# Structural and Functional Studies of C-Nucleoside Formation Enzymes




Wenbo Li

St Hugh's College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Michaelmas 2023

# Abstract

Throughout the course of human medicinal history, natural products have played a pivotal role in both the promotion of good health and the treatment of diseases. In the 20th century, industrialisation and scientific knowledge saw a boom in active single compounds isolated from nature and used to treat disease directly or after relatively minor chemical modification. However, by the end of the 20th century this approach had fallen out from favour. The difficulty in the *de novo* design of biologically active chemical scaffolds coupled to revolutions in genomic sequencing and chemical biology, has sparked a renaissance of interest in natural products. Nucleosides are characterized by their structurally privileged backbone, and analogues with antiviral, anti-cancer, and antibacterial activities are well known. Conventional molecules with C-N linkages are labile, meaning breakdown *in vivo* can be an issue for some therapies. The substitution of C-C glycosidic bonds holds the promise of preserving the drug properties of the parent compound while rendering the compound essentially undegradable.

The thesis overviews the history of natural product discovery, tracing its evolution from ancient folk traditions to the contemporary application of advanced genetics and chemical modifications for biosynthesis. The introduction chapter details some particularly well-known natural products and their socio-economic impacts. Finally, it emphasizes how structural biology can elucidate the mechanisms of enzyme-catalysis and aid in developing natural product analogies.

Two chapters focus on the structural, chemical, and biophysical studies of a C-nucleoside formation enzyme, ForT, from the formycin (a C-C nucleoside) biosynthetic pathway. The atomic resolution structures for both binary, ternary, and product complexes are described.

Biophysical measurements, computational approaches and kinetics data allow for a detailed description of the mechanism. Chemical synthesis permitted an analysis of the substrates' scope of the enzyme.

The thesis builds towards the study the enzyme mechanism using X-ray Free Electron Laser (XFEL). A critical part of this work is the development of photo-activated cage substrates, these will allow synchronization of the catalytic mechanisms. We have successfully designed and synthesized potential photo-caged substrates. Isothermal titration calorimetry (ITC) identified which compounds bound, whilst liquid chromatography-mass spectrometry showed the molecules were inactive. Co-complex crystal structure showed that the caged compounds, which were identified by ITC, did bind in the same orientation as APDA. These structures allow for the rationalization of what molecules can bind at the active site. We established crystallization conditions which yield microcrystals, a necessary preparation for XFEL time-resolved experiments.

The final experimental chapter describes the structural and functional characterization of a protein related (to ForT) from the methanopterin pathway. This enzyme $\beta$-RFAPS synthesizes a C-nucleoside utilizing PRPP and an aromatic compound, 4-aminobenzoic acid. During purification of a modified $\beta$-RFAPS enzyme and crystal structural analysis indicated the presence of unidentified molecule at the enzyme's active site. We collaborated with Prof. Carol Robinson's group employing native mass spectroscopy and identified the molecule guanosine tetraphosphate (ppGpp). We concluded that ppGpp must be tightly bound, only by treatment with formic acid were we able to remove it. Once ppGpp was removed from the enzyme, the enzyme showed binding affinity of approximately 100 nM. We were able to demonstrate robust initial enzyme kinetic assay.

# Contents

**ABSTRACT** ..... **2**

**CONTENTS** ..... **4**

**ACKNOWLEDGMENT** ..... **7**

**1. CHAPTER ONE** ..... **9**

**INTRODUCTION** ..... **9**

1.1. Natural Product History ..... 9
*1.1.1. Natural Products from Plants* ..... *11*
*1.1.2. Natural Products Originated from Microorganisms* ..... *14*
1.2. Natural Product Drug Discovery ..... 17
1.3. Nucleoside Natural Products ..... 20
1.4. Structural Biology and Time Resolution ..... 25
*1.4.1. Structural Biology History* ..... *25*
*1.4.2. X-ray Free Electron Lasers (XFELs) and Time Resolution* ..... *28*

**2. CHAPTER TWO** ..... **32**

**FORT INITIAL DATA** ..... **32**

2.1. Introduction ..... 32
2.2. Material and Methods ..... 34
*2.2.1. General* ..... *34*
*2.2.2. Cloning and Construct of forT* ..... *34*
*2.2.3. Site Directed Mutagenesis* ..... *34*
*2.2.4. Over-Expression of ForT and Variants* ..... *37*
*2.2.5. Purification and Optimization* ..... *37*
*2.2.6. Crystallization and Co-Crystallization* ..... *38*
*2.2.7. LC-MS Assay of WT ForT Activity* ..... *39*
*2.2.8. Isothermal Titration Calorimetry* ..... *39*
*2.2.9. MIMS Assay of Catalytic Activity* ..... *40*
*2.2.10. Data Collection Processing and Refinement* ..... *40*
2.3. Results ..... 41
*2.3.1. Protein Expression and Purification* ..... *41*
*2.3.2. Crystallization Optimization* ..... *43*
*2.3.3. Overall Structural Analysis* ..... *44*
*2.3.4. Chemical Synthesis of 4-amino-1H-3,5-pyrazole dicarboxylic acid (APDA)* ..... *48*
*2.3.5. LC-MS Activity Assay and MIMS Assay* ..... *50*
*2.3.6. ITC Binding Affinity with PRPP* ..... *53*
2.4. Discussion ..... 57

**3. CHAPTER THREE** ..... **60**

**STRUCTURAL AND FUNCTIONAL STUDIES** ..... **60**

3.1. Introduction ..... 60
3.2. Material and Methods ..... 65
*3.2.1. General ..... 65*
*3.2.2. Site-Directed Mutagenesis ..... 65*
*3.2.3. ITC Binding Assay with PRPP, APDA, and Other Substrates ..... 67*
*3.2.4. LC-MS Assay for Determining the Catalytic Activity of ForT and its Variants ..... 67*
*3.2.5. General screening procedure for establishing the substrate scope of ForT ..... 69*
*3.2.6. ForT-catalysed synthesis of 5-amino-1-(β-d-ribofuranosyl) pyrazole-4-carboxylate 5'-monophosphate ..... 69*
*3.2.7. Structural Biology ..... 70*
3.3. Results ..... 71
*3.3.1. Biophysical Measurements. ..... 71*
*3.3.2. Structural Biology ..... 81*
*3.3.3. Computational Modelling by Collaborators ..... 89*
*3.3.4. Substrate Scope of WT ForT ..... 94*
3.4. Discussion ..... 99

**4. CHAPTER FOUR ..... 105**

**TIME RESOLUTION APPROACH ..... 105**

4.1. Introduction ..... 105
4.2. Material and Methods ..... 108
*4.2.1. Synthesis of Cage Compounds ..... 108*
4.2.1.1. General Synthesis of Intracyclic Photocages ..... 108
4.2.1.2. General Synthesis of Extracyclic Photocages ..... 108
*4.2.2. ITC Binding Assay ..... 109*
*4.2.3. LC-MS Turnover Assay with Cage Compounds ..... 109*
*4.2.4. Laser Activation Assay ..... 110*
*4.2.5. Co-Crystallization with Cage Compounds ..... 111*
*4.2.6. Data Collection and Solve Structure with Cadge Compounds ..... 112*
*4.2.7. Microcrystal Formation ..... 112*
4.3. Results ..... 114
*4.3.1. Binding Affinity of Different Cage Compounds ..... 114*
*4.3.2. Caged Compounds are Not Substrates for ForT ..... 117*
*4.3.3. Laser Activation Experiments with Caged Compounds ..... 118*
*4.3.4. Quantum Yield of Cage Compounds by Laser Assay ..... 121*
*4.3.5. Structure of Cage Compounds in Complex with ForT ..... 124*
*4.3.6. Micro Crystallisation Approach ..... 127*
4.4. Discussion ..... 131

**5. CHAPTER FIVE ..... 136**

**β-RFAPS STRUCTURE AND FUNCTIONS ..... 136**

5.1. Introduction ..... 136
5.2. Material and Methods ..... 138
*5.2.1. General ..... 138*
*5.2.2. Cloning and Construct of β-RFAPS and N-terminal Extension ..... 138*
*5.2.3. Over-Expression of Protein ..... 138*
*5.2.4. Protein Purification ..... 139*
*5.2.5. Crystallization ..... 140*
*5.2.6. Structural Biology ..... 140*
*5.2.7. LCMS End Point Assay ..... 141*
*5.2.8. Native Mass Identification of Ligands as Guanosine Tetraphosphate (ppGpp) ..... 141*
*5.2.9. ITC Binding Assay with PRPP ..... 142*
*5.2.10. LCMS Kinetic Assay ..... 142*
5.3. Results ..... 143
*5.3.1. Protein Expression and Purification ..... 143*
*5.3.2. Crystallization ..... 145*

*5.3.3. Ligand Identification* ..... *146*
*5.3.4. Structural Biology* ..... *148*
*5.3.5. ITC Binding Assay with PRPP* ..... *156*
*5.3.6. LC-MS Assay* ..... *158*
5.4. DISCUSSION ..... 160
**6. CHAPTER SIX ..... 164**
**SUMMARY AND FUTURE RESEARCH ..... 164**
6.1. SUMMARY ..... 164
6.2. FUTURE RESEARCH PLANS ..... 167
**7. BIBLIOGRAPHY ..... 169**

# Acknowledgment

First and foremost, I extend my deepest gratitude to Professor James Naismith for offering me the invaluable opportunity to pursue my Ph.D. in his esteemed laboratory. As my principal mentor, Jim has been an unwavering pillar of support and guidance throughout my doctoral journey. His encouragement and enthusiastic affirmations have been a source of strength, echoing in my mind through every challenge and obstacle that I encountered. His belief in me laid the foundation for all my achievements during this period.

The collaboration with Professor Nigel Richard from the University of Cardiff has been a cornerstone of this project, providing invaluable insights and broadening its impact significantly. I am deeply indebted to Dr. Georgina Girt and Dr. Sisi Gao, whose expertise and collaborative spirit have been pivotal to our team's success. Special appreciation goes to Dr. Haigang Song for his exceptional assistance with crystallization and native mass spectrometry, and to Dr. Huanting Liu for his generosity in supplying vectors and sharing his expertise in cloning and expression techniques.

My heartfelt thanks to the entire Structural Biology Division at university of Oxford and the Rosalind Franklin Institute for their support and camaraderie.

