plate was incubated at 37°C for the duration of the measurements. The excitation wavelength is 360 nm while the emission wavelength is 450 nm, eight readings were performed per well for the same duration of the measurements. Analysis of the results and plotting were performed using Graphpad Prism 6 software.

2.5. Protein overexpression

2.5.1. His6-tagged proteins

The transformed strains were grown at 37°C in 1 L LB broth containing carbenicillin (50 µg/mL). After having reached an OD₆₀₀ ~ 0.5/0.6, the protein expression was induced by addition of IPTG, to 1 mM final concentration. After 2 h of induction, the cells were pelleted in a Sigma® 4-16K centrifuge at 5,170 g for 10 min. The pellets were resuspended on ice in 60 mL of lysis buffer (Table 10) containing an EDTA-free protease inhibitor cocktail from Roche®. From there, the samples were sonicated 8 rounds 30 sec on, 30 sec off, on full power. The cell debris was sedimented by centrifugation using a Beckman Coulter Avanti J26-XP1 centrifuge (14,636 g, 20 min, 4°C). Supernatants were applied to an Ni-NTA column (Qiagen) (2-ml packed resin bed volume), equilibrated with lysis buffer. After passing the supernatant at a 1 mL/min pace through the column, the column was washed with 500 mL of lysis buffer to eliminate contaminants. The sample was eventually eluted in 10 mL of elution buffer. Following that, the sample was dialyzed with at least four batches of dialysis buffer (see Table 10 for buffer composition). Finally, the samples were concentrated using Vivaspin columns, MWCO 10,000 Da (Sartorius) in an Eppendorf 5810R centrifuge (3,220 g, 20 min, 4°C).
Table 10: Buffers for His$_6$-tagged proteins expression

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Components</th>
<th>per Litre</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysis</td>
<td>NaCl 5 M</td>
<td>60 mL</td>
<td>300 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.5 1 M</td>
<td>50 mL</td>
<td>50 mM</td>
</tr>
<tr>
<td></td>
<td>Imidazole 3 M</td>
<td>3.3 mL</td>
<td>10 mM</td>
</tr>
<tr>
<td></td>
<td>Glycerol 100%</td>
<td>50 mL</td>
<td>5% v/v</td>
</tr>
<tr>
<td></td>
<td>6-Mercaptoethanol</td>
<td>350 µL</td>
<td>5 mM</td>
</tr>
<tr>
<td>Elution</td>
<td>NaCl 5 M</td>
<td>60 mL</td>
<td>300 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.5 1 M</td>
<td>50 mL</td>
<td>50 mM</td>
</tr>
<tr>
<td></td>
<td>Imidazole 3 M</td>
<td>100 mL</td>
<td>300 mM</td>
</tr>
<tr>
<td></td>
<td>Glycerol 100%</td>
<td>50 mL</td>
<td>5% v/v</td>
</tr>
<tr>
<td>Dialysis</td>
<td>NaCl 5 M</td>
<td>60 mL</td>
<td>100 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.5 1 M</td>
<td>50 mL</td>
<td>25 mM</td>
</tr>
<tr>
<td></td>
<td>Glycerol 100%</td>
<td>100 mL</td>
<td>10% v/v</td>
</tr>
<tr>
<td></td>
<td>EDTA 0.5 M</td>
<td>2 mL</td>
<td>1 mM</td>
</tr>
<tr>
<td></td>
<td>DTT</td>
<td>154 mg</td>
<td>1 mM</td>
</tr>
</tbody>
</table>

2.5.2. His$_6$-MBP-tagged proteins

The strains containing the plasmids for the expression of ICD, IDH, ICL and AceK were grown at 37°C in glucose-supplemented LB broth (0.2% w/v glucose) and selected with carbenicillin (50 µg/mL). When an OD$_{600}$ ~ 0.5/0.6 was reached, the protein expression was induced with IPTG, to 0.3 mM final concentration (manufacturer’s guidelines suggest a lower concentration of IPTG for better results). After another 2 h of induction, the cells were harvested in a Beckman Coulter Avanti J26-XP1 centrifuge (7,741 g, 15 min, 4°C). The pellets were resuspended on ice in 60 mL of column buffer containing an EDTA-free protease inhibitor cocktail. The samples were then sonicated for 8 rounds of 30 sec on, 30 sec off, on full power and the cell debris was sedimented by centrifugation using the same Beckman Coulter Avanti J26-XP1 centrifuge (14,636 g, 20 min, 4°C).

The supernatants were then applied to an amylose resin column (NEB) (2-ml packed resin bed volume) equilibrated with 200 mL of column buffer (Table 11). After passing the supernatants at a 1 mL/min pace to allow binding of the MBP-tag, the column was washed using 500 mL of column buffer. Finally, the samples were eluted in 10 mL of elution buffer 1. A first dialysis is necessary to lower the salt concentration in the sample prior to Factor Xa protease cleavage. The sample was treated with four batches of dialysis buffer 1, two batches with 100 mM of NaCl then another two batches with 50 mM. The samples were subsequently concentrated to reduce the volume to 2 mL using Vivaspin columns, MWCO 10,000 Da in an Eppendorf 5810R centrifuge (3,220 g, 20 min, 4°C). The concentration of the sample was estimated by measuring its absorbance at 280 nm using an Eppendorf BioSpectrometer. The estimation of the protein solution concentration permits to calculate the correct amount of Factor Xa protease to add in the next step.
All four proteins ICD, IDH, ICL and AceK were cleaved in order to remove the His<sub>6</sub>-MBP tag using Factor Xa protease applying a ratio 1:100 (1 µg of Factor Xa cleaves 100 µg of fusion protein). The cleavage was conducted for 24 h at 6°C with rotation. To stop the cleavage, 200 µL of p-aminobenzamidine agarose was poured in a new tube and washed several times with Factor Xa cleavage buffer before the sample was applied. The sample was mixed, then centrifuged 1 min at 16,873 g in a benchtop centrifuge in order to save the supernatant containing the His<sub>6</sub>-MBP tag and the native protein.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Component</th>
<th>per Litre</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column</td>
<td>NaCl 5 M</td>
<td>40 mL</td>
<td>200 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.4 1 M</td>
<td>20 mL</td>
<td>20 mM</td>
</tr>
<tr>
<td></td>
<td>EDTA 0.5 M</td>
<td>2 mL</td>
<td>1 mM</td>
</tr>
<tr>
<td></td>
<td>DTT</td>
<td>154 mg</td>
<td>1 mM</td>
</tr>
<tr>
<td>Elution 1</td>
<td>NaCl 5 M</td>
<td>40 mL</td>
<td>200 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.4 1 M</td>
<td>20 mL</td>
<td>20 mM</td>
</tr>
<tr>
<td></td>
<td>EDTA 0.5 M</td>
<td>2 mL</td>
<td>1 mM</td>
</tr>
<tr>
<td></td>
<td>Maltose 500 mM</td>
<td>20 mL</td>
<td>10 mM</td>
</tr>
<tr>
<td>Dialysis 1</td>
<td>NaCl 5 M</td>
<td>20 mL then 10 mL</td>
<td>100 then 50 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.5 1 M</td>
<td>25 mL</td>
<td>25 mM</td>
</tr>
<tr>
<td></td>
<td>Glycerol 100%</td>
<td>100 mL</td>
<td>10%</td>
</tr>
<tr>
<td>Factor Xa cleavage</td>
<td>NaCl 5 M</td>
<td>10 mL</td>
<td>50 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 6.5 1 M</td>
<td>20 mL</td>
<td>20 mM</td>
</tr>
<tr>
<td></td>
<td>CaCl&lt;sub&gt;2&lt;/sub&gt; 500 mM</td>
<td>2 mL</td>
<td>1 mM</td>
</tr>
<tr>
<td>Lysis 2</td>
<td>NaCl 5 M</td>
<td>60 mL</td>
<td>300 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.5 1 M</td>
<td>50 mL</td>
<td>50 mM</td>
</tr>
<tr>
<td></td>
<td>Glycerol 100%</td>
<td>50 mL</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>β-Mercaptoethanol</td>
<td>350 µL</td>
<td>5 mM</td>
</tr>
<tr>
<td>Dialysis 2</td>
<td>NaCl 5 M</td>
<td>20 mL</td>
<td>100 mM</td>
</tr>
<tr>
<td></td>
<td>TRIS-HCl, pH 7.5 1 M</td>
<td>25 mL</td>
<td>25 mM</td>
</tr>
<tr>
<td></td>
<td>Glycerol 100%</td>
<td>100 mL</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>EDTA 0.5 M</td>
<td>2 mL</td>
<td>1 mM</td>
</tr>
<tr>
<td></td>
<td>DTT</td>
<td>154 mg</td>
<td>1 mM</td>
</tr>
</tbody>
</table>

The supernatants were then applied to an Ni-NTA column, previously equilibrated with 500 mL of lysis buffer 2. A loop was created by connecting the top and bottom of the column for an hour to optimize the binding of the His part of the tag, leaving the native (cleaved) protein in the flowthrough. The flowthrough was then collected in 15 mL of lysis buffer 2. A second dialysis was performed with dialysis buffer 2 (see Table 11 for buffer composition) followed by concentration of the samples in Vivaspin columns, MWCO 10,000 Da in an Eppendorf 5810R centrifuge (3,220 g, 20 min, 4°C) to reduce the volume to 2 mL. Finally, all samples were aliquoted in 50 µL and snap frozen in liquid nitrogen.
2.5.3. Protein concentration determination

2.5.3.1. Direct determination

Using an Eppendorf BioSpectrometer, the absorbance of the samples was measured at 280 nm and the concentration was determined after computing the extinction coefficient $\varepsilon$ in Expasy ProtParam tools. The concentration of the samples could be determined based on the extinction coefficient and molecular weight of the proteins. See below the Table 12 compiling the relevant parameters.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Extinction coefficient (M$^{-1}$.cm$^{-1}$)</th>
<th>Molecular weight (g.mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD</td>
<td>57870</td>
<td>45577.40</td>
</tr>
<tr>
<td>His$_6$-ICD</td>
<td>57870</td>
<td>46400.25</td>
</tr>
<tr>
<td>His$_6$-MBP-ICD</td>
<td>124220</td>
<td>92779.88</td>
</tr>
<tr>
<td>IDH</td>
<td>82280</td>
<td>81634.12</td>
</tr>
<tr>
<td>His$_6$-IDH</td>
<td>82280</td>
<td>82594.11</td>
</tr>
<tr>
<td>His$_6$-MBP-IDH</td>
<td>148630</td>
<td>128973.74</td>
</tr>
<tr>
<td>ICL</td>
<td>54320</td>
<td>58886.70</td>
</tr>
<tr>
<td>His$_6$-ICL</td>
<td>54320</td>
<td>59709.55</td>
</tr>
<tr>
<td>His$_6$-MBP-ICL</td>
<td>120670</td>
<td>106089.18</td>
</tr>
<tr>
<td>AceK</td>
<td>79760</td>
<td>66760.82</td>
</tr>
<tr>
<td>His$_6$-AceK</td>
<td>79760</td>
<td>67583.67</td>
</tr>
<tr>
<td>His$_6$-MBP-AceK</td>
<td>146110</td>
<td>113963.30</td>
</tr>
</tbody>
</table>

2.5.3.2. Bradford assay

Alternatively, the Bradford assay was used in order to determine the concentration of protein samples. Following the method from Olson et al.\cite{202}, the preparation of the standard curve included $\gamma$-globulin, NaOH and Bradford reagent (BioRad). The preparation of the samples to determine the concentration followed a similar procedure by adding NaOH and Bradford reagent to the solution. All measurements were performed using an Eppendorf BioSpectrometer at a wavelength of 595 nm.

2.6. SDS-PAGE analysis

After having conducted purification with MBP or Ni-NTA resins, 5 µL of each sample was saved at each step for SDS-PAGE analysis. Protein samples were prepared in 4X sample buffer and boiled at 95°C for 5 min. Finally, they were loaded onto a 12 % acrylamide gel and electrophoresed in running buffer for 15 min at 120 V, then 1 h at 150V. Precision Plus Protein marker (BioRad) was used as an indicator of molecular mass. Table 13 shows the buffers used for the preparation of the gels.
2. Materials and methods

Table 13: Buffers used for preparation and running gels

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Components</th>
<th>per Litre</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X upper pH 6.8</td>
<td>TRIS</td>
<td>60.55 g</td>
<td>0.5 M</td>
</tr>
<tr>
<td></td>
<td>10% SDS w/v</td>
<td>50 mL</td>
<td>0.5% w/v</td>
</tr>
<tr>
<td>5X lower pH 8.8</td>
<td>TRIS</td>
<td>151.42 g</td>
<td>1.25 M w/v</td>
</tr>
<tr>
<td></td>
<td>10% SDS w/v</td>
<td>50 mL</td>
<td>0.5%</td>
</tr>
<tr>
<td>10X running pH 8.3</td>
<td>Glycine</td>
<td>144 g</td>
<td>1.92 M</td>
</tr>
<tr>
<td></td>
<td>TRIS</td>
<td>30 g</td>
<td>0.25 M</td>
</tr>
<tr>
<td></td>
<td>SDS</td>
<td>10g</td>
<td>35 mM</td>
</tr>
</tbody>
</table>

2.7. Gel filtration

Analytical gel filtration with multi-angle light scattering was performed at a flow rate of 0.5 mL/min through a 300 × 7.8 mm TSK-Gel G3000 SWXL column (Toso Haas) equilibrated with 20 mM Tris-HCl pH 8.0, 150 mM NaCl. The column eluate was monitored in line with a Mini-DAWN light scattering detector at 690 nm, Quasi-elastic light scattering detector, differential refractometry and absorption detector at 280 nm. Molecular masses were calculated using Astra Software (Wyatt Technologies) and the Debye fit method. A gel Filtration Marker Kit for Protein Molecular Weights 29,000-700,000 Da from Sigma-Aldrich® was used to calibrate the column.

2.8. Analytical Ultra Centrifugation (AUC)

The technique used to determine the oligomeric assembly of the enzymes in this study was sedimentation velocity. This technique is a classical method for the characterization of interactions of purified proteins in dilute solutions. Conceptually, it consists simply of the application of a centrifugal force, the observation of the macromolecular redistribution through the chamber, and the eventual quantitative analysis of the data. Sedimentation techniques provide information on self-association properties of a protein, enabling to distinguish higher-order quaternary structure (dimer, trimer, tetramer for example). The starting point in most situations is calculating the sedimentation coefficient distributions c(s), which extracts information on purity, number of species, their relative abundance and low-resolution shapes. So, sedimentation velocity permits the observation of the separation of proteins due to their different rates of migration in the centrifugal field.

Sedimentation velocity analysis was conducted on ICD and IDH. The samples were first dialysed with 100 mM NaCl and 25 mM TRIS to remove all traces of glycerol and DTT that could affect the readings. AUC was performed using a Beckman Optima XL-I with absorbance and interference optical detection system and an An60Ti rotor. Each protein was tested at three different concentrations: 0.1, 0.5 and 1 mg/mL. Epon double-sector centrepieces were filled with 400 µL of sample solution or PBS (control), respectively. All four samples were centrifuged at a rotor speed of 11,648 g and at a rotor temperature of 20°C. Absorbance data were acquired at a wavelength of 280 nm, and in time intervals of 2 min; interference scans were taken at time intervals of 1 min.
Firstly, buffer viscosity, protein partial specific volumes and frictional ratios were calculated using the software SEDNTERP. Secondly, data were analysed using SEDFIT based on the parameters determined by SEDNTERP.

### 2.9. Enzyme kinetic measurements

#### 2.9.1. Isocitrate dehydrogenase assay

##### 2.9.1.1. Basic kinetics and regulator screening

All enzyme kinetics were performed by incubating the purified enzyme in 50 mM TRIS pH 7.5 and 5 mM MgCl₂ at 37°C. The initial rates were measured across a range of substrate (+)-potassium D₃-threo-isocitrate monobasic (0 to 600 µM) at a set concentration of 200 µM NADP⁺, then across a range of NADP⁺ (0 to 600 µM) at a set concentration of 200 µM of substrate. Absorbance measurements of NADPH formation were recorded at 340 nm. The values were converted to millimolar per minute assuming a molar extinction of 6,220 M⁻¹.cm⁻¹ for NADPH. The allosteric regulator assays were carried out in the same conditions (50 mM TRIS pH 7.5, 5 mM MgCl₂, 200 µM NADP⁺ and 200 µM D₃-threo-isocitrate) screening 33 potential regulators at a fixed concentration of 1 mM (see Table 14 for a full list). All assays were conducted in triplicate.

<table>
<thead>
<tr>
<th>Table 14: List of regulators tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>acetate</td>
</tr>
<tr>
<td>acetyl-CoA</td>
</tr>
<tr>
<td>ADP</td>
</tr>
<tr>
<td>AMP</td>
</tr>
<tr>
<td>ATP</td>
</tr>
<tr>
<td>citrate</td>
</tr>
<tr>
<td>coenzyme A</td>
</tr>
<tr>
<td>3-deoxy-2-keto-6-phosphogluconate (KDPG)</td>
</tr>
<tr>
<td>D-glucose-6-P</td>
</tr>
<tr>
<td>D-glyceraldehyde</td>
</tr>
<tr>
<td>fructose-1,6-biphosphate</td>
</tr>
<tr>
<td>fumarate</td>
</tr>
<tr>
<td>GDP</td>
</tr>
<tr>
<td>glycolic acid</td>
</tr>
<tr>
<td>glyoxylic acid</td>
</tr>
<tr>
<td>GTP</td>
</tr>
</tbody>
</table>

##### 2.9.1.2. Competitive kinetics

Following the screening of regulators, more kinetic assays against isocitrate were performed to characterize the activity of the regulators to determine activators, inhibitors and the competitive, uncompetitive or non-competitive effect of these molecules. To do so, the enzyme was incubated in 50 mM TRIS pH 7.5 and 5 mM MgCl₂ at 37°C with 200 µM NADP⁺ and 1 mM of the regulator. This
was tested against a range of concentrations of isocitrate from 0 to 600 µM. All measurements were performed in triplicate and monitored at 340 nm.

2.9.1.3. Effect of nucleotides

From previous work done in the lab, ATP seemed to have an effect on ICD activity. Further investigation was then performed to define with greater details how ATP affects ICD activity.

2.9.1.3.1. Simple study of the effect of nucleotides

As described above in section 2.9.1.2, the effect of ATP and other nucleotides was studied by incubating ICD$_{Pa}$ in 50 mM TRIS pH 7.5, 5 mM MgCl$_2$, 200 µM of substrate, 200 µM of NADP$^+$ and 1 mM of ATP. The samples were incubated at 37°C and analysed at t$_{0\text{min}}$ and t$_{30\text{min}}$. Similar conditions were used to study the effect of other nucleotides (AMP, ADP, GDP and GTP) and all measurements were recorded at 340 nm.

2.9.1.3.2. Study of several concentrations of ATP

The effect of ATP was further investigated by looking at the effect of several concentrations of ATP. The assay was designed to incubate ICD$_{Pa}$ in 50 mM TRIS pH 7.5, 5 mM MgCl$_2$, 200 µM of substrate, 200 µM of NADP$^+$ and increasing concentrations of ATP (0.025, 0.05, 0.1, 0.5, 1, 5, 10, 15, 25 and 50 mM). Again, the experiment was done in triplicate.

2.9.1.3.3. Time-dependent effect of ATP

Refined analysis of the effect of ATP over time was done in order to see the effect of a longer incubation time on the rate of inhibition by ATP. Samples of 200 µL containing 50 mM TRIS pH 7.5, 5 mM MgCl$_2$, 5 mM of substrate, 5 mM of NADP$^+$ and ATP (either 0.025, 0.05, 0.1, 0.5, 1, 5, 10, 15, 25 or 50 mM) were incubated at 37°C. From these samples, 5 µL were taken out at 0, 5 min, 10 min, 20 min, 30 min, 45 min and dispensed in a reaction mixture with 50 mM TRIS pH 7.5, 5 mM MgCl$_2$, 200 µM of substrate, 200 µM of NADP$^+$ and the corresponding concentration of ATP. The formation of NADPH was monitored at 340 nm.

2.9.1.3.4. Competitive kinetics against ATP

Finally, the competitive kinetics against isocitrate or NADP$^+$ were performed using the same design as for the competitive kinetics against a regulator. The enzyme was incubated at 37°C in 50 mM TRIS pH 7.5, 5 mM MgCl$_2$ with a 200 µM fixed concentration of D-threo-isocitrate and another fixed concentration of ATP (either 0.1, 0.5, 1, 5, 10 mM), following which a range of NADP$^+$ concentration was tested from 0 to 600 µM. The reverse experiment was also done, incubating the enzyme in the same conditions with a 200 µM fixed concentration of NADP$^+$ and tested across a range of concentration of isocitrate from 0 to 600 µM. The formation of NADPH was monitored at 340 nm.
2.9.1.4. Investigation of AceK on ICD and IDH

Isocitrate dehydrogenase kinase/phosphatase was incubated with either ICD or IDH in order to investigate a potential effect on the activity of each enzyme. This is an indirect assay monitoring the loss of activity with a decrease in the formation of NADPH at 340 nm. The assay was conducted by mixing in a 1 mL total volume 50 mM TRIS pH 7.5, 5 mM MgCl$_2$, 200 µM NADP$^+$, 200 µM D$_3$-isocitrate, 1 mM ATP, 2.19 nM ICD or 0.612 nM IDH protein and 10 nM of AceK protein. The control samples consisted of the same reaction mixture minus ATP and AceK; this was to assess the stability of the enzyme over time. Control tubes were prepared then incubated, and ICD or IDH enzyme was added immediately before measurement. The samples tubes were prepared with all the reagents except the substrate. Isocitrate was added at the moment of the measurement at t$_{0min}$ and t$_{30min}$. Incubation at 37°C was done in a heat block over the course of the experiment.

2.9.2. Isocitrate lyase assay

2.9.2.1. Direct assay

Basic kinetics were performed using a modified direct assay$^{145}$. The enzyme was incubated in 25 mM imidazole pH 6.8, 5 mM MgCl$_2$, 1 mM EDTA, 4 mM phenylhydrazine and across a range of substrate concentrations (0 to 600 µM). All samples were incubated at 37°C before and during the assay, monitoring the formation of phenylhydrazine-glyoxylate complex at 324 nm.

The majority of the regulators were tested using this method with a fixed concentration of 80 µM of substrate observing the effect of 1 mM of regulator. All the values were converted to millimolar per minute assuming a molar extinction of 16,800 M$^{-1}$.cm$^{-1}$ for the phenylhydrazine-glyoxylate complex formed during the reaction.

2.9.2.2. Coupled assay

The coupled assay employed lactate dehydrogenase (LDH) from porcine heart$^{207}$. The necessity for the coupled assay is due to the fact that phenylhydrazine reacts with the aldehyde group of glyoxylate forming the phenylhydrazine-glyoxylate complex. Conceptually, phenylhydrazine reacts with any ketone or aldehyde. However, the chemicals shown in Table 15 below already possess either one of these, making the direct assay unsuitable. The coupled assay overcomes this issue by using a secondary reaction with LDH that mirrors the response of ICL to the presence of the regulator.

Thus, the ICL was incubated with 50 mM MOPS-NaOH pH 7.3, 15 mM MgCl$_2$, 5 mM DTT, 1 mM EDTA, 400 µM NAD$^+$ and 60 U of LDH. The assay was first used to screen a few regulators that could not be assayed with the direct assay (see Table 15) in the presence of a fixed concentration of substrate at 200 µM challenged by 1 mM of regulator. From there, competitive kinetics at a fixed concentration of regulator against a range of substrate concentration (0 to 400 µM) was performed to study the type of response. The values were converted to milli-micromoles of NADH per minute assuming a molar extinction of 6,220 M$^{-1}$.cm$^{-1}$ for NADH.
Table 15: Regulators showing an impaired response in the direct assay

<table>
<thead>
<tr>
<th>Succinate+glyoxylate</th>
<th>α-ketoglutarate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glyoxylate+oxaloacetate</td>
<td>Glyoxylate</td>
</tr>
<tr>
<td>Malate+oxaloacetate</td>
<td>D-glyceraldehyde</td>
</tr>
<tr>
<td>Pyruvate</td>
<td>D-glucose-6-phosphate</td>
</tr>
<tr>
<td>3-deoxy-2-keto-6-phosphogluconate</td>
<td>6-phospho-gluconate</td>
</tr>
<tr>
<td>Oxaloacetate</td>
<td>Fructose-1,6-bisphosphate</td>
</tr>
<tr>
<td>Methylglyoxal</td>
<td></td>
</tr>
</tbody>
</table>

2.9.2.3. Competitive kinetics

The screening of regulators resulted in showing the potency of a few molecules, but these molecules do not have any ketone or aldehyde group. The competitive kinetics were conducted using the direct assay with phenylhydrazine. In that case the enzyme was incubated with 25 mM imidazole pH 6.8, 5 mM MgCl₂, 1 mM EDTA, 4 mM phenylhydrazine, 1 mM of the regulator and tested across a range of substrate concentrations (0 to 200 µM). Similar to kinetic experiments conducted to determine \( K_m \) and \( V_{max} \), the samples were incubated at 37°C during the experiment and monitored at 324 nm for the formation of phenylhydrazine-glyoxylate complex.

2.9.3. Kinetics plots and calculations

All data generated by the kinetics experiments were plotted and analysed using Graphpad Prism 6. The parameters such as \( K_m \) and \( V_{max} \) were calculated according to the Michaelis-Menten equation:

\[
Y = V_{max} \times \frac{X}{K_m + X}
\]

in which \( V_{max} \) is the maximum enzyme velocity, and \( K_m \) is the Michaelis-Menten constant. \( X \) is the variable concentration of substrate or co-factor tested.

\( K_{1/2} \) was calculated following a sigmoidal fitting equation:

\[
Y = V_{max} \times \frac{X^h}{K_{1/2}^h + X^h}
\]

where \( h \) is the Hill slope. The curve is sigmoidal due to positive cooperativity. \( K_{1/2} \) is the concentration of substrate that produces a half-maximal enzyme velocity. \( V_{max} \) and \( X \) are the same as described above.

Finally, \( k_{cat} \) was derived from either:

\[
Y = E_t \times k_{cat} \times \frac{X}{K_m + X}
\]

\[
k_{cat} = \frac{V_{max}}{E_t}
\]

where \( E_t \) is the concentration of enzyme used in the experiments.
Calculations carried out using known concentrations of enzyme for each assay ($E_i$). The optimal enzyme concentrations were found to be:

- ICD: 2.19 nM
- IDH: 0.612 nM
- ICL: 42.4 nM

Regarding competitive kinetics, a Lineweaver-Burk plot was used to characterize the type of competitive, non-competitive or uncompetitive profile of the different regulators tested. It is a simple double-reciprocal of the data where

$$X = \frac{1}{[S]}$$

and

$$Y = \frac{1}{V_{max}} \times \left(1 + \frac{K_m}{[S]}\right)$$

in which $[S]$ is the range of concentration of substrate tested against the regulators that displayed a statistically significant response over the screening assay.

Statistical significance of the response regarding the effect of regulators was determined using a one-way ANOVA test with $p$-value < 0.05 (*) unless otherwise stated.

### 2.10. Thermal shift assay

Thermal shift was performed on IDH and ICL as both enzymes responded to a selection of regulators. The assay allows to quantify the change in thermal denaturation of the protein under varying conditions. In this case, the binding of low molecular weight ligands can increase the thermal stability of the protein as described by Koshland. This assay was conducted to confirm some findings from the kinetic assays of IDH and ICL, the binding of some regulators would help determine the affinity of the regulators with the protein.

To monitor protein unfolding, the fluorescent dye Sypro orange was used. Sypro orange is a highly sensitive dye, commonly used for thermal shift assay as its fluorescence properties ($\lambda_{ex}$ 470 nm / $\lambda_{em}$ 570 nm) are compatible with filter sets found on real-time PCR instruments. The unfolding process following the raising of the temperature during the assay exposes the hydrophobic regions of proteins and results in an increase in fluorescence. The dye is released from the hydrophobic cavities of the protein and is eventually used to monitor the protein-unfolding transition.

### 2.10.1. Testing regulators

All experiments were done in Hard-Shell® 96-Well PCR Plates (Biorad) at a 25 µL total reaction volume in each well. Regarding ICL, the analysis was carried out in the presence of 50 mM MOPS, 15 mM MgCl$_2$, 400 µM of regulator and 5 µM of protein pre-mixed with Sypro orange dye at
a 1:250 ratio of protein:Sypro orange. IDH analysis was similar: the reaction mixture included 50 mM TRIS pH 7.5, 5 mM MgCl₂, 400 µM of regulator and 5 µM of pre-dyed protein.

2. Materials and methods

2.10.2. Equipment settings

The thermal shift assay was conducted in a Biorad CFX-Connect Real Time Detection System, originally designed for PCR. The system contains a heating/cooling device for accurate temperature control and a charge-coupled device (CCD) detector for simultaneous imaging of the fluorescence changes in the wells of the microplate. The plate was heated from 25 to 95°C with a heating rate of 0.5°C/min. The Bio-Rad CFX Manager software was used to operate the equipment.

2.10.3. Analysis of thermal shift data

The fluorescence imaging data were fitted to the sigmoidal Boltzmann equation which assumes a two-state model (before and after transition). It analyses the Relative Fluorescence Unit (RFU) as a unit of the readings for the assay:

\[ F_x = F_A + \frac{(F_B - F_A)}{1 + e^{-\frac{r}{k_1}}} \]

as a function of temperature, supported by Graphpad Prism 6 software. The equation includes the following parameters: \( F_A \) is the initial or pre-transition reading (the protein is correctly folded at low temperatures). \( F_B \) is the late or post-transition reading (after passing a certain threshold, the stability of the protein breaks down and it unfolds, liberating the fluorescent dye and thus a higher reading). \( T_m \) represents the actual melting or transition temperature: it is the x-value equivalent to half of \( F_B \). Finally, \( k_1 \) is the slope greater than one of the derived curve during the transition phase.

The thermal shift itself is simply derived from the difference between the values obtained for the samples with regulators vs. the protein alone as a control:

\[ \Delta T_m = T_{m_{regulator}} - T_{m_{control}} \]

Usually, adding a low molecular weight molecule enhances the stability of the protein, since the transition temperature is higher than the control temperature.

2.11. Crystallisation and X-ray experiments

2.11.1. Approach to crystallisation

2.11.1.1. Crystallisation screening and optimization

Crystallization trials were performed using the vapor diffusion sitting-drop technique in 96-well MRC 2-drop plates (Molecular Dimensions). Screening and viewing of all proteins were performed using a Rock Imager 1000 (Formulatrix) operated by Rock Maker software. The plates were stored at 19°C for the duration of the experiment and imaging was automatically set to take pictures of the
crystal growth after 1, 2, 3, 5, 8 (+ UV image to characterize protein crystal vs. salt crystal), 13, 21 and 34 days of incubation.

Screening plates tested were as follow:

- From Molecular Dimensions: JCSG+, Midas, PACT suite, Wizard I and II
- From Qiagen: Classics suite, PEGs I suite, PH Clear I suite

In cases where the screening was successful, but the crystals were too small, optimization was pursued to refine the conditions of the screening. Optimization plates were usually prepared using the Dragonfly® crystal robot (TTP Labtech Ltd) which dispensed microliters of solution in a horizontal or vertical gradient and added the substrate or the co-factor to some rows of the plate. Protein and mother solution dispensing was performed by the Mosquito® crystal robot (TTP Labtech Ltd) mixing 200 nL of each solution.

### 2.11.1.2. Crystallisation conditions for ICD

The first isocitrate dehydrogenase crystallised in the presence of sodium acetate 0.1 M pH 4.6 with 30% v/v PEG 300. The protein concentration was 10 mg/mL. These conditions were successful from the PEGs suite screening plate.

### 2.11.1.3. Crystallisation conditions for IDH

The second isocitrate dehydrogenase crystallised after optimization in the presence of 21.5% w/v PEG 3350, NaH$_2$PO$_4$ 0.2 M, glycerol 5% and NADP$^+$ 150 µM at a protein concentration of 17.5 mg/mL.

### 2.11.1.4. Crystallisation conditions for ICL

The last enzyme crystallised after optimization with HEPES 0.1 M pH 5, CaCl$_2$ 0.1 M, 20% w/v PEG 6000, glycerol 5%, glyoxylate 1 mM and 2% thymol as additive. The protein concentration of 18 mg/mL permitted crystallisation.

### 2.11.2. Crystal “fishing” and storage of the samples

Fully grown crystals were “fished” using CrystalCap™ SPINE HT (Hampton Research) mounted loops and vials. The crystals were cryo-protected after immersion in a drop containing the crystallization condition solution supplemented with 26% v/v ethylene glycol and then snap-frozen in liquid nitrogen. All crystals were stored in their vials and kept in liquid nitrogen before being sent to the synchrotron.

### 2.11.3. Sending samples and Synchrotron “shooting”

The X-ray diffraction data collection was performed at Diamond Light Source, beamline I04-1 or I02 (Oxford, UK). The crystallographic data collection statistics are summarised in table 16 for all three proteins.
2. Materials and methods

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th>ICD</th>
<th>IDH</th>
<th>ICL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation source</td>
<td>Diamond (UK), I04-1</td>
<td>Diamond (UK), I04-1</td>
<td>Diamond (UK), I02</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9282</td>
<td>0.9282</td>
<td>0.9795</td>
</tr>
<tr>
<td>Space group</td>
<td>P121</td>
<td>C222,</td>
<td>I222</td>
</tr>
<tr>
<td>Cell dimensions a, b, c (Å)</td>
<td>88.88, 95.55, 104.01</td>
<td>126.46, 149.02, 201.13</td>
<td>80.94, 116.02, 128.53</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 99, 90</td>
<td>90, 90, 90</td>
<td>90, 90, 90</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>2.70-47.78 (2.70-2.79)</td>
<td>2.71-29.62 (2.71-2.78)</td>
<td>1.88-29.49 (1.88-1.93)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>304691 (17419)</td>
<td>940727 (66819)</td>
<td>614357 (27689)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>44233 (2978)</td>
<td>51799 (3682)</td>
<td>48986 (3287)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>6.9 (5.8)</td>
<td>18.2 (18.1)</td>
<td>12.5 (8.4)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.0 (89.4)</td>
<td>99.6 (97.2)</td>
<td>99.4 (92.1)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>17.9 (2.0)</td>
<td>17.3 (3.7)</td>
<td>21.7 (3.0)</td>
</tr>
<tr>
<td>R-merge</td>
<td>0.068 (0.826)</td>
<td>0.176 (0.891)</td>
<td>0.086 (0.656)</td>
</tr>
<tr>
<td>R-pim</td>
<td>0.028 (0.348)</td>
<td>0.059 (0.304)</td>
<td>0.036 (0.340)</td>
</tr>
<tr>
<td>CC-half</td>
<td>0.99 (0.79)</td>
<td>0.99 (0.92)</td>
<td>0.99 (0.78)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rwork (%)</td>
<td>24.56</td>
<td>20.80</td>
<td>18.32</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>28.37</td>
<td>26.67</td>
<td>21.17</td>
</tr>
<tr>
<td>No. of molecules per ASU</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>No. of total atoms (no H)</td>
<td>12832</td>
<td>11551</td>
<td>3798</td>
</tr>
<tr>
<td>atoms for ligands</td>
<td>n/a</td>
<td>58</td>
<td>6</td>
</tr>
<tr>
<td>atoms for waters</td>
<td>36</td>
<td>197</td>
<td>182</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran plot analysis, number of residues in:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Favoured regions (%)</td>
<td>92.19</td>
<td>93.91</td>
<td>96.90</td>
</tr>
<tr>
<td>Allowed regions (%)</td>
<td>7.03</td>
<td>5.54</td>
<td>2.69</td>
</tr>
<tr>
<td>Disallowed regions (%)</td>
<td>0.78</td>
<td>0.55</td>
<td>0.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B-factor (Å²)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average/Wilson</td>
<td>114.21/68.00</td>
<td>52.31/43.36</td>
<td>35.10/26.42</td>
</tr>
<tr>
<td>Ligands</td>
<td>n/a</td>
<td>28.88</td>
<td>52.86</td>
</tr>
<tr>
<td>Solvent</td>
<td>38.04</td>
<td>35.87</td>
<td>32.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMS deviations</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond lengths (Å)</td>
<td>0.003</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.602</td>
<td>1.52</td>
<td>1.43</td>
</tr>
<tr>
<td>PDB entry</td>
<td>5M2E</td>
<td>6G3U</td>
<td>6G1O</td>
</tr>
</tbody>
</table>

2.11.4. Crystal structure determination, model building and refinement

All crystal structures were solved by molecular replacement (MR) using templates already deposited in the Protein Data Bank (PDB). ICD, IDH and ICL were solved using the templates 1BL5, 4ZDA (to be published) and 3I4E (to be published) from Escherichia coli, Mycobacterium smegmatis and Burkholderia pseudomallei respectively. The choice of these templates was guided after a BLAST analysis of each P. aeruginosa protein sequence against the PDB. The results with the highest identity scores were extracted for molecular replacement work.
All MR calculations were performed using PHASER\textsuperscript{214} as part of the PHENIX\textsuperscript{215} crystallographic software package. The obtained models were alternatively subjected to manual rebuilding using the molecular graphic suite COOT\textsuperscript{216} and crystallographic refinement with PHENIX software or CCP4 package.
3. Isocitrate dehydrogenase (ICD)

3.1. Introduction

In this chapter, I will be discussing the first isocitrate dehydrogenase, ICD. I will start with the results of a bioinformatic analysis of the genomic context of icd gene and the analysis of ICD protein with respect to its evolution and conservation among pathogenic bacteria. Overall, phylogenetic analysis reveals that the use of nicotinamide adenine dinucleotide phosphate (NADP+) by prokaryotic ICD arose around the time eukaryotic mitochondria first appeared, about 3.5 billion years ago. The switch of the coenzyme specificity of prokaryotic ICD from nicotinamide adenine dinucleotide (NAD+) to NADP+ appears to be an ancient adaptation to anabolic demand for NADPH during growth on acetate\textsuperscript{181,217}. This evolutionary aspect demonstrates how important ICD/IDH are in all prokaryotic and eukaryotic cells. ICD is particularly well conserved among bacteria as the enzyme of the TCA cycle that decarboxylates isocitrate. It is present in all bacteria studied here for the purpose of understanding the importance of the isocitrate branchpoint, and how carbon flux is rerouted depending on the carbon source available in the environment. ICD has been thoroughly studied in \textit{E. coli}\textsuperscript{218} since the 1970s regarding its kinetics, expression profile and structure. However, little is known about it in \textit{Pseudomonas}.

In this chapter, I investigate the structure and kinetics of ICD in \textit{P. aeruginosa}. Firstly, this study includes all the steps of the purification and analysis of the oligomeric status with gel filtration and analytical ultracentrifugation. I will then discuss the gene expression profile of icd gene when the cells are grown in different carbon sources. Most of the results I present here are related to the kinetics of the enzyme and its regulation via phosphorylation by isocitrate dehydrogenase kinase/phosphatase. I also solved the crystal structure of ICD and I will then complete the results with a thorough description of the structure. This investigation includes the protein-protein interface analysis, a comparison with the \textit{E. coli} ICD structure and a detailed analysis of the active site.

3.2. Bioinformatic analysis

In the \textit{P. aeruginosa} genome, the icd gene is positioned at 2.965 Mbp. This gene is composed of 1,257 base pairs (figure 6A). icd is located next to the idh gene (the second isocitrate dehydrogenase); both genes share the same 539 bp-long intergenic region and are divergently transcribed. Upstream of icd is cspD, which encodes a cold-shock protein member of the CspA cold shock stress adaptation family of proteins\textsuperscript{219}. icd orthologues are also present in other \textit{Pseudomonas} species (figure 6B). In other species of \textit{Pseudomonas}, the icd sequence is conserved. \textit{P. fluorescens}, \textit{P. putida} and \textit{P. stutzeri} have a gene that is the same size (1,257 bp), although the annotated icd gene in \textit{P. syringae} is much longer at 2,223 bp. Further investigation showed that icd in \textit{P. syringae} in fact corresponds to idh. In all species of \textit{Pseudomonas}, the predicted sub-cellular location of the
translated protein is cytoplasmic. Furthermore, the overall genomic context in each species shows the same pattern: cold-shock protein (cspD) or ATP-binding protease component (clpA) adjacent to icd.

Figure 6: Genetic context of the icd gene in P. aeruginosa PAO1 and other Pseudomonas species (A) The highlighted icd annotated gene is 1,257 bp encoding a 419 amino acid-long protein (stop codon included). idh gene is located immediately adjacent to the icd gene and shares a single intergenic region of 539 bp (B) icd orthologues in other Pseudomonas species display a similar genomic organisation, with idh nearby and cold shock protein or ATP-binding protease component upstream.
3. Isocitrate dehydrogenase (ICD)

S. aureus

-------- MTAEKIT-QGETLNVPHPIIFIEGIDGPIIDWKAAAASVIDAAVEKA 48

H. pylori

MAYNKLIQKPEGRTIK-DNKLHVPNNHPPIIFIEGIDGIDSITPAMAVKDAVEQA 59

P. aeruginosa

-----N-GYQIKQPAPDTQKTRYMDLSVSVPKIPPIIEGIDGIDSPVMVIRDAVEKA 58

E. coli

------------MSKVVVPAGKKITVDAGQKLVPHNPIIFIEGIDGVDTPAMINVDAAVEKA 56

Y. pestis

---------YNGKVVVPAGKKKTVDAQGKLVPHNPIIFIEGIDGVDTPAMINVDAAVEKA 57

\*

S. aureus

YNKGRKIVEFKEVQALMQAKDAF--------T-GWLEWQETLDTNYEXILIADVQPLTQP 99

H. pylori

YNGKRIKAIYEVFRGVEICYQKSFEDKYLSSPEEGWLLPDTEAINHVKSYQKPT 119

P. aeruginosa

YNGKRIKAIYEVFRGVEICYQKSFEDKYLSSPEEGWLLPDTEAINHVKSYQKPT 119

E. coli

YNKGRKIVEFKEVQALMQAKDAF--------T-GWLEWQETLDTNYEXILIADVQPLTQP 99

Y. pestis

YNKGRKIVEFKEVQALMQAKDAF--------T-GWLEWQETLDTNYEXILIADVQPLTQP 99

\*

S. aureus

GFLSNVALLQOELDLICTPVRWFKGVPDVPVRKDVPQVDMFPFWEDYAGIEFKGT 159

H. pylori

GFLSNVALLQOELDLICTPVRWFKGVPDVPVRKDVPQVDMFPFWEDYAGIEFKGT 159

P. aeruginosa

GFLSNVALLQOELDLICTPVRWFKGVPDVPVRKDVPQVDMFPFWEDYAGIEFKGT 159

E. coli

GFLSNVALLQOELDLICTPVRWFKGVPDVPVRKDVPQVDMFPFWEDYAGIEFKGT 159

Y. pestis

GFLSNVALLQOELDLICTPVRWFKGVPDVPVRKDVPQVDMFPFWEDYAGIEFKGT 159

\*

S. aureus

TEVKKVDFLQWFGMLGATHRHPFTSGICIGPKVQEGTCLVRSAIQHAIYINRKS 219

H. pylori

AEAEKLIHLQWFGMLGATHRHPFTSGICIGPKVQEGTCLVRSAIQHAIYINRKS 219

P. aeruginosa

PAEAVKFIKFELQWFGMLGATHRHPFTSGICIGPKVQEGTCLVRSAIQHAIYINRKS 238

E. coli

ADAIEVKKFLREAGMLGATHRHPFTSGICIGPKVQEGTCLVRSAIQHAIYINRKS 238

Y. pestis

ADAIEVKKFLREAGMLGATHRHPFTSGICIGPKVQEGTCLVRSAIQHAIYINRKS 238

\*

S. aureus

NGHNTGFQGQWGDALSLFEGFQDGTDIQYDEVENUEGARDDAAANQAEGEKEG 279

H. pylori

NGHNTGFQGQWGDALSLFEGFQDGTDIQYDEVENUEGARDDAAANQAEGEKEG 279

P. aeruginosa

NGHNTGFQGQWGDALSLFEGFQDGTDIQYDEVENUEGARDDAAANQAEGEKEG 279

E. coli

NGHNTGFQGQWGDALSLFEGFQDGTDIQYDEVENUEGARDDAAANQAEGEKEG 279

Y. pestis

NGHNTGFQGQWGDALSLFEGFQDGTDIQYDEVENUEGARDDAAANQAEGEKEG 279

\*

S. aureus

IKSMTADFLQIRLPAEDVTVMTMLNGYISDALAAQVQGIGIAPGAINYETCAHI 339

H. pylori

IKSMTADFLQIRLPAEDVTVMTMLNGYISDALAAQVQGIGIAPGAINYETCAHI 339

P. aeruginosa

IKSMTADFLQIRLPAEDVTVMTMLNGYISDALAAQVQGIGIAPGAINYETCAHI 339

E. coli

IKSMTADFLQIRLPAEDVTVMTMLNGYISDALAAQVQGIGIAPGAINYETCAHI 339

Y. pestis

IKSMTADFLQIRLPAEDVTVMTMLNGYISDALAAQVQGIGIAPGAINYETCAHI 339

\*

S. aureus

FEATHTAGYKAGLWNVPSWSSLVMLMHLNWQEDAKIDTIREDPTASAKKTVYDPA 399

H. pylori

FEATHTAGYKAGLWNVPSWSSLVMLMHLNWQEDAKIDTIREDPTASAKKTVYDPA 399

P. aeruginosa

FEATHTAGYKAGLWNVPSWSSLVMLMHLNWQEDAKIDTIREDPTASAKKTVYDPA 399

E. coli

FEATHTAGYKAGLWNVPSWSSLVMLMHLNWQEDAKIDTIREDPTASAKKTVYDPA 399

Y. pestis

FEATHTAGYKAGLWNVPSWSSLVMLMHLNWQEDAKIDTIREDPTASAKKTVYDPA 399

\*

S. aureus

RLMDGAEVSTSAFADELIRLKNK 422

H. pylori

RLMDGAEVSTSAFADELIRLKNK 422

P. aeruginosa

RLMDGAEVSTSAFADELIRLKNK 422

E. coli

RLMDGAEVSTSAFADELIRLKNK 422

Y. pestis

RLMDGAEVSTSAFADELIRLKNK 422

\*

Figure 7: ClustalOmega alignment of ICD amino acid sequence (yellow: phosphorylation site, green: substrate binding, blue: magnesium binding). ICD protein length is slightly variable from one species to another, the average length is around 420 amino acids. However, it is remarkable how well the enzyme is conserved. The later adaptation of ICD in species such as E. coli or P. aeruginosa does not seem to have affected the amino acid sequence.
sequences. The same is true when it comes to the residues that bind the magnesium ion (highlighted in blue) necessary for the catalytic reaction to take place. Finally, and most importantly with regard to regulation, the serine residue (highlighted in yellow) that is phosphorylated by the isocitrate dehydrogenase kinase-phosphatase seems to be conserved.

3.3. Purification of ICD

3.3.1. Preliminary results

In order to study P. aeruginosa ICD (namely ICD\textsubscript{Pa} in the rest of this dissertation), I first purified the enzyme. Firstly, the icd gene was PCR amplified from PA01 gDNA and cloned into either pQE80 or a modified version of pMAL-c2x. The pQE80 vector that carries icd was introduced into E. coli DH\textsubscript{5}α for overexpression. The cloning was done by a previous member of the laboratory. Initially, I purified ICD\textsubscript{Pa} with a His\textsubscript{6} tag attached to the N-terminus of the protein (encoded in the pQE80 vector) using a Ni\textsubscript{NTA} column. However, tagged proteins can be experimental hurdles in crystallisation, I thus decided to produce a new un-tagged protein to circumvent these hurdles. I started using a modified version of the pMAL-c2x vector. A His\textsubscript{6} tag was added to the N-terminus of the already present MBP-tag in the vector (work performed by Dr. Martin Welch). The vector was then renamed His-pMAL-c2x. A Factor Xa cleavage site located between the C-terminus of the MBP and N-terminus of the encoded protein was already part of the initial vector, allowing the eventual purification of an un-tagged protein following Factor Xa digestion. I then cloned the ICD encoding gene into His-pMAL-c2x, transformed the construct into the same E. coli DH\textsubscript{5}α to overexpress and purify the product. Before cleavage, the construct would be a His\textsubscript{6}-MBP-ICD protein. Briefly, the purification workflow follows three steps. The first step uses an amylose resin to purify the construct via affinity of the MBP tag for the resin. The second step uses Factor Xa to cleave the His\textsubscript{6}-MBP-tag from the N-terminus of ICD\textsubscript{Pa}. Finally, the third step uses a Ni-NTA resin to retain the His\textsubscript{6}-MBP-tag while the cleaved, native ICD\textsubscript{Pa} in solution flows through and is collected.

The figure 8 presented above shows that the molecular mass of the purified ICD is between 37 and 50 kDa. This corresponds nicely to the molecular mass predicted from the ICD amino acid sequence, which is 45,577 Da. This step demonstrates that the purification of the protein is complete, the molecular mass of the product is consistent with the predicted mass. The product of the purification after overexpression from E. coli DH\textsubscript{5}α containing the icd pQE80 construct was similarly successful, the molecular mass of the concentrated solution of protein was consistent.
with its prediction. The SDS-PAGE analysis results after migration of the His$_6$-tagged ICD showed the presence of additional bands (see appendix 2), probably degradation products that did not affect the continuation of the work. It is noticeable that the downstream treatment of the sample with Factor Xa resulted in a single band instead but with a much lower yield than the simple purification conducted on the His$_6$-tag version of ICD.

### 3.3.2. Gel filtration analysis

Besides SDS-PAGE analysis, a complementary technique to estimate the oligomeric status of ICD$_{Pa}$ is gel filtration. A sample of 40 µg of concentrated ICD$_{Pa}$ solution were injected into a TSK-Gel G3000 SWXL column equilibrated following the guidelines described in chapter 2. The native molecular mass was determined to be approximately 90,800 Da (see the results header of figure 9). Furthermore, the narrow peak observed in the figure 9 graph demonstrates that the sample, which is the product of the purification from the icd pQE80 construct, eventually has very little contaminants (see appendix 3 for the full results). These results show that the native ICD from *P. aeruginosa* is a dimeric enzyme. This result is in accordance with ICD in *E. coli*, which is also a dimer$^{183}$. It demonstrates how much both enzymes are similar, the high amino acid identity (79%), and the oligomeric status.

### 3.3.3. AUC data

Analytical ultracentrifugation with a set of sedimentation velocities (SV) was used to further examine the oligomeric status of ICD$_{Pa}$. SV is a complementary technique to gel filtration that provides an information-rich characterization of many aspects of protein behaviour in solution, including the protein mass and size, density, hydrodynamic shape, size-distribution and purity. In the case of ICD, two hundred boundary profiles were recorded at three different concentrations (0.1, 0.5 and 1 mg/mL). The data were analysed as specified in chapter 2 using SEDNTERP and SEDFIT. The profiles recorded and presented in figure 10 show that the frictional ratio for ICD is 1.33. The frictional ratio (calculated as $f/f_0$) is a parameter of AUC that changes depending on the shape of the protein. On one hand, $f$ (for frictional coefficient) is a measure of the resistance to movement of a molecule; this resistance is a function of both the size and the shape of the molecule. It is measured
by its rate of sedimentation. On the other hand, \( f_0 \) corresponds to the frictional coefficient of a spherical particle from the Stokes equation. This means that a frictional ratio closer to 1 suggests that the protein is more globular. With a frictional ratio of 1.33, the results for ICD\(_{Pa}\) support the formation of a dimer with a moderately extended shape.\(^{220}\)

Furthermore, the calculated molecular mass was 91,500 Da, which agrees nicely with the data from gel filtration. This leads to the conclusion that ICD is a stable dimer in solution, again comparable in _P. aeruginosa_ with ICD from _E. coli_.

### 3.4. Gene expression of _icd_

Two carbon sources, acetate and glucose, were tested to compare the responses of the cells as an adaptation to their environment. In order to investigate the effect of carbon source on the expression of _icd_, the promoter of the gene was fused with a promoter-less LacZ in pLP170. Theoretically, in the case of access to acetate only, and because it has only two carbons, the use of the glyoxylate shunt instead of the TCA cycle would be the appropriate response to save these two atoms from the decarboxylating steps of the TCA cycle. By doing so, the cells are capable of producing...
3. Isocitrate dehydrogenase (ICD)

the intermediates for other biochemical pathways, gluconeogenesis and to grow biomass. On the other hand, glucose is a rich carbon source. If the cells only have access to such a nutrient source, their metabolism should be rewired to promote the TCA cycle only. Glucose is a ring of six carbons, the loss of two molecules of \( \text{CO}_2 \) does not affect the continuation of the TCA cycle. So, as a summary that means that acetate vs. glucose is in fact glyoxylate shunt vs. TCA cycle. The results of the growth of \textit{Pseudomonas} in both media in figure 11 show that there is no impairment of the growth of the bacteria whether they have access to acetate or glucose. There even seems to be an overall better growth on single carbon sources for the strain containing the \textit{Picd::lacZ} construct (blue line) compared with the wild type containing the empty vector (black line). I have no explanation for this since the two strains should differ only in the presence of \textit{Picd} in the PAO1 strain transformed with \textit{Picd::lacZ} construct.

The measurements from the fluorescent \( \beta \)-galactosidase assay are plotted as a bar chart in both panels of the figure 11. They show that there is a slight expression of the \textit{lacZ} encoding gene present in the promoter-less vector, as seen between 7 and 10h of experiment in the panel B (MOPS + glucose). This suggests that the supposedly promoter-less plasmid responds to induction and still produces a very small amount of \( \beta \)-galactosidase, although this is negligible compared with the \textit{Picd::lacZ} construct.

The response of the construct is of interest. When the cells are grown in acetate (panel A), the results from the \( \beta \)-galactosidase assay highlight a strong increase of the signal over the exponential phase between 6 and 7h of experiment; followed by a decrease as the bacteria reach stationary phase (8-10h). The maximum value is around 3,000 nmol.min\(^{-1}.\text{A}_{600}^{-1}\).

On the other hand, when the cells are grown in glucose, the \( \beta \)-galactosidase activity seems more consistent along the time course of the experiment. The results show a slightly steadier expression of \textit{lacZ} when fused with the \textit{icd}
3. Isocitrate dehydrogenase (ICD)

So, growth of PAO1 in a shorter carbon medium (acetate) seems to trigger the expression of icd more as the β-galactosidase activity values are higher; this trend seems consistent with a study performed on E. coli\(^{221}\). In this study on E. coli, the Picd::lacZ gene expression varied depending on whether glucose or acetate was used for cell growth, the β-galactosidase activity measured after the cells were grown in acetate was 2.3-fold higher than when the cells where grown on glucose in aerobic conditions. This is a similar pattern seen in P. aeruginosa; the β-galactosidase activity measured from the cells grown in acetate was 2.4-fold higher than in the cells grown in glucose. The fact that icd is still expressed in limited nutrient source demonstrates that the TCA cycle is still operational. The main hypothesis is that the production of α-ketoglutarate and co-production of NADPH is vital to combat oxidative stress. NADPH is the ultimate reductive force required to neutralize reactive oxygen species (ROS) as it maintains anti-oxidative systems in reduced/active states. ICD is the major NADPH-producing enzyme in ROS-exposed P. fluorescens when grown on simple carbon source\(^{222}\).

3.5. Kinetic analyses

3.5.1. Michaelis-Menten data

ICD kinetic profile was determined following the established assay assessing the formation of NADPH at 340 nm by mixing a solution of purified untagged ICD with its substrate, co-factor and magnesium ion\(^{222}\). The design of the experiment is set to investigate the K\(_{m}\) of the enzyme by varying its substrate concentration while maintaining the co-factor at a saturating concentration. Then, the affinity of the enzyme for the co-factor was investigated by varying NADP\(^+\) concentration while maintaining the substrate at a saturating concentration. This yielded K\(_{isocitrate}\) = 25 µM and K\(_{NADP^+}\) = 32 µM (see figure 12A-B). The V\(_{max}\) for isocitrate was 0.08 ΔmM.min\(^{-1}\) and 0.09 ΔmM.min\(^{-1}\) for NADP\(^+\). The Michaelis-Menten equation fits the data correctly. Despite the fact that ICD\(_{PA}\) is a dimer, and as such is a potential allosteric enzyme consisting of multiple subunits and active sites, it is safe to infer that the enzyme follows a steady-state ordered mechanism. This has been extensively studied in Pseudomonas nautica\(^{224}\), Escherichia coli\(^{225}\), R. sphaeroides\(^{226}\). Finally, the first order rate constant k\(_{cat}\) was calculated based on the enzyme concentration used in the assay (2.19 nM). The values of k\(_{cat}\) for isocitrate and NADP\(^+\) were 36,529 min\(^{-1}\) (609 s\(^{-1}\)) and 41,095 min\(^{-1}\) (684 s\(^{-1}\)) respectively. These values are high; however they are not atypical for a central metabolism enzyme.

It has been suggested that enzymes operating in central metabolism are under stronger selective pressures to increase their rates, hence their higher k\(_{cat}\)\(^{227}\).

ICD has not been studied in Pseudomonas aeruginosa before, but kinetic parameters have been reported for ICD from M. tuberculosis\(^{49}\) and E. coli\(^{228}\). For the Mycobacterium enzyme, K\(_{isocitrate}\) = 50 µM and K\(_{NADP^+}\) = 15 µM, while for the Escherichia enzyme, K\(_{isocitrate}\) = 11 µM and K\(_{NADP^+}\) = 17
µM. In my hands, purified ICDEc yielded $K_{i}$iso = 20 µM and $K_{i}$NADP⁺ = 15 µM, i.e., values falling within the same range as the published ones. Furthermore, $k_{cat}$ was determined for the M. tb and E. coli enzymes at respectively, 33 s⁻¹ and 80 s⁻¹. This highlights a variability in the kinetics of this enzyme in different pathogenic species. ICD activity in Pseudomonas aeruginosa is similar, overall, when compared to the E. coli model.

### 3.5.2. Screening of potential regulators

I wanted to investigate the potential regulation of ICD activity by a subset of molecules that the protein might encounter in the cell. This includes a selection of 31 products from central metabolism including the TCA cycle and the Entner-Doudoroff pathway. The enzyme was incubated with saturating concentrations of isocitrate and NADP⁺ along with magnesium and a fixed concentration of each regulator. Similar to the kinetic studies discussed in the previous section, I monitored the formation of NADPH at 340 nm as the enzymatic reaction progresses in the presence of each regulatory candidate. The results were analysed with a one-way ANOVA (p < 0.05 and n = 3) (see figure 13A), and statistically significant results are marked with an asterisk.

Firstly, the product of the reaction (α-ketoglutarate) has a slight (23%) yet significant negative effect on the enzyme activity indicative of product inhibition. The literature reports another case of ICD product inhibition in Brevibacterium flavum in which the isocitrate dehydrogenase is inhibited to 36% of its activity229. This indicates that ICDpa is less sensitive to product inhibition than the B. flavum enzyme. However, further investigation about the effect of this product inhibition was performed to understand the type of inhibition and will be discussed later.

Secondly, neither glyoxylate alone nor oxaloacetate alone affected the activity of ICD. However, there was a noticeable inhibitory effect when an equimolar (1 mM) mixture of both compounds was present, yielding 93% inhibition of ICDpa. As seen in figure 13A, the bar chart shows a large decrease in ICDpa activity when compared with the uninhibited control. Such an inhibition has already been reported in the literature for ICD from E. coli225, T. thermophilus230 and
mammals cells\textsuperscript{231}. The different investigating groups found that the mixture of both compounds caused 25-75\%, 42\% and 40\% inhibition respectively. In this context, ICD\textsubscript{Pa} seems remarkably affected with only 7\% of activity remaining. For the \textit{E. coli} enzyme, the mechanism that has been proposed to explain this inhibition is the formation of a condensation product with a very strong affinity for ICD. Using HPLC analysis, Nimmo et al. demonstrated that mixtures of glyoxylate and oxaloacetate form, firstly oxalomalate and then 4-hydroxy-2-oxoglutarate. Oxalomalate is likely the active compound inhibiting the enzyme due to its similarity with isocitrate\textsuperscript{225}.