On a more personal note, the bedrock of my journey has undoubtedly been my family. To my father, Mr. Enshan Li, a beacon of resilience and wisdom, whom I owe my foundational principles and work ethic. His life lessons, often shared in quiet moments, have imbued me with the tenacity to pursue my dreams relentlessly. To my mother, Mrs. Chunhua Liu, whose nurturing and compassionate spirit has been my constant source of comfort and courage. Her

unwavering faith in my abilities, coupled with her boundless love and sacrifices, has empowered me to navigate the challenges of this journey with perseverance.

To my wife, Shiqi Liu, your enduring support and companionship have been my sanctuary. Your understanding, patience, and faith in me, even in the face of daunting challenges, have been a wellspring of motivation. Your presence in my life is a testament to the power of love and partnership in overcoming obstacles and achieving dreams.

This thesis stands as a tribute not only to my academic endeavours but also to the unwavering support and love of my family. Their contributions to my journey are etched into every page of this work, and I am eternally grateful for their presence in my life.

# Chapter One

# Introduction

## 1.1. Natural Product History

Throughout human history, natural products (NP) and secondary metabolites have played crucial roles in treating a variety of diseases including cancer, infectious diseases, cardiovascular disease, and multiple sclerosis [1–10]. Despite decreasing interest in its utilization in modern drug discovery and development, natural products still offer great scaffold diversity and intricate structural complexity [6]. Bioactive natural products share several common features such as: higher molecular weight, a few $sp^3$ carbon atoms, several oxygen atoms, fewer nitrogen and halogen atoms, large number of hydrogen-bond donors and acceptors, and greater molecular rigidity [1,11–14]. During evolution, NPs are optimized for different biological activities such as tackling protein – protein interactions or involvement in cell cycle regulation, which are potential drug target for cancer and infectious diseases [15].

Comparing to conventional synthetic chemistries, natural products (NPs) possess unique characteristics that present both advantages and challenges in the drug discovery process. As less than 10 percent of the Earth's biodiversity has been studied for potential bioactivity, yet NPs provide great potentials in drug discovery to enrich the synthetic compounds libraries [8].

Clay tablets found in cuneiform from Mesopotamia dating back to 2600 B.C represent one of the earliest known records of using natural products as treatment reagents. These tablets

described the use of *Cupressus sempervirens* and *Commiphora* oil for treating coughs, colds, and inflammation, which are still used today [8]. The Ebers Papyrus found in Egypt which is dated to 1550 B.C. was written in Ancient Egypt with records of more than 700 drugs originated from different plants [16]. Similarly, Chinese traditional medicine is renowned for using herbal remedies for treating different diseases. ShenNong Herbal was written around 100 B.C. which recorded 365 plant-based drugs and Tang Herbal was documented 659 A.D. with 850 natural products-derived treatment drugs [16]. The Greek physician, Pedanius Dioscorides (90 A.D.), father of pharmacognosy, made great contributions in collection and usage of medicinal herbs that have been widely known for more than 1,500 years [16]. The other great example of Greek natural scientist pioneer and philosopher, was using medicinal herbs for treatments around 300 B.C. Western Europe particularly in Germany, France, England, and Ireland started developing natural treatments with knowledge originated from Churches and Monasteries during the Middle Ages. Owing to geological advantages, the Arabs combined the knowledge from both Greco-Roman and Chinese herbs together with their own experiences to create natural product treatments. Ibn Sina, commonly known as Avicenna, who is not only a pharmacist but also a philosopher, physician, and poet, established the first privately owned pharmacy in the world around the 9th century. His contribution towards medicine and pharmacy was recorded in *The Book of healing*, and *The Canon of Medicine*. He has been recognised as the father of early modern medicine [8].

Natural product treatments exist in different forms, such as herbs, concoctions, and extracts [16]. Treatments contain one or more key bioactive chemicals which until recently remained unidentified [16]. The utilization of natural products for disease treatment derived from centuries of experiential trials, as modern scientific research techniques were unavailable [17]. Natural products are now known to be biosynthesized during secondary metabolism (hence the other name secondary metabolite), are produced for purposes including organism defence mechanisms, signalling or environmental adaption [16]. Unlike primary metabolites, secondary metabolites are generally not essential for organism growth and development [18]. The biosynthesis of secondary metabolites encompasses a wide range of chemical structures and biological functionalities [16].

### 1.1.1. Natural Products from Plants

Natural product can be sourced from a variety of different origins. Plants account for majority of the biomass on earth and have existed for millions of years. They have been evolving and adapting across different environments consequently producing a variety of secondary metabolites (natural products). These have been used for medicinal purposes throughout history [19].

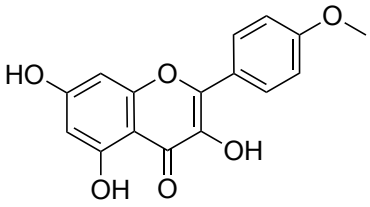
For example, Northern Americans and Northern Europeans consumed the plant *Ligusticum scoticum* as a means to treat stomach flatulence and protect against infections [16]. The liquid secreted from the stem of *Alhagi maurorum*, has been commonly documented and used by the Ayurvedic people to treat fever, anorexia, epistaxis, constipation, obesity, leprosy, and dermatosis. People in India and Israel use its root extract to treat asthma and bloody diarrhoea [16]. Not all plant-based treatments are safe though, *Atropa belladonna* is one of the most toxic plants found in Europe, North Africa, and Western Asia, which was used as an anaesthetic until 19$^{th}$ century [16].

Approximately 80% of pharmaceutical drugs derived from plants continue to be linked to their historical pharmacological functions [20]. For primary or self-administered plant-based medicines remain the most common treatment [21]. In addition, they can provide insight into modern biochemistry and metabolism, as well as serving as starting points for new medicines.

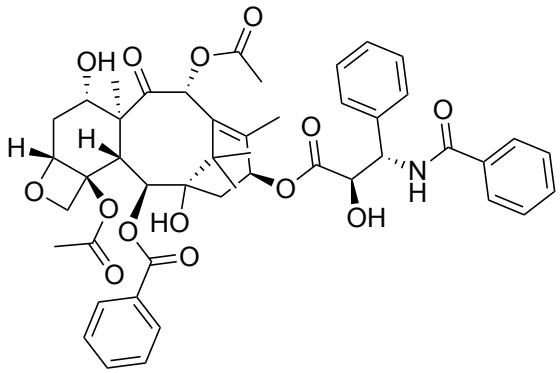
Artemisinin (**1**) is one the most famous example of traditional herbal medicine translated into modern drug treatment. Prof. Youyou Tu was award with Nobel prize in Physiology or Medicine in 2015 for her contribution in discovery of using artemisinin (**1**) as malaria treatment [22]. Artemisinin (**1**) was originally found in the plant *Artemisia annua* and documented in Chinese traditional medicine for treating fevers and chills. It was introduced as first-line malaria treatment in 2000, as a result, between 2000 and 2015, the global malaria incidence rate reduced by 25 percent and mortality rate reduced by 42 percent [3,22]. Paclitaxel (Taxol®) (**2**), one of the most widely used drug for cancer chemotherapy, has been discovered and isolated from the bark of *Taxus brevifolia* by United States Department of Agriculture (USDA) in 1962 [23]. It can induce cell death through reducing microtubules dynamicity and

prevent depolymerization [24]. Due to the relatively low natural extraction yield of paclitaxel from trees, the drug is now produced synthetically from its precursor baccatin III (**3**) that exists in needles of *Taxus brevifolia* [16,25]. The FDA approved the use of Taxol® in 1992 for treatment of breast cancer and expanded to use in other cancers afterwards [26].

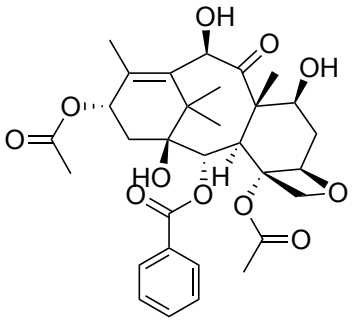
Calanolide A (**4**), the first plant-based compound exhibiting anti-HIV activity, was discovered and isolated from the leaves and twigs of *Calophyllum lanigerum* during the anti-HIV screening programme conducted by the National Cancer Institute (NCI) in 1980s [27]. Calanolides are a class of non-nucleoside reverse transcriptase inhibitors (NNRTI) that not only showed significant anti-HIV activity, but also exhibited potential in anticancer, antiparasitic, and antimicrobial[27]. The phase I clinical trial for assessing safety and effectiveness of calanolide A (**4**) started in 2013 and was completed in 2016 [27]. In 2017, a structural analogue of calanolide A, F18 (10-chloromethyl-11-demethyl-12-oxo-calanolide A) has shown more potent anti-HIV activity than calanolide A [28,29]. Although calanolide A has not yet been approval by the FDA or other drug regulatory bodies, calanolides itself and its structural derivatives, serve as an investigation starting points for new class of anti-HIV and anti-cancer NNRTI drugs [27].



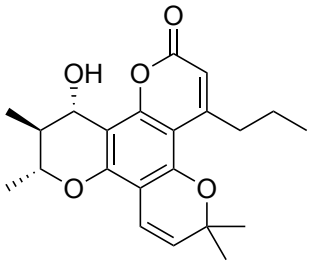

Artemisinin (**1**)




Paclitaxel (**2**)




Baccatin III (**3**)




Calanolide A (**4**)

**Figure 1.1.** Examples of famous natural products from plants used as treatments.

### 1.1.2. Natural Products Originated from Microorganisms

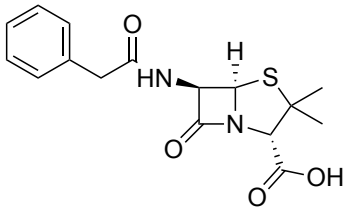
Microorganisms are second to plants for biomass on earth [30]. They not only inhabit nearly all ecological niches, but also form intricate relationship with other organisms [31]. Microorganisms biosynthesize a large range of primary and secondary metabolites [32,33].