The inhibition of ICD\textsubscript{Pa} by \(\alpha\)-ketoglutarate was further investigated. A solution of purified enzyme was incubated across a range of isocitrate concentrations with a saturating concentration of NADP\(^+\) and a fixed concentration of \(\alpha\)-ketoglutarate (1 mM) to monitor the activity of ICD\textsubscript{Pa}. By doing so, I determined how \(\alpha\)-ketoglutarate competes against isocitrate. Analysis of the resulting kinetic data using a Lineweaver-Burk plot (figure 13B) shows that the control and \(\alpha\)-ketoglutarate lines intersect the y-axis (1/\(V_{\text{max}}\)) at the same point but their x-intersection (-1/\(K_m\)) is different. This indicates that \(\alpha\)-ketoglutarate is a competitive inhibitor of isocitrate binding to ICD. The \(K_m\) value in presence of \(\alpha\)-ketoglutarate is higher than the control (64 \(\mu\M\)) whereas \(V_{\text{max}}\) is 0.09 \(\Delta\text{mM.min}^{-1}\), so within the range of the \(V_{\text{max}}\) of the control (see section 3.5.1). The scenario of a competitive inhibition supposes that the inhibitory compound(s) are capable of binding to a free active site of the enzyme, which then prevents binding of the substrate.

The affinity of the inhibitor for the enzyme can be inferred from the \(K_i\) value. The \(K_i\) represents the concentration of inhibitor at which, under saturating substrate conditions, the reaction
rate is half of the maximum reaction rate \( (V_{\text{max}}) \). The \( K_i \) value was determined from the modified Michaelis-Menten equation:

\[
K_{m(\text{obs})} = K_m \times \left(1 + \frac{[I]}{K_i}\right)
\]

in this experiment, \([I] = 1000 \mu\text{M}\). The \( K_i \) obtained for \( \alpha\text{-ketoglutarate} \) was 665 \( \mu\text{M} \). This is a rather high value for \( K_i \) but it is consistent with another study conducted on \( \text{Brevibacterium flavum} \) ICD. For \( B.\text{flavum} \) ICD, the \( K_i \) is 1.1 mM\(^2\). A \( K_i \) in the high \( \mu\text{M} \) range is rather high, indicating that \( \alpha\text{-ketoglutarate} \) is a weak inhibitor.

### 3.5.3. Effect of nucleotides

#### 3.5.3.1. Effect of ATP

The regulation of ICD by central metabolism compounds did not reveal strong regulatory candidates. The results were significant for some molecules but overall, there is no obvious trend that would explain the rerouting of carbon at the TCA Glyoxylate Branchpoint (TGB). This indicates that there must be other regulatory mechanisms controlling \( ICD_{Pa} \) activity. One mechanism is the effect of AceK (isocitrate dehydrogenase kinase/phosphatase) to inactivate ICD. This is based on the extensive knowledge of ICD in \( E.\text{coli} \) for which we know that AceK is the major regulator. In \( E.\text{coli} \), AceK phosphorylates ICD rendering it inactive, thus pushing \( \text{isocitrate} \) through the glyoxylate shunt. I investigated this mechanism in \( P.\text{aeruginosa} \) and this is discussed in the section 3.5.4. However, I also examined the effect of ATP (and other related nucleotides). In \( E.\text{coli} \), AceK phosphorylates ICD in presence of ATP, so I decided to see if ATP alone has an intrinsic effect on \( ICD_{Pa} \). The ubiquity of ATP and its association with AceK, makes the molecule a good regulatory candidate.

A brief study of the effect of AceK on \( ICD_{Pa} \) has already been performed by a PartII student in 2014. She concluded that \( ICD_{Pa} \) is completely inhibited by ATP after 2 h of incubation. Interestingly, the effect of ATP on bacterial \( \text{isocitrate} \) dehydrogenase has not been widely studied. A few publications report such an effect in the \( \text{Crithidia fasciculata} \) enzyme\(^{232,233} \) and the \( \text{Salmonella typhimurium} \) enzyme\(^{234} \). To investigate this further, I incubated the enzyme with saturating concentrations of substrate and co-factor and a fixed concentration of 1 mM ATP. In the \( \text{Pseudomonas aeruginosa} \) enzyme, there was a significant inhibition of ICD activity in presence of 1 mM of ATP (figure 14A). This effect is measurable after an incubation time of up to 30 min. The loss of activity in the \( P.\text{aeruginosa} \) enzyme (28%) \( C.\text{fasciculata} \) nor in \( S.\text{typhimurium} \) in which the inhibition accounted for 55% and 60% of activity loss respectively. However, in \( Pseudomonas \), ICD is not the only enzyme affected by ATP. Another metabolic enzyme of the Entner-Doudoroff pathway exhibits inhibition by ATP: \( \text{glucose-6-phosphate dehydrogenase (G6PD)} \). In that case, ATP seems to inhibit G6PD activity. The enzyme performs at 15% of its capacity, the inhibition is significantly stronger than for ICD. The effect of a single concentration 1 mM ATP on \( ICD_{Pa} \) is significant, thus,
I decided to investigate further and studied the effect of lower and higher concentrations of the same nucleotide ATP.

Other concentrations of ATP had a similar inhibitory effect on \( \text{ICD}_{P_K} \) as seen in figure 14B. Higher concentrations of ATP greatly affected ICD activity. From 25 mM and beyond, there was total loss of activity when ATP is incubated with the enzyme. Concentrations less than 1 mM or equal to 1 mM do not show inhibition. This corroborates the results from figure 14A, inhibition by 1 mM ATP was significant after 30 min of incubation. So, the length of incubation seems to be important in the case of low concentrations of ATP, I decided to investigate further. The purified enzyme was incubated in substrate and co-factor saturating concentrations and with low concentrations of ATP (0.025 to 1 mM) for up to 30 min. Higher concentrations of ATP (5 to 20 mM) were also tested.

The results were plotted as a bar chart (figure 14C), and all the values were first analysed with a two-way ANOVA analysis \((p < 0.05 \text{ and } n = 3)\). Even the very low concentration of 0.025 mM ATP had a slight, yet significant, inhibitory effect. At higher concentrations i.e. from 0.05 to 5 mM, the pattern was similar. The initial rate of the enzyme is lowered by 0.02 mM.min\(^{-1}\) when incubated with 0.025 mM ATP; while the effect of 5 mM ATP inhibits \( \text{ICD}_{P_K} \) by 0.06 mM.min\(^{-1}\). Higher concentrations of ATP (10 mM and more) show an immediate inhibition of the enzyme when compared with control and nearly total inhibition after incubation. This indicates that \( \text{ICD}_{P_K} \) is affected by ATP in two ways. On the one hand, high concentrations of ATP immediately inhibit the enzyme; on the other hand, for low concentrations of ATP, inhibition is time-dependent. For the latter, a greater effect is seen after 30 min of incubation. The need of incubation to observe the effect of ATP may explain why no inhibition could be detected in figure 13A.
I have demonstrated that ATP inhibits ICD. The next step in this study is to understand what type of inhibition ATP yields. To analyse ATP inhibition profile, purified ICD<sub>Pa</sub> was incubated with saturating concentration of NADP<sup>+</sup> and fixed concentrations of ATP (either 0.1, 0.5, 1, 5 and 10 mM) across a range of isocitrate. Then vice versa, the enzyme was incubated with saturating concentration of isocitrate with fixed concentrations of ATP and across a range of NADP<sup>+</sup>. The concentration of ATP in the cell can reach up to 10 mM, so the inhibition may be physiologically meaningful. Marr et al. studied the inhibitory activity of ATP and other nucleotides on ICD from *Salmonella typhimurium*<sup>234</sup> and demonstrated a competitive inhibition by ATP with respect to isocitrate and NADP<sup>+</sup>. In *Pseudomonas aeruginosa* however, as indicated in figure 15A-B, inhibition by ATP with respect to both isocitrate and NADP<sup>+</sup> is of a non-competitive type. The lines on both graphs intersect the x-axis (-1/K<sub>m</sub>) at the same point. Additionally, the lower concentrations of ATP (0.1, 0.5 and 1 mM) do not show a large response as the plots are very similar to the control line. This is consistent with the previous results seen in figure 15B, low concentrations of ATP have little effect on ICD<sub>Pa</sub> activity. However, at higher concentrations (5 and 10 mM) the effect is more obvious.

Further studies of this inhibition in ICD from *Vigna radiata*<sup>237</sup> (mung bean), suggests that as in *S. typhimurium*, ATP is a competitive inhibitor with respect to isocitrate. On the other hand, studies performed on ICD from another eukaryote (*Crithidia fasciculata*, a trypanosomatid species of mosquitoes) showed that ATP is a non-competitive inhibitor with respect to both the substrate and co-factor of the enzyme<sup>231</sup>. Additionally, there is one hypothesis that has been advanced to explain how ATP inhibits ICD. It is thought that the inhibition of ICD by ATP could be correlated with the decrease of free Mg<sup>2+</sup> in the reaction mixture<sup>238</sup> as the nucleotide might act as a chelator<sup>239,240</sup>. Nearly 90% of the magnesium in a bacterial cell is chelated with either ribosomes, DNA, RNA, membranes elements and nucleotide triphosphates<sup>241</sup>. Indeed, the Mg-ATP complex is the true substrate of numerous enzymes (mostly kinases)<sup>242,243</sup>. In *Pseudomonas aeruginosa*, since ATP is a non-competitive inhibitor it can bind to either the free enzyme or to the enzyme-substrate complex during the reaction. The magnesium depletion mechanism hypothesized
would explain the inhibition. If ATP chelates magnesium, the enzyme no longer has access to an essential catalytic element and the reaction is aborted. The time-dependent inhibition by lower concentrations of ATP might support this hypothesis. In the reaction mixture, it is possible to infer that the enzyme and ATP will compete for the free magnesium, eventually abolishing ICD activity. In this regard, the higher concentrations of ATP (10 mM and more) greatly exceed the concentration of magnesium in the reaction mixture (5 mM). To investigate this further, the nearby conserved tryptophan residue (Trp135) could be used as a fluorescent reporter to monitor metal ion binding in the active site. To do so any prebound Mg\(^{2+}\) should be removed from the purified protein. The demetallized protein can then be incubated with Mg\(^{2+}\) then Mg\(^{2+}\) and ATP. A change in the relative tryptophan fluorescence intensity will indicate that the environment around the tryptophan residue(s) changes upon binding of Mg\(^{2+}\) compared to the control (demetallized protein only) and further changes if ATP complexes free Mg\(^{2+}\).

3.5.3.2. Effect of other nucleotides

After seeing the inhibitory effect of ATP, I decided to see if this was transferable to other nucleotides. I therefore tested AMP, ADP, GDP and GTP at a single concentration (1 mM) after an incubation time of 30 min. All nucleotides demonstrate inhibition of ICD as shown in figure 16. AMP has less effect on the enzyme than ADP, GDP or GTP. AMP only inhibits ICD by 27% whereas ADP and GDP inhibit by 55% and 58% respectively. GTP falls in between with an inhibition of 38%. Overall, the results with P. aeruginosa ICD are similar to the results obtained with S. typhimurium ICD, in which ADP and GTP inhibited by 60% AMP by 10% only. Once again, the inhibition by these nucleotides is significant after a period of incubation of 30 min which is why it was not detected during the screening of the regulators.

Similar to the hypothesis previously discussed of ICD inhibition due to the formation of a Mg-ATP complex, Mg is also chelated by other nucleotides; AMP, ADP, ATP and GTP. The time-dependent inhibition of ICD by these nucleotides under the conditions described for this experiment, may explain the results. As the nucleotide concentration is of 1 mM competing against isocitrate for Mg\(^{2+}\) at 5 mM, it is possible to attribute the delay observed to this primary competition for magnesium. The Mg-ATP complex formation and the Mg-isocitrate complex formation is the foremost competition before binding to ICD\(_{Pa}\). Further investigation would study the demetallized purified protein incubated with i.e. Mg\(^{2+}\), or Mg\(^{2+}\) and isocitrate, or Mg\(^{2+}\) and ATP, or Mg\(^{2+}\), isocitrate and ATP. The relative tryptophan fluorescence
3. Isocitrate dehydrogenase (ICD)

Intensity and fluorescence quenching (using Trp135 as a fluorescent reporter) would indicate the conformational changes of the active site upon binding of Mg$^{2+}$ or Mg-isocitrate complex.

### 3.5.4. Effect of AceK

The third and last mechanism investigated for the regulation of ICD is isocitrate dehydrogenase kinase/phosphatase (AceK) in *Pseudomonas aeruginosa*. AceK is the most recognized regulator of ICD activity and helps to coordinate carbon flux through the glyoxylate shunt in *E. coli* by inhibiting isocitrate dehydrogenase. This has been thoroughly studied in *E. coli*, and AceK remains the first example of prokaryotic phosphorylation identified. It is known that in *E. coli*, AceK phosphorylates its substrate, ICD, on Ser113. In *P. aeruginosa*, Ser115 corresponds perfectly to the phosphorylation site as it matches the 3D structure conformation. This will be discussed in the next section regarding the ICD$_{Pa}$ crystal structure.

In order to test the effect of AceK on ICD from *Pseudomonas aeruginosa*, both enzymes were incubated with isocitrate and NADP$^+$. Upon addition of ATP, AceK would phosphorylate ICD and the remaining ICD activity would be monitored by measuring the reduction of NADP$^+$ to NADPH. The loss of ICD activity, if any, accounts for AceK phosphorylation. The incubation, up to 30 min, reflects previous work done in the laboratory showing that there might be a time-dependent factor acting for significant loss of activity of the enzyme. The results in figure 17 show that after addition of ATP, there is initially no significant decrease in ICD activity in the presence of AceK. However, after 30 min of incubation, ICD activity is decreased by 71%. Note that the inhibition is not total, as ICD remains up to 29% active. This could suggest that even though flux through the glyoxylate shunt is promoted upon inactivation of ICD, the TCA cycle is not completely shut down. This suggests that even in *vivo*, when the cells are growing on a limited nutrient source (aka. acetate) and rewire their carbon flux through the glyoxylate shunt, there is still a substantial flux through the TCA cycle. Saving two atoms of carbon is essential for the bacteria by rerouting carbon flux through the glyoxylate shunt. However, the cells sustain the production of NADPH by ICD and NADH by α-ketoglutarate dehydrogenase. The NADPH/NADH-generating reactions coupled to central carbon metabolism ensure that the cellular redox balance is maintained.$^{246-248}$

Overall, these results highlight the apparent similarities between ICD$_{Pa}$ and the other isocitrate dehydrogenases, particularly that from *E. coli*. In *P. aeruginosa*, ICD is sensitive to competitive product inhibition. It is greatly inhibited by a mixture of glyoxylate and oxaloacetate.
3. Isocitrate dehydrogenase (ICD)

But more importantly, the principle means of controlling the activity of ICD is by AceK-dependent phosphorylation.

3.6. Crystal structure of ICD

3.6.1. Structure description and interface analysis

The three-dimensional structure of ICD<sub>P</sub> was solved by molecular replacement using ICD<sub>E</sub> (PDB 1BL5) as a template. The structure was refined to 2.7 Å. The analysed crystal contains two dimeric ICD molecules in the asymmetric unit. Each dimer is composed of 836 amino acid residues each, along with 36 water molecules (see figure 18C). The structure presented is an apo structure as it has been crystallised without any specific ligand. Overall, the electron density map allowed largely unambiguous tracing of all molecules in the unit cell. The structure is available in the PDB under the accession code 5M2E and the appendix 5 compiles a few of the validation elements including Ramachandran plot and Polygon analysis.

![Figure 18: ICD crystal structure](image)

The ICD dimeric molecule is roughly kidney shaped, with an intricate clasp domain between the two subunits as shown in figure 18A-B. A PISA analysis indicates that with 26 interfaces calculated, the assembly is indeed a dimer. The binding between the two subunits is made of a close
interaction of two α-helices located on top of the structure as in figure 18A. Underneath are two β-strands criss-crossing each other. Additional loops and helices strengthen the interaction along the inner side of the structure, where four α-helices create a channel-like sub-structure inside the dimer. Such a conformation might explain a potential clasp-like effect when the substrate and co-factor are bound. The whole structure is comprised of 14 α-helices and 20 β-strands per monomer. In addition, 92.2% of residues are in the most preferred regions of the Ramachandran plot.

The residues comprising the interface between the two subunits of ICD were defined using MacPyMOL with the InterfaceResidues script. This script finds interface residues between two proteins or chains, using the following concept. It evaluates interface residues based on the difference between the complex-based areas and the chain-only-based areas. Such a difference isolates the residues that are overlapping and in contact to form the dimer. The cut-off used to identify the residues was of 0.5 Å², which is the difference in area over the residues considered as interface residues. Residues whose ΔASA (difference in accessible surface area) from the complex to a single chain is greater than this cut-off are kept. In addition to the size and shape of the interface, the ΔASA remains a powerful method that may eventually provide a measure of the binding strength through the correlation between the hydrophobic free energy of transfer from polar to a hydrophobic environment and the solvent ASA.

![Figure 19: Analysis of the interface of ICD (A) Space-filling representation of the residues from both chains forming the interface (B) Intermolecular contact map, a black dot is present at the cross-over of residues i and j, belonging to molecule 1 and molecule 2, respectively, if any pair of atoms belonging to the two residues is closer than the cut-off distance (default value is 8 Å) (C) Distance range map reporting inter-molecular contacts at increasing distances, as coloured dots. Red, yellow, green and blue indicate contacts within 7 Å, 10 Å, 13 Å and 16 Å, respectively](image)

A more thorough analysis by COCOMAPS indicates that most of the residues of each chain involved in the interface are between positions 150 and 210 and then again a few more residues around position 300. Both tools concluded that there are in total 75 residues from chain A and 74 residues from chain B that form the interface. The charts in figure 19 show the distribution of the residues forming the interface, and the distance range with a lower cut-off of 7 Å is shown in red. Both charts overlap showing that there are more residues around position 200 which corresponds to the top part.
3. Isoctitate dehydrogenase (ICD)

of the interface. This correlates with the inner contact of the two α-helices and the β-strands that form the clasp on top of the structure. The remaining residues scatter along the four α-helices making the core of the dimer. With regard to the ΔASA in homodimers generally, it varies widely from 368 Å² to 4,746 Å², and the results for ICD fall into that category with an average of 3,136.7 Å². This also means that 16% of the overall surface of the protein is buried upon the complex formation. The polar vs. non-polar surfaces between the two subunits show that the mean polar interface area is 2,224.8 Å² whereas the non-polar is 911.9 Å². Finally, deeply buried atoms in the core of the protein and interfacial residues in protein binding complexes tend to have lower B-factors in comparison to the rest of the protein. This is the case for most of the 149 residues making the interface of ICD. There is indeed a trend of reduced B-factors, in comparison to the outer edge of the protein. The trend is even more obvious when comparing to the loops that tend to show more disorder and flexibility.

3.6.2. Structural comparison with E. coli

Dimeric ICD of P. aeruginosa and E. coli share 79% of amino acid identity. There is a slight difference between the two, as E. coli is two-residues shorter, but the overall structure is extremely similar. The residues involved in the substrate, co-factor and cation binding are all conserved but with a shift of two residues. The superimposition of two apo structures of ICD from both species (E. coli PDB 3ICD) demonstrates a tenuous difference in the bending of the structure indicated by a rmsd of 1.98 Å (figure 20). There are also a few differences in the organisation of the structure. The large β9-strand present in P. aeruginosa (black arrow) from Cys196 to Ser204 is absent in E. coli. However, three small β-strands are missing in P. aeruginosa (β1, β2 and β11 E. coli numbering) (see appendix 4 for a detailed alignment of the amino acid sequence and secondary structure elements). Overall, the secondary structure is comparable; most of the α-helices overlap while the loops are more variable. The 12 β-sheets forming the core of the enzyme are persistent in both structures. When looking at the dimer, the E. coli structure seems to be marginally more closed whereas the P. aeruginosa structure has a slightly “wider” conformation.

The E. coli model used for residue comparison (PDB 4AJB) is a complex of ICDEc (a K100M mutant) crystallised with isocitrate, Mg²⁺ and thio-NADP⁺ (a non-hydrolysable analogue of NADP⁺). The rmsd value after overlaying the two proteins was 1.94 Å. Overall, the secondary
structure elements of the cleft around the isocitrate binding site are tighter in the ICD<sub>Ec</sub> structure, consistent with the presence of an activated form of the enzyme that closes its active site upon binding of the substrates. The binding of the substrate and co-factor clearly affects the overall bending of the active site and the dimer. The superposition of the two structures in figure 21A demonstrates how deep isocitrate is embedded into the active cleft. It sits at the very back of the cleft surrounded by α4 on the right, β6-7 at the back and α9 (from chain A) and α8 (from chain B) on the left. Refer to appendix 4 for a sequence and secondary structure alignment. Crystallographic analyses of the superposition of both structures yield considerable insight into the molecular basis of catalysis. When overlaid with the E. coli model, the residues binding the substrate in P. aeruginosa (Ser115, Asn117, Arg121, Arg131, Arg155, Tyr162 and Lys232*) (residues marked with * belong to the opposite chain) exhibit a different conformation (figure 21B): they all point outward from the active site. The residue Tyr162 shows a difference of 2.83 Å in the placement of the side chain away from Tyr160 in E. coli bound to the substrate. Additionally, the most remarkable shift is Arg121 which is displaced by about 7.41 Å outside the active site in comparison with Arg119 bound to isocitrate in E. coli. This displacement is further increased by the shift of 2.63 Å at the Cα of Arg121. Another strongly affected residue is Ser115, the distance between the two Cα between ICD<sub>Pa</sub> and ICD<sub>Ec</sub> is 3.58 Å. The hydroxyl group of this side chain points nearly 180° in the opposite direction when the substrate is not bound. This suggests isocitrate binding changes the conformation of the catalytic site. Finally, the binding of magnesium and its coordination with isocitrate is promoted by Asp285* and Asp309. However, the conformation of these residues is unaffected by Mg<sup>2+</sup>.

Following up the substrate, NADP<sup>+</sup> binds more on the rim of the cleft as seen in figure 21C-E. It engages with a much more flexible section of the enzyme. Most of the residues are located on a loop running between β6 and α10; a few more are part of α12 from chain A. More residues belong to the same α8 in chain B. This establishes a domain closure by a hinge motion of the loop from chain A that comes capping the active site cleft during catalysis<sup>276</sup>. The binding of the co-factor can be divided into two categories: the binding of the adenosine moiety on one hand and the nicotinamide moiety on the other. Similar to the substrate binding, residues from both chains participate in the binding of NADP<sup>+</sup>. Generally, the number of residues required to bind NADP<sup>+</sup> is greater than for the substrate as the molecule is much bigger and is part of a complex network with water molecules and isocitrate. Previous work<sup>185,245,257-262</sup> highlighted different residues for co-factor binding, and shed some light understanding the multiples stages of activation of the enzyme and the consequences on the flexibility of the active cleft. The template used to study NADP<sup>+</sup> binding is 4AJ3, a wild type ICD<sub>Ec</sub> crystallised with isocitrate, Ca<sup>2+</sup> and NADP<sup>+</sup>. The adenosine moiety is directly or indirectly bound to ten residues from both chains, creating a complex network with water molecules. As seen in figure 21D, the main ones are His341, Gly342, Thr343, Asn354, Tyr347, Tyr393, Gln289*, and Arg294* accompanied by Gln290* interacting with water. In P. aeruginosa, all these residues are conserved. The superposition of the two active sites reveals a large shift of the β6/α10 loop between ICD<sub>Pa</sub> in yellow which is in an open conformation and ICD<sub>Ec</sub> after closure of the cleft initiated by
NADP⁺ binding. The top part of the loop shifts by 4.40 Å calculated between the two Cα of ICD_{Ec} His339 and ICD_{Pa} His341. The bottom part of the loop displaced even more (by 7.33 Å) between ICD_{Ec} Tyr345 and ICD_{Pa} Tyr437. The α12 helix that carries Tyr393 and Arg397 in *P. aeruginosa* is displaced by 4.20 Å on average away from the active site cleft. The residues from chain B are better conserved in their orientation. The Cα are closer between the two species, however the side chains are dissimilar. For example, the side chain of Arg294* is 4.74 Å away from its equivalent Arg292* bound to the phosphate group. The nicotinamide moiety interacts mainly with the short loop located
between α4 and β5, in particular Lys100-Thr107. This forms an anchor for the nicotinamide ring binding. Two additional residues, Asn234* and Glu338, promote the interaction with isocitrate and water molecules. The superposition of those residues in figure 21F shows that the loop in *P. aeruginosa* increasingly shifts from 4.06 Å (Pro102 vs. Pro104) to 7.55 Å (Thr107 vs. Thr105). The increasing shift of the loop corresponds to closing motion promoted by NADP+ binding, the “front” of the loop (see figure 18A for a frontal view of the dimer) must cover more distance to interact with the co-factor. Overall, the binding residues are conserved, but their conformation is dictated by the binding of the ligands, this also promotes the hinge motion that operates after binding isocitrate and NADP+.

### 3.7. Discussion

The most common isocitrate dehydrogenase among bacterial species is a remarkable example of conservation across evolution. The alignment and phylogenetic tree discussed in section 3.2.2 proved that ICD is particularly well conserved among Gram-negative bacteria but also between Gram-negative and Gram-positive bacteria. Based on the work performed to investigate the evolutionary distinction of eubacterial NADP-dependent isocitrate dehydrogenase it was concluded that the enzyme is monophyletic, and it diverges near the branchpoints of the eukaryotic NAD+ and NADP+ dependent isocitrate dehydrogenases. This seems to indicate that NADP+ dependence evolved very early, concomitant with the appearance of eukaryotes estimated around 2-3.5 billion years ago. Due to its major role in carbon metabolism, ICD shows strong similarity across a wide range of species as seen in the previous sections. The length of the enzyme is as conserved across most bacterial species.

The overexpression and purification of ICD in *Pseudomonas aeruginosa* yielded a concentrated solution of protein. Further investigation of the oligomeric status of the enzyme in this microorganism confirmed that it has an homodimeric quaternary structure which is consistent with other species harbouring the same enzyme. The gel filtration and sedimentation velocity experiments both yielded nearly the same calculated molecular mass (90 kDa). These findings corroborate the conservation status of ICD in *Pseudomonas aeruginosa* compared with the well-studied ICD from *E. coli*. Despite the fact that the glyoxylate shunt was first discovered in *Pseudomonas* strain KB1, little is known about the TGB in this microorganism nor any characteristics of ICD.

Here, I present the crystal structure and kinetics of ICD from *Pseudomonas aeruginosa*. A crucial objective of the work was to characterize this enzyme and establish differences in behaviour that may account for understanding how *P. aeruginosa* partitions carbon flux in central metabolism. Structural analyses and sequence alignments conducted with CCP4MG, PyMol and Jalview highlighted the similarity of ICDPa with ICDEc, especially in the residues that are thought to be in contact with either the substrate or the co-factor. In this study, ICDPa is an apo structure with no ligand or co-factor in the electron density map. The superposition of ICDPa and ICDEc structures shows conformational displacements of the residues lining the substrate and co-factor binding sites in ICDPa. Further structural analyses and alignments have revealed that the catalytic site and the overall...
shape are highly conserved, including the positioning of the residues that participate in the substrate, ion and co-factor binding (figure 21). The alteration in the overall bending of the active site cleft in \textit{P. aeruginosa} could be attributed to the fact that the model used for comparison includes the substrate, co-factor and metal, thus triggering a more “closed” conformation of the enzyme.

From a general point of view, the catalytic site in ICD\textsubscript{Pa} lies in the same configuration as in ICD\textsubscript{Ec}, displaying a pocket at the front of each subunit divergently oriented from each other. The phosphorylation site (Ser115) is on the helix α14 at the front of the pocket, the residue overlays with Ser113 in \textit{E. coli} which indicates that it is capable of binding the substrate and probably involved in AceK-mediated phosphorylation. In the case of inactivation by phosphorylation, the negative charge of the phosphate leads to complete inactivation of the enzyme in \textit{E. coli} as the serine can no longer bind to the substrate\textsuperscript{188}. However, when binding the substrate, there is a conformational change, the side-chain is reoriented inward as seen in figure 21B. Finer-Moore et al.\textsuperscript{267} have already suggested a conformational change of ICD\textsubscript{Ec} in presence of AceK. AceK promotes a cascade of conformational changes making Ser113 (in \textit{E. coli}) available for phosphorylation. This has been confirmed by the study of a S113 mutant\textsuperscript{188,268–270}. All substitutions for that serine resulted in lowered activity when compared to the wild type enzyme. Further investigation on ICD from \textit{P. aeruginosa} exploring the effect of S115 directed mutagenesis. Such experiments help understanding how enzymes discriminate between ligands such as substrates and inhibitors for the development of rationally designed drugs for example. These experiments would also focus on AceK phosphorylation capacity on ICD and how other residues replacing Ser115 could implement or diminish the phosphorylation-mediated inactivation.

---

**Figure 22**: Proposed catalytic mechanism of isocitrate dehydrogenase. ICD mechanism follows a general acid-base chemistry supplemented by two main catalytic residues Lys232 (chain B) and Asp307 (chain A). Figure adapted from Aktas and Cook\textsuperscript{270}
3. Isocitrate dehydrogenase (ICD)

The active site is well defined: seven residues bind to isocitrate but interestingly three residues in particular explain the catalytic mechanism of ICD. Aktas and Cook proposed a catalytic mechanism based on a general acid-base chemistry involving the lysine-tyrosine pair\(^{271}\) (figure 22). The similarity of the active site with other bacterial species such as *E. coli*\(^{256}\) and *M. tuberculosis*\(^{184}\) supports evidences for this mechanism. This is a two-step mechanism with deprotonation followed by decarboxylation. Firstly, Lys232\(^*\) would initiate dehydrogenation by uncoupling the proton from the hydroxyl on C(2) of isocitrate. This is permitted after Asp309 (metal binding residue) aids the lysine to act as a base by deprotonating it. NADP\(^+\) is then reprotonated and eventually forms NADPH. On the other hand, oxalosuccinate, an enol intermediate, is the intermediate reaction product. From there, electron rearrangement occurs leading to decarboxylation at C(3) of isocitrate. In the meantime, Tyr162 approaches C(3) for protonation concomitant with Lys232\(^*\) second deprotonation at C(2) hydroxyl to yield the α-ketoglutarate product. Further down the line, the triad Asp309-Lys232\(^*\)-Tyr162 might work the same in the second isocitrate dehydrogenase IDH. The remarkable similarity between the two active sites helps to anticipate a comparable mechanism.

The *K_m* of ICD for both isocitrate and NADP\(^+\) is comparable to other bacterial species. The affinity of ICD\(_{Pa}\) for isocitrate and NADP\(^+\) is half of the affinity observed for *Mycobacterium* ICD. Furthermore, the affinity for NADP\(^+\) is 2-fold greater than ICD\(_{Mc}\) or ICD\(_{Ec}\). The overall conclusion is that ICD\(_{Pa}\) is similar to that from *E. coli* or *M. tuberculosis* regarding substrate and co-factor affinity. Additionally, the results presented in this work suggest that the product of the reaction, α-ketoglutarate, has a negative feedback on the enzyme. However, the effect on ICD\(_{Pa}\) does not seem to be a strong inhibition. The competitive inhibition profile indicates that α-ketoglutarate binds to the active site, probably due to its similarity with the substrate. This is a typical non-allosteric or non-regulatory enzyme. The weak inhibitory activity (23%) from α-ketoglutarate does not support carbon re-routing through the glyoxylate shunt instead of the TCA cycle. The competition of ICD with IDH and ICL for isocitrate supposes another regulatory mechanism capable of adjusting ICD activity based on metabolic cues. Furthermore, the presence of AceK in *P. aeruginosa* and Ser115 in ICD suggests that there is indeed such a mechanism. On the other hand, oxalomalate demonstrated a very significant inhibition of the enzyme. This has been extensively studied in the past, and it is now explicit that oxalomalate interacts with the isocitrate-binding site.

<table>
<thead>
<tr>
<th></th>
<th><em>K_m</em> (µM)</th>
<th><em>V_satur</em> (ΔmM/s)</th>
<th><em>k_cat</em> (s(^{-1}))</th>
<th><em>k_cat/K_m</em> (s(^{-1})/µM)</th>
<th><em>K_i</em> (µM)</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>isocitrate</td>
<td>25.58</td>
<td>0.0013</td>
<td>608.83</td>
<td>23.80</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>α-ketoglutarate</td>
<td>63.99</td>
<td>0.0016</td>
<td>745.81</td>
<td>11.66</td>
<td>665</td>
<td>competitive</td>
</tr>
<tr>
<td>NADP(^+)</td>
<td>32.26</td>
<td>0.0015</td>
<td>684.93</td>
<td>21.23</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

The large loss of activity of ICD observed after incubation with AceK however, means that the phosphorylation-mediated inhibition is the main switch to decrease carbon flux through the TCA.
cycle. Work performed by a Part II student in the laboratory demonstrated the inhibition of ICD in the presence of AceK. AceK-mediated phosphorylation of ICD is presumed to be the regulatory mechanism by which *P. aeruginosa* is capable of inactivating ICD to partially divert carbon flux through ICL and malate synthase. However, ICD is not totally inactivated by AceK; this means that flux through the TCA cycle is biologically relevant when the cells grow on limited carbon source and oxidative stress conditions (accumulation of ROS upon cell growth). Keeping a small percentage of the flux through the Kreb’s cycle allows the bacteria to keep the NADPH-generating steps active and sustain an essential reducing agent as it plays an important role in various biological processes and is the driving force behind numerous biosynthetic reactions\textsuperscript{272}. In order to be active, AceK requires ATP and it has been demonstrated in this study that ATP alone has a significant inhibition on isocitrate dehydrogenase. It is suggested that, after incubation, ATP interacts with the enzyme-substrate complex in a non-competitive way by altering the V\textsubscript{max}. This appears to be a new aspect about isocitrate dehydrogenase that has not been shown before. The literature previously described a non-competitive inhibition in different species. The uncommon behaviour of ICD\textsubscript{Pa} towards ATP is of interest regarding the regulation by AceK and the potential effect on the whole isocitrate branchpoint. Additionally, considering the concentrations of ATP in the cytoplasm of the cell, this inhibition could play a role *in vivo* but we do not know about it. This is a question to address in further research.
4. Isocitrate dehydrogenase (IDH)

4.1. Introduction

In bacteria, there are two isocitrate dehydrogenases. The most commonly found is the dimeric ICD discussed in the previous chapter. Some bacteria encode a second isocitrate dehydrogenase isozyme, IDH. In microorganisms in which IDH has been identified, it is described as a monomer (i.e. *Azotobacter vinelandii*, *Corynebacterium glutamicum*, *Mycobacterium tuberculosis*). IDH catalyses the same reaction as ICD, using the same substrate and co-factor. Interestingly, *Pseudomonas aeruginosa* is one of the few species that has the two isocitrate dehydrogenases. In this chapter, I will explore the structure and kinetics of the *P. aeruginosa* IDH.

I will start with the bioinformatic analysis of *idh* gene and its encoded product, IDH. IDH was successfully overexpressed and purified, gel filtration analysis and AUC with sedimentation velocity indicate its oligomeric status. The effect of carbon substrates on the *idh* gene expression profile was investigated. I will then review the enzymology of IDH. The kinetics parameters (*K_m*, *V_max*, *k_cat*) and regulation of IDH have been characterised, it is worth noting here that AceK is ineffective in regulating this enzyme. Finally, I will present the first *P. aeruginosa* crystal structure of IDH. The investigation includes a comparison with multiple structures from other species, a detailed analysis of the clasp domain and the active site, and an investigation to understand why AceK is ineffective.

4.2. Bioinformatic analysis

The *idh* gene is located beside the *icd* gene and is positioned at 2.967 Mbp in the PAO1 genome. Both genes share an intergenic region of 539 bp and are divergently transcribed. *idh* is composed of 2,226 bp and encodes a 741-residue long protein (figure 23A). Downstream of *idh* is PA2625, encoding a conserved hypothetical protein. The next gene (*trmU*) encodes a tRNA methyltransferase. *trmU* is predicted to be included in an operon with the other PA2625, PA2627 and PA2628 upstream and downstream of it, which likely play a role in transcription, RNA processing and degradation\(^{[7]}\). These metabolic genes (*icd* and *idh*) are not operonic but scattered within the genome, flanked by a group of genetic information processing related genes and nucleotide metabolism.

Other species of *Pseudomonas* also encode an *idh* gene in their genome. *P. fluorescens*, *P. putida*, *P. stutzeri* and *P. syringae* show a similar size of *idh* and cytoplasmic location of IDH (the corresponding genes are coloured in red). As shown in figure 23B shows that *P. aeruginosa*, *P. fluorescens*, *P. putida* and *P. stutzeri* possess both isocitrate dehydrogenases in the same organization (opposite strands of DNA and direction of transcription). However, *P. syringae* only has one isocitrate dehydrogenase. The downstream genomic context of *idh* is similar in *P. aeruginosa* and *P. syringae*
as both species exhibit an operon of tRNA processing with *trmU*. The other three species (*P. fluorescens*, *P. putida* and *P. stutzeri*) show another operonic organisation with a tRNA processing centralised around *mnmA*, a tRNA-specific 2-thiouridylase.

![Diagram of genetic context of the *idh* gene in *P. aeruginosa* PAO1 and other *Pseudomonas* species](image)

Figure 23: Genetic context of the *idh* gene in *P. aeruginosa* PAO1 and other *Pseudomonas* species (A) The highlighted *idh* annotated gene is 2,226 bp encoding a 742 amino acid-long protein (stop codon included). *icd* is located immediately adjacent to the *idh* gene sharing a single intergenic region of 539 bp (B) *idh* orthologues in other *Pseudomonas* species display a similar genomic organisation with *icd* nearby and cold shock protein or ATP-binding protease component. The coloured legend indicates the subcellular localization of the encoded gene.
4. Isocitrate dehydrogenase (IDH)

The continuation of the bioinformatic study includes the study of the IDH protein sequence by BLAST analysis. This analysis excluded the *Pseudomonas* genus to focus only on all other bacteria that have been identified with an IDH protein. Here again, the results show a remarkably high sequence conservation ranging from 99 to 70%. The average length of the enzyme in all bacteria extracted from the BLAST analysis is similar; there is at most a shift of two or three residues with very little divergence in IDH. More interestingly, IDH has been characterized in two psychrophilic bacteria: *Colwellia psychrerythraea* and *C. maris*\(^{171}\) (*Colwellia maris* was previously described as *Vibrio* sp. strain ABE-1\(^{171}\)). The enzyme in these microorganisms is well adapted to low temperatures. Additionally, these bacterial species exhibit both isocitrate dehydrogenases; in this case, it could be reasonable to assume that IDH is selected to continue the TCA cycle reaction at low temperature since ICD is not functional below 25°C. Further investigation showed that a genetic element is responsible for low-temperature-inducible expression of IDH in marine bacteria. A 35 bp transcriptional silencer was identified in the *idh* promoter region as 5′-GTTATACCATAACGGACCTTAATTCTTTACGTAACA-3′. A Clustal Omega alignment of the silencer region with the *idh* promoter region in *P. aeruginosa*, as seen in figure 24 shows that the silencer is reasonably conserved. However, it is split into two and thus does not appear to be functional as IDH activity is significantly reduced at low temperatures in *P. aeruginosa*. Indeed, previous work done by a PartII student in the laboratory indicated that IDH activity is 2-fold decreased at temperatures below 27°C.

The amino acid sequence alignment generated with Clustal Omega in figure 25 shows that the IDH sequence is well conserved among diverse bacteria. All bacteria aligned exhibit consistent length of the protein around 740 residues. Furthermore, the residues involved in the binding of the substrate (highlighted in green) and magnesium cation (highlighted in blue) are the same and consistently belong to well-conserved regions of the amino acid sequence. Moreover, it is interesting to note that all the binding residues are an exact copy of the ones found in ICD. The evolution of IDH from ICD supposes the conservation of the active site features in IDH.

---

**Figure 24**: Alignment of *idh* promoter and transcriptional silencer. The alignment proves that it is conserved in *P. aeruginosa* in terms of sequence but it cannot be functional due to its separation.
Figure 25: ClustalOmega alignment of IDH amino acid sequence (highlighted residues in green: substrate binding, or blue: magnesium binding). IDH protein sequence is highly conserved between these four examples of divergent species. All substrate binding residues are similar to the ones found in ICD.
4.3. Purification of IDH

4.3.1. Preliminary results

The investigation of IDH in \textit{P. aeruginosa} (IDH\textsubscript{Pa}) started with the purification of the enzyme. Similar to ICD, the \textit{idh} gene was first PCR amplified from PAO1 genome then cloned into either pQE80 (yielding His\textsubscript{6}-tagged protein) or His-pMAL-c2x (yielding native protein after Factor Xa cleavage). The vector was then introduced into \textit{E. coli} DH5\textsubscript{α} for overexpression. The extraction and purification of the protein followed the same steps as already discussed in the previous chapter (section 3.3.1). I then worked with two types of protein, on the one hand is a His\textsubscript{6}-tagged IDH and on the other hand cleaved, native IDH.

The SDS-PAGE analysis after the purification using the \textit{idh} His-pMAL-c2x system is shown in the figure 26. The sample was initially denatured by boiling in 4X SDS-PAGE loading buffer supplemented with 200 mM of DTT for 5 min. From there, the sample was loaded into the gel before migration (see chapter 2 for migrating conditions). As shown in the figure 26, there is a clear single band for the purified IDH sample with a molecular mass slightly greater than the 75,000 Da ladder band. Alternatively, the Expasy toolbox predicted, from the amino acid sequence, a molecular mass of 81,634 Da. Given the margin between the 75,000 Da and 100,000 Da ladder bands, I concluded that the molecular mass of the concentrated solution of protein was consistent with its prediction. This means that the overexpression and purification processes are successful in producing the correct IDH enzyme from \textit{P. aeruginosa}. Furthermore, the initial denaturation of the sample before SDS-PAGE analysis results in a single band corresponding to a monomeric-state of IDH. The oligomeric status of the enzyme is investigated in the next sections 4.3.2 and 4.3.3. Regarding the His\textsubscript{6}-tagged version of IDH\textsubscript{Pa}, similar to ICD again, the SDS-PAGE analysis showed the presence of additional bands, probably degradation products (see appendix 2). The treatment of the IDH solution extracted from the \textit{idh} His-pMAL-c2x expressing system with Factor Xa systematically delivered a product with a single band but a much lower yield. The simple purification conducted on the His\textsubscript{6}-tagged IDH however, yielded a more concentrated solution.

4.3.2. Gel filtration analysis

Besides SDS-PAGE analysis, another complementary technique to estimate the oligomeric status of IDH\textsubscript{Pa} is gel filtration. A sample of 40 µg of concentrated IDH\textsubscript{Pa} solution was injected into a
TSK-Gel G3000 SWXL column previously equilibrated following the guidelines described in chapter 2. The molecular mass was estimated to be approximately 235,500 Da (see the molar mass moments values in figure 27). Furthermore, the narrow peak demonstrates that the sample, which was from the idh pQE80 expression construct, has few contaminants (see appendix 6 for the full results). These results show that IDH\(_{Pa}\) is assembled in a higher order oligomeric status. The molecular mass indicated from the gel filtration data would suggest that IDH is a trimer or an elongated dimer as gel filtration does not permit the detection of the shape of the protein. However, these results allow us to assert with confidence that in P. aeruginosa, IDH is not a monomer. IDH in other bacteria has usually been described as monomeric\(^{182,277-279}\). Two other species have been defined with either a dimeric or tetrameric IDH in M. tuberculosis\(^{167}\) and T. maritima\(^{280}\) respectively. P. aeruginosa seems to join a small group of bacteria that has two isocitrate dehydrogenases, in which IDH is not monomeric.

### 4.3.3. AUC data

Analytical ultracentrifugation with a set of sedimentation velocities (SV) was used to further examine the oligomeric status of IDH\(_{Pa}\). The results from gel filtration argued in favor of a non-monomeric IDH, but further investigation was required to better understand IDH oligomeric status. To do so, two hundred boundary profiles were recorded at three different concentrations (0.1, 0.5 and 1 mg/mL). The data were analysed as specified in chapters 2 and 3. The absorbance results obtained with the second concentration were computed, only thirty boundaries were used to fit the Stoke equation as seen in figure 28.

The sedimentation coefficient graph in figure 28 shows a clear peak corresponding to a molecular mass of 273 kDa with a sedimentation coefficient of 11 Sv. This suggests that IDH is at least a trimer. However, the frictional ratio of 1.148 is particularly low. Indeed, a frictional ratio lower than 1.2 for a complex quaternary structure, for a metabolic enzyme is unusual. Further discussion argued that the value is underestimated. This could be the product of interference from glycerol, potentially present in the sample solution despite a thorough dialysis prior the experiment. So, how is the frictional ratio underestimated and what are the consequences? Firstly, AUC is defined as matrix-free, macromolecular separation method in which molecules are separated based on their size and shape using a strong centrifugal field. That means that larger molecules will sediment faster than
4. Isocitrate dehydrogenase (IDH)

smaller molecules. Furthermore, hydrodynamic shapes, such as spheres or ellipses, will sediment faster than shapes with more friction, such as random coils or rods. In sedimentation velocity experiments, the sample is centrifuged at high speed to analyse the boundaries formed during sedimentation. In that case the sedimentation coefficient is defined Svedberg equation:

\[ s = \frac{M(1 - \bar{\nu}\rho_s)}{N_A f} \]

where \( s \) is the sedimentation coefficient, in Svedberg (1 Svedberg = \( 10^{-13} \) s), \( M \) is the molar mass, \( \bar{\nu} \) is the partial specific volume, \( \rho_s \) is the solvent density, \((1 - \bar{\nu}\rho_s)\) is the buoyancy term, \( N_A \) is Avogadro’s constant and \( f \) is the friction coefficient. The Svedberg equation complements the Stokes-Einstein equation (this equation details the diffusion coefficient of a particle undergoing motion in a fluid at uniform temperature) as such:

\[ s = \frac{M(1 - \bar{\nu}\rho_s) D}{RT} \]

\( D \) is the translational diffusion coefficient, \( R \) is the gas constant and \( T \) is the absolute temperature. So, this is the basis, it is evident that a higher molecular mass will result in a higher sedimentation coefficient meaning that the particle migrates faster through the solution column.

Figure 28: SV-AUC data of IDH protein. The figure shows 30 absorbance scans recorded at 280 nm for a protein sample of 0.5 mg/mL. The rmsd of the scans is 0.0082, within the noise rmsd. The most prominent, almost, single peak with a molecular mass of 273 kDa suggests that IDH protein is a trimer or potentially a tetramer. The frictional ratio in this experiment may have been underestimated due to the presence of glycerol.
4. Isocitrate dehydrogenase (IDH)

Sedimentation coefficients of macromolecules are usually obtained in a buffered solution to aid their solubility. However, buffer salts affect the solvent viscosity and density, which therefore affect the rate at which the boundary sediments. That means that sedimentation coefficients are corrected for solvent conditions. The sedimentation velocity analyzed by SEDFIT relies on the algorithm called c(s)- continuous distribution of sedimentation coefficients. The diffusion is eventually corrected by finding the frictional ratio:

\[
\frac{f}{f_0} = \frac{M(1 - \bar{v}\rho_s)}{N_A 6\pi \eta_0} \left(\frac{4\pi N_A}{3\bar{v}M}\right) \frac{1}{s}
\]

where \(f/f_0\) is the frictional ratio and \(\eta_0\) is solvent dynamic viscosity. The frictional ratio is the ratio of the drag of the macromolecule and the drag of a perfect sphere of equal molar mass and anhydrous volume. The relation between the frictional ratio and solvent dynamic viscosity supposes that in presence of glycerol, the viscosity increases. As a consequence, the frictional ratio decreases leading to an underestimated to a value of 1.148. However, despite these unsettling results, IDH certainly has a higher order quaternary structure more likely to be a tetramer or a dimer-tetramer equilibrium in solution.

4.4. Gene expression of idh

Similar to the work performed on ICD, the gene expression profile of idh was investigated after the cloning of the gene in pLP170 (a promoter-less lacZ fusion vector) and its introduction into PAO1. The cells were grown in minimal medium supplemented with either acetate or glucose. As detailed in section 3.4, the two carbon sources used should promote the partitioning of the carbon flux between either the glyoxylate shunt (acetate) or the TCA cycle (glucose). The study of the cell growth shows the effect of carbon source on PAO1 containing the Pidh::lacZ fusion compared with the control strain containing the promoter-less vector.

As shown in figure 29 (green line), the growth rate of the cells in minimal media supplemented with either acetate or glucose, containing the Pidh::lacZ fusion is not affected when compared with the control (black line). The growth rate of the transformed PAO1 with the Pidh::lacZ fusion is even slightly better in single carbon source in comparison with the control line. I have no explanation for this since the two strains differ only in the presence of Pidh in the PAO1 containing Pidh::lacZ construct. Pidh-PAO1 and WT-PAO1 contain the metabolic machinery that is the TCA cycle and glyoxylate shunt, they are capable to utilize any carbon source available in the environment.

As in the previous chapter, the values from the β-galactosidase assay are plotted as a bar chart in both panels of the figure. These values reflect the expression of Pidh::lacZ when the cells were grown in minimal medium supplemented with acetate (figure 29A) or glucose (figure 29B). The data show that the expression resulting from Pidh::lacZ construct (green) transformed into PAO1 is systematically slightly lower than for the Picd::lacZ construct (blue). In the case of cell growth in acetate, the maximum is attained at 7-8h of experiment with a peak of 1,300 nmol.min⁻¹.A₆₀₀⁻¹. After
this stage, there is a decrease of the signal detected until 10h of experiment. Overall, the profile is similar to what I have discussed with the Picd::lacZ construct, the signal obtained with the Pidh::lacZ remains lower.

When the cells are grown in minimal medium supplemented with glucose, there is a sharp increase of the production of 4-methylumbelliferone from 5h of growth. The maximum is attained after 7h of growth when the signal peaks at 890 nmol.min⁻¹.A₅₄₀⁻¹ corresponding to the end of the exponential phase and before entering the stationary phase. The increase and decrease of the Pidh::lacZ construct expression is similar to the Picd::lacZ construct expression. Even when the cells are grown in acetate, icd and idh promoters initiate transcription, the TCA cycle is not completely suppressed. The subsequent synthesis of ICD and IDH supposes that there is permanent carbon flux through the TCA cycle.

Suzuki et al. did a similar experiment to test the expression of icd and idh genes from Colwellia maris cloned into pIS102 or pIS202 respectively and introduced into E. coli DEK2004 (a mutant strain defective in icd). C. maris is another one of the few species that has both ICD and IDH. So, the expression profile of both encoding genes was investigated via Northern blot analysis, the chosen E. coli DEK2004 was grown in minimal medium supplemented with either acetate or glucose. The results obtained show similar expression of icd and idh. The monomeric isocitrate dehydrogenase (IDH equivalent) is consistently expressed at a lower level compared with the dimeric one (ICD equivalent) on rich (glucose) or limited (acetate) carbon sources after 2h of growth. These results highlight the complexity of the competition at the TGB branchpoint. These observations in C. maris and P. aeruginosa indicate that the two isocitrate dehydrogenases are actively expressed even when the bacteria are grown on a short-chain carbon source. This suggests that in an effort to circumvent oxidative stress, prokaryotic cells upregulate NADP⁺-dependent ICD and IDH to produce NADPH,
a reducing equivalent required to regenerate antioxidants, and α-ketoglutarate, a powerful antioxidant.

4.5. Kinetic analyses

4.5.1. Michaelis-Menten data

Since the specificity for the substrate and co-enzyme is the same as for ICD, untagged IDH was studied using the same method by monitoring the formation rate of NADPH at 340 nm. The kinetic studies at variable concentrations of NADP⁺ or isocitrate revealed that $K_{\text{isocitrate}} = 18 \, \mu$M and $K_{\text{NADP}^+} = 34 \, \mu$M (figure 30A-B). The $V_{\text{max}}$ for isocitrate is 0.027 mM.min⁻¹ and 0.031 mM.min⁻¹ for NADP⁺. Additionally, the oligomeric status of the protein described previously suggests that with multiple subunits and hence active sites, the enzyme is probably allosterically regulated.

The first order rate $k_{\text{cat}}$ for isocitrate, calculated as monomeric IDH, was 11,020 min⁻¹ (183 s⁻¹). This is consistent with a metabolic enzyme performing at high rate.

The monomeric IDH has been studied in multiple species, although kinetic data have been reported for only a few species. In *M. tuberculosis*, IDH is known to have a $K_{\text{isocitrate}} = 20 \, \mu$M and $K_{\text{NADP}^+} = 19.6 \, \mu$M. In *C. glutamicum* and *Rhodobacter sphaeroides*, IDH has a greater affinity with $K_{\text{isocitrate}} = 5 \, \mu$M and $K_{\text{NADP}^+} = 4 \, \mu$M, and $K_{\text{isocitrate}} = 3.3 \, \mu$M and $K_{\text{NADP}^+} = 2.5 \, \mu$M respectively for each microorganism. IDH in *C. glutamicum* is the only enzyme of the TCA cycle yielding α-ketoglutarate from isocitrate. The organisation of the TGB in *Corynebacterium* is similar to that in *E. coli*, IDH has a greater affinity ($K_m = 5 \, \mu$M) for isocitrate than ICL ($K_m = 280 \, \mu$M). The competition between ICD and ICL is limited, carbon flux is primarily directed through the TCA cycle. However, *P. aeruginosa* has ICD and IDH, the $K_m$ of the enzymes is similar with 25 µM and 18 µM respectively.
and they compete with ICL for isocitrate. This means that the TGB in *P. aeruginosa* is more complex, the mechanisms that regulate the partitioning of metabolite fluxes at the bifurcation of the oxidative TCA cycle and the glyoxylate shunt have to balance anaplerotic fluxes (glyoxylate shunt) and fluxes generating energy and biosynthetic precursors (TCA cycle).

### 4.5.2. Non-effect of AceK on IDH

ICD is clearly regulated through phosphorylation by isocitrate dehydrogenase kinase/phosphatase. This has been thoroughly studied in *E. coli* and demonstrated in *P. aeruginosa* in the previous chapter. IDH is functionally similar to ICD, and the serine phosphorylated in ICD is also present in IDH. An important question to address is therefore whether IDH is also phosphorylated by AceK. In order to investigate the effect of AceK on IDH, I incubated the two enzymes in the conditions described in chapter 2.

Figure 31 shows that after incubation, ICD is clearly inhibited by AceK. In comparison, IDH is not significantly inhibited by AceK even after 30 min. There was no statistical difference after analysing the data using a one-way ANOVA test based on a triplicate and *p* < 0.05. This means that the regulation of IDH is not phosphorylation-mediated. The regulation (if any) of IDH must be via another mechanism.

### 4.5.3. Screening of potential regulators

After confirmation of the lack of phosphorylation-driven regulation, I investigated the potential regulatory effect of organic acids and intermediates from the TCA cycle. This pursues the hypothesis of allosteric regulation of IDH (figure 32). I found that several compounds either activate or inhibit the activity of IDH. Like ICD, IDH is strongly inhibited by an equimolar mixture of glyoxylate and oxaloacetate. The mechanism of inhibition is through condensation of the two molecules to form oxalomalate. This has been reported in multiple organisms (*E. coli*, *C. glutamicum*, *R. vannii*, *B. flavum*) in which the mixture affects the dimeric as well as the monomeric isocitrate dehydrogenase. It is speculated that oxalomalate could form under physiological conditions and that this could play an important role in regulation of the tricarboxylic acid cycle.

IDH was also strongly activated by several compounds. Oxaloacetate, pyruvate and glyoxylate activate IDH by 188%, 194% and 243% respectively. The activation by pyruvate has been reported in *P. fluorescens* IDH; however in this microorganism IDH activity is increased by 35% only even though *P. aeruginosa* and *P. fluorescens* IDH share 66% identity. In *M. smegmatis*, glyoxylate was recently reported to stimulate IDH activity in a dose-dependent manner, enabling
4. Isocitrate dehydrogenase (IDH)

“rheostatic” control of flux through the glyoxylate shunt. A mixture of malate and oxaloacetate also increases IDH activity by 158%. However, this could be the effect of oxaloacetate alone, as malate did not show any effect by itself. Finally, there is no product inhibition of IDH by α-ketoglutarate. Despite the similarities in both active sites, the two enzymes respond differently to the same compound.

The effect of oxaloacetate, glyoxylate and pyruvate suggests that the regulatory mechanism of IDH involves a metabolite-mediated enzyme activation to achieve a balanced bifurcation of fluxes between the TCA cycle and the glyoxylate shunt. At that stage, in P. aeruginosa, I demonstrated that ICD is inhibited by AceK, but IDH is not. The latter is then activated by a subset of metabolites that act to maintain flux through the TCA cycle. Oxaloacetate and pyruvate emerge as potent activators of IDH, they signal to the cell that there are sufficient gluconeogenic precursors for biomass production. Overall, the effect of activators has not been reported for P. aeruginosa IDH before. These findings challenge the established regulatory paradigm in E. coli. This constant activation of IDH would ensure carbon flux through the TCA cycle to sustain secondary pathways related to oxidative stress and generate energy and biosynthetic precursors²⁸⁶,²⁸⁷.

4.5.4. Activation profile kinetics

The double-reciprocal plots for oxaloacetate and pyruvate in figure 33 show the activation effect of these molecules on IDH. As shown in figure 33B, oxaloacetate and pyruvate are uncompetitive activators, as the x-intersect and y-intersect are different from the control which also means that the regulator binds only to the complex formed between the enzyme and the substrate. So, these data indicate that both molecules bind to the enzyme-substrate complex and promote catalysis to produce α-ketoglutarate. Additionally, it is clear that in presence of the activators, the classical Michaelis-Menten curve has a sigmoidal profile (figure 33A). This demonstrates positive cooperative binding; the binding of the substrate or ligand at one binding site affects the affinity of other sites for their substrates. In that case, all three compounds had a pronounced effect on $V_{\text{max}}$, and thus $k_{\text{cat}}$ with only a small impact on $K_{\text{m}}$ (alternatively $K_h$). The kinetic parameters in the presence of 1 mM oxaloacetate are $K_h = 39$ µM and $V_{\text{max}} = 0.178$ mM.min⁻¹, and in presence of pyruvate $K_h = 42$ µM and $V_{\text{max}} = 0.207$ mM.min⁻¹. There is a 7-fold increase of $V_{\text{max}}$ compared with

Figure 32: Screening of potential regulators on IDH. Oxaloacetate, pyruvate and glyoxylate strongly activate IDH activity. Malate and oxaloacetate mixture activates IDH due to the oxaloacetate. IDH is also inhibited by glyoxylate and oxaloacetate forming oxalomalate. Values marked with * are significantly different (p<0.05 and n=3)
control. This means that $k_{\text{cat}}$ is similarly increased by a 6-fold factor ($k_{\text{cat}} = 1210$ s$^{-1}$ and $1408$ s$^{-1}$ in presence of oxaloacetate and pyruvate respectively whereas $k_{\text{cat}} = 183$ s$^{-1}$ for isocitrate).

Glyoxylate also displays a very strong sigmoidal response compared to control (figure 33A). The corresponding Lineweaver-Burk plot (figure 33B) suggests that glyoxylate acts as a non-competitive activator. The molecule can either bind to the free enzyme or the enzyme-substrate complex. The $V_{\text{max}}$ is significantly increased by a factor of 5 to $0.136$ mM.min$^{-1}$ in presence of glyoxylate compared with $0.027$ mM.min$^{-1}$ for isocitrate alone. Equally, $k_{\text{cat}}$ increases to $925$ s$^{-1}$ for glyoxylate.