Fungi, have been widely used in folklore and in traditional medicine. *Fomitopsis betulina*, previously as known as *Piptoporus betulinus*, is a fungus grown on the bark of birch trees that has been valued as disinfectant agent and used to control bleeding [16]. The field mushroom, *Agaricus campestris* which is widely distributed and cultivated over the world, has been known for treating moth ulcers and sore throat [16]. In Scotland, a thin slice of *A. campestris* is applied to heal scalds and burns. *A. campestris* contains bioactive natural products which can enhance insulin secretion *in vitro* by unclear mechanisms [34]. Lichen is well documented as a raw material in cosmetics and medicine in ancient China and Egypt [16]. While no drug product derived from lichens has been approved yet, they continue to play an important role in the dye industry [35]. Although microorganisms have been part of human history for entire civilization, recent advances and developments in microbiology and synthetic biology are progressively impacting enzymology, pharmacology and drug discovery[16].

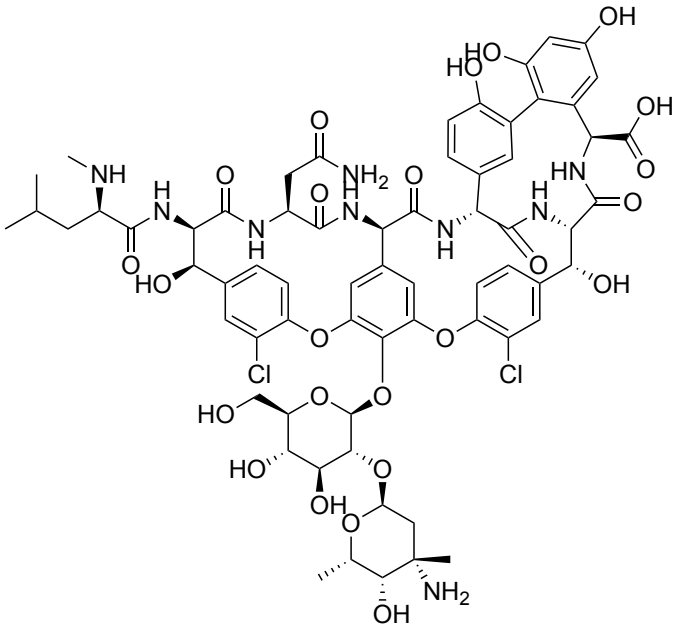
The discovery of penicillin (**5**) from the fungus *Penicillium notatum* by Alexander Fleming in 1929 is perhaps the most celebrated discovery of a natural product drug [36,37]. The 1945 Nobel Prize in Physiology and Medicine was awarded to Chain, Florey, and Fleming for their work in not only finding but also mass production of penicillin (**5**) that have saved countless lives. The isolation and commercialization of biosynthetic penicillin started a revolution of drug discovery [37–40]. Penicillin (**5**) contains a beta-lactam ring that can inhibit the enzyme beta-lactamase thus preventing the formation of bacteria cell wall and causing cell death [41]. During the 1950 to 70s, the beta-lactam core containing nature products was intensively studied and other beta-lactam antibiotics identified including nocardicin A (**6**), imipenem (**7**), and aztreonam (**8**) [42–44]. During this period other new classes of antibiotic were isolated and used including the glycylcyclines (derived from tetracycline).

Another well-known natural product is the so called “last resort” antibiotic vancomycin (**9**), a glycopeptide isolated from *Amycolatopsis orientalis* by Edmund Kornfeld in 1953. It exhibits broad range antibacterial activities against both gram-negative and gram-positive bacteria such as *Streptococci* and *Staphylococci*. Vancomycin inhibits the process of bacterial cell wall biosynthesis by inhibiting a transpeptidase preventing elongation of peptidoglycan [45] leading to cell death. The FDA has approved the use of vancomycin to treat patients with severe infection, multi-drug resistance bacterial infection or hypersensitivity to penicillin [10]. The biosynthetic pathway of vancomycin has been well described, although next generation total synthesis of vancomycin has been reported, industrial manufacture of vancomycin is still purified from fermentation without undergoing further chemical modification[46–48].

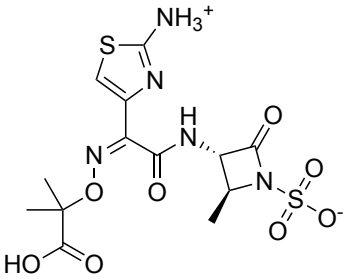
Doxorubicin (**10**) is a derivative of amrubicin (**11**) that was isolated from *Streptomyces peucetiu*s and can disrupt topoisomerase-II-mediated DNA replication process. This increases quinone-type free radical product which induces apoptosis (programmed cell death) [49]. Doxorubicin, under the trade name Adriamycin® is used to treat breast cancer, bladder cancer, lung cancer, thyroid cancer, acute leukaemia, and lymphomas [10,50].



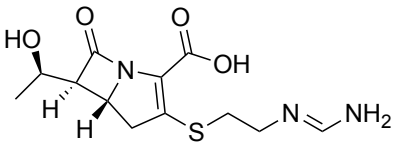
Penicillin (**5**)



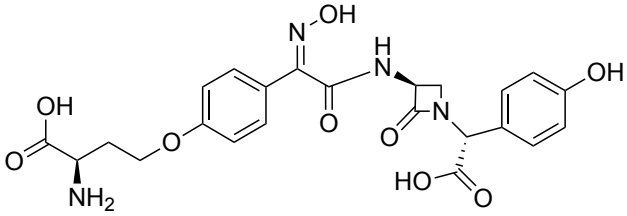
Vancomycin (**9**)



Aztreonam (**8**)

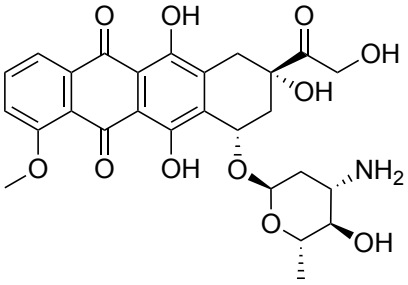


Imipenem (**7**)

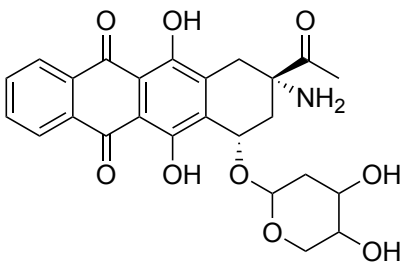


Nocardicin A (**6**)

**Figure 1.2.** Examples of natural products developed antibiotics.



Doxorubicin (**10**)



Amrubicin (**11**)

**Figure 1.3.** Examples of natural products developed anti-cancer treatments.

## 1.2. Natural Product Drug Discovery

The foundation of natural product drug discovery methods relies on bioactivity guided natural product isolation. The process begins by identifying organisms or materials that are bioactive then extracting and concentrating the activity, discarding other compounds before finally reaching a single compound. (Figure 1.4)

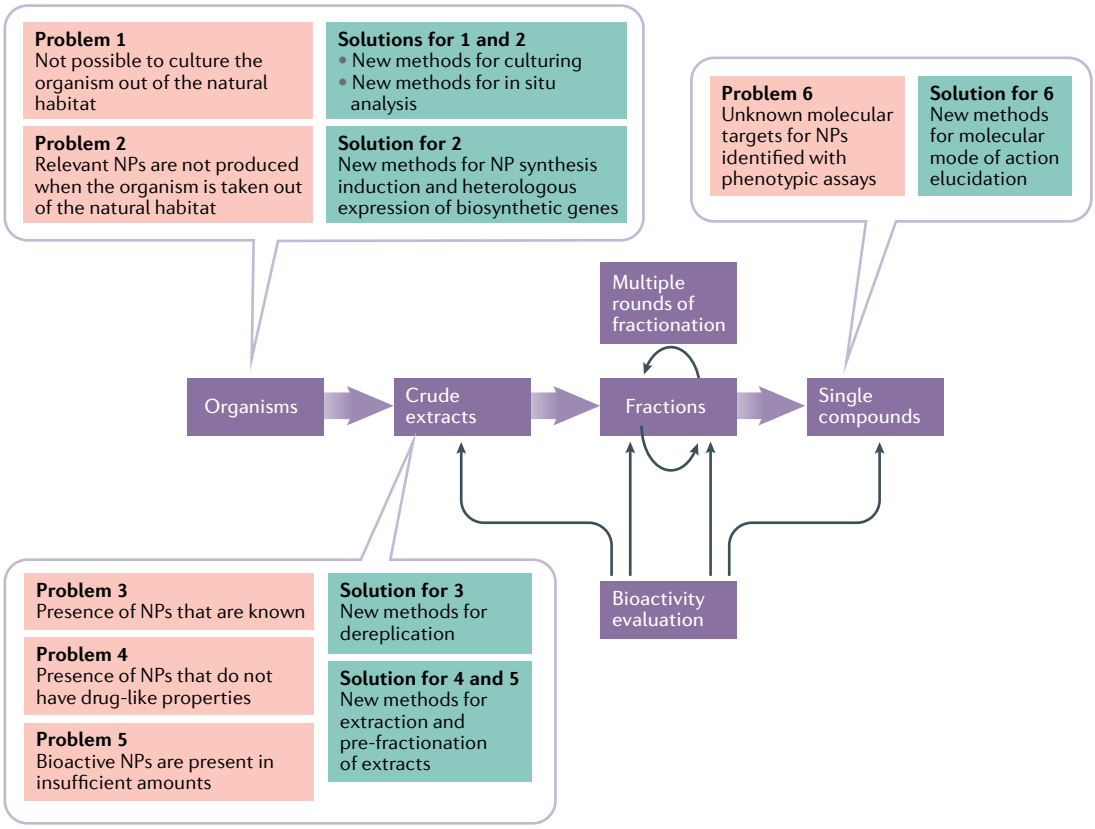



**Figure 1.4.** Workflow chart of natural product drug discovery process. (Figure modified from [15])

It was noted that by changing environmental culturing conditions such as pH, temperature, and nutrient sources resulted different NPs from the same organism [51,52]. This has been called "One Strain Many Compounds" (OSMAC) [53]. OSMAC has been widely used for bioactive compounds discovery and identification. However, it is sometimes challenging to predict and mimic the complexity of natural culturing environments [54,55]. One solution to these problems is by introducing co-culturing systems with adding "helper strains'' to the microorganisms to

stimulate their natural interactions enabling its targeted metabolic programmes for natural product production and identification [56]. Once example is the co-culture of fungi with *Streptomyces* species for the development of novel antibiotic NPs [57,58]. More recently, the iChip platform and omics techniques have been developed to screen and select the native environment for microorganism culturing and even can predict the specific needed environment before culturing by metagenomics analysis. With the advance of these technologies, the successful development of new antibiotic teixobactin and isolation of sacchairbacterium species with no previous culturing information were achieved [59–61].