However, the double-reciprocal plot for the co-factor shows a different profile (figure 33C). The $y$-intercept shows that oxaloacetate and pyruvate were non-competitive activators with respect to NADP$^+$. This indicates that oxaloacetate and pyruvate bind to the enzyme-substrate complex. The $K_m$ is not affected upon binding of the activator. Only the $V_{\text{max}}$ is changed by the binding oxaloacetate or pyruvate, and both show the exact same profile as both lines overlap. The $V_{\text{max}}$ is similarly increased in presence of oxaloacetate or pyruvate with values of $0.179$ mM.min$^{-1}$ and $0.203$ mM.min$^{-1}$ respectively, compared with $0.031$ mM.min$^{-1}$ for the control. Consequently, $k_{\text{cat}}$ is increased. There is a 6-fold increase of $k_{\text{cat}}$ upon addition of oxaloacetate or pyruvate ($k_{\text{cat}} = 1217$ s$^{-1}$ or $1280$ s$^{-1}$ in presence of oxaloacetate and pyruvate respectively whereas $k_{\text{cat}} = 210$ s$^{-1}$ for NADP$^+$).

With respect to NADP$^+$, glyoxylate is an uncompetitive activator as the $x$-intersect and $y$-intersect differ from the control. This means that the kinetics in presence of glyoxylate exhibits a different trend. The $K_m$ and $V_{\text{max}}$ in presence of glyoxylate are significantly increased upon binding of the activator, $K_m = 160$ µM and $V_{\text{max}} = 0.228$ mM.min$^{-1}$. This

---

**Figure 33**: Activators of IDH (A) Isocitrate dependency of IDH kinetics changes from hyperbolic to sigmoidal in presence of glyoxylate, oxaloacetate and pyruvate B) Effect of the regulators with respect to isocitrate shows non-competitive or uncompetitive activation (C) The same trend appears with respect to NADP$^+$. 

74
difference in the $V_{\text{max}}$ influences a 7-fold increase of the $k_{\text{cat}}$ up to 1544 s$^{-1}$ in presence of glyoxylate whereas $k_{\text{cat}} = 210$ s$^{-1}$ for isocitrate alone.

These results suggest that there is cooperativity in substrate binding, which is consistent with IDH adopting a higher-order structure in solution. Isocitrate and NADP$^+$ bind in the active pocket, so glyoxylate or oxaloacetate or pyruvate can bind to another site and increase the activity of the enzyme. The *in vitro* activity of IDH is under the control of metabolic intermediates that activate the enzyme, potentially to push isocitrate through the TCA cycle to generate sufficient energy and biosynthetic precursors that remain essential during catabolism of C$_2$-units nutrients source.

### 4.5.5. Thermal shift assay

I conducted a thermal shift assay (TSA) to study further the enzyme mechanism. Isocitrate and NADP$^+$ bind to IDH active site, while the regulators bind to subsites. The assay helps estimating the correlation of protein stabilization and affinity of binding. The thermal shift assay allows an efficient and reliable study of such interactions between a protein and regulators. The principle is based on the binding of low molecular weight ligands that can increase the thermal stability of the protein. The method involves a small amount of protein mixed with some SYPRO Orange dye. The dye binds to hydrophobic patches/denatured protein and fluoresces. As the temperature increases and the protein unfolds, it is easy to monitor the increase in fluorescence and determine a melting temperature. The temperature and fluorescence monitoring are done using a qPCR machine. It is then straightforward to derive the difference in the melting temperatures, $\Delta T_m$, between the control and the protein in presence of regulators.$^{288,289}$

The control (protein alone) serves as a baseline to determine if there is indeed a shift in the denaturation temperature of the complex by adding a ligand. From the figure 34, the only notable response is seen when the substrate is present with a positive shift of 6°C. So, the binding of the substrate strongly stabilizes the enzyme. However, a difficulty of fluorescence techniques in analysis
of protein–ligand interactions is to define an adequate threshold. It is common in compounds library screening to use a cut-off of $1-2^°C_{212}$.

As seen in the figure, the rest of the compounds tested did not affect protein stability: neither pyruvate nor oxaloacetate or even the mixture of glyoxylate and oxaloacetate. Only the substrate is significantly different compared to control. Isocitrate binds to the active site, but the activators are of a non-competitive or uncompetitive type, thus they bind to another site on the enzyme. The design of the experiment needs to be implemented, I tested each regulator alone, without the substrate in the mixture. Since oxaloacetate and pyruvate are non-competitive activators, they will not bind to IDH in the absence of isocitrate. The results showed that there is indeed no regulator binding without isocitrate present, meaning that the $\Delta T_m$ is barely affected. Further study of IDH stability and IDH mechanism supposes a TSA with the substrate and regulator in the reaction mixture.

4.6. Crystal structure of IDH

4.6.1. Structure description

I solved the structure of IDH by molecular replacement using the structure of *Mycobacterium smegmatis* IDH (PDB 4ZDA) as well as an I-TASSER-generated model as templates. *Pseudomonas* and *Mycobacterium* IDH share 68% identity, making the *M. smegmatis* structure a suitable template. However, due to difficulties in treating the diffraction data (unit cell content), the I-TASSER model was also used to calculate the phasing. The structure was refined to 2.7 Å, with $R_{cryst}$ 20.80% and $R_{free}$ 26.67%. The analysed crystal contains two non-symmetrical molecules of IDH (figure 35A); the chain B (light green) is tilted and showed regions of high flexibility due to static disorder. Comparison of the structure of the chains A and B reveal large conformational differences. As shown in the figure 35B, the overall core of the enzyme is consistent between the two chains, but the smaller domain in chain B is shifted. This is especially noticeable on the helices and loops represented at the front, in particular $\alpha_{26-28}$. During model building of the two chains of figure 35A, this was particularly obvious as extended sections of these regions had to be redrawn and relocated within the electron density. Despite having two closely juxtaposed molecules present in the asymmetric unit, a PISA analysis indicated that there is no specific interaction that could result in the formation of a quaternary structure.

A single molecule of IDH resembles the dimeric ICD (figure 35C). An extensive study of *Azotobacter vinelandii* IDH proved that the substrate binding residues are absolutely identical and that they share a common folding topology with ICD$_{Ec}$. Crystal structure analysis of ICD$_{Ec}$ has shown that two identical subunits are related by a crystallographic 2-fold axis. A similar pseudo-2-fold axis exists in the larger domain of IDH$_{Av}$ and IDH$_{Pa}$. Indeed, the larger domain cannot be separated into two equivalent regions, because of the insertions of four $\alpha$-helices ($\alpha_{10-13}$ in *P. aerugiñosa*). Crystal structure of ICD$_{Pa}$ and IDH$_{Pa}$ revealed that the tertiary structures is as conserved. Such a fusional structure suggests that IDH is the result of protein monomerization by a domain duplication from
ICD, IDH originated from the partial duplication of the shared ancestral gene, rather than the convergent evolution from different genes\textsuperscript{185}.

I crystallized \textit{P. aeruginosa} IDH with NADP\textsuperscript{+} but only chain A showed electron density suitable to fit a molecule of the co-factor (figure 35D, ball and stick representation). Additionally, despite the fact that no other ligand was added in the crystallisation conditions, I was able to detect electron density for a molecule of \(\alpha\)-ketoglutarate deeper in the active site pocket. The arrangement of the two ligands fits nicely with the structure solved for \textit{M. tuberculosis} IDH (PDB 5KVU, data not published). In the \textit{M. tuberculosis} structure, a small C\textsubscript{4} molecule (malate or succinate) was modelled deep inside the pocket behind a molecule of NADP\textsuperscript{+}. This supports the kinetic data for an ordered sequential mechanism in which isocitrate binds first. This process causes a conformational change required for binding the co-factor, NADP\textsuperscript{+}, which caps the active site cleft.

Finally, each chain can be divided into two domains displayed in orange and purple in the figure 35D. The smaller domain (orange) contains the N- and C-terminus; it is composed of fourteen \(\alpha\)-helices (\(\alpha\)1-6 and \(\alpha\)21-28) along with four \(\beta\)-strands (\(\beta\)1-3 and \(\beta\)24). The larger domain (purple)
comprises thirteen α-helices (α7-20) and nineteen β-strands (β4-23), including the eight β-strands forming the clasp-like domain (refer to appendix 7 for a sequence/secondary structure alignment).

The NADP⁺-unbound chain B structure parameters reveal high-temperature factors, especially for the small domain, which suggests static disorder in the crystal in lower resolution of the diffraction data. This means that the polypeptide chain B was resolved for 737 of its 741 residues, the first four residues had no corresponding electron density to be modelled. Overall, this reinforces the importance of NADP⁺ to stabilise IDH during crystallisation and how the conformation of IDH is altered upon binding of the co-factor.

### 4.6.2. Structural comparison and domain movement

There are fewer crystal structures of IDH than that of ICD in the protein databank. Further study of *P. aeruginosa* IDH compares the crystal structure with IDH of *A. vinelandii*, *C. glutamicum*, *M. smegmatis* and *M. tuberculosis*. The superposition of them demonstrates that they have a very similar conformation. They overlay exactly the same secondary structure features despite a slight shift of some secondary structures. As seen in figure 36, the superposition of the ribbon representation of all structures demonstrates that despite the obvious differences in the length of the sequences, the overall shape of the protein is conserved in these microorganisms.

The rmsd value after superposing the different structures ranges from 0.89 Å to 3.53 Å. The results of the superposition can be classified in three categories from three averages of rmsd of 1 Å, 2 Å and over 3 Å. The superposition of the structures of *P. aeruginosa* with the other structures available crystallised with NADP⁺ shows the closest average rmsd (less than 1Å). The rmsd between IDH<sub>Pa</sub> and IDH<sub>Mt</sub> (PDB 5KVU, Mycobacterium tuberculosis) is of 0.89 Å while IDH<sub>Pa</sub> and IDH<sub>Av1</sub> (PDB 1J1W, Azotobacter vinelandii) shows 0.97 Å of difference. This is demonstrated in figure 36A in which the tight superposition of the three structures is obvious. The outer α-helices overlay nicely, with some loops showing a little bit more flexibility. Both structures are in complex with NADP⁺; *M. tuberculosis* has additional molecules of either malate or succinate thus a lower rmsd. This might explain why it closes the active pocket and tightens the monomer in the presence of either one or both of the substrates. This is also consistent with the model in *P. aeruginosa* in which α-ketoglutarate is present. With the molecule of NADP⁺ sitting at the entrance of the active site pocket, it is reasonable to think that it closes the pocket more tightly. The second category with an average rmsd of 2 Å compiles two different species crystallised with the same ligands, isocitrate and an atom of manganese. These entries, IDH<sub>Av2</sub> (PDB 1ITW, Azotobacter vinelandii) and IDH<sub>Ms</sub> (PDB 4ZDA, Mycobacterium smegmatis) show a rmsd of 1.89 Å and 2.02 Å, respectively, when overlaid onto IDH<sub>Pa</sub>. As seen in figure 36B, the difference remains subtle when comparing elements of the secondary structures in the core of the enzyme. However, the loops and α-helices on the left and bottom-right of the overlay show some more obvious deviation. The main difference is the obvious presence of the substrate in the active site pocket. The location of isocitrate is also very different: being more buried into the active site, isocitrate might not trigger the same closing effect on the two sub-domains of
4. Isocitrate dehydrogenase (IDH)

Because there is no NADP+ on top of isocitrate that would pull together the monomer, the active site pocket remains semi-open leading to a greater difference when overlaying the models. Finally, the third category merges the two structures of Corynebacterium glutamicum IDH, IDH\textsuperscript{Cg1} and IDH\textsuperscript{Cg2}, respectively, (PDB 3MBC and 2B0T) against Pseudomonas aeruginosa IDH. The rmsd of IDH\textsubscript{Pa} vs. IDH\textsuperscript{Cg1} and IDH\textsuperscript{Cg2} is the most different with values of 3.42 Å and 3.53 Å respectively. This deviation is especially noticeable in figure 36C where the arrow points out the shift of a major α-helix in the small domain of IDH. The crystallisation conditions for IDH\textsuperscript{Cg1} and IDH\textsuperscript{Cg2} were different. IDH\textsuperscript{Cg1} co-crystallised with NADP(H) and an atom of magnesium, whereas IDH\textsuperscript{Cg2} co-crystallised with only magnesium. However, it is clear that the higher rmsd is probably due more to an intrinsic difference in the overall conformation of the protein than to an effect of the presence or absence of ligands. The molecule of NADP+ alone does not pull together the active pocket as seen previously. The comparison of bacterial IDH structures in the PDB and the new IDH\textsubscript{Pa} proved that the crystallisation conditions play a major role in the conformation of the enzyme. This also suggests that the binding of the substrates is accompanied by structural rearrangements.

4.6.3. The “clasp” domain

The phenomenon of monomerization forming IDH from ICD has been extensively studied in Azotobacter vinelandii and there are many aspects to it. Firstly, with respect to the binding of the substrate, all the residues are very well conserved (this will be discussed in the next section). An
interesting feature of the monomerization of IDH from ICD is the evolution of the clasp domain in ICD to a Greek key motif in IDH. Initially, the clasp domain is the combination of an α-helix (α5) and two β-strands (β8-9) from each chain as seen in figure 37C. From the superposition of IDH and ICD, this clasp is replaced by eight β-strands (β6-9 and β16-20) linking the larger domain to the smaller domain in a similar way the clasp is linking the two protomers of ICD. Each domain of IDH has four β-strands that form a double Greek key motif (figure 37A-B). Such a stereotypic β-sandwich consists of two pleated β-strands packed against each other; as a result, the cross-section has an elliptic shape290. The Greek key motif is an evolutionarily conserved super-secondary protein structural fold that offers structural compactness and high intrinsic stability against stress291.

Structure-based sequence alignment in figures 37D-E shows the residues forming the two Greek key motifs. The β6-9 section, Glu186-Ser228 (figure 37D), represents either the purple or yellow ribbon in the panels A and B. Of the 42 amino acids that make the motif, the alignment shows seven conserved residues. The second β16-19 section, Asp424-Lys462 (figure 37E), is much more conserved with thirteen residues included in the four β-strands.

Figure 37: Equivalent of the clasp-domain in IDH (A) IDH₀, the same chain goes back and forth from the large to the small domain creating two Greek key motifs sitting on top of the structure (B) IDH₆, shows another Greek key motif (C) ICD₀, each chain supplies one α-helix and two β-strands to form the clasp domain (D) Clustal alignment of the first motif Glu186-Ser228 (E) Clustal alignment of the second motif Asp424-Lys462

4.6.4. The active pocket and the binding configuration

The electron density corresponding to a molecule of NADP⁺ (present in the crystallisation conditions) was detectable. However, only one monomer showed density to fit such a molecule. The overall flexibility and disorganisation of chain B paired with poorer electron density could not support the modelling of a molecule of NADP⁺ per chain. Another surprise resulted from correcting the arrangement of NADP⁺, shifting the nicotinamide ring into an anti-configuration instead of a syn-configuration liberated enough unmodeled matrix to fit a C₅ molecule (probably α-ketoglutarate). Isocitrate would not fit as it has an extra carboxylic acid function on C(3), while α-ketoglutarate was accepted after rounds of refinement. NADP⁺ was located in the active cleft between the two domains, whereas α-ketoglutarate was located deeper in the cleft (closer to the β-sandwich). As shown in the
4. Isocitrate dehydrogenase (IDH)

Figure 35D, both molecules are located in the active site as a ball and stick representation with the electron density grid. The ligands lie perfectly at the junction between the two domains. The molecule of NADP⁺ depicted in a beige grid shows through the electron density with a contouring of 1.7σ, and the same contouring was applied to the molecule of α-ketoglutarate in light blue. The electron density for NADP⁺ is very clear whereas the electron density proposed for α-ketoglutarate is more difficult to interpret. The decision to add this ligand was based on: 1) the length of the five-carbon long chain for the space given in the electron density, 2) the overall orientation in the active site pocket and 3) improved refinement statistics.

A more detailed exploration of the active site and the neighbourhood of NADP⁺ and α-ketoglutarate gives insights into the molecular mechanisms employed by the enzyme. Similar to ICD, the residues for the binding of the substrate and co-factor are well characterized in IDH. There are seven residues involved in binding the substrate. The figure 38 shows an overlay of the following species: A. vinelandii IDHₐᵥ (PDB 1ITW or 1J1W), M. smegmatis IDHₘₙ (PDB 4ZDA) and M.
**4. Isocitrate dehydrogenase (IDH)**

*P. aeruginosa* IDH₄ₛ (PDB 5KVU) with the same colour coding. IDH₄ₛ1ITW co-crystallised with Mn²⁺ and isocitrate, IDH₄ₛ1J1W co-crystallised with NADP⁺, IDH₆ₛ co-crystallised with Mn²⁺ and isocitrate and IDH₆ₛ co-crystallised with succinate or malate and NADP⁺. Crystal structure analysis of the active sites of IDH co-crystallised with isocitrate (IDH₄ₛ, IDH₆ₛ) or a C₄ molecule (α-ketoglutarate IDH₆ₛ or succinate/malate IDH₆ₛ) or NADP⁺ indicates that the residues involved in binding are very well conserved. The aspartate residues binding magnesium or manganese (when present in the crystallisation conditions) overlay with the same fidelity.

As seen in **figure 38A**, α-ketoglutarate is embedded deep into the active site cleft. The active site comprises secondary structure elements such as α6, α20, β4 and flexible loops. The residues in IDH₆ₛ, Ser133, Asn136, Arg140, Arg146, Lys255, Tyr420 and Arg547 are the equivalent of the residues Ser115, Asn117, Arg121, Arg131, Lys232*, Tyr162 and Arg155 in ICD₆ₛ (see **figure 39B**). Crystal structure comparison of IDH₆ₛ, IDH₄ₛ, IDH₆ₛ and IDH₆ₛ in **figure 38B** shows that the residues involved in binding the substrate (alternatively α-ketoglutarate in IDH₆ₛ) are similar. The most distinct difference is between the serine residues of *P. aeruginosa* IDH and *M. smegmatis* IDH. The distance between the Cα of the residues is 2.28 Å. The rest of the residues barely show any difference in their orientation, angle or bond length. Overall, IDH₄ₛ and IDH₆ₛ superpose very tightly; apart from the distinct difference of the serine residue, the average discrepancy between residues is 1.3 Å. This applies to the following residues: Asn136, Arg140, Arg146, Tyr420 and their equivalent in the other structures. The underpinning element is the slightly more open conformation of the active site in IDH₆ₛ compared with IDH₄ₛ or IDH₆ₛ. The presence of α-ketoglutarate instead of isocitrate seems to have released some of the tension and the enzyme is ready to liberate the product. This could be further demonstrated by the different angle of Lys257 in *Mycobacterium tuberculosis* that is 5.42 Å away from Lys255 in *Pseudomonas aeruginosa*. The fact that IDH₆ₛ has been co-crystallised in presence of a C₄ molecule supposes that the residue is too far away and cannot interact with the ligand. This also indicates that IDH₆ₛ active site is an intermediate state. Ser133 and Arg140 are displaced outward of the active site, as α-ketoglutarate is ready to exit the active site. The superposition of these similar models highlights the conformational intermediate states of the active site operating upon binding of isocitrate or α-ketoglutarate.

Further investigation on the co-factor binding indicates that the residues binding NADP⁺ are similarly conserved (see **figure 38D**). Similar to the substrate binding, the active site pocket configuration is maintained across the species studied. The identification of the residues is based on the analysis of the neighbourhood of the NADP⁺ ligand using the dedicated CCP4mg tool with a threshold of 4 Å. The results were cross-checked with the few publications addressing NADP⁺ binding study in bacterial species. Of all the structures available in PDB, only two of them had NADP⁺ embedded in their active sites, namely *Azotobacter vinelandii* (IDH₄ₛ PDB 1J1W) and *Mycobacterium tuberculosis* (IDH₆ₛ PDB 5KVU). Before investigating the binding residues, the ligand itself is first analysed. A superposition of the three molecules of NADP⁺ from the three structures show that the rmsd between IDH₆ₛ and IDH₄ₛ is 0.75 Å while it is only 0.29 Å between NADP⁺ in
4. Isoctirate dehydrogenase (IDH)

IDH_{Pa} and IDH_{Mt}. In general, the known residues involved in binding the co-factor are Asn85, Ser87, the triad Gly584-Ser585-Ala586, His589, another triad Arg600-Trp601-Asp602 and finally Arg649. The majority of these residues are an almost perfect match in terms of superposition. There is barely any significant change in the orientation, and angle of the main or side chains. Similar to ICD, some residues are dedicated to binding the ADP moiety while others are binding the nicotinamide ring. It seems that the adenine moiety is close to the side chains of His589 and Trp601 and the main chain of Asp602, which means that hydrophobic interactions are primarily involved. Arg649 plays a major role in binding the phosphomonoester, even though the actual oxygen bound is different depending on the structure. On the other hand, the nicotinamide ring interacts with Asp87 and Ser89 even though there is no apparent hydrogen bond formation. The proximity analysis with a 4 Å threshold highlighted these residues and this is consistent with the study of IDH_{Ac}^{186}.

4.6.5. Why is AceK inefficient on IDH?

I have concluded that despite the fact that the serine phosphorylated in ICD is conserved in IDH, this alone is not sufficient to allow AceK-dependent inhibition of IDH. A structural study of ICD_{Pa} and IDH_{Pa} based on the knowledge extracted from a crystallised complex of ICD and AceK in Escherichia coli (PDB 3LCB) illustrates the protein-protein contact between ICD and AceK and why IDH is not phosphorylated by AceK. The remarkable conservation of the active site demonstrates that the secondary structure along the substrate binding residues is very well preserved. Figure 39A shows that the secondary structure elements making the cleft of ICD and IDH are identical. The multiple loops, α-helices and β-strands overlay nicely. Furthermore, all the binding residues are remarkably conserved in the active site pocket in a similar conformation, including the phosphorylation site (Ser115 in ICD vs. Ser133 in IDH) (figure 39B). The Cα displacement between ICD and IDH residues is the result of conformational changes upon binding of α-ketoglutarate in IDH. The active site in IDH is in a closed conformation, i.e. Arg140 (IDH) is displaced by 3.6 Å compared with Arg121 (ICD) that does not bind to a ligand.

The recognition of ICD by AceK has been extensively studied^{190,292}. AceK has a Substrate Recognition Loop (or SRL) (Glu484-Pro510) which extends deeply into the active site cleft of ICD. Further analysis of the contact between the two enzymes from E. coli shows that there are two recognizable motifs in ICD: the P-loop (phosphorylation loop, Leu103-Leu114) which carries Ser113 and the ARS (AceK Recognition Segment, Gly254-Asn268). These two segments are present in P. aeruginosa ICD in a similar conformation advocating for the AceK-mediated phosphorylation (figure 39C). The two loops have the same orientation in E. coli ICD and P. aeruginosa ICD and the phosphorylation site, Ser113 in E. coli, Ser115 in P. aeruginosa is conserved. The P-loop seems flexible depending on AceK presence. In the case of ICD_{Pa}, it remains disengaged and in a more open configuration. In ICD_{Ec}, the same motif is closer to the SRL; this seems to seal the active site cleft before phosphorylation. This flexibility might have a major role in the configuration of the active site depending on the presence or absence of substrates and clearly for phosphorylation/dephosphorylation.
4. Isocitrate dehydrogenase (IDH)

purposes. However, and despite the fact that IDH is structurally similar to ICD overall, inhibition by phosphorylation does not happen in IDH. A superimposition study of the two structures (ICD$_{Pa}$ and IDH$_{Pa}$) shows that the ARS and the P-loop are missing in IDH. The P-loop is replaced by two sections that include four α-helices ($\alpha$2-$\alpha$3 and $\alpha$4-$\alpha$5), while the ARS is also replaced by a thread of five α-helices ($\alpha$10-14) (figure 39D). This change in the active site cleft makes it impossible for AceK, and in particular the SRL, to make close contact with its substrate. Additionally, there is a strong steric effect that prevents AceK from being able to engage into contact. IDH is by nature bulkier than ICD, making it impossible for AceK to approach the active site. Further, it has already been demonstrated that even a slight alteration of either the ARS or P-loop prevents AceK from recognizing ICD$^{190}$.

4.7. Discussion

IDH remains poorly studied compared with ICD. The E. coli model used in microbiology allowed a thorough investigation of dimeric ICD. IDH however, has triggered more attention, in particular in Mycobacterium tuberculosis, studying the carbon fluxes between the TCA cycle and glyoxylate shunt during infections. Pseudomonas aeruginosa is one of the few micro-organisms that possess the two isocitrate dehydrogenases, and I studied both isozymes in this work. The distribution
of IDH within pathogenic bacterial species demonstrates that the enzyme is less prevalent than ICD. For example, out of the 43 species selected to construct all phylogenetic trees, only 13 of them encode IDH. In species that do encode IDH, the enzyme is particularly well conserved, with notably several psychrophilic species. IDH is known in psychrophilic bacteria to be well-adapted to cold, which gives the advantage of a functional TCA cycle in a cold environment.

To start this work on IDH, I successfully cloned and overexpressed the enzyme either with a N-terminal His\textsubscript{6}-tag or in the native form (after Factor Xa digestion). However, the gel filtration and AUC data showed a discrepancy in the native molecular mass. The molecular mass obtained by gel filtration supported the presence of a trimer or elongated dimer. In contrast, the AUC data were consistent with a tetramer in solution. Overall, both approaches suggest that IDH has a higher-order quaternary structure. It is no longer possible to label it as a monomer, at least not in \textit{P. aeruginosa}. The uncertainty in the oligomeric status requires further investigation. The addition of the substrate or a regulator did not refine the molecular mass. I recommend the production of another solution of IDH before any new AUC-SV experiment, furthermore the addition of the substrate and a non-competitive regulator could improve the results.

The gene expression of \textit{Pidh::lacZ} alongside \textit{Picd::lacZ} indicated the effect of carbon sources on the transcription of the genes and the equilibrium between the expression of \textit{idh} and \textit{icd}. I grew the cells in acetate to redirect carbon flux through the glyoxylate shunt, or in glucose to promote the TCA cycle. The two strains containing the isocitrate dehydrogenase promoter constructs showed better growth in minimal medium supplemented with single carbon source compared with the control. I have no explanation for this since these are not mutant strains, the control and \textit{lacZ} reporter fusion strains should express a similar level of metabolic enzymes. The β-galactosidase assay indicated that the \textit{Picd::lacZ} construct is consistently transcribed at a slightly higher level compared with the \textit{Pidh::lacZ} construct when the cells are grown in acetate or glucose. This would suggest that ICD remains the preferred enzyme used in the TCA cycle to yield \textit{α-ketoglutarate}. IDH appears to be less dominant as it is less transcribed during bacterial growth, but nonetheless, essential even during growth on limited carbon sources as ICD and IDH are consistently transcribed. This indicates that ICD and IDH are vital for fuelling the nutritional requirements of persistent \textit{P. aeruginosa}, however these metabolic enzymes emerge as mediators of antioxidant defence. Hypothetically, ICD, IDH and \textit{α-ketoglutarate} dehydrogenase, as part of a stratagem to combat oxidative stress, respond to an oxidative insult with an increased production of NADPH. The latter acts as a reductive force required to nullify the oxidative environment triggered by the oxidation of NADH, maintains anti-oxidative systems in reduced/active states; while the former produces ROS via the respiratory chain. A reductive environment promoted by NADPH is essential for normal cellular activity. So, the creation of a reductive environment mediated by the enhanced formation of NADPH is instrumental to the survival of \textit{P. aeruginosa} in an oxidative environment and the manipulation of metabolic pathways and enzymes involved in NADH/NADPH metabolism is pivotal to ensure that ROS levels remain in the non-toxic range. Although ROS-detoxifying enzymes, such as catalase and SOD, are important
in nullifying oxidative stress, the efficacy of these enzymes depends on the reductive environment promoted by NADPH. To survive, high NADPH counters oxidative stress and low NADH limits ROS formation from the respiratory chain.\textsuperscript{222,293}

The kinetic parameters reveal a $K_m$ for isocitrate and NADP$^+$ that are very similar to ICD. This means that ICD and IDH compete in the same range of affinity for the substrate and the cofactor. In bacteria that have IDH or ICD only, the affinity for isocitrate is greater, ensuring carbon flux through the TCA cycle. However, in Pseudomonas aeruginosa, the similar affinity of ICD and IDH for isocitrate addresses the question of the regulation of IDH aside of the regulation of ICD. The mechanisms that regulate the partitioning of metabolite fluxes at the bifurcation of the oxidative TCA cycle and the glyoxylate shunt emerge. Strict regulation of fluxes at the TGB is essential during catabolism of C$_2$-units derived from fatty acids in order to balance anaplerotic fluxes (glyoxylate shunt) and fluxes generating energy and biosynthetic precursors (TCA cycle). I demonstrated that ICD (like in other species, particularly E. coli) is regulated by the AceK-mediated phosphorylation. Inactivation of ICD by phosphorylation secures a rerouting of the carbon flux through the glyoxylate shunt instead of the TCA cycle. In the case of IDH, AceK does not have any effect on its activity. The incubation of the two enzymes with ATP did not result in any loss of activity of IDH over time. The screening of regulators became then of particular interest to investigate any allosteric effect of metabolic intermediates.