Analytical techniques are essential for bioactivity guided natural products drug discovery. Through the advancement of chromatography, spectrometry, and computational approaches, the identification and structural elucidation of bioactive natural products (NPs) have become possible [62–64]. Once the phenotypic assay has identified the potential bioactive target, the metabolite profiling of NPs from crude extracts can be analysed by NMR spectroscopy or mass spectrometry with combination of liquid chromatography [65,66]. Liquid chromatography separates the crude extract into fractionated samples that can distinguish different NPs and its isomers [67,68]. Although conventional NMR spectroscopy has limitations due to poor sensitivity, it can analyse both unfractionated and fractioned sample to obtain quantitative and detailed structural information [69]. High-resolution mass spectrometry combined with liquid chromatography is currently the standard for metabolite profiling only requiring a small amount of sample [70]. Structural information of NPs can be determined by higher-field NMR spectroscopy and more sensitive probes [71,72]. More recently microcrystal electron diffraction (MicroED) has yielded high resolution structure data from very small amounts of material natural product drug discovery research [73,74].

This traditional process is laborious and has multiple failure points. Once the organisms have been successfully confirmed and cultured, there might be insufficient amounts of bioactive NPs present in the crude extracts, or the NPs do not exhibit drug-like properties. These problems can be mitigated by optimizing methods for extraction, dereplication and pre-fractionation of extracts. Ultimately, the approach is limited by the sensitivity of assay methods if one is trying to isolate an unknown active compound. The molecular biology

revolution has led to more innovative means of production through the cloning of the genes required natural product biosynthesis.

Bioactive natural products usually contain polyketide or peptide cores that are biosynthesized by enzymes like polyketide synthase (PKSs) and non-ribosomal peptide synthetases (NRPSs). The biosynthetic enzymes are encoded on conserved genome clusters [75]. Genome mining uses biosynthetic knowledge and bioinformatic tools to predict the chemical scaffold of metabolic products from identified NPs biosynthetic gene cluster [76–78]. This brings significant benefits for dereplication purposes, and when combined with biophysical / analytical techniques, it can significantly accelerate the process of identification, determination, and characterization of NPs [74,79–81]. Figure 1.5 shows the workflow of genome mining based approach natural products drug discovery [15]. Natural products provide diverse chemical scaffold and various bioactivities for developing new drugs or pharmacokinetic core of novel drugs. To enhance the pharmacological properties of NPs that exhibit suboptimal characteristics, two strategies are commonly employed: 1. Total chemical synthesis and chemical derivatization. 2. Biosynthetic engineering to produce target analogues. A combination of both approaches can be employed.

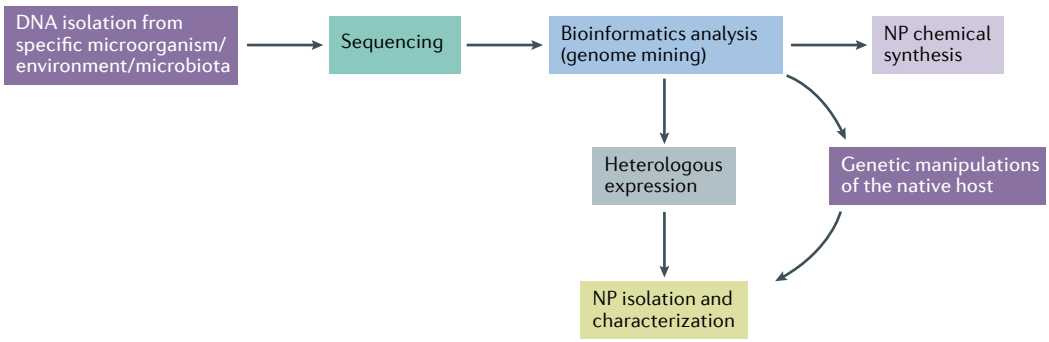



**Figure 1.5.** Workflow chart shows the genome mining based approach in natural products drug discovery. (Figure modified from [15])

## 1.3. Nucleoside Natural Products

Natural products have evolved to privilege certain scaffolds in order to cross cell membranes and perform biological activities. By 2014, based on Newman's statistics number of total approved new drugs, more than 60% of the drugs are directly or indirectly developed from natural products [3]. However, since 1990s, pharmaceutical industries have demonstrated decreased interest in NPs. The declining trend in registering NPs as new drugs can be explained by the challenges posed by their often-complex stereochemical structures, the challenges of *de novo* chemical synthesis, and the limited availability of these compounds [4,82–84].

Nucleoside analogues have a chemical scaffold that consists of a five-carbon sugar ring and a nucleobase analogue. The nucleotides adopt a common structure consisting at least one phosphate or phosphate like group [85]. The most prevalent nucleobase moieties include analogues of adenine (**12**), thymine (**13**), cytosine (**14**), guanine (**15**), and uracil (**16**). With this structural mimicry, the nucleoside analogue is recognised easily by cellular or viral enzymes [86,87]. Many nucleoside analogues have demonstrated antiviral and anticancer activity, including several now in use in the clinic gemcitabine (**17**), fludarabine (**18**) and vidarabine (**19**) [88,89]. Nucleoside analogues enter the cell through different specific nucleoside transporters located on the cell membrane, which is phosphorylated by nucleoside kinases before utilisation by DNA/RNA polymerases. This can suppress DNA/RNA replication indirectly through the accumulation of mutations in the genome or directly via a chain termination mechanism[88,90]. The same mechanism operates in anti-cancer and anti-viral compounds.

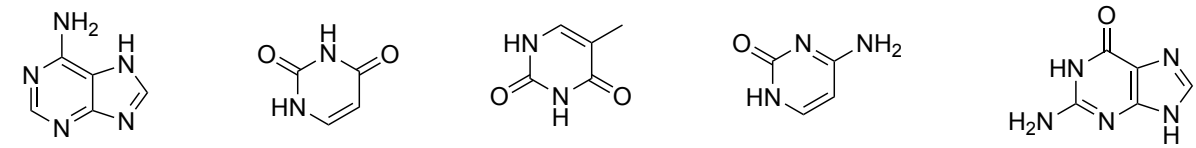



**Figure 1.6.** Structure of the five nucleobases encountered in DNA and RNA.

A drawback of nucleoside analogue drugs is the *in vitro* instability of the glycosidic bond. Both cells and viruses have their own mechanisms in protecting its genetic integrity, one strategy is by removing the damaged or modified nucleobases with DNA/RNA glycosylases [91]. Bacteria have evolved pathways to improve the stability of these molecules and prevent the hydrolysis of the nucleotide bond, replacing the nitrogen atom on the nucleobase ring with carbon atom to create C-nucleosides [92–94]. Showdomycin (**20**), a uridine analogue C-nucleoside, the first C-nucleoside with antibacterial activities was discovered by Nishimura *et al*. in 1964 from *Streptomyces showdonesis* [95–97]. Formycin A (**21**) and pyrazofurin (**22**) are the two other naturally existing C-nucleoside compounds that display important biological activities. Formycin demonstrates both antibiotic as well as antiviral properties (against HIV), additionally anticancer (leukaemia) activity has been reported [98–101]. Pyrazofurin (**22**) is more specific as an antiviral compound against RNA viruses [102].

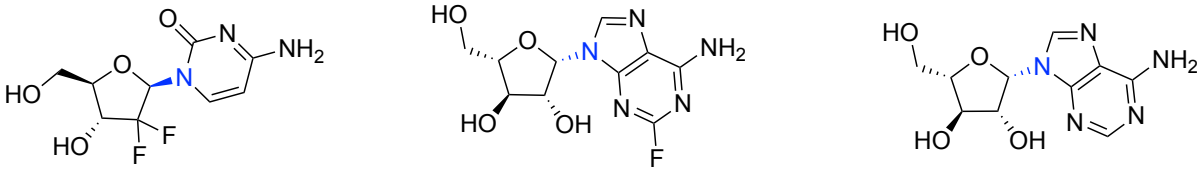



**Figure 1.7.** N-nucleoside analogue drugs.

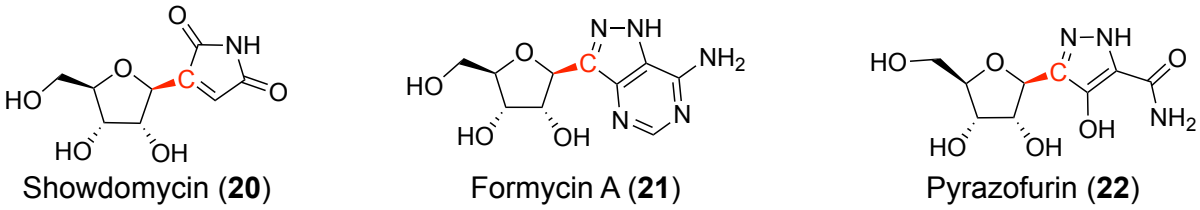



**Figure 1.8.** Naturally existing C-nucleosides and its analogues.

Immucillin-H (**23**) and BCX4430 (Galidesivir) (**24**) were inspired and derived from formycin; where in both molecules the ribose sugar is replaced with iminosugar, creating an additional amine group [103–105]. The iminosugar contains an extra NH group that acts as a hydrogen bond donor in the enzyme binding site [106]. Immucillin-H (**23**) is a potent molecule in leukemia treatment and works by inhibiting purine nucleoside phosphorylase. BCX4430 (**24**) demonstrates promising results against different viruses such as Ebola, Yellow Fever, Zika and Coronaviruses [103,107–109]. Remdesivir (**25**) (GS5734) is another C-nucleoside prodrug derived from formycin demonstrating promising effectiveness against human coronavirus, including the SARS-CoV-1 and the pandemic-related Covid-19 (SARS-CoV-2)[110]. The biosynthetic pathways of formycin and pyrazofurin can be used for C-nucleoside drug development. Gaining a comprehensive understanding of the molecular mechanisms underlying biological synthetic tools is essential for the rational engineering of enzymes involved in the pathway. This understanding paves the way for producing novel C-nucleosides using these engineered enzymes.

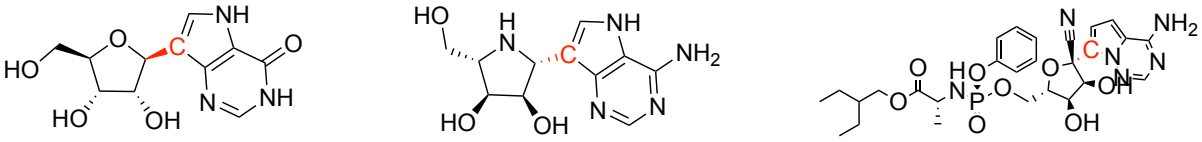



**Figure 1.9.** Clinical application of C-nucleoside drugs.

The *de novo* organic synthesis of formycin's side chain requires approximately 15 steps [111,112]. This process involves the introduction of multiple protecting groups and the formation of numerous undesired side products[111]. Figure 1.10 summarises the chemical synthetic route of formycin A from commercially available compound 2,3,5-tri-O-benzyl-D-ribofuranose. In contrast, *in vivo* biosynthesis via fermentation is much more efficient and is used commercially for production[113,114]. The formycin pathway has been identified from the gene cluster of *Streptomyces kaniharaensis* SF-557 [115,116].

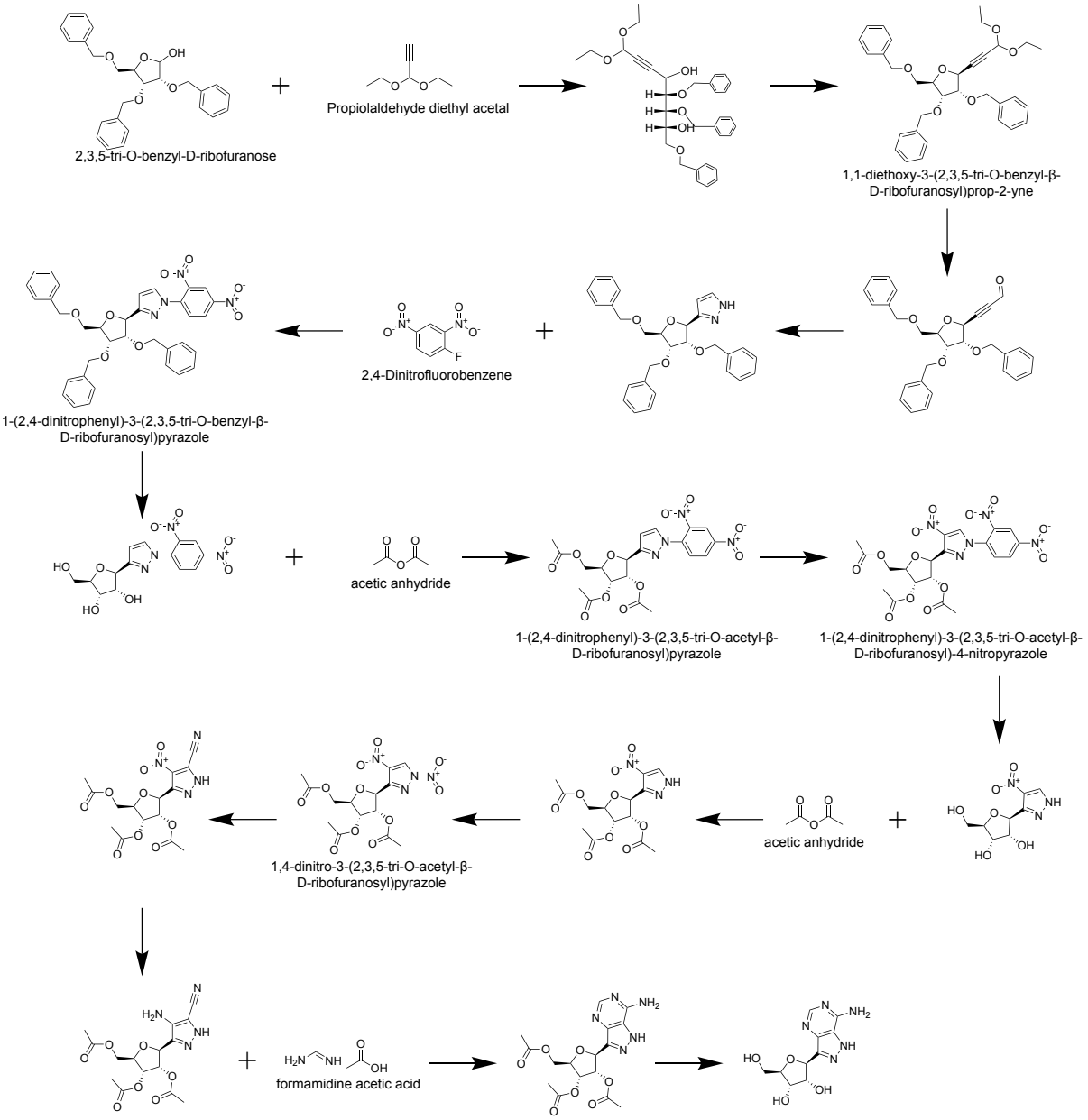



**Figure 1.10.** Overview of the chemical synthetic pathway of Formycin A from commercially available 2,3,5-tri-O-benzyl-D-ribofuranose. Chemical synthesis requires more than 15 steps to synthesize the side chain of formycin, which involves introducing many different protection groups.

In the biosynthetic pathway, atoms in formycin are thought to be derived from glutamine and lysine [117]. In the current model, the enzymes ForJ, ForK and ForL are thought to catalyse hydrazine formation as part of the pyrazole ring [115,118]. The reaction is initiated by a flavoprotein $N^6$-monooxygenase (ForJ) using lysine and NADPH acts as a cofactor to form the N-N moiety of the hydrazine structure [102]. Formation of the pyrazole side chain on formycin is followed by a pyridoxal 5'-phosphate (PLP) dependent aminotransferase (ForK) that aminates the hydroxyl group into amine, creating the pathway intermediate 4-amino-1*H*-pyrazole-3,5-dicarboxylic acid (APDA) (**26**) [119]. Formation of C-C glycosidic bond occurs between phosphoribosyl pyrophosphate (PRPP) (**27**) and 4-amino-1*H*-pyrazole-3,5-dicarboxylic acid (APDA) (**26**), where PRPP contributes its ribose moiety using hydrolysis of the high energy pyrophosphate as energy source to drive the reaction [115,120]. This reaction is catalysed by a beta-ribofuranosylaminobenzene 5'-phosphate (beta-RFA-P) synthase (named ForT) to produce the important intermediate carboxyaminopyrazole ribonucleotide (CAPR) (**28**), releasing pyrophosphate and carbon dioxide [121] rendering the reaction irreversible (Figure 1.11). The key intermediate C-nucleoside CAPR (**28**) is further modified by the enzymes in the synthetic pathway to close the ring structure adjacent to pyrazole ring to produce formycin A [102].

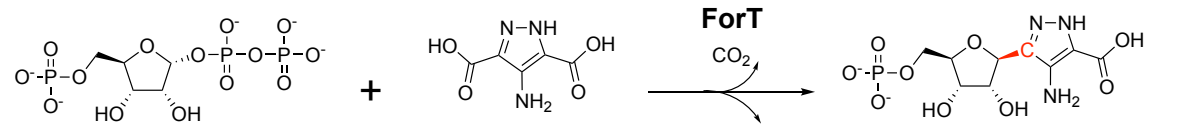



**Figure 1.11.** C-C bond formation reaction catalysed by ForT between PRPP (**27**) and APDA (**26**).

C-nucleosides natural products are fascinating molecules with diverse structures and functions. They are of interest not only for their biological activities but also for their potential as pharmaceutical tools for drug development. The formation of C-nucleosides typically involves enzymes known as C-nucleoside biosynthetic enzymes or C-nucleoside synthases. It is important to understand the how nature evolves to solve the puzzle for high energy C – C bond formation.

## 1.4. Structural Biology and Time Resolution

### 1.4.1. Structural Biology History

The strong correlation between structure and function, makes the structural description of protein molecules a fundamental pillar of biochemistry and biology. Dated back to 1600s, light microscope was the main power tool for studying structures of biological question until twentieth century. In 1873, Ernst Abbe defined the resolution limit of light microscopy to be the wavelength of visible light itself. The wavelength of visible light is around 0.5 µm, and the highest resolution that any illuminating source can reach (in theory) is half the illuminating wavelength. A bacteria cells is around 1 µM, allows individual bacteria to be imaged. Other cells such as those from animals and plants are much larger around 10-100 µm in diameter. Their large subcellular organelles including nuclei and mitochondria can be resolved (with difficulty) [122]. Studies at molecular level are impossible with light microscopy, DNA, RNA, protein, carbohydrate, and lipid macromolecules are typically around 0.002 – 0.05 µm in diameter, and the atomic level insights requiring 1Å (0.1 nm) in resolution [123].

X-rays are a form of electromagnetic radiation with a wavelength of around 0.1 nm much shorter than visible light. The theoretical resolution for X-ray imagining allows the resolution of atomic details. A simple X-ray microscopy capable of visualising atoms does not exist, because a lens that could focus the X-rays sufficiently does not exist [123,124]. In 1912 Max Von Laue in Germany showed that crystals scattered X-ray radiation, won the 1914 Nobel Prize. William and his son Laurence Bragg described the mathematical relationship between wavelength and the crystal planes, the Bragg equation, won the 1915 Nobel Prize. The intensity and phase of the diffracted beam are functions of the atoms within the diffracting plane. Detection technology then and now does not allow direct measurement of the phase. Although it was possible to use diffraction to determine the structure of simple salts such as sodium chloride, the arrangement of atoms within the crystals of more complex molecules remained extremely challenging, relying on Patterson methods and heavy atom substitution [125]. This method did allow the determination of the structure of penicillin [126] and vitamin B12 [127,128], work for which Dorothy Hodgkin was awarded the Nobel Prize in 1964. A major technological breakthrough came with the invention of direct methods by Jerome Karl [129] and

Herbert Hauptman [130], which were awarded the 1985 Nobel Prize in chemistry. This method permitted calculation of the phases from experimental data, and thus enabling the solution of the crystal structure. The continuing developments in methods and computational power, had allowed by the 1950s, X-ray crystallography to determine the structure of protein crystals and DNA fibres. By the end of the 1980s the molecular biology revolution permitted easy access to pure protein, coupled to routine access to synchrotrons and high-power computing, leading to a massive increase in the number. During this period, NMR, electron microscopy, electron diffraction, and neutron diffraction were all used to determine macromolecular structures to resolutions allowing placement of atoms [131–134]. Most recently artificial intelligence approaches AlphaFold and RosettaFold have revolutionised structural biology. AlphaFold is cutting-edge computational method that developed by DeepMind, using deep learning techniques, specifically deep convolutional neural networks, and recurrent neural networks to analyse amino acid sequences to give remarkable accurate prediction of the 3D coordinates of a protein. RossettaFold is also a computational approach in structural calculation, developed for over two decades by University of Washington and UCLA. It combines physics-based modelling with bioinformatics techniques to predict possible protein conformations and energy states [135–138]. Both AlphaFold and RossettaFold had significant impacts on structural biology, complementing traditional methods like X-ray crystallography and cryo-electron microscopy to advance understanding of protein functions and facilitating drug discovery.