### Table 18: Kinetics parameters of IDH

<table>
<thead>
<tr>
<th></th>
<th>$K_m$ (mM)</th>
<th>$V_{max}$ ($\Delta$mM/s)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_m$ (s$^{-1}$/µM)</th>
<th>$K_i$ (mM)</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>isocitrate</td>
<td>17.70</td>
<td>0.0005</td>
<td>183.67</td>
<td>10.38</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>glyoxylate</td>
<td>20.67</td>
<td>0.0023</td>
<td>925.17</td>
<td>44.76</td>
<td>n/a</td>
<td>non-competitive</td>
</tr>
<tr>
<td>oxaloacetate</td>
<td>39.16</td>
<td>0.0030</td>
<td>1210.88</td>
<td>30.92</td>
<td>n/a</td>
<td>non-competitive</td>
</tr>
<tr>
<td>pyruvate</td>
<td>41.95</td>
<td>0.0035</td>
<td>1408.16</td>
<td>33.57</td>
<td>n/a</td>
<td>non-competitive</td>
</tr>
<tr>
<td>NADP$^+$</td>
<td>34.48</td>
<td>0.0005</td>
<td>210.88</td>
<td>6.12</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>glyoxylate</td>
<td>160.40</td>
<td>0.0038</td>
<td>1544.22</td>
<td>9.63</td>
<td>n/a</td>
<td>uncompetitive</td>
</tr>
<tr>
<td>oxaloacetate</td>
<td>29.00</td>
<td>0.0030</td>
<td>1217.69</td>
<td>41.99</td>
<td>n/a</td>
<td>non-competitive</td>
</tr>
<tr>
<td>pyruvate</td>
<td>49.46</td>
<td>0.0034</td>
<td>1380.95</td>
<td>27.92</td>
<td>n/a</td>
<td>non-competitive</td>
</tr>
</tbody>
</table>

I demonstrate that IDH activity is mediated by metabolite-based regulation. The most intriguing result from the screening of regulators is the absence of identified physiological inhibitors (apart from oxalomalate). Interestingly though, three molecules have a potent activatory effect on IDH. Glyoxylate, oxaloacetate and pyruvate are very effective activators as they increase IDH activity by up to 2.4-fold. None of these molecules showed a competitive effect against either isocitrate or NADP$^+$, instead they showed a non-competitive or uncompetitive effect indicating that the binding of the compound occurs in a site other than the active pocket. All three compounds changed the isocitrate dependency of IDH kinetics from hyperbolic to sigmoidal. These three compounds had a pronounced effect on $k_{cat}$, so initially $V_{max}$, but only a small impact on $K_m$ (even in the case of non-competitive activators). However, most allosteric enzymes are K-type, which means that the $K_m$ is
altered. Only a small subset of enzymes are V-type, with a pronounced change of \( V_{\text{max}} \). In the case of IDH, the significant increase in \( k_{\text{cat}} \) exemplified the V-type trend of IDH. The latter category represents only \( \sim 1\% \) of all studied enzymes\(^{294}\).

Activation of IDH mediated by glyoxylate, is ideally suited because glyoxylate is a product of the ICL-catalyzed cleavage of isocitrate to glyoxylate and succinate. According to this scheme, if glyoxylate levels rise, the corresponding increase in IDH activity ensures that carbon flux will shift towards the TCA cycle, leading to a decrease in glyoxylate levels. If glyoxylate levels fall, the corresponding decrease in ICD activity ensures that carbon flux will shift back towards the glyoxylate shunt, leading to a recovery of glyoxylate levels. The glyoxylate-mediated activation of IDH functions as a molecular rheostat to maintain the proper balance of fluxes between the TCA cycle and the glyoxylate shunt\(^{295}\). Activation of IDH mediated by gluconeogenic substrates oxaloacetate and pyruvate is similarly suited. Oxaloacetate is the product of the oxidation of malate by malate dehydrogenase with reduction of NAD\(^+\) to NADH. According to this scheme, if oxaloacetate levels rise, so do NADH levels, the consequent increase of IDH activity ensures carbon flux through the TCA cycle, leading to increased levels of NADPH. This secures a NADH/NADPH homeostasis to circumvent oxidative stress. If oxaloacetate levels fall, NADH levels fall and NADPH production is no longer required to nullify oxidative stress. Pyruvate is the end product of glycolysis and the product of the dephosphorylation of PEP by pyruvate kinase. Increase in pyruvate levels is a metabolic cue suggesting growth on media with glucose, so catabolism of C\(_6\)-units pushes flux through the TCA cycle as there is no risk in losing two carbons in the form of CO\(_2\). The corresponding increase in IDH activity fulfils the cataplerotic functions of the TCA cycle producing the gluconeogenic precursors to serve energy and to grow biomass.

I solved the first X-ray structure of IDH from *Pseudomonas aeruginosa*. The structure presented shows two non-related protomers of IDH. The co-factor NADP\(^+\) crystallised in the IDH active site pocket, as did a molecule of \( \alpha \)-ketoglutarate. Comparison of IDH\(_{Pa}\) with several bacterial and fungal IDH indicates that the structure of IDH is well conserved. Further investigation of the active site highlighted that the residues involved in substrate, co-factor and metal binding are conserved. The remarkable aspect of the work on IDH\(_{Pa}\) is the binding of only one molecule of NADP\(^+\) out of the two protomers represented. The bound protomer of IDH\(_{Pa}\) with NADP\(^+\) and \( \alpha \)-ketoglutarate demonstrates that the active site is in an intermediate state of energy. The rmsd value from the superposition of IDH\(_{Pa}\) with IDH co-crystallised with isocitrate or IDH\(_{Pa}\) with IDH co-crystallised with NADP\(^+\) demonstrated that binding of the ligands triggers conformational changes. The hinge-motion of the smaller domain of IDH closes the active site of the enzyme. The presence of NADP\(^+\) and \( \alpha \)-ketoglutarate in IDH\(_{Pa}\) activates this hinge-motion, but only partially as the C\(_5\) compound disengages binding with several residues, leaving it ready to exit the active site.

Work done on IDH from *Azotobacter vinelandii* investigating the evolutional pathway from ICD to IDH highlighted the similarity of the active site between the two enzymes. In *P. aeruginosa*, the substrate binding residues are identical between ICD and IDH, including the phosphorylation
site. I demonstrated that IDH is not inhibited by phosphorylation. The recognition motifs (ARS and P-loop) lining the ICD active site and recognised by AceK are not present in IDH. ICD and IDH have a similar affinity for isocitrate and are independently regulated via two mechanisms. The presence of these two isozymes remains unclear. The architecture of the regulatory circuit that controls bifurcation of fluxes at the branchpoint for the TCA cycle indicates that ICD and IDH are essential for growth on C₂-units and C₆-units. Phosphorylation-mediated inhibition of ICD but cross-activation of IDH by glyoxylate, oxaloacetate and pyruvate suggests that metabolic cues maintain a molecular rheostat to balance fluxes between the TCA cycle and the glyoxylate shunt, in conjunction with NADPH-NADH homeostasis. Further investigation with Δicd, Δidh or Δicd-idh mutant strains of PAO1 (strains produced by Dr. Stephen Dolan in Dr. Martin Welch laboratory) would address the question if ICD and IDH are essential for growth of P. aeruginosa in an acetate supplemented growth medium.
5. Isocitrate lyase (ICL)

5.1. Introduction

Isocitrate lyase is the first enzyme of the glyoxylate shunt and is encoded by aceA. ICL catalyses the cleavage of isocitrate to yield glyoxylate and succinate. Then, the glyoxylate is condensed with an acetyl-CoA unit to yield malate in a reaction catalysed by malate synthase (MS). ICL therefore competes for the same substrate as ICD and IDH. During growth on fatty acids, carbon flux enters the central carbon metabolism as acetyl-CoA units. A fraction of the TCA cycle is diverted into the glyoxylate shunt to replenish biosynthetic intermediates by anaplerotic reactions. The glyoxylate shunt is vital for survival of P. aeruginosa during infection scenarios for reasons that are not yet clear. Furthermore, ICL plays an important role in ensuring optimal T3SS expression (PcrV, PopN, ExoS and ExsD expression are reduced in an aceA mutant) when P. aeruginosa is grown in oxygen-limited conditions\textsuperscript{201}. This indicates that ICL impinges upon virulence of P. aeruginosa during infection scenarios such as the conditions encountered in the oxygen-limited environment of chronic CF lung infections\textsuperscript{297}.

In this chapter, I investigate the structure and kinetics of ICL from Pseudomonas aeruginosa. The SDS-PAGE analysis of ICL assesses the quality of the enzyme after purification; followed by the gel filtration and analytical ultracentrifugation studies indicating the oligomeric status of ICL in solution. I will continue by presenting the gene expression profile of aceA during growth of the cells in two different carbon sources (acetate and glucose). Most of the results presented refer to the kinetics of the enzyme and its allosteric regulation by a subset of metabolic intermediates. Lastly, I investigate the crystal structure of ICL. I will present a comparison of ICL\textsubscript{Pa} to other bacterial and fungal ICL enzymes, a protein-protein interface analysis and a study of the active site cleft.

5.2. Bioinformatic analysis

The aceA gene, PA2634, is positioned at 2.977 Mbp in the PAO1 genome. Upstream of the gene is a 595-bp long denoted intergenic region that contains the promoter and gene regulatory elements. The Shine-Dalgarno sequence is located 10 bases upstream of the start codon to recruit the ribosome during translation. aceA is composed of 1,596 bp encoding a 531-amino acid long protein as seen in figure 40A.

The upstream and downstream regions of aceA are uncharacterized. Upstream of aceA, the genes numbered PA2630 to PA2633 form an operon (computationally predicted) and are annotated as hypothetical proteins. PA2631 shows a high probability to be an acetyl transferase. Further studies on gene ontology indicate that PA2633 may be involved in protein secretion process. Downstream of aceA, PA2635 and PA2636, predicted to form another operon, are related to a probable arginine
translocation pathway. However, unlike in \textit{E. coli}\textsuperscript{45}, the glyoxylate shunt enzymes in \textit{P. aeruginosa} do not comprise a single operon.

Figure 40: Genetic context of the \textit{aceA} gene \textit{P. aeruginosa} PAO1 and other \textit{Pseudomonas} species. (A) The highlighted \textit{aceA} annotated gene is 1,596 bp encoding a 531 amino acid-long protein. Two uncharacterized operons frame \textit{aceA}, they are computationally predicted to be involved in protein secretion (upstream of \textit{aceA}) and in arginine translocation (downstream of \textit{aceA}). (B) \textit{aceA} orthologues in other \textit{Pseudomonas} species indicate that ICL is cytoplasmic (red box)
5. Isocitrate lyase (ICL)

aceA orthologues are also present in other Pseudomonas species (figure 40B). In these other Pseudomonas species, the aceA sequence is conserved, however the length of the gene sequence is more variable. The examples in figure 40B indicate a predominant length of 1,326 bp, although the annotated aceA gene in P. aeruginosa is 1,596 bp-long and in P. stutzeri, aceA is composed of 1,644 bp. The majority of other Pseudomonas species carry a shorter aceA gene, the longer aceA in P. aeruginosa is an exception encoding 100 additional residues that must give an advantage either in catalysis or assembly. These residues represent multiple gaps corresponding to Glu156-Ala169, Gln271-Phe288, Gln298-Lys311 and Ser407-Ala443 in P. aeruginosa.

A BLAST analysis of the ICL amino acid sequence demonstrates the prevalence of ICL in most bacteria. ICL sequence identity within those species containing an ICL-encoding gene ranges from 91% (A. vinelandii) to 25% (B. pertussis) compared with the P. aeruginosa enzyme. However, some bacteria do not have a glyoxylate shunt. This is the case for the pathogens Campylobacter jejuni, Clostridium difficile, Legionella pneumophila, Listeria monocytogenes, Staphylococcus aureus for example. Absence of the shunt may relate to the specific growth conditions in which these species flourish. Indeed, common infections caused by C. jejuni, C. difficile, L. monocytogenes or S. aureus are associated with the blood-stream or gut. These two environments provide a high concentration of glucose ranging from 5 mM to 50 mM available for infective bacteria. Intracellular bacteria, such as L. pneumophila, which reside within host-derived vacuoles find a potential rich source of carbon in cytosolic glycogen.

The alignment of five examples in the figure 41 shows the missing 100 residues in E. coli, C. glutamicum and M. tuberculosis in comparison with P. aeruginosa and B. cepacia. There are three prominent gaps between residues 144-163, 271-306 and 397-433 (Pseudomonas numbering). The boxes in figure 41 highlight in green the residues binding the substrate, in pink the catalytic triad and in blue the magnesium binding residues.
5. Isocitrate lyase (ICL)

B. cepacia
MSQYQDDIKAVA-GLKENHSASWAIPEYARMAQ---NKFKTLGDIAYKXIMADN 57

P. aeruginosa
MSAQYIKEKIVA-ALGKEGWSANIPEYAMRQ---NKFKTLGDIAYKXIMADN 57

E. coli
H---MRXTRQGQLEELQKEM---TFPRMEGTRPFAEDTVKNSVLPSEFLQGALQGRKRLWRL 57

C. glutamicum
MSVCGPQESQPEQMDTNPQKVDRTYRTSEADTVLAQGTVSEHETLARSGEELDAD 60

M. tuberculosis
MSVCGPQESQPEQMDTNPQKVDRTYRTSEADTVLAQGTVSEHETLARSGEELDAD 60

* : * : * : * : *

B. cepacia
AAYADAPAKTYQSLGCMNGQIQKKQSMIKKSLHEGTSRYLQEGMAVALLSE-GLP 116

P. aeruginosa
AAYADAPAKTYQSLGCMNGQIQKKQSMIKKSLHEGTSRYLQEGMAVALLSE-GLP 116

E. coli
H---GEKGRYINSGLALITFQAOAQ---QA----KAGEAVLQGSOVAADLHASYL 107

C. glutamicum
T---GR--GDYTNALGCTONGAQaa---Q---RALGKAVLQGSOVAADLHASYL 109

M. tuberculosis
R---DE---LWTVNLATGCTONGAQaa---Q---RALGKAVLQGSOVAADLHASYL 107

* : * : * : * : * : * : * : *

B. cepacia
DQSMEKTLSVALLHETTLPGRADARELGFLRELOTADA---KAAJQKEIDNHTV 173

P. aeruginosa
DQSMEKTLSVALLHETTLPGRADARELGFLRELOTADA---KAAJQKEIDNHTV 173

E. coli
DQSLYPAVPVPVAIVRINTFRRAQDQI---WASAGIEPDFPQYVDR 149

C. glutamicum
DQSLYPAVPVPVSRVIRANNALSDEIA---RT---EGDTSVDN 147

M. tuberculosis
DQSLYPAVPVPVSRVIRANNALSDEIA---RT---EGDTSVDN 145

* : * : * : * : * : * : *

B. cepacia
HVPITAIADAGNQAEATYLLAKTFIEACGCIQEVNSDEQCGQDGKTVPHHEDF 233

P. aeruginosa
HVPITAIADAGNQAEATYLLAKTFIEACGCIQEVNSDEQCGQDGKTVPHHEDF 233

E. coli
YFLPITVADEAPGQOLMALWIKAMIEAGAVAEVFQSLAQXKLQGMKGVKLTQEA 209

C. glutamicum
WVPPITVADEAPFQQLMVLVQKMAAQAAGAVTQSLAQXKINTERNKHGKVLPITQEA 207

M. tuberculosis
WVPPITVADEAPFQQLMVLVQKMAAQAAGAVTQSLAQXKINTERNKHGKVLPITQEA 205

* : * : * : * : * : * : *

B. cepacia
IAKRIAIYFALEFLVVDQIGYATDLSTLSQAGKQLMTYVDAKQGQNAFLGEEASA 293

P. aeruginosa
IAKRIAIYFALEFLVVDQIGYATDLSTLSQAGKQLMTYVDAKQGQNAFLGEEASA 293

E. coli
IQRLVAALADDVTQV-DLVRATDADADLITS---CDC---246

C. glutamicum
IRTLSARLADAVNTM-TPVIAITDAAATLITS---DVC---244

M. tuberculosis
IRTLSARLADAVNTM-TPVIAITDAAATLITS---DVC---244

* : * : * : * : * : * : *

B. cepacia
DLQGNGVDVIEEIKRGKLLPEPLQNSFQPRACGGEVRLDCTTSILQNADDLWITTEKF 353

P. aeruginosa
SELNGGVDVIEEIKRGKLLPEPLQNSFQPRACGGEVRLDCTTSILQNADDLWITTEKF 356

E. coli
------------------------PYSDFIEITGERTGFRTH--AGEIAQISRG-LAYAPYDVLWCTST 293

C. glutamicum
------------------------ERQDFIPITGERTGAEYHYVE--NGLEPIETR-ASYAYAM1MWDGCT 291

M. tuberculosis
------------------------ERQDFIPITGERTGAEYHYVE--NGLEPIETR-ASYAYAM1MWDGCT 289

* : * : * : * : * : * : *

B. cepacia
HIAIQGGMVSEIRKVIFNPALKVYNYNSPSNWFTLNFQAFDTDMAKGDVSYDAKALOMS 413

P. aeruginosa
HVQIKAMVNIKLPEVNFPAKNLYNSPSNWFTLNFQAFDTDMAKGDVSYDAKALOMS 416

E. coli
DELARRPAFAQAINYKPQKLAYNSPSNWQMNLDRTTIA----334

C. glutamicum
DELAARKQFEAVGSEPFQDQGSYSCPSFWSLALDADIA---332

M. tuberculosis
DELAARKQFEAVGSEPFQDQGSYSCPSFWSLALDADIA---330

* : * : * : * : * : * : *

B. cepacia
VEYDETAEKALADKREIRTQADOASREARGIFHILILTPTYTAALSTDNLAIKEDFGQQLM 473

P. aeruginosa
VEYDETAEKALADKREIRTQADOASREARGIFHILILTPTYTAALSTDNLAIKEDFGQQLM 476

E. coli
------------------------FQDFQI-EKMGFCITLTAGLFHLNLYMFDLXYAG-ECMT 373

C. glutamicum
------------------------FQDFQI-EKMGFCITLTAGLFHLNLYMFDLXYAG-ECMT 373

M. tuberculosis
------------------------FQDFQI-EKMGFCITLTAGLFHLNLYMFDLXYAG-ECMT 373

* : * : * : * : * : * : *

B. cepacia
GYVAGKQRKIE---ROGCIAVHYQHSMGSGDDDHKEYFSG---EAAKAGKDNTHMQFQ 527

P. aeruginosa
AYVKGQKRQFEL---ROGCIAVHYQHSMGSGDDDHKEYFSG---EAAKAGKDNTHMQFQ 527

E. coli
HVYVEKQPPFEAINAXYKPFVSHQQVEQTVGFYDKVITIIQGQTSVATLSTGSEEOFQ 434

C. glutamicum
SFDLQNRKEFFAEEBERGFTAVKQHREYVAGFQDIATTVPD-NNTTLKLGEGEQFQH 434

M. tuberculosis
AYVEGQREFA---AEEERTTKRHREYVAGFQDIATTVPD-NNTTLKLGEGEQFQH 428

* : * : * : * : * : * : *

Figure 41: ClustalOmega alignment of ICL amino acid sequence (green: substrate binding, pink: catalytic triad, blue: magnesium binding). ICL protein is variable between a "short" (430 residues) and a "long" (530 residues) sequence. Very few regions are highly conserved, but the highlighted residues are systematically part of small clusters of fully conserved residues.
5.3. Purification of ICL protein

5.3.1. Preliminary results

Similar to ICD and IDH described in the previous chapters, I purified ICL. The aceA gene was PCR-amplified from the PAO1 genome and cloned into pQE80 (work carried out by a Part II student preliminary work done in 2014) or His-pMAL-c2x vector. The resulting construct was introduced into E. coli DH5α to overexpress the protein. The purification of His₆-tagged ICL (aceA pQE80 cloned vector) and His₆-MBP-tagged ICL (aceA His-pMAL-c2x cloned vector) used a Ni-NTA resin or a multi-step procedure with amylose resin (first His₆-MBP-tag cleaved from the construct) and Ni-NTA resin (His₆-MBP-tag is retained, native ICL is collected).

The SDS-PAGE analysis (figure 42) revealed that the purified ICL protein had a monomeric molecular mass of predicted by Expasy Prot Param is 58,886 Da for untagged ICL. The band on the gel migrates with an apparent molecular mass of between 50 kDa and 75 kDa. The SDS-PAGE analysis of the purified His₆-tagged ICL (see appendix 2) indicated that the molecular mass was around 60 kDa, a larger band migrated with an apparent molecular mass of between 50 kDa and 75 kDa. ICL purified from E. coli DH5α aceA pQE80 and E. coli DH5α aceA His-pMAL-c2x expression systems were therefore successfully expressed and purified to homogeneity.

5.3.2. Gel filtration analysis

The molecular mass of ICL in solution, as determined by gel filtration through a TSK-Gel G3000 SWXL column calibrated with the protein markers albumin (mol.mass. 66,000 Da), alcohol dehydrogenase (mol.mass. 150,000 Da), β-amylase (mol.mass. 200,000 Da) and apoferritin (mol.mass. 443,000 Da), was around 232,000 Da (figure 43). This indicates that ICL is a tetramer in solution. This is consistent with the isocitrate lyase characterized from other microorganisms including E. coli, Bacillus sp. and M. tuberculosis.
5.3.3. AUC data

In addition to gel filtration, ICL was analysed using AUC with sedimentation velocity (SV-AUC). This was carried out on samples containing three different concentrations of the protein (0.1, 0.5 and 1 mg/mL). The results presented in figure 44 were obtained with the lowest concentration of 0.1 mg/mL. The fitting of the Stokes equation was successful for all dataset, with very little variability between the three samples. The data were analysed as specified in chapter 2 using SEDNTERP and SEDFIT.

The profiles recorded and presented in figure 44 show a single protein peak with a sedimentation coefficient of 10.5 S. The calculations of the frictional ratio indicate that the isocitrate lyase molecule is more globular, with a frictional ratio of 1.17. Finally, the calculated molecular mass was 231,000 Da, which agrees well with the data from gel filtration. This leads to the conclusion that ICL is a stable tetramer in solution, consistent with other bacteria previously reported, such as E. coli, Bacillus sp. and M. tuberculosis. These parameters can be compared with those of the purified isocitrate lyase of Pseudomonas indigofera. The molecular mass of the P. indigofera enzyme is 222,000 Da and its frictional ratio was calculated to be 1.39.

![Figure 44: SV-AUC data of ICL protein.](image)

Figure 44: SV-AUC data of ICL protein. The figure shows 40 absorbance scans recorded at 280 nm for a protein sample of 0.1 mg/mL. The rmsd of the scans is 0.0069, significantly higher than the noise rmsd. The single peak with a molecular mass of 231 kDa indicates that ICL is a tetramer. The frictional ratio (1.17), closer to 1, suggests a more globular assembly.
5.4. Gene expression of aceA

The expression of aceA was investigated by cloning the promoter of the gene into pLP170 (PaceA::lacZ reporter gene construct). The construct was then introduced into PAO1. As with the Picd::lacZ and Pidh::lacZ reporter constructs described in earlier chapters, PAO1 cells were grown in minimal medium supplemented with acetate or glucose as a carbon source. The expression of aceA was assessed using a fluorescent β-galactosidase assay. Growth of the cells was monitored as OD$_{600}$.

As shown in figure 45 (pink line), the growth rate of the cells containing the PaceA::lacZ reporter construct is not affected whether the cells were grown in minimal medium supplemented with acetate or glucose. When compared with the control line in black, the growth rate is similar. This confirms that the PaceA::lacZ construct introduced into PAO1 does not impair growth of the cells. One noticeable aspect is the difference in the results of the optical density measurements between the two media: growth in acetate was lower as the highest value is 0.2 OD$_{600}$ while the highest value in glucose is 0.6 OD$_{600}$. Glucose remains the preferred carbon source over acetate.

The data in figure 45A show that aceA expression is up-regulated during growth in minimal medium supplemented with acetate. After 5h, the levels of β-galactosidase activity increase significantly. The highest β-galactosidase activity for PaceA::lacZ construct reaches 5,000 nmol.min$^{-1}$.A$_{600}^{-1}$ before entering stationary phase. Then, β-galactosidase activity decreases throughout the stationary phase. In figure 45B, the β-galactosidase assay reveals low expression of aceA during growth in minimal medium supplemented with glucose. The β-galactosidase activity of the Pacea::lacZ construct remains low throughout the entire time of growth. The maximum value attained was 300 nmol.min$^{-1}$.A$_{600}^{-1}$ after 6h of growth. Taken together, these
data indicate low expression of the aceA gene during growth on glucose, and increased expression of the aceA gene during growth on acetate.

In conclusion, when the cells are grown in acetate, the glyoxylate shunt is activated by transcriptional induction of the gene encoding isocitrate lyase. This demonstrates that P. aeruginosa growing on C2-units sources use the glyoxylate shunt for anaplerosis of the TCA cycle while maintaining oxidative TCA cycle activity for the production of energy and gluconeogenic precursors as the expression of ICD and IDH persists\textsuperscript{307}. However, the growth rate of the strain containing the PaceA::lacZ construct is different compared with the strain containing the Picd::lacZ construct or the Pidh::lacZ construct. I speculate that there is another form of transcriptional regulation of icd and idh but not of aceA. A candidate effector is the CbrA-CbrB two-component system. Indeed, studies performed by Nishijyo et al. found that in a cbrB mutant PAO4455, glucose was used with reduced efficiency but little or no growth defect was observed with acetate\textsuperscript{34}. The expression of the TCA cycle enzymes is regulated by Crc (section 1.2.1.2) and a connection between the Cbr system and the catabolite repression exerted through Crc has been reported recently in P. aeruginosa\textsuperscript{32}. It indicates that CbrB involvement in regulation of carbon catabolism may be extended to central carbon catabolic pathways. So, this adaptation to a C2-units nutrient source is comparable to E. coli in which there is a switch to the glyoxylate shunt in case of limited carbon source\textsuperscript{145,308}. In E. coli, there is essentially no expression of the glyoxylate shunt enzymes during growth on glucose, whereas in P. aeruginosa, aceA is always expressed at low levels in these conditions. Interestingly, fungi have the same ability to rewire their metabolism as the cells overexpress ICL when acetate is the sole carbon source\textsuperscript{309,310}.

5.5. Kinetic analyses

5.5.1. Michaelis-Menten data

Isocitrate lyase is a co-factor-independent enzyme, the direct assay used monitors the formation rate of the glyoxylate-phenylhydrazine complex at 324 nm. The kinetic studies at variable concentrations of isocitrate revealed that $K_m$ is 12 µM and $V_{max}$ is 0.048 mM.min\textsuperscript{-1} (figure 46). The tetrameric quaternary structure of ICL indicates that the $k_{cat}$ value is 1,128 min\textsuperscript{-1} (19 s\textsuperscript{-1}). The catalytic efficiency of ICL\textsubscript{Pa} is $k_{cat}/K_m = 1.55$ s\textsuperscript{-1}.µM\textsuperscript{-1}. In comparison to ICL in E. coli, there is a noticeable difference as the latter has a
Isochoritate lyase (ICL)

\[ \frac{k_{\text{cat}}}{K_m} = 0.37 \text{ s}^{-1} \text{ M}^{-1} \]

The catalytic efficiency of ICL\(_{Pa}\) is greater than that of ICL\(_{Ec}\), which is consistent with the difference of \(K_m\) between ICL\(_{Pa}\) and ICL\(_{Ec}\).

These values are remarkably different from the well-studied *Escherichia coli* enzyme. An extensive study of the branchpoint in this microorganism clearly established carbon flux partitioning. In *E. coli*, the affinity of ICL for isocitrate is \(K_m = 604 \text{ M}^{-1}\); hence a lower \(k_{\text{cat}}/K_m\). Other microorganisms have a similar disparity with \(K_m\) ranging from 23 M, 50 M to 280 M for *Chlorella sp.*, *Neurospora crassa* and *Corynebacterium glutamicum* respectively. *M. tuberculosis* encodes two isocitrate lyases jointly required for Mtb survival, ICL1 and ICL2 have a lower affinity for isocitrate. The \(K_m\) are respectively 130 M and 1.3 M for ICL1 and ICL2. In *Pseudomonas aeruginosa*, ICL has a slightly lower \(K_m\) for isocitrate than the isocitrate dehydrogenases. The low \(K_m\) of ICL for isocitrate means that flux through the TCA cycle enzymes does not need to be significantly decreased to get flux through the glyoxylate shunt. This also indicates that the control of flux through the TGB in *P. aeruginosa* is far more finely balanced. The coordination of metabolic flux in *P. aeruginosa* includes key regulators which affect IDH and subsequently ICL activity to redirect carbon flux through the TCA cycle or glyoxylate shunt. These compounds feature relatively little in the control mechanisms reported for other microorganisms characterised to date.

The activity of ICL is tightly linked to the presence of magnesium cation. In *Mycobacterium tuberculosis* for example, absence of magnesium leads to negligible ICL activity. It is supposed that the actual substrate in *Pseudomonas* ICL is a complex of isocitrate with magnesium and that magnesium acts as a non-essential activator. ICL is therefore to be considered, at least kinetically, a metal-ion-activated enzyme. The importance of magnesium is primordial for the cleavage of isocitrate; as long as magnesium is complexed, the non-activated and the activated enzyme forms are catalytically active.

**5.5.2. Screening for regulators**

The effect of gluconeogenic precursors on IDH activity revealed a new regulation of the enzyme in *P. aeruginosa*. I investigated the effect of the same subset of metabolic intermediates on ICL of *P. aeruginosa*. However, note that for molecules containing a ketone or aldehyde group I needed to employ the coupled assay with LDH to circumvent the formation of a phenylhydrazine complex that would interfere with the readings.

No activators of ICL activity were identified. However, I found that several compounds inhibit ICL activity (figure 47). Only glyoxylate activates ICL; however glyoxylate is also a substrate of LDH (used in the coupled assay). ICL was effectively inhibited by citrate, coenzyme A, glycolate, itaconate, maleate, malate, malonate, 3-nitropropionate, oxaloacetate, PEP, pyruvate, succinate, malate + oxaloacetate, glyoxylate + oxaloacetate and succinate + glyoxylate. Itaconate, maleate, 3-nitropropionate, oxaloacetate and pyruvate most potently inhibit ICL activity. The percentage of inhibition varies from 93 to 98% compared with the control. Some of these inhibitors are known from studies on ICL from other organisms, and are synthetic compounds. For example, itaconate, 3-
5. Isocitrate lyase (ICL)

Iso citrate lyase (ICL) and 3-bromopyruvate are all well-established inhibitors of ICL. These compounds are analogs of succinate (itaconate, 3-nitropropionate) or glyoxylate (3-bromopyruvate). The effect of 1 mM of itaconate has been extensively studied in ICL from *Pseudomonas indigofera* as it inhibits ICL activity by 96% which is similar to my observations here with the ICL of *P. aeruginosa*. Maleic acid, oxaloacetate and pyruvate are also known inhibitors of ICL from *Escherichia coli*, *Neurospora crassa*, *Brevibacterium flavium* and *Pseudomonas indigofera*. However, the magnitude of the inhibitory effect is variable. Oxaloacetate 1 mM, for example, has been reported to inhibit ICL activity by 70% in *Brevibacterium flavium*, whereas pyruvate (1 mM) has a more limited effect on ICL with an inhibition of 7% only.