The advancements in structural biology have catalysed breakthroughs in various fields such as biochemistry, biotechnology, bioinformatics, and synthetic biology, which have been awarded with Nobel Prize (Table 1.1).

**Table 1.1** Summary of major Nobel prizes for Structural biology

| Year | Name | Area | Achievement |
|---|---|---|---|
| 1914 | Max von Laue | Physics | Discovery of the diffraction of X-rays by crystals |
| 1915 | Sir William Henry Bragg and William Lawrence Bragg | Physics | Analysis and solved crystal structures by X-rays |
| 1954 | Linus Pauling | Chemistry | Structural elucidation of complex substance including alpha helix and beta strand |
| 1962 | Max Ferdinand Perutz and John Cowdery Kendrew | Chemistry | Structural determination of globular proteins |
| 1962 | Francis Harry Compton Crick, James Dewey Watson, and Maurice Hugh Ferderick Wilkins | Physiology or Medicine | Discovery molecular structure of nucleic acids |
| 1964 | Dorothy Crowfoot Hodgkin | Chemistry | Structural determination of penicillin and vitamin B-12 by X-rays |
| 1982 | Aaron Klug | Chemistry | Development of crystallographic electron microscopy to elucidate nucleic acid protein complexes |
| 1985 | Herbert A. Hauptman and Jerome Karle | Chemistry | Development of direct methods for X-ray crystallography |
| 1988 | Johann Deisenhofer, Robert Huber and Hartmut Michel | Chemistry | Determination of 3D structure of a photosynthetic reaction centre |
| 1991 | Richard R. Ernst | Chemistry | Development the methodology of high resolution nuclear magnetic resonance spectroscopy |
| 1994 | Bertram N. Brockhouse and Clifford G. Shull | Physics | Development of neutron scattering techniques for structural studies |
| 1996 | Paul D. Boyer, John E. Walker, and Jens C. Skou | Chemistry | Structural and mechanism elucidation of adenosine triphosphatase |
| 2002 | John B. Fenn, Koichi Tanaka, and Kurt Wuthrich | Chemistry | Development of mass spectrometric analysis and nuclear magnetic spectroscopy in 3D structural determination |
| 2003 | Peter Agre and Roderick MacKinnon | Chemistry | Structural determination of water and ion channels |
| 2006 | Roger Kornberg | Chemistry | Structural determination of DNA-dependent RNA polymerase |

| | | | |
|---|---|---|---|
| 2009 | Venkatraman Ramakrishnan, Thomas A. Steitz, and Ada E. Yonath | Chemistry | Structural and functional studies of the ribosome |
| 2012 | Robert J. Lefkowitz and Brian K. Kobilka | Chemistry | Structural determination of G protein-coupled receptors |
| 2017 | Jacques Dubochet, Joachim Frank, and Richard Henderson | Chemistry | Developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution |

### 1.4.2. X-ray Free Electron Lasers (XFELs) and Time Resolution

The most common structural techniques such as X-ray crystallography and Cryo-Electron microscopy determine the static structure of the protein. The use of lasers and Laue diffraction had been used to try to probe dynamics of enzymes inside crystals. Ilme Schlicting has been pioneered to study time-resolved enzymatic mechanism through serial femtosecond crystallography. She demonstrated photo activation methods to determine time-resolved dynamic structure of the active site for a photosensitive enzyme in picosecond timescale [139]. In contract, NMR does determine a dynamic structure, allowing motions (typically large domain movements) on the microsecond timescale to be determined. The arrival of the X-ray free electron lasers, which produce X-rays suitable for high resolution diffraction, allowed the possibility of time-resolved macromolecular crystallography (TRX). XFELs produce ultra-brilliant X-rays (three orders of magnitude more brilliant than synchrotrons) [140,141] that allow diffraction before destruction [142,143]. In practice by using a short pulse (< 100 fs) “pink” (a narrow range of wavelengths) beam a diffraction image can be recorded from a very small (2µm x 2µm x 2µm) crystal [144]. The intense beam ionises atoms in the crystals, the resulting positive charges repel each other leading to the so-called columbic explosion. Crucially this event is slower than the formation of the diffraction image. By determining structures at the nanosecond timescale, the possibility to experimentally determine the chemistry of enzymatic reactions [145–147]. This would allow understanding of not only the dynamics of the reaction but also observing the predicted short-lived key intermediates from enzyme catalytic reactions [148,149].

Due to the "collect before destroy" XFELs experiments require many (1,000's) microcrystals to obtain a full data set (Figure 1.12). In the very first experiments crystal slurries were sprayed in a supersonic jet into the beam, relaying on probability to achieve an interaction between crystal and beam [150–152]. In order to reduce the amount of sample required in XFELs experiments, different sample delivery strategies have been developed. For example, 'toothpaste' injector was designed for membrane proteins like GPCRs and rhodopsins [153–155]. For soluble proteins, protein microcrystals are incorporated into viscous liquids first. This controlled extrusion of crystals into X-ray pulses for diffraction, allowing a gradual process [156–159]. Fixed target approach has also been used for XFELs shooting microcrystals. This has been achieved by placing the protein microcrystals onto a thin film or a grid and moving X-ray beam to diffract each crystal on the grid [160,161]. In addition, tape drive delivery system, although not as commonly used as some other previously mentioned sample delivery methods, mounts microcrystals onto a thin, continuous tape that slowly fed through the XFEL beam path for data collection [162].

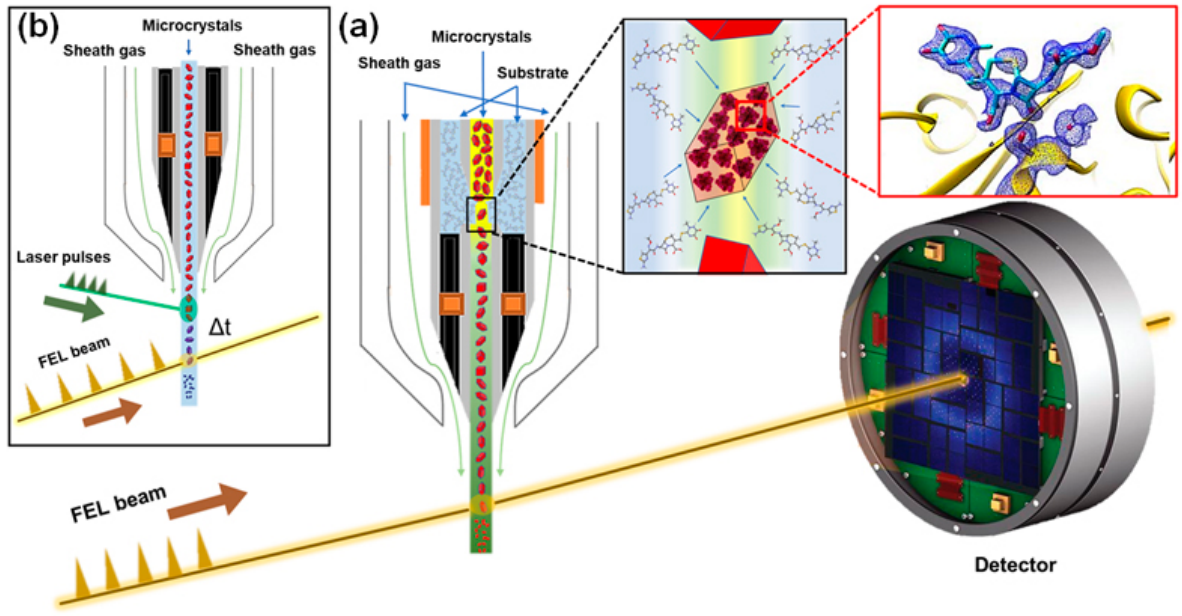



**Figure 1.12.** XFEL experiment. The ultra-bright single pulse X-ray beam shot on single microcrystal gives diffraction pattern before destroying of the crystals. Figure modified from [141]. (a) Set up of time resolved mix and inject experiment, where the substrate and the enzyme are mixed in the injector before data collection. (b) Set up of time resolved experiment for

photo activation, where the laser can active the reaction, and the *Δt* can be used to calculate the structure in time domine. Figure modified from [163].

In order to study the mechanisms and dynamics of non-electron dependant enzymes, the ideal system will have the ability to initiate the reaction at the same precise point in every enzyme in the crystal. By varying the time between initiation and measurement, it is then possible to observe the reaction in the crystal. Of course, these methods rely on the crystal remaining ordered during the reaction and the stochastic nature of reactions does limit synchronicity [145]. For electronic and photo reactions, lasers can be used to initiate the reaction. Successful studies on photocycle of photoactive yellow protein (PYP) has been reported, where the enzyme is naturally activated by light in visible wavelength range [164].

For enzyme reactions that require chemical bond formation / breakage more complex methods to initiate the reaction are required. A rapid-mixing strategy has been applied to the reaction initiation in microfluidics-based and fixed target sample delivery system, where small chemical molecules can diffuse into microcrystals to activate the reaction [165,166]. There are two challenges for this approach. The first is technical, the active site of the enzyme will have to be accessible to the compound being soaked and that binding does not disrupt the lattice, this depends on the crystal packing and can be hard to control. The second issue that if the diffusion time is significantly longer than the chemical reaction being studied then the diffraction pattern will correspond to a mixture of states governed by the diffusion wave.

In principle, the most attractive method for reaction initiation remains photoactivation by a laser, all enzymes in the crystal are illuminated at the same time and thus reactions synchronised in the crystal. The problem is biology has very few reactions that naturally work in this manner. Experimentally the only method is through the use of substrates that are modified with a light-sensitive group that inactivates them (known as photocaged compounds) [167]. Upon exposure to light of the desired wavelength, the light sensitive group breaks of the substrate leaving it active for the desired reaction. In XFELs, the laser flash is synchronised with an accelerator X-ray source that allows calculation of time delay in reaction initiation [168]. The combination of time resolution information and ultra-fast X-ray snapshot diffractions allows the atomic resolution molecular “movie” to be generated, to study the dynamic

enzyme structure in real time at atomic level. There are two major drawbacks to this approach. At a technical level, the synthesis of suitable photocaged compounds is demanding. The cage has to be activated with a power and wavelength that do not damage the crystal, it has to switch on and off substrate reactivity, and it has to be synthetically tractable. The more fundamental challenge is that enzymes evolve to operate on their desired substrate. The addition of the photocage changes the substrate molecule, and in many cases the modified substrate no longer binds.

# Chapter Two

# ForT Initial Data

## 2.1. Introduction

Natural products with biological activities have evolved to enter cells, and such structures are known as privileged scaffolds[169]. An important branch of chemistry is harnessing such scaffolds for further chemical development. Nucleoside-based drugs have played a key role in antiviral and anticancer treatments. To date, the majority of approved nucleotide- based drug have a C-N linkage between sugar and nucleobase, which is easily decomposed by exposure to mild acid or via action of nucleoside phosphorylases catalysed transformations. As a consequence, C-N nucleoside drugs lack biological stability in cellular environment. This issue could be resolved by replacing the C-N linkage by a C-C bond to give a C-nucleoside compound. These C-C compounds cannot be degraded *in vivo* by the same mechanism as their C-N analogues. Formycin and pyrazofurin are two C-nucleoside natural products that have been shown to exhibit both antiviral and anticancer activities. As discussed, these C-nucleosides require a complex chemical synthesis strategy. In contrast, biological synthesis provides a route where engineered enzymes can be used as synthetic tools to give higher yield and purity products.

Our group using the published gene cluster for formycin biosynthesis from *Streptomyces Kaniharaensis*[115] cloned, purified and solved the high-resolution crystal structures of two enzymes in the pathway[119,170]. This thesis extends the work by studying the key C-C bond

formation enzyme ForT, which catalyses the reaction between 5-phospho-α-D-ribosyl diphosphate (PRPP) (**27**) and 4-amino-1*H*-pyrazole-3,5-dicarboxylic acid (APDA) (**26**) to give C-nucleoside product [171] (Figure 2.1).

*β*-RFAPS is a related C-nucleoside forming enzyme from the methanopterin biosynthetic pathway in *M. jannashii*. This enzyme couples 4-hydroxybenzoic acid with PRPP to form C-C nucleoside product with $CO_2$ and PPi release driving the reaction irreversible [172] (Figure 2.1). Sequence analysis ForT (and *β*-RFAPS) indicate they represent a new class of C-C bond forming enzymes.

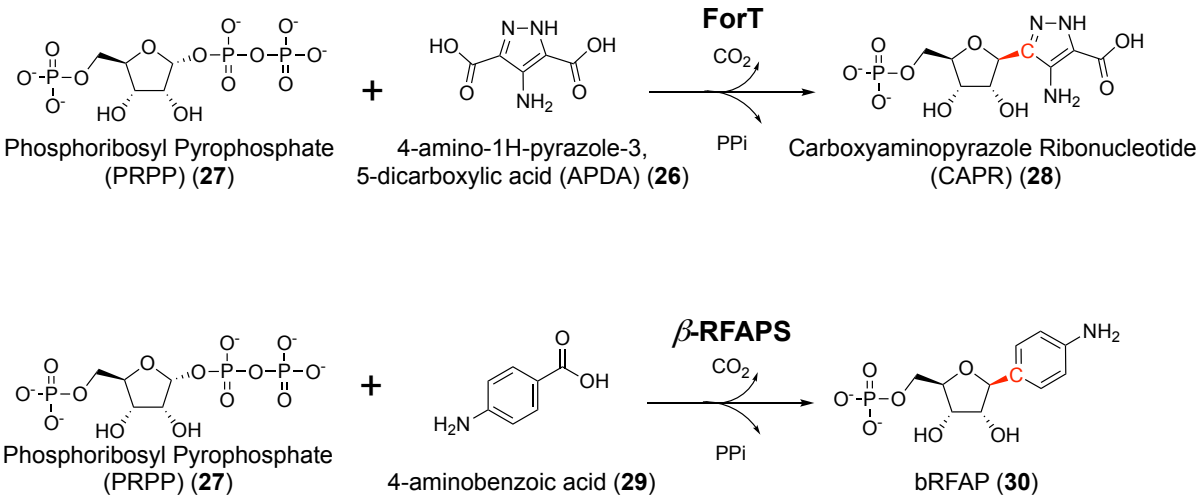



**Figure 2.1.** C-glycosidic bond formation reaction. ForT catalyses the C-glycosidic bond formation between PRPP (**27**) and APDA (**26**), where *β*-RFAPS uses 4-aminobenzoic acid (*p*ABA) (**29**) as second substrate to form bRFAPS (**30**) [121,172].

## 2.2. Material and Methods

### 2.2.1. General

Reagents used were purchased from Sigma-Aldrich or Fisher Scientific unless specified. Oligonucleotides were sourced from Integrated DNA Technologies. dNTP and DNA polymerase were part of the EMD Millipore Novagen KOD Hot Start DNA Polymerase kit. DNA sequencing was carried by Source Bioscience.

### 2.2.2. Cloning and Construct of *forT*

The encoded gene *forT* was isolated and cloned into pEHISTEV vector with N terminal $His_6$-tag followed by a TEV protease cleavage site as previously reported [173].

### 2.2.3. Site Directed Mutagenesis

Mutations around the active site on ForT have been guided by structural information and achieved by using previously developed site-directed mutagenesis protocol [174]. Primer pairs were designed with mutation sites following established rules and synthesized by Integrated DNA Technologies. Both forward and reverse primers used in mutagenesis have shown in table 2.1. In the total volume of 50 µL PCR reaction tube, it contains 50 ng of ForT template, 1 µM of both forward and reverse primers, 200 µM dNTPs and 1 unit of Q5 High-Fidelity DNA Polymerase (New England Biolabs). The PCR program was set for 94 degrees, 7 min for initial denaturation, 12 x 95 ˚ C, 1 min denaturation followed by calculated melting temperature $T_m$ - 5 ˚ C 1 min annealing and 68/72 ˚ C 5.5 min extension, 68/72 ˚ C 10 min final extension.

The PCR product was digested by Dpnl for 1 hour at 37 ˚ C and 5 µL of the digested product was used for transformation into *Escherichia Coli* DH5a cells by heat shock at 42˚ C for 40 seconds followed by cooling down on ice for 5 minutes. 1 mL of overgrowth media was added to the culture and incubated at 37 ˚ C for an hour before transferred to the LB agar plate containing 50 µg/ml kanamycin. The agar plate was incubated at 37 ˚ C for further 20 hours.

A single colony was picked from the agar plate and inoculated into 10 ml LB media containing 50 μg/ml kanamycin and incubated at 37 ˚ C with 200 rpm for 20 hours. Cells were harvested by spin down at 4000 rpm with benchtop centrifuge, plasmid was extracted using QIAprep Spin MiniPrep Kit (QIAGEN) and the concentration was measured by Thermo Scientific NanoDrop. All mutations were confirmed by DNA sequencing using Source Bioscience service.

**Table 2.1.** Primers used for the site-directed mutagenesis of ForT

| Primers | Sequences (5'-3') |
|---|---|
| R19A-f | TGCTCCAGCAGCGCTTAGCTTCACGCTGATTAGCCTTGAC |
| R19A-r | GCTAAGCGCTGCTGGAGCACTGACCCGAACCGCGC |
| R19K-f | TGCTCCAGCAAAACTTAGCTTCACGCTGATTAGCCTTGACGG |
| R19K-r | GCTAAGTTTTGCTGGAGCACTGACCCGAACCGCGCCG |
| S21A-f | CAGCAAGACTTGCGTTCACGCTGATTAGCCTTGACGGCAGTTCC |
| S21A-r | GCGTGAACGCAAGTCTTGCTGGAGCACTGACCCGAACCGC |
| H99A-f | CCTTCCCCAAGCGAGCGGTTTTGGATCTAAAACTAGTACCTTGTTG |
| H99A-r | CCGCTCGCTTGGGGAAGGGCCTCAAGTACGTCGACACG |
| G101P-f | CAACACAGCCCGTTTGGATCTAAAACTAGTACCTTGTTGGCTG |
| G101P-r | GATCCAAACGGGCTGTGTTGGGGAAGGGCCTCAAGTACG |
| S104A-f | GGTTTTGGAGCGAAAACTAGTACCTTGTTGGCTGTAGGACATGCC |
| S104A-r | CTAGTTTTCGCTCCAAAACCGCTGTGTTGGGGAAGGGCCTCAA |
| T106V-f | GGATCTAAAGTGAGTACCTTGTTGGCTGTAGGACATGCCTAC |
| T106V-r | CAAGGTACTCACTTTAGATCCAAAACCGCTGTGTTGGGGAAG |
| G134P-f | CGTACCTTACCGCGGGGTCGTACCTCAGGTGCTAGTACAG |
| G134P-r | CCCCGCGGTAAGGTACGAGCAAGCTCGCGCAGGTCAGG |
| R135A-f | CCTTAGGTGCGGGTCGTACCTCAGGTGCTAGTACAGGC |
| R135A-r | TACGACCCGCACCTAAGGTACGAGCAAGCTCGCGCAG |
| T138V-f | GGTCGGGGTCGTGTGTCAGGTGCTAGTACAGGCCTTTCCGCCTATG |
| T138V-r | AGCACCTGACACACGACCCCGACCTAAGGTACGAGCAAGCTC |
| S139A-f | GTCGTACCGCGGGTGCTAGTACAGGCCTTTCCGCCTATGG |
| S139A-r | CTAGCACCCGCGGTACGACCCCGACCTAAGGTACGAGCAAG |
| S142A-f | AGGTGCTGCGACAGGCCTTTCCGCCTATGGCGGATTTCTG |
| S142A-r | AAAGGCCTGTCGCAGCACCTGAGGTACGACCCCGACCTAA |
| S174A-f | CGCCCGGCGCGTTTCGCACAGCAAGTAGCTCCGCCTAAG |
| S174A-r | GCGAAACGCGCCGGGCGTAAATACTTTTGTGGAGCCTCTGCG |
| S174L-f | CGCCCGCTGCGTTTCGCACAGCAAGTAGCTCCGCCTAAG |
| S174L-r | GCGAAACGCAGCGGGCGTAAATACTTTTGTGGAGCCTCTGCG |

### 2.2.4. Over-Expression of ForT and Variants

ForT and its variants shared the same transformation and expression protocol. 2 µL of plasmid DNA at a concentration of 100 ng/µl were transformed into *E.Coli* BL21 (pLysS) commercial competent cells which are subsequently plated on to LB agar plate containing 50 µg/mL kanamycin and 25 ng/mL chloramphenicol. The plate was incubated at 37 ˚ C overnight. A small-scale culture was started by picking single colony and inoculating it into 150 mL LB media with the same concentration of kanamycin and chloramphenicol. The pre-culture were left overnight in 37 ˚ C and 200 rpm shaking speed incubator, where 10 mL of the pre-culture is inoculated into 1 L of large culture with LB and the same concentration of correlated antibiotics on the second day. Cultures were incubated at 37 ˚ C and 200 rpm until optical density at 600 nm wavelength ($OD_{600}$) reaches 0.6, and then IPTG was added at final concentration of 0.4 mM to induce the protein expression. Incubation temperature was then adjusted to 20 ˚ C for further 20 hours with shaking speed at 200 rpm. The cells were harvested by centrifugation at 8000 rpm for 20 minutes, and cell pellets were collected and kept at -80 ˚ C for further purification.