Another subset of milder inhibitors includes citrate, coenzyme A, glycolate, malate, malonate, PEP and succinate. Citrate, glycolate, malate, PEP and succinate are known inhibitors of ICL from *E. coli*, *P. indigofera*, *Chlorella pyrenoidosa* and *Brevibacterium flavium*. Citrate inhibition of ICL is mild in other organisms, ranging from 6% to 16% in *Chlorella* and *Brevibacterium* respectively.

In *P. aeruginosa* however, it has a more pronounced effect with 50% inhibition. Glycolate inhibition of ICL in *P. indigofera* (35%) and *P. aeruginosa* (34%) is similar. Malate inhibition of ICL in *P. aeruginosa* is more pronounced with a drop of ICL activity to 50%. In *Chlorella* and *Brevibacterium* ICL, malate shows a milder inhibition by 5 to 24% respectively. PEP and succinate are known inhibitors of ICL from *E. coli*, *C. pyrenoidosa*, *P. indigofera* and *B. flavum*. The inhibition of ICL by PEP in these species ranges from 1 to 20%, the effect on ICLPa is more pronounced with a loss of activity of 72%. Succinate has a stronger effect on ICL activity with a reported inhibition ranging from 13 to 56% in *Chlorella* and *Brevibacterium* respectively. In *P. aeruginosa*, succinate decreases ICL activity by 85%. Coenzyme A and malonate have not been reported as ICL inhibitors before. They are potent inhibitors, coenzyme A decreases ICL activity by 66%, while malonate reduces the enzyme activity by 75%. It is unsurprising that *P. aeruginosa* ICL activity is modulated by certain metabolites. The potent effect of these demonstrates that ICL regulation is more nuanced; the glyoxylate shunt regulation responds to metabolic cues. Oxaloacetate and pyruvate, two emerging

![Figure 47: Results of screening of potential regulators on ICL. Several compounds inhibit ICL activity compared with control. Glyoxylate activatory effect is a false positive as glyoxylate is also a substrate for LDH used in the coupled assay. Values marked with * are significantly different compared with control (p<0.05 and n=3). All compounds were tested at a 1 mM concentration.](image-url)
key regulators, reciprocally regulate IDH and ICL. By activating IDH and inhibiting ICL, this ensures flux through the TCA cycle.

5.5.3. Inhibition kinetic study

The inhibitors of ICL fell into three categories: non-competitive inhibitors, uncompetitive inhibitors and ‘mixed’ inhibitors (figure 48, see appendix 2D-G for the corresponding Michaelis-Menten plots). Citrate, coenzyme A, glycolate, malate, oxaloacetate and pyruvate non-competitively inhibit ICL from *P. aeruginosa*. This means that the inhibitor is capable of binding to the free enzyme or more likely to the enzyme-substrate complex. The $K_m$ is not affected by the presence of the inhibitor, but the $V_{max}$ is altered. Table 19 compiles the kinetic parameters for the inhibitors of ICL, all had a pronounced effect on $k_{cat}$ with little effect on $K_m$. The $V_{max}$ varies depending on the potency of that compound.

For example, citrate, malate and coenzyme A have a moderate effect. More potent inhibitors such as glycolate, oxaloacetate and pyruvate feature a more reduced $V_{max}$ of ICL. This is consistent with the double-reciprocal plot in which the lines for citrate, malate and coenzyme A derive further from the control. In the case of non-competitive inhibition, the $K_i$ was calculated according to the following equation:

$$K_i = \frac{[I]}{(V_{max}/V_{max,obs})} - 1$$

with $[I] = 1$ mM. The strongest physiological inhibitors were oxaloacetate and pyruvate. They inhibit ICL with a $K_i$ of 1.41 mM and 1.68 mM respectively. The reported intracellular concentrations of pyruvate and oxaloacetate are up to 10 mM\textsuperscript{325} for the former and 0.03 mM\textsuperscript{326} for the latter. A more moderate inhibitor, citrate, has a $K_i$ of 15.79 mM while the intracellular concentration can be up to 2 mM in glucose-fed, exponentially growing *E. coli*\textsuperscript{327}.

| Table 19: Kinetic parameters of ICL in presence of inhibitors |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | $K_m$ (µM)      | $V_{max}$ (ΔmM/min) | $k_{cat}$ (s\(^{-1}\)) | $k_{cat}/K_m$ (s\(^{-1}\)/µM) | $K_i$ (mM) | Effect        |
| isocitrate     | 18.29           | 0.048            | 18.81            | 1.55          | n/a            | n/a            |
| citrate        | 18.80           | 0.045            | 17.69            | 0.94          | 15.79          | non competitive|
| coenzyme A     | 22.56           | 0.038            | 14.94            | 0.66          | 3.85           | non competitive|
| glycolic acid  | 18.20           | 0.031            | 12.15            | 0.67          | 1.82           | non competitive|
| malic acid     | 16.61           | 0.039            | 15.33            | 0.92          | 4.41           | non competitive|
| oxaloacetate   | 18.42           | 0.028            | 11.01            | 0.60          | 1.41           | non competitive|
| pyruvate       | 18.22           | 0.030            | 11.79            | 0.65          | 1.68           | non competitive|
| itaconic acid  | 6.27            | 0.022            | 8.67             | 1.38          | 0.92           | uncompetitive   |
| maleic acid    | 3.57            | 0.016            | 6.29             | 1.76          | n/a            | mixed          |
| malonic acid   | 12.01           | 0.031            | 12.19            | 1.01          | 2.85           | uncompetitive   |
| nitropropionic acid | 6.86   | 0.019            | 7.47             | 1.09          | 0.96           | uncompetitive   |
| PEP            | 6.04            | 0.023            | 8.88             | 1.47          | n/a            | mixed          |
| succinate      | 7.82            | 0.018            | 7.09             | 0.91          | 0.92           | uncompetitive   |
5. Isocitrate lyase (ICL)

Figure 48: Kinetic analysis of ICL in presence of inhibitors (A) Non-competitive inhibition by milder regulators citrate, coenzyme A, glycolate (B) Potent non-competitive inhibition by malate, oxaloacetate, pyruvate (C) Uncompetitive inhibition by synthetic compounds includes itaconate and 3-nitropropionate (D) Uncompetitive or mixed profile exhibited by stronger inhibitors such as maleate, PEP and succinate

Itaconate, maleate, malonate, 3-nitropropionate, PEP and succinate show an uncompetitive or mixed inhibition of ICL. Here, the inhibitor binds to a subsite of the enzyme-substrate complex. In the presence of uncompetitive inhibitors, the $K_m$ and $V_{max}$ decrease upon binding of the compounds. As seen in Table 19, the values of $K_m$ and $V_{max}$ vary greatly depending on the inhibitor. Due to the uncompetitiveness of these compounds, $K_i$ is replaced by $K_i'$. $K_i'$ was determined from the y-intersect of the Lineweaver-Burk plot:

$$K_i' = \frac{[I]}{(Y_{intersect} \times V_{max}) - I}$$

with $[I] = 1 \text{ mM}$. Itaconate and 3-nitropropionate (synthetic compounds analog of succinate) are similar; $K_i'$ is on average 0.93 mM. Malonate, a more moderate inhibitor, has a higher $K_i'$ of 2.85 mM. Potent inhibitors such as itaconate and 3-nitropropionate show parallel lines increasingly shifted (figure 48C) compared with control. On the other hand, maleic acid and PEP show a mixed inhibition. PEP has been previously reported to be an uncompetitive inhibitor of isocitrate lyase from *E. coli* with a $K_i'$ value of 0.89 mM$^{329}$. The mixed inhibition of ICL from *P. aeruginosa* by maleate and PEP supposes that the compound has an unequal affinity for the free enzyme and the enzyme-substrate complex, hence a mixture of competitive and uncompetitive phenotypes. The rate equation considers $K_{ic}$ (competitive part) and $K_{iu}$ (uncompetitive part). However, in order to determine these two parameters, further investigation would require the kinetic analysis of ICL with variable concentrations of substrate and inhibitor.

### 5.5.4. Thermal shift assay

To investigate the binding of these inhibitors further, I conducted a thermal shift assay. This approach is predicated on the assumption that if a low molecular weight ligand binds to the enzyme, this leads to a ‘tightening up’ of the structure and consequent stabilization against thermal
denaturation. The TSA determines a $\Delta T_m$ that indicates the increased temperature stability of ICL with the regulator bound compared with the control.

The bar chart in figure 49 compiles the results from the assay. The binding of isocitrate demonstrates a significant shift of 4.5°C. However, all the compounds tested also affected protein stability. Non-regulatory molecules, for example acetate, ADP, D-glucose-6-P, have a $\Delta T_m$ similar to the substrate. I have no explanation for this since these molecules do not influence ICL activity, so do not bind to ICL. Uncompetitive as well as non-competitive inhibitors would be expected to show larger $T_m$ shifts in the presence of substrate as the effect would be dependent upon the additivity of the individual binding. However, oxaloacetate and pyruvate, two potent non-competitive regulators have a similar $\Delta T_m$ than the substrate. I have no further explanation for this as the TSA experiment investigated the enzyme and the regulator alone, no substrate was added to the experimental conditions. So uncompetitive and non-competitive inhibitors would not be able to bind to a subsite.

5.6. Crystal structure of ICL

5.6.1. Structure description

The three-dimensional structure of ICL was solved by molecular replacement using the *Burkholderia pseudomallei* structure (PDB 3I4E) as a template. The structure was refined to 1.9 Å with a $R_{\text{cryst}}$ of 18.32% and a $R_{\text{free}}$ of 21.17%. The analysed crystal contains one molecule in the asymmetric unit, the polypeptide chain contains 486 residues (ICL is 531-residue long) along with 182 water molecules. No electron density was detected for the last 45 residues at the C-terminus of the protein. The core of the enzyme is very well conserved and provides molecular insights regarding the active site pocket. The active site of the protein contained a calcium ion coordinated with glyoxylate and three water molecules (red circle figure 50A).

Each monomer of ICL contains 17 $\alpha$-helices and 14 $\beta$-strands, together with a small number of helical turns (figure 50A). The N-terminal region of the polypeptide chain comprises three consecutive $\alpha$-helices, with the first two lying on the periphery of the molecule ($\alpha_1$ and $\alpha_2$). The structure then adopts a TIM-barrel like structure composed of 8 $\beta$-strands ($\beta_1$-4, $\beta_6$, $\beta_{12}$-14), alternating with either one or two $\alpha$-helices. The five remaining $\beta$-strands ($\beta_7$-11) form a third
5. Isocitrate lyase (ICL)

domain, the head-domain (figure 50A) (see appendix 10 for a structure/sequence alignment). ICL polypeptide is related by a crystallographic two-fold axis to yield a tetrameric structure (figure 50B) with extensive contact between each protomer.

![Figure 50: ICL crystal structure](image)

**5.6.2. Structural comparison and contact analysis**

Investigation on ICL lead to the characterization of the crystal structure in known bacterial pathogens (*E. coli*<sup>195</sup>, *M. tuberculosis*<sup>196</sup>, *B. pseudomallei* (to be published), *B. melitensis* (to be published) and *Y. pestis* (to be published)). However, these are shorter than ICL from *P. aeruginosa*. Only one microorganism (*Aspergillus nidulans*) has an enzyme comparable to ICL<sub>Pa</sub> with 538 residues, whereas there are 531 residues in *P. aeruginosa* ICL.

![Figure 51: Superposition of ICL<sub>Pa</sub> with other species](image)

The superposition of ICL<sub>Pa</sub>, ICL<sub>Bp</sub>, ICL<sub>Ec</sub> and ICL<sub>Mt</sub> (figure 51A) demonstrates that despite the shorter sequences of ICL<sub>Ec</sub>, ICL<sub>Mt</sub> and ICL<sub>Bp</sub>, the rmsd is similar for the three with 1.35 Å, 1.30 Å and 1.31 Å respectively. Most of the secondary structure elements are conserved. The core of the
enzyme is structurally conserved, including the TIM-barrel like structure. However, alignment of the amino acid sequence of ICL from *P. aeruginosa*, *E. coli* and *M. tuberculosis* indicates differences between the longer and the shorter ICL enzymes (see section 5.2.2). The shorter version of ICL (e.g. ICL{sub}Ec and ICL{sub}Mt) lacks blocks of sequence that are present in the sequence of ICL{sub}Pa as Leu145-Ala163 (green in figure 51A), Ile272-Ile306 (yellow figure 51A) and Ala398-Arg433 (magenta figure 51A). The superposition of ICL{sub}Pa with other ICL from microorganisms (ICL{sub}Ec, ICL{sub}Mt and ICL{sub}As) shows that the overall fold of ICL{sub}Pa is distinct from the other species. The first two α-helices at the N-terminus are flipped by a 180° angle relative to the axis of α3 as seen in figure 51A. The first missing block of sequence in ICL{sub}Ec (Leu145-Ala163 in *P. aeruginosa*) is the extension of α6 by an additional four helical turns and of α7 by an additional helical turn. The second block of sequence (Ala398-Arg433 in *P. aeruginosa*) is the extension of α13 and α14. These two α-helices extend away from the core of the protomer, giving the tetramer a star-like shape. The superposition of ICL{sub}Pa with ICL{sub}As (figure 51B) shows a similar conservation of the core of the enzyme. The rmsd is 1.43 Å. The most obvious difference in *P. aeruginosa* ICL compared with all other ICL relates to the third block of sequence (Ile272-Ile306 in *P. aeruginosa*). This forms a relatively unstructured “head-domain” in ICL{sub}Pa reminiscent of the “head-domain” present in *A. nidulans*\(^\text{229}\) (figure 51B). The head-domain in *A. nidulans* is somehow bulkier with five α-helices, whereas in *P. aeruginosa* this domain is comprised of four β-strands and loops. The extensive interaction of these “head-domains” generates intimate contacts for the tetramer formation. Figure 51A shows that the shorter ICL{sub}Ec and ICL{sub}Mt do not have this “head-domain”. ICL{sub}Pa therefore features traits from both prokaryotic and eukaryotic ICL.

Protein-protein interactions play crucial roles in many biological functions. The analysis of the characteristics of protein-protein interfaces is to search for the factors that contribute to the affinity and specificity of protein-protein interactions. Further investigation on ICL{sub}Pa using QtPISA revealed the residues comprising the interfaces between the four subunits. This analysis was further supported by COCOMAPS and InterfaceResidues script for PyMOL. The QtPISA interface summary concluded that there are six potential interfaces with the following chains pairing: A-B, A-C, A-D, B-C, B-D and C-D.

The first pairing between chains A and B produces similar results when analysing it with COCOMAPS and InterfaceResidues. There are 50 residues forming the interface between the two chains as seen in figure 52A. Its symmetrical counterpart between chains C and D shows the same results flipped 180° on the x-axis (figure 52B). Most of the residues are located around position 300. Figure 52C, which is the distance map resulting from the analysis, shows a strong cluster of contact around position 300 corresponding to the “head-domain”. Three smaller clusters appear at the intersection of positions 100 and 200. The overall ΔASA (accessible surface area) is the same for both A-B and C-D interfaces and is of 1,854.5 Å\(^2\). However, the distribution of the interface residues between polar and non-polar favourably tilts towards polar residues with an ΔASA of 1,114.65 Å\(^2\) against 739.85 Å\(^2\) for the non-polar residues.
Another pairing is between chains A and C and again its symmetry between chains B and D (figures 52D-E). According to the analysis conducted by InterfaceResidues and COCOMAPS, there are 38 residues involved in the formation of the interface. The distance map (figure 52F) shows two clusters formed by 23 residues situated in the first 50 of chain A interacting with the remaining 15 residues situated around position 250 of chain C. The reverse happens between the first residues in the 50\textsuperscript{th} of chain C interacting with position 250 of chain A. The same occurs between chain B and
D as the distance map is exactly the same (see appendix 11 for complete results). With fewer residues, the parameters of the interface are reduced and show that indeed the $\Delta$ASA is reduced to 1,368.45 Å². Same results are observed regarding the B-D interface. With respect to the balance between polar and non-polar residues, there are more polar residues with an ASA of 904.05 Å² while non-polar residues represent 464.4 Å².

Finally, the last set of interfaces is between the chains A and D and the symmetry B and C. The number of residues differs between InterfaceResidues for PyMOL, COCOMAPS and QtPISA with 72, 70 and 76 residues respectively. Figures 52G-H show the symmetrical view of the two interfaces relative to the x-axis of the tetramer (frontal view figure 52J). The arrangement of the interfaces as seen in figure 52I highlights four clusters. There is contact of a subset of 27 residues scattered around position 100 starting from Gln69 to Gln139. Another cluster is visible along the truncated C-terminus, composed of 35 residues between Gln436 and Gln485. The last two clusters are symmetrical to each other as the N-terminal region of one chain interacts with the C-terminal region of the other. This includes the remaining 8 residues between positions 100 and 450 of each chain. Such a larger interface has a wider $\Delta$ASA of 2,953.1 Å² which represents 12.23% of the protomer surface. Polar residues are preponderant with a $\Delta$ASA of 2,105.8 Å², while non-polar $\Delta$ASA is 847.35 Å².

In conclusion, the protein-protein interface of ICL$_{Pa}$ is complex as the enzyme is a tetramer in solution. The reconstruction of the multiple contacts between the four chains in figures 52J-K demonstrates the complexity of the interface. There are in total 316 residues involved in stabilizing the tetramer. The total $\Delta$ASA is of 12,352.1 Å². The assembly of oligomers ideally responds to a power law relationship $^{330}$: $ASA = 5.3 \times M^{0.70}$ where ASA is in Å² and M is the molecular mass in Daltons$^{331}$. Considering the truncated ICL$_{Pa}$ solved here (M = 54,000 Da estimated by Expasy), the expected ASA for a monomer would be 20,936 Å². The formation of the tetramer is presumably based on two dimers coming together. Each dimer (AD and BC interfaces) is characterised by an inter-subunit helix-swapping of $\alpha$12 and $\alpha$13 between two crystallographically related subunits$^{332-334}$. So, if that protein forms dimers and tetramers, their expected ASA will be 35,455 Å² for a dimer (5.3 x (2x54,000)$^{0.70}$) and 60,042 Å² for a tetramer (5.3 x (4x54,000)$^{0.70}$). The expected area buried in the formation of the tetramer is 2 x 35,455 – 60,042 = 10,868 Å². This estimation is similar to the observed total ASA which corroborates the results presented.

5.6.3. Active site analysis of ICL$_{Pa}$

The crystallisation conditions of ICL included CaCl$_2$ supplemented with 1 mM glyoxylate. The data collection and solving was successful in finding suitable electron density for both ligands (figure 53A). The coordination of Ca$^{2+}$ with glyoxylate and three water molecules was validated using CheckMyMetal$^{335}$. In the active site of the structure I present here, glyoxylate and Ca$^{2+}$ are displaced. This is reflected by a significant shift of the glyoxylate moiety by 5.7 Å towards the $\beta$4-$\beta$5 loop (Glu213-Val231) and the Ca$^{2+}$ is also shifted by 1.8 Å.
5. Isocitrate lyase (ICL)

Table 20: Residues comparison of ICL active site

<table>
<thead>
<tr>
<th>Substrate binding</th>
<th>Mg$^{2+}$ binding</th>
<th>β4-5 loop</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICL&lt;sub&gt;Pa&lt;/sub&gt;</td>
<td>Y99 S101 W103 R260 E352</td>
<td>N380 T451 D184 D186 E213-V231</td>
</tr>
<tr>
<td>ICL&lt;sub&gt;Bm&lt;/sub&gt;</td>
<td>Y82 S84 W86 E281 N309 T343 D147 E149 E176-I194</td>
<td></td>
</tr>
<tr>
<td>ICL&lt;sub&gt;Bp&lt;/sub&gt;</td>
<td>Y86 S88 W90 E286 N314 T348 D154 E156 E183-V201</td>
<td></td>
</tr>
<tr>
<td>ICL&lt;sub&gt;Mt&lt;/sub&gt;</td>
<td>Y89 S91 W93 E289 N317 T351 D149 E159 E186-V204</td>
<td></td>
</tr>
<tr>
<td>ICL&lt;sub&gt;Ec&lt;/sub&gt;</td>
<td>Y89 S97 W99 E395 N423 T457 D168 D170 E197-V215</td>
<td></td>
</tr>
</tbody>
</table>

Further investigation of the binding mechanism in the active site of ICL<sub>Pa</sub> involved superposition with ICL<sub>Bm</sub> (Brucella melitensis PDB 3P0X, to be published) as the latter was co-crystallised with isocitrate. Despite the low amino acid sequence identity between the two enzymes (28%), ICL<sub>Pa</sub> and ICL<sub>Bm</sub> overlay with a low rmsd of 1.31 Å. The core of the protein and the TIM-barrel like structure are similar. The active site is located at the entrance of the TIM-barrel. It is comprised of two distinct parts. Most of the residues are located on the β-strands and loops around the rim of the TIM barrel, and there is a flexible loop between β4 and β5. The binding of isocitrate involves interaction of the substrate with the triad Tyr99-Ser101-Trp103, supported by Arg260, Glu352, Asn380 and Thr451. Catalysis is possible only with the participation of another triad Cys222-Glu223-His224 as part of the β4-β5 loop. The latter is a very flexible structure that caps the active site.
site upon binding of the substrate. Two additional residues, Asp184 and Asp186 might bind to the magnesium before catalysis. This preliminary investigation with *B. melitensis* also included ICL from *E. coli*<sup>199</sup>, *M. tuberculosis*<sup>196</sup> and *B. pseudomallei* (to be published) as these structures display different state of activation upon the presence of the substrate or products. As seen in **figure 53B**, it is evident that all of the residues discussed above are conserved. The **table 20** aligns the binding residues and catalytic residues of all species. The Mg<sup>2+</sup> binding residue Asp186 in ICL<sub>Pa</sub> differs from other bacterial species but is similar to Asp170 in *A. nidulans* ICL. The conformation of the active site is comparable, the TIM-barrel-like structure provides a rigid anchor for binding, while the flexible loop flutters depending on the presence of *isocitrate*.

Of the several residues involved in binding *isocitrate*, three are of particular interest: Cys222-Gly223-His224. They constitute the end of the β4-β5 loop (Glu213-Val231) in ICL<sub>Pa</sub>. This loop is well conserved in all *isocitrate* lyases, ICL<sub>Pa</sub>, ICL<sub>Bm</sub>, ICL<sub>Bp</sub>, ICL<sub>Ec</sub>, ICL<sub>Mt</sub> and ICL<sub>As</sub> (**figure 53B-C**). The amino acid sequence alignment of this section demonstrates that out of the 19 residues, 9 are fully conserved and 5 are of strongly similar categories (**figure 53D**). For example, that loop corresponds in *P. aeruginosa* (ICL-Pa in **figure 53D**) to Glu213-Val231 or in *E. coli* (ICL-Ec in **figure 53D**) to Glu186-Val204. The loop itself has been extensively studied in *M. tuberculosis* ICL and is known to play a major role in the mechanism of action of ICL<sup>196,336</sup>. As seen in **figure 53B-C**, the loop is systematically present in bacterial species, and its spatial disposition depends on the crystallisation conditions. In the case of ICL<sub>Bm</sub> and ICL<sub>Mt</sub>, the loop is in a closed configuration since ICL<sub>Bm</sub> is bound to *isocitrate* whereas ICL<sub>Mt</sub> is bound to the reaction products, succinate and glyoxylate. On the other hand, in ICL<sub>Bp</sub>, ICL<sub>Ec</sub> and ICL<sub>Pa</sub>, the loop points away from the active site. ICL<sub>Bp</sub> is an apo structure. ICL<sub>Ec</sub> was co-crystallised with pyruvate, the latter does not apparently faithfully mimic the true substrate or product of the reaction and therefore resembles the apo structure. Similarly, in ICL<sub>Pa</sub>, glyoxylate is also not recognised which means that the loop stays in this open conformation. When the substrate is embedded in the active site, the motion of the loops becomes evident; it swings inwards to cap the active site, as determined in *Mycobacterium tuberculosis*<sup>196</sup>. The cysteine and histidine residues are essential for the catalytic activity of the enzyme as part of the KKCGH sequence motif. The superposition of the same active site residues of *P. aeruginosa* with *A. nidulans* (**figure 53C**) demonstrates the conservation of the substrate binding residues. Despite the low identity between the two amino acid sequences (26%), the *isocitrate* binding residues and more importantly the catalytic loop conservation indicates that the mechanism of action of ICL is apparently similar in prokaryotes and eukaryotes.

### 5.7. Discussion

Fatty acids or acetate metabolism is an important feature of the pathogenicity of *Pseudomonas aeruginosa* during infection. Consumption of fatty acids, converted into acetyl-CoA, involves regulation of carbon flux bifurcation between the oxidative TCA cycle and the glyoxylate shunt. The mechanisms that regulate the partitioning of metabolite fluxes at the bifurcation of the
TCA cycle and glyoxylate shunt were unknown in *Pseudomonas aeruginosa*. The regulation maintaining carbon flux at the TGB is essential during catabolism of C$_2$-units in order to balance anaplerotic fluxes and to conserve atoms of carbon (glyoxylate shunt aka. ICL), and to conserve energy-generating fluxes (TCA cycle aka. ICD/IDH). The glyoxylate shunt becomes centrally-important for biomass growth of *Pseudomonas aeruginosa*. The driver behind this work is that *P. aeruginosa* is incapable of surviving solely on C$_2$-units catabolism without a functional glyoxylate shunt. Indeed, a double mutant of *P. aeruginosa* in which ICL and MS (malate synthase) are absent is cleared from a mouse pulmonary infection model in 48h\textsuperscript{65}. This suggests that no other mechanism can circumvent the loss of the glyoxylate shunt, indicating the great potential for therapeutic agents which target both enzymes of the shunt in *P. aeruginosa*. Furthermore, as presented in section 5.1, an aceA mutant of *P. aeruginosa* showed a pronounced loss of T3SS-related virulence factors. This means that therapeutic agents targeting the glyoxylate shunt would also tackle virulence, a powerful tool against the proliferation of *Pseudomonas aeruginosa*. Investigation of the TGB, and in this chapter ICL, provides a better understanding of the enzymology of the branchpoint which is essential before any step further towards drug-design is to be made.

ICL is a tetramer in solution. This is consistent with other bacterial species in which ICL has been characterized. Beforehand, however, I investigated the effect of acetate and glucose as sole carbon sources on the expression of aceA. The growth of PAO1 carrying the PaceA::lacZ construct was not impaired when the cells were grown in acetate-supplemented or glucose-supplemented minimal medium. This revealed a net up-regulation of ICL transcription when PAO1 is grown on acetate. This suggests that C$_2$-unit catabolism efficiently promotes central carbon metabolism to divert isocitrate through the glyoxylate shunt instead of the TCA cycle. However, there is also a remarkable low aceA expression when PAO1 is grown on glucose, suggesting a more graded metabolic response. This ability of *P. aeruginosa* to rewire its gene expression profile is a remarkable trait for bacterial survival during infection as intermediates of the TCA cycle must be replenished by anaplerotic reactions\textsuperscript{66,151,164}.

In *Pseudomonas aeruginosa*, three enzymes compete for isocitrate. ICL affinity for isocitrate is similar, even slightly better, than its TCA cycle counterparts, ICD and IDH. ICL K$_m$ is 12 µM, while ICD K$_m$ is 25 µM and IDH K$_m$ is 18 µM. The enzymology in *P. aeruginosa*, however, is more complex as it also involves AceK. The regulation at the TGB of *Pseudomonas* is managed by AceK-dependent phosphorylation of ICD, a partly rheostatic regulation of IDH and a more finely balanced regulation of ICL. A number of metabolic intermediates have a significant effect on ICL activity, and more importantly, all compounds are inhibitors (figure 54). None of the compounds show competitive inhibition of ICL which suggests that there is allosteric regulation of the enzyme. Non-competitive, uncompetitive and mixed inhibition of ICL are the three profiles seen in *P. aeruginosa*. Two key regulators emerged, oxaloacetate and pyruvate, which reciprocally inhibit ICL and activate IDH. They finely coordinate metabolic flux partitioning between the TCA cycle and the glyoxylate shunt. When these two intermediates are abundant, suggesting that there are enough gluconeogenic
precursors for biomass production, IDH is activated while ICL is inhibited leading to the rerouting of flux through the TCA cycle. In contrast, when the levels of oxaloacetate and pyruvate decrease, IDH becomes deactivated and ICL is disinhibited leading to greater carbon flux through the glyoxylate shunt. The same paradigm can be applied to PEP; higher levels of the compound suggests there are sufficient gluconeogenic precursors, so ICL is inhibited to redirect flux through ICD and IDH. The architecture of the regulatory circuit that controls partitioning of fluxes at the TGB is strikingly different in *P. aeruginosa* compared with the established paradigm in *E. coli*. In the latter, carbon flux partitioning is mediated by AceK-dependent inhibition of ICD, which diverts flux from the TCA cycle to the glyoxylate shunt. This is reinforced by the higher affinity of ICD from *E. coli* with a $K_m$ of 11 µM while ICL affinity for isocitrate is 604 µM.

The preliminary steps of the ICL structure solving (Matthews coefficients calculations, phasing and molecular replacement) were straightforward. With a resolution of 1.9 Å, the ligands (glyoxylate and calcium) were detectable. However, the coordination of the ligands in presence of water did not match the usual conformation seen in other ICL structures. A typical coordination geometry for Ca$^{2+}$ is octahedral, however in this case CheckMyMetal validated a square antiprism geometry. The shorter bond lengths characteristic of calcium, and the requirement of at least six

![Figure 54: Schematic of all ICL inhibitors compared with isocitrate. Colour coding: non-competitive, uncompetitive and mixed inhibitors. Coenzyme A is a distinct inhibitor of ICL, as a non-competitive regulator, it binds to a subsite of the enzyme, however there is no structural homology with any other compound. All other inhibitors have a similar structure indicating allocation to a subsite of ICL.](image-url)
5. Isocitrate lyase (ICL)

oxygen atoms to coordinate it, significantly shifted the location of the ligands in the active site.\textsuperscript{237,338} The core of the ICL structure showed great similarity with other bacterial and fungal species characterized, as the TIM-barrel like fold is preciously conserved. There are several differences, however, that make ICL\textsubscript{Pa} unique. ICL\textsubscript{Pa} is 100-residue longer than the common “shorter” bacterial ICL, this means a few $\alpha$-helices are elongated giving ICL\textsubscript{Pa} tetramer its star-like shape and, more importantly, there is a “head-domain” increasing contact between each protomer of the tetramer. Comparing ICL\textsubscript{Pa} and ICL\textsubscript{An} (similar length, but eukaryotic ICL\textsubscript{An}), most of the differences in ICL\textsubscript{Pa} concentrate in the “head-domain”, which places ICL\textsubscript{Pa} as an intermediate enzyme between prokaryotic ICL and eukaryotic ICL. The quaternary structure however remains unchanged and the protein-protein contact between the four subunits is a complex arrangement.

Further investigation on the active site of ICL from \textit{P. aeruginosa} helps understanding the mechanism of action of the enzyme. All residues binding isocitrate are conserved in all ICL enzymes whether it is a bacterial or a fungal species; it is common to observe constant conservation of the catalytic residues and binding residues among enzymes even though they have low sequence identity. A detailed study of the active site highlighted the presence of a flexible catalytic loop (the $\beta$4-$\beta$5 loop). The phenomenon of catalytic flexibility is also presented as plasticity. Conformational changes are frequently observed as part of enzyme mechanisms, according to which substrate and specific ligand binding is associated with a conformational change. Loops participating in the active sites of enzymes are an exception. These loops are often referred to as lids, turns or flaps, and their location at the entry of the active site plays a major role in substrate selectivity, and the recognition and facilitation of substrate binding into the binding cleft. Upon substrate binding, the disordered, solvent-exposed loop adopts a more compact and ordered conformation making interaction with subsites of the ligand and/or other residues of the protein. The closing of the cap the ligand in the active site cleft and protects the ligand from the aqueous environment. Moreover, access of the active site to other molecules is prevented.\textsuperscript{339,340} This aspect of ICL catalysis structurally features the uncompetitive and non-competitive regulation of the enzyme in \textit{Pseudomonas aeruginosa}. In the presence of the substrate and inhibitor, the substrate might bind first, closing the active cleft. The only option for the inhibitor is to bind to a secondary location on the enzyme and stop isocitrate cleavage.

The exact mechanism by which ICL converts isocitrate into glyoxylate and succinate is not fully understood yet. It has been inferred that the ICL reaction mechanism involves a Claisen-type condensation via the formation of an enolic intermediate.\textsuperscript{196} The first step involves deprotonation of the isocitrate hydroxyl group followed by fragmentation of the isocitrate to form glyoxylate and succinate.\textsuperscript{341,342} This has been studied in \textit{Mycobacterium tuberculosis}\textsuperscript{349,336}, but could reasonably be applied to \textit{Pseudomonas aeruginosa} considering the high conservation of the active site. The catalytic residue Cys222 serves as a general acid to aid the formation of the succinate co-product, whereas the residue His224 interacts with the catalytic Cys222 aiding the formation of the aci-carboxylate intermediate that forms upon C(2)-C(3) bond cleavage. The reaction could not take place without a
5. IsoCitrate Lyase (ICL)

general base essential to deprotonate the C(2) hydroxyl group of the isoCitrate substrate. This would be Arg260 as it is within range and already forms hydrogen bonds with the hydroxyl group. The mechanism of action would follow this step-by-step process as shown in figure 55 (adapted from Bhusal et al.199). In addition, magnesium is important for the catalysis. IsoCitrate binds to ICL by chelating the Mg$^{2+}$ via the C(1) carboxylic group and C(2) hydroxyl group. It has been demonstrated that Mg$^{2+}$ is required for optimal activity of the enzyme. Replacement by Mn$^{2+}$ does not support the same catalytic efficiency and other divalent cations would simply abolish the activity of the enzyme313.

![Figure 55: Suggested catalytic mechanism of isoCitrate lyase. Figure adapted from Bhusal et al., 2017199.](image)

To conclude, isoCitrate lyase is of major interest in Pseudomonas aeruginosa. The enzyme is vital for replenishing intermediates of the TCA cycle using anaplerosis. Without a functional glyoxylate shunt, P. aeruginosa is unable to grow on fatty acids or acetate that are the main nutrient sources during infection scenarios. The regulation mechanism proposed here involves metabolite-mediated enzyme activation (IDH) and inhibition (ICL) to achieve a balanced bifurcation of fluxes between two pathways. In future, it will be important to establish the binding affinity of the regulators of ICL to further determine the “druggability” of isoCitrate lyase. The current well-studied inhibitors in Mycobacterium tuberculosis are 3-nitropropionate, itaconate and 3-bromopyruvate. Although these compounds are potent inhibitors of M. tuberculosis ICL and P. aeruginosa ICL, they are nonselective and possess high toxicity, rendering them unsuitable as potential drug candidates. Effects on the growth of rats were reported against itaconate343, 3-nitropropionate caused neurotoxicity344 and 3-bromopyruvate seems to be an energy blocker (glycolytic pathway enzymes in mammalian cells)345.
Therefore, these inhibitors were often used as control experiments in ICL inhibitors studies only. The interest of the research in isocitrate lyase became more decisive with the first structure solved from <i>M. tuberculosis</i>. Since then, it has been suggested as an alluring drug-target not only in bacteria (principally <i>P. aeruginosa</i> and Mycobacteria) but also in fungi<sup>163</sup>. The research I did on <i>P. aeruginosa</i> ICL should provide a preliminary ground for downstream development.
Conclusion

The adaptation of *Pseudomonas aeruginosa* to its environment has proven to be important during infections. The limited access to rich nutrient sources such as glucose, notably in cystic fibrosis lungs, obliges the microorganism to a metabolic rewiring. Increased transcription of isocitrate lyase encoded by *aceA* in *P. aeruginosa* CF isolates indicates the importance of the glyoxylate shunt typically used during growth on fatty acids and acetate as sole carbon sources\(^{36}\). The use of these substrates involves metabolic regulatory mechanisms to ensure a balance between metabolite oxidation for energy gain and metabolite conservation for biomass production. To conserve metabolites for biomass production, *Pseudomonas* is capable of redirecting the carbon flux through the glyoxyllate shunt to bypass the two CO\(_2\)-forming steps and utilizes this anaplerotic pathway to replenish gluconeogenic precursors. This is where the pivotal isocitrate branchpoint between the TCA cycle and the glyoxylate shunt is of particular interest. At this point, three enzymes compete for the same substrate, two isocitrate dehydrogenases (ICD and IDH) and isocitrate lyase (ICL). In comparison with the well-studied model of *E. coli*, in which there are only two competitors for isocitrate and a detailed understanding of fluxes partitioning, the molecular mechanisms that control partitioning of carbon fluxes between the TCA cycle and glyoxylate shunt were poorly studied in *P. aeruginosa*.

In *P. aeruginosa*, all three enzymes have a similar affinity for the substrate. All three K\(_m\) are within the same range with 18 µM, 25 µM and 12 µM for ICD, IDH and ICL respectively. This diverges significantly from what we know in *E. coli*. In enteric bacteria, when the cells evolve in an acetate rich environment, the glyoxylate shunt is activated by transcriptional induction of the catabolite-repressed ICL-encoding gene and MS-encoding gene. After such transcriptional activation, the flux ratio between the TCA cycle and the glyoxylate shunt is controlled by post-translational regulation mediated by AceK. The architecture of the *E. coli* TGB is simple, in case of adaptation to acetate in the environment, phosphorylation-driven inactivation of ICD redirects isocitrate through the glyoxylate shunt. This means that the bacteria are able to conserve two atoms of carbon. If glucose is reintroduced into the environment, AceK dephosphorylates ICD that becomes active, fluxes through the glyoxylate shunt is reduced to be redirected towards the TCA cycle. The K\(_m\) of ICD and ICL in *E. coli* reinforce this understanding as ICD has a higher affinity for isocitrate with a K\(_m\) of 11 µM while ICL has a K\(_m\) of 604 µM. In *P. aeruginosa*, this AceK-dependent inactivation of ICD is functional, even though it does not fully abolish ICD activity. However, the multiplicity of isocitrate dehydrogenases and the highly competing isocitrate lyase in *P. aeruginosa* undermine the *E. coli* model and supposes a more finely tuned architecture of the TGB.

I have demonstrated that ICD is regulated by AceK-mediated phosphorylation but IDH is not. Despite the similarity of ICD and IDH active sites, AceK is incapable of accessing and recognizing the phosphorylation site Ser133 in IDH. Here, I show that branchpoint partitioning at the TGB is
controlled by metabolite-mediated activation of IDH alongside inhibition of ICL. Investigation on potential regulators of IDH found that glyoxylate, oxaloacetate and pyruvate are potent activators of IDH. In contrast, ICL is inhibited by oxaloacetate and pyruvate. This means that the latter two compounds emerge as key regulators tuning an elegant balance of carbon fluxes at the *P. aeruginosa* TGB. Glyoxylate, as a unique product of isocitrate lyase, functions as a “rheostat” of IDH activity. So, according to this paradigm, when glyoxylate levels increase, IDH is activated while ICL is inhibited, this ensures carbon fluxes through the TCA cycle, leading to a decline of glyoxylate levels but also securing replenishment of the NADPH/NADH homeostasis to circumvent the increasing levels of ROS. Alternatively, when glyoxylate levels decrease, IDH becomes deactivated, ICL becomes disinhibited, and carbon flux is driven towards the glyoxylate shunt, leading to a recovery of glyoxylate levels. Oxaloacetate and pyruvate finely regulate the balance of carbon fluxes between the TCA cycle and glyoxylate shunt. When the levels of these two compounds increase, this means that the cells have sufficient gluconeogenic precursors for biomass production, activating IDH and inhibiting ICL to ensure again flux through the TCA cycle. By doing so, the cells also secure flux through the NADPH-forming step that is isocitrate dehydrogenase as NADPH supplies the reductive power necessary to quell the oxidative potential of ROS. This is consistent with a study performed on *Pseudomonas fluorescens* that revealed the importance of NADH/NADPH homeostasis to 1) provide ATP via the process of oxidative phosphorylation and 2) nullify the oxidative environment after the former phosphorylation\textsuperscript{248}. When the levels of oxaloacetate and pyruvate decrease, IDH becomes deactivated and ICL becomes disinhibited which restores fluxes through the glyoxylate shunt. Sufficient levels of another gluconeogenic precursor, PEP, inhibits ICL which is consistent with the previous paradigm. PEP is a product of glycolysis, and by inhibiting ICL, PEP redirects fluxes towards the TCA cycle.

This elegant regulatory mechanism of the TGB in *P. aeruginosa* is in contrast with the well-studied *E. coli* and *Mycobacterium smegmatis*. In the former, post-translational AceK-mediated inhibition of ICD redirects fluxes through the glyoxylate shunt after up-regulation of ICL-encoding gene transcription. Work performed by Dr. Stephen Dolan (Martin Welch laboratory) indicates that in *E. coli*, AceK phosphatase activity is stimulated by α-ketoglutarate, pyruvate and oxaloacetate. This reinforces activation of ICD and greater flux through the TCA cycle as gluconeogenic precursors levels are sufficient for biomass production. In the latter species, *M. smegmatis*, the rheostatic control of IDH by glyoxylate ensures increased flux through the TCA cycle as levels of glyoxylate increase. In contrast, decreased levels of glyoxylate deactivate IDH to ensure that flux go through the glyoxylate shunt again regaining levels of glyoxylate. The architecture of the bifurcation in *P. aeruginosa* partially conserves this rheostat that is glyoxylate on IDH. *P. aeruginosa* also conserves the effect of AceK on ICD, phosphorylation of ICD deactivates ICD and dephosphorylation by the same AceK activates ICD. However, IDH and ICL are remarkably regulated by a subset of compounds that both activate and inhibit the two enzymes. *P. aeruginosa* mechanistic regulation at the bifurcation between the TCA cycle and the glyoxylate shunt resembles aspects seen in other species,
however it remains strikingly different compared with the established paradigm in *E. coli* (figure 56 adapted from Crousilles et al.\(^{347}\)).

The structural data of all three enzymes are a major step to characterise the branchpoint in *Pseudomonas*. Despite the fact that the glyoxylate shunt was first discovered in *Pseudomonas* strain KB1, we knew very little about the enzymology of the bifurcation between the TCA cycle and the glyoxylate shunt. I presented here three novel structures of ICD, IDH and ICL. ICD is a very conserved enzyme among bacteria. Structure solving of *P. aeruginosa* ICD based on *E. coli* ICD demonstrated high level of similarity between the two enzymes. The active site is similar, and superposition work of ICD\(_{Pa}\) with ICD\(_{Ec}\) (co-crystallised with isocitrate and thio-NADP\(^+\)) demonstrated that binding of the substrate and co-factor triggers a hinge motion that closes the active site. This also demonstrated how AceK recognises its substrate (ICD); because the active site of ICD\(_{Pa}\) and ICD\(_{Ec}\) is conserved, the two motifs recognized by AceK (P-loop and ARS) are present in ICD\(_{Pa}\) and there is indeed phosphorylation-mediated regulation of ICD activity. IDH\(_{Pa}\) was solved with NADP\(^+\) and \(\alpha\)-ketoglutarate in one of the two chains present in the asymmetric unit. The crystallographic data demonstrated high levels of similarity of IDH\(_{Pa}\) compared with IDH from other

\[ \text{Figure 56: Comparison of the TGB architecture in *E. coli*, *M. smegmatis* and *P. aeruginosa*. Interactions that activate the target enzyme are shown in red, interactions that inhibit the enzyme are shown in blue. The rheostat mechanism of glyoxylate is present in *P. aeruginosa* as well as the effect of AceK. However, the architecture of the TGB in *P. aeruginosa* is remarkably different. Figure adapted from Crousilles et al.,2018\(^{347}\).} \]
species (i.e. *Mycobacterium tuberculosis, Corynebacterium glutamicum*). Despite the low identity with the mycobacterial species, the secondary structure and active site are similar. Like in ICD, IDH substrate and co-factor binding residues are similar across several species. When IDH binds isocitrate and NADP⁺, there is a hinge motion that pulls the smaller domain of the enzyme to close the active site. In contrast to the first chain of IDH bound to α-ketoglutarate and NADP⁺, the smaller domain of the second chain of *P. aeruginosa* IDH (not bound to the ligands) showed flexibility and static disorder. This suggests that binding of the ligands greatly stabilizes the protein during crystallisation. IDH is not phosphorylated by AceK, although Ser133 in IDH resembles Ser115 in ICD (phosphorylation site). It is interesting to capture the evolution of IDH from ICD, rendering it immune to AceK as the substrate recognition loop of AceK cannot anchor to the IDH site. Finally, ICL in *P. aeruginosa* presented in this work features some remarkable structural observations. The TIM-barrel like structure of *P. aeruginosa* ICL is similar to several bacterial species; however, a significant number of additional residues manifest themselves as structural features projecting away from the globular core of the enzyme. One speculation is that they are involved in protein-protein interactions tightening the tetramer. One of the more remarkable observation is the presence of a catalytic loop in ICL that swings inwards upon binding of the substrate. I speculate that the non-competitive and uncompetitive inhibition of ICL is aided by this mechanism. In presence of isocitrate, the regulator cannot bind to the active site, hence an allostery-mediated inhibition of ICL.

In conclusion, I have shown that the bifurcation between the TCA cycle and the glyoxylate shunt in *P. aeruginosa* has a complex enzymology profoundly different to that in all other organisms characterised to date. The driver behind the study of TGB is its potential as a drug-target. Ongoing research work on *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa* demonstrated the importance of the glyoxylate shunt to survive in a murine lung model of infection. Furthermore, an alternative screening using a glyoxylate shunt-promoting medium to grow bacteria lead to the identification of eight different compounds that had impressive MICs against multiple clinically-relevant Gram-negative pathogens. Furthermore, metabolism and virulence are intrinsically related. The down-regulation of expression of T3SS components in a Δacea mutant PAO1 is an example. The physiological importance of the shunt in infection is not solely due for the survival of the bacteria but also for the virulence and ability of the bacteria to colonize a niche. Therapeutic agents with the ability to inhibit flux through the glyoxylate shunt or to redirect flux through the TCA cycle, would potentially deliver a “double punch” by targeting both metabolic defect and virulence. Considering that *P. aeruginosa* is listed by the WHO in the most critical group of pathogens that require new antibiotics, targeting a core component of the cell that is central metabolism is surely worth more research.
References


Valentini, M., Gonzalez, D., Mavridou, D. A. & Filloux, A. Lifestyle transitions and adaptive

Microbiol.

Smith, R. S. & Iglewski, B. H. virulence factors and host factors in the establishment of urinary tract infections.


Valentini, M., Gonzalez, D., Mavridou, D. A. & Filloux, A. Lifestyle transitions and adaptive


118. Jurtshuk, P. in *Medical Microbiology* (University of Texas Medical Branch at Galveston, 1996).

122


1775 (2013).


Doyle, S. A., Beernink, P. T. & Koshland, D. E. Structural basis for a change in substrate specificity: Crystal structure of S113E isocitrate dehydrogenase in a complex with isopropylmalate,


310. Scholer, A. & Schuller, H.-J. A carbon source-responsive promoter element necessary for activation of the isocitrate lyase gene ICL1 is common to genes of the gluconeogenic pathway in the yeast
352. Huang, K., Li, Z., Jia, Y., Dunaway-Mariano, D. & Herzberg, O. Helix swapping between two α


Appendixes

Appendix 1: Michaelis-Menten plots
Appendix 2: SDS-PAGE analysis of His₆-tagged proteins

The analysis after the purification of His₆-tagged ICD, IDH and ICL indicates that the products correspond to the estimated molecular mass. ICD band is between the 50 kDa and 37 kDa ladder bands, coherent with an estimated molecular mass of 46 kDa. IDH molecular mass is estimated to be 82 kDa, the band on the gel seems slightly underestimated, however the previous *idh* insert sequencing was correct. ICL has an estimated molecular mass of 59 kDa, the third band on the gel is also coherent compared with the molecular mass ladder. Finally, AceK is predicted to be 67 kDa, which is coherent with the band seen on the gel. Overall, the proteins purified are correct and the purification yielded a highly concentrated solution of each protein. The smearing bands below the main ones were indicative of degradation products that did not interfere in any downstream experiment.
Appendix 3: Full gel filtration results of IID

ASTRA 4.90.08 summary Report for ICD 051114_01

File : C:\Program Files\WTC\Astra 4.90.08\enzymes\ICD 051114_01.MDF
Sample ID : ICD prot
Operator : VM/VK

COLLECTION INFORMATION

Collection time : Wed Nov 05, 2014 04:11 PM GMT Standard Time
Instrument type : miniDAWN
Cell type : KS
Laser wavelength : 690.0 nm
Solvent name : water
Solvent RI : 1.330
Calibration constants DAWN : 7.7250e-06
Flow rate : 2.8350e-05
AUX1 : 2.8350e-05
Flow rate : 0.500 mL/min

QELS Information
Instrument : Wyatt QELS
Solvent viscosity : 8.9451e-03 @ 25.0 °C (gm cm²/sec)

PROCESSING INFORMATION

Processing time : Wed Nov 05, 2014 04:37 PM GMT Standard Time
DAWN/AUX1 delay : 0.217 mL
Fit method / model : Zimm
Calculation method : dn/dc + AUX Constant
QELS delay time range : 1.10e-06 - 1.00e+00 sec
Detectors used : 1 2 3

RESULTS

Table:

<table>
<thead>
<tr>
<th>Volume (mL)</th>
<th>PEAK #1</th>
<th>PEAK #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.550 - 8.833</td>
<td>10.192 - 10.667</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slices</th>
<th>35</th>
<th>58</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>A2 (mol mL/g²)</th>
<th>0.000e+00</th>
<th>0.000e+00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit degree</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Injected Mass (g)</td>
<td>4.0000e-05</td>
<td>4.0000e-05</td>
</tr>
<tr>
<td>Calc. Mass (g)</td>
<td>4.8947e-05</td>
<td>5.6701e-06</td>
</tr>
<tr>
<td>dn/dc (mL/g)</td>
<td>0.185</td>
<td>0.185</td>
</tr>
<tr>
<td>Polydispersity(Mw/Mn)</td>
<td>1.001±0.010 (1.0%)</td>
<td>1.024±0.126 (12%)</td>
</tr>
<tr>
<td>Polydispersity(Mw/Mn)</td>
<td>1.002±0.017 (1.7%)</td>
<td>1.049±0.223 (21%)</td>
</tr>
</tbody>
</table>

Molar Mass Moments (g/mol)

| Mn | 9.070e+04 (0.7%) | 1.129e+05 (8%) |
| Mw | 9.079e+04 (0.7%) | 1.156e+05 (8%) |
| Mz | 9.087e+04 (1.6%) | 1.184e+05 (19%) |

R.M.S. Radius Moments (nm)

| Rn | 9.3 (27%) | 27.4 (49%) |
| Rw | 9.3 (27%) | 28.7 (36%) |

C:\Program Files\WTC\Astra 4.90.08\enzymes\ICD 051114_01.MDF
ASTRA for Windows 4.90.08

page 1

Thu Aug 13, 2015 03:37 PM
Appendix 4: ESPript representation of \(P. \text{aeruginosa} \text{ICD}\)
Appendix 5: Extracted information from the validation report of ICD PDB deposition
The percentile ranks, Ramachandran plot and Polygon chart show the final results of structure solving of ICD after deposition to the PDB.
Appendix 6: Full gel filtration results of IDH

<table>
<thead>
<tr>
<th>File</th>
<th>C:\Program Files\WTC\Astra 4.90.08\enzymes\IDH 20141106_01.MDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample ID</td>
<td>IDR prot</td>
</tr>
<tr>
<td>Operator</td>
<td>EN/VK</td>
</tr>
</tbody>
</table>

**COLLECTION INFORMATION**

- Collection time: Thu Nov 06, 2014 03:06 PM GMT Standard Time
- Instrument type: miniDAWN
- Cell type: M5
- Laser wavelength: 690.0 nm
- Solvent name: water
- Solvent RI: 1.330
- Calibration constants:
  - DAWN: 7.7250e-06
  - AUX1: 2.8350e-05
- Flow rate: 0.500 mL/min
- QELS Information:
  - Instrument: Wyatt QELS
  - Solvent viscosity: 6.9451e-03 @ 25.0 °C (gm cm²/sec)

**PROCESSING INFORMATION**

- Processing time: Thu Nov 06, 2014 03:36 PM GMT Standard Time
- DAWN/AUX1 delay: 0.217 mL
- Fit method / model: Zimm
- Calculation method: dn/dc + AUX Constant
- QELS delay time range: 1.10e-06 - 1.00e+00 sec
- Detectors used: 1 2 3

**RESULTS**

<table>
<thead>
<tr>
<th>PEAK #1</th>
<th>PEAK #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume (mL)</td>
<td>7.167 - 7.400</td>
</tr>
<tr>
<td>Slices</td>
<td>29</td>
</tr>
<tr>
<td>A2 (mol mL/g')</td>
<td>0.000e+00</td>
</tr>
<tr>
<td>Fit degree</td>
<td>1</td>
</tr>
<tr>
<td>Injected Mass (g)</td>
<td>4.0000e-05</td>
</tr>
<tr>
<td>Calc. Mass (g)</td>
<td>1.2148e-05</td>
</tr>
<tr>
<td>dn/dc (mL/g)</td>
<td>0.185</td>
</tr>
<tr>
<td>Polydispersity(Mw/Mn)</td>
<td>1.000±0.015 (1.5%)</td>
</tr>
<tr>
<td>Polydispersity(Mw/Mz)</td>
<td>1.000±0.025 (2.5%)</td>
</tr>
<tr>
<td>Molar Mass Moments (g/mol)</td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td>2.355e+05 (1.0%)</td>
</tr>
<tr>
<td>Mw</td>
<td>2.356e+05 (1.0%)</td>
</tr>
<tr>
<td>Mz</td>
<td>2.356e+05 (2.3%)</td>
</tr>
<tr>
<td>R.M.S. Radius Moments (nm)</td>
<td></td>
</tr>
<tr>
<td>Rn</td>
<td>10.4 (32%)</td>
</tr>
<tr>
<td>Rw</td>
<td>10.5 (32%)</td>
</tr>
</tbody>
</table>
Appendix 8: Extracted information from the validation report of IDH PDB deposition

The percentile ranks, Ramachandran plot and Polygon chart show the final results of structure solving of IDH after deposition to the PDB.
Appendix 9: Full gel filtration results of ICL

ASTRA 4.90.08 summary Report for ICL 051114_01

File : C:\Program Files\WTC\Astra 4.90.08\ensys\ICL 051114_01.MDF
Sample ID : ICL prot
Operator : EN/VK

**Peak ID - ICL 051114_01**

**COLLECTION INFORMATION**

Collection time : Wed Nov 05, 2014 04:47 PM GMT Standard Time
Instrument type : miniDAWN
Cell type : K5
Laser wavelength : 690.0 nm
Solvent name : water
Solvent RI : 1.330
Calibration constants
DAWN : 7.7250e-06
> AUX1 : 2.8350e-05
Flow rate : 0.500 mL/min
QELS Information
Instrument : Wyatt QELS
Solvent viscosity : 8.9451e-03 @ 25.0 °C (gm cm²/sec)

**PROCESSING INFORMATION**

Processing time : Wed Nov 05, 2014 05:10 PM GMT Standard Time
DAWN/AUX1 delay : 0.217 mL
Fit method / model : Zimm
Calculation method : dn/dc + AUX Constant
QELS delay time range : 1.10e-06 - 1.00e+00 sec
Detectors used : 1 2 3

**RESULTS**

<table>
<thead>
<tr>
<th>PEAK #1</th>
<th>PEAK #2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Volume (mL)</strong></td>
<td>7.550 - 7.775</td>
</tr>
<tr>
<td>Slices</td>
<td>28</td>
</tr>
<tr>
<td>A2 (mol mL/g²)</td>
<td>0.000e+00</td>
</tr>
<tr>
<td>Fit degree</td>
<td>1</td>
</tr>
<tr>
<td>Injected Mass (g)</td>
<td>4.0000e-05</td>
</tr>
<tr>
<td>Calc. Mass (g)</td>
<td>2.2188e-05</td>
</tr>
<tr>
<td>dn/dc (mL/g)</td>
<td>0.185</td>
</tr>
<tr>
<td>Polydispersity (Mw/Mn)</td>
<td>1.001±0.009 (0.9%)</td>
</tr>
<tr>
<td>Polydispersity (Mz/Mn)</td>
<td>1.002±0.015 (1.5%)</td>
</tr>
<tr>
<td>Molar Mass Moments (g/mol)</td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td>2.318e+05 (0.6%)</td>
</tr>
<tr>
<td>Mw</td>
<td>2.321e+05 (0.6%)</td>
</tr>
<tr>
<td>Mz</td>
<td>2.323e+05 (1.4%)</td>
</tr>
<tr>
<td>R.M.S. Radius Moments (nm)</td>
<td></td>
</tr>
<tr>
<td>Rn</td>
<td>6.4 (52%)</td>
</tr>
<tr>
<td>Rw</td>
<td>6.4 (51%)</td>
</tr>
</tbody>
</table>
Appendix 10: ESPript representation of ICL
Appendix 11: Full results from COCOMAPS analysis on ICL
Equivalent interfaces AB = CD, AC = BD, AD = BC

The first column of graphs presents a black dot at the crossover of two residues i and j, belonging to Molecule 1 and Molecule 2, respectively, if any atom of the two residues are closer than the cut-off distance chosen by the user (default value being 8 Å). The second column of graphs with each contact colored according to the physico-chemical nature of the two interacting residues, violet = hydrophilic-hydrophilic, green = hydrophobic-hydrophobic, yellow = hydrophilic-hydrophobic. The corresponding Accessible Surface Area (ASA) Table for each mirrored interface specifies all the characteristics of the said interface.

<table>
<thead>
<tr>
<th>Interface AB or CD</th>
<th>Buried area upon the complex formation (Å)</th>
<th>Buried area upon the complex formation (%)</th>
<th>Interface area (Å)</th>
<th>Interface area MOL1 (%)</th>
<th>Interface area MOL2 (%)</th>
<th>POLAR Buried area upon the complex formation (Å)</th>
<th>POLAR Buried area upon the complex formation (%)</th>
<th>POLAR Interface area (Å)</th>
<th>NO POLAR Buried area upon the complex formation (Å)</th>
<th>NO POLAR Buried area upon the complex formation (%)</th>
<th>Residues at the interface, TOT (n)</th>
<th>Residues at the interface, MOL1 (n)</th>
<th>Residues at the interface, MOL2 (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buried area upon the complex formation (Å)</td>
<td>3709.0</td>
<td>Buried area upon the complex formation (%)</td>
<td>7.68</td>
<td>Interface area (Å)</td>
<td>1854.5</td>
<td>Interface area MOL1 (%)</td>
<td>7.68</td>
<td>Interface area MOL2 (%)</td>
<td>7.68</td>
<td>POLAR Buried area upon the complex formation (Å)</td>
<td>2229.3</td>
<td>POLAR Buried area upon the complex formation (%)</td>
<td>60.11</td>
</tr>
<tr>
<td>Buried area upon the complex formation (Å)</td>
<td>2736.9</td>
<td>Buried area upon the complex formation (%)</td>
<td>5.67</td>
<td>Interface area (Å)</td>
<td>1368.45</td>
<td>Interface area MOL1 (%)</td>
<td>5.67</td>
<td>Interface area MOL2 (%)</td>
<td>5.67</td>
<td>POLAR Buried area upon the complex formation (Å)</td>
<td>1808.1</td>
<td>POLAR Buried area upon the complex formation (%)</td>
<td>66.06</td>
</tr>
<tr>
<td>Buried area upon the complex formation (Å)</td>
<td>5906.2</td>
<td>Buried area upon the complex formation (%)</td>
<td>12.23</td>
<td>Interface area (Å)</td>
<td>2953.1</td>
<td>Interface area MOL1 (%)</td>
<td>12.23</td>
<td>Interface area MOL2 (%)</td>
<td>12.23</td>
<td>POLAR Buried area upon the complex formation (Å)</td>
<td>4211.6</td>
<td>POLAR Buried area upon the complex formation (%)</td>
<td>71.31</td>
</tr>
</tbody>
</table>
Appendix 12: Extracted information from the validation report of ICL PDB deposition

The percentile ranks, Ramachandran plot and Polygon chart show the final results of structure solving of ICL after deposition to the PDB.