### 2.2.5. Purification and Optimization

Cell pellets were thawed and resuspended in lysis buffer [500 mM NaCl, 50 mM sodium phosphate (pH 7.5), 20 mM imidazole, and 3 mM beta-mercaptoethanol (BME), with addition of EDTA-free protease inhibitor tablets from Roche (1 tablets per 50 mL lysis buffer), chicken egg lysozyme at final concentration 1 mg/mL, and bovine DNase at 0.4 mg per gram of wet cell pellet. The re-suspended cells were lysed by using cell disruptor at 30K psi at 4 ˚ C twice. The lysate was further centrifuged at 17,000 rpm, 4 ˚ C for 20 min to remove cell debris. The supernatant was filtered before loading onto Ni Sepharose 6 FF column (GE Healthcare) equilibrated with lysis buffer. The column was washed three column volumes of washing buffer [500 mM NaCl, 50 mM sodium phosphate (pH 7.5), 30 mM imidazole, and 3 mM BME]. The target protein with 6 X His-tag was eluted with elution buffer [500 mM NaCl, 50 mM sodium phosphate (pH 7.5), 250 mM imidazole, and 3 mM BME]. The elution was collected and desalted through a desalting column (16/10 Desalting, GE Healthcare) with desalting

buffer [100 mM NaCl, 50 mM sodium phosphate (pH 7.5), 3 mM BME]. Fractions containing target protein were collected and TEV protease was added at mass ratio of 10:1. EDTA was added to a concentration of 2 mM to eliminate any remaining Ni ion in the solution. The protein was digested for 3 hours at 20 ˚ C before being loaded into second nickel column which has been equilibrated with desalting buffer. The flow though was collected and concentrated to 5 mL using Vivaspin® 30 kDa concentrator. The concentrated solution was then applied to Superdex 200 (GE Healthcare) gel filtration column that has been preequilibrated with gel filtration buffer [150 mM NaCl, 10 mM HEPES (pH 7.4), and 1 mM TCEP]. The gel filtration peak was analyzed by SDS PAGE to determine its purity, and the pure proteins were collected, concentrated, aliquoted and flash frozen with liquid nitrogen for storage in -80 ˚ C freezer.

### 2.2.6. Crystallization and Co-Crystallization

ForT crystallisation was first screened using purified ForT (43mg/ml) premixed with 10 mM PRPP. The best screening hit was obtained in JCSG screen containing 40 mM KH2PO4, 20 % (v/v) glycerol, 16 % (w/v) PEG 8000 with 43 mg/ml protein at 20 °C. The hit conditions were further optimised by adding 20 mM Tris-HCl (pH 7.5-9), 20 mM HEPES (pH 7-8), 20 mM Bis-Tris (pH 6-7) or 20 mM Bis-Tris propane (pH 6-8). Bis-tris propane as buffer gave the best crystals, and we did not see any variation in quality between pH 6 and 8.

ForT crystals were further optimised and finally grown in the hanging drop plates (EasyXtal 15-well DG-Tool X-Seal) using 1 µL of the protein solution (43 mg/ml) mixed with 1 µL of reservoir solution in the hanging drop. The crystals were mounted in a cryo-loop (Molecular Dimensions) and cryo-protected in solutions containing mother liquor plus 200 mM PRPP, immediately frozen in liquid nitrogen and stored in a cryogenic Dewar. X-ray data were collected at Diamond Light Source I04-1.

The original hit was further optimised and co-crystallized with 50 mM PRPP in the hanging drop plates (EasyXtal 15-well DG-Tool X-Seal) using 1 µL of the protein solution (25 mg/ml) and 1-2 µL of reservoir solution in the hanging drop. The protein was incubated with 50 mM

PRPP on ice before setting up crystallization plates. The crystals were mounted in a cryo-loop (Molecular Dimensions) and cryo-protected in solutions containing mother liquor plus 200 mM PRPP. The crystals were then frozen by plunging them into liquid nitrogen and sent in a cryogenic Dewar to Diamond Light Source I04-1 for data collection.

### 2.2.7. LC-MS Assay of WT ForT Activity

Based on previous study carried out by Liu's group [121], 4-amino-*1H*-3,5-pyrazole dicarboxylic acid (APDA) (**26**) (2.5 mM) and PRPP (**27**) (5 mM) were incubated with WT ForT (25 µM) in sodium phosphate buffer [50 mM sodium phosphate (pH 7.5), 100 mM NaCl, and 50 mM $MgCl_2$] for 1 hour at room temperature with gentle rolling. The reaction was quenched by boiling the reaction tube at 95 ˚ C in a heat block for 1 minute. The precipitated protein was removed by centrifugation at 10,000 g for 10 minutes before the supernatant was analysed by LC-MS. This sample was loaded onto a Kinetex 2.6 µm XB-C18 100A, 100 x 4.6 mm LC column using a 15 minutes gradient of 5-100% acetonitrile and 5 mM ammonium carbonate at a flow rate of 1 mL/min. The mass spectrum was set for positive detection where the product carboxyaminopyrazole ribonucleotide (CAPR) (**28**) has a m/z = 340 corresponds to the $[M+H]^+$.

### 2.2.8. Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) tests equilibrium binding between ForT and PRPP using a MicroCal PEAQ-ITC (Malvern) instrument. Both ForT and its variants were buffer exchanged into gel filtration buffer [150 mM NaCl, 10 mM HEPES (pH 7.5), and 1 mM TCEP]. PRPP was dissolved into gel filtration buffer to give final concentration of 2.5 mM as a titration solution. The cell chamber of ITC experiment contains 250 µM ForT or its variants. The experiments were carried out at a stirring speed of 750 rpm at 25 ˚ C with using a reference power of 5 µcal/s. Control experiment was achieved by a buffer-to-buffer titration to establish a baseline. Data was processed by using MicroCal PEAQ ITC analysis software (Malvern) to obtain a nonlinear one-site binding model to calculate enzyme equilibrium binding information.

### 2.2.9. MIMS Assay of Catalytic Activity

Membrane inlet mass spectrometry (MIMS) (HDR-20, Hiden) was used to monitor real-time $CO_2$ formation in ForT-PRPP-APDA enzymatic turnover reaction [175]. The reaction chamber has a volume of 2 mL and contains 50 mM TES buffer at pH 6.8, 1 mM APDA (**26**), 2 mM $MgCl_2$ and 6.6 μM ForT variants. The reaction was initiated by adding freshly prepared solution of PRPP (**27**) to give final concentration of 2 mM, and $CO_2$ emission was detected in real time by monitoring the ion current at 44 m/z. The concentration of $CO_2$ was calculated from ion current signal by using established standard curve [176].

### 2.2.10. Data Collection Processing and Refinement

The data sets of ForT native structure were collected previously by Dr. Gao at Diamond Light Source I04-1 and processed through Diamond online processing software DIALS and XIA2 [177,178]. The phase was determined, and structure was solved by using molecular replacement with its analogue protein β-RFAPS as starting model through PHASER in CCP4i2 programme suit [179,180].

The data set of ForT-PRPP complex structure were collected at Diamond Light Source I04 and processed through Diamond online processing software DIALS and XIA2 [177,178]. The phase was determined, and structure was solved by suing molecular replacement with ForT native structure as starting model through PHASER in CCP4i2 programme suit [178,180]. The PRPP ligand was placed guided by the Fo-Fc and 2Fo-Fc electron density maps in COOT, and the model was refined by Refmac5 in CCP4i2 programme suit [181].

## 2.3. Results

### 2.3.1. Protein Expression and Purification

ForT expression and purification conditions have been established by previous work led by Dr. Sisi Gao. However, the yield of ForT native protein remained low. To improve the yield during purification process, 10 mM of imidazole has been introduced into the reverse nickel column purification. This can help to reduce the non-specific binding between cleaved ForT and nickel resin. The yield of ForT product increased significantly, with the UV absorbance for final step gel filtration peak to more than 800 mAU compared to 200 mAU previously with same amount of elution with same conditions. The gel filtration profile for ForT showed a symmetry peak, with elution volume around 80 to 90 ml from Superdex S200 16/600 column, consistent with previous experiments (Figure 2.2 A). SDS-PAGE gel analysis confirmed each step of the purification process from cell lysate to nickel affinity purification and size exclusion chromatography (Figure 2.2 C). Figure 2.2 D shows the SDS-PAGE gel from the elution peak of the gel filtration process and further confirmed the purity of the protein that collected for crystallization.

A simpler purification process was introduced for functional studies, which has skipped the TEV protease cleavage and reverse nickel column. This can significantly reduce the time of purification process and allow more experiments for the mutants' functional studies. Figure B shows the gel filtration profile of ForT native protein right after nickel affinity column. ForT with $His_6$ tag seems formed some aggregates and eluted around 40 – 50 ml, the corrected folded protein still eluted between 80 to 90 ml with symmetrical peak. The purity of the protein was confirmed by SDS-PAGE shown in Figure 2.2 E.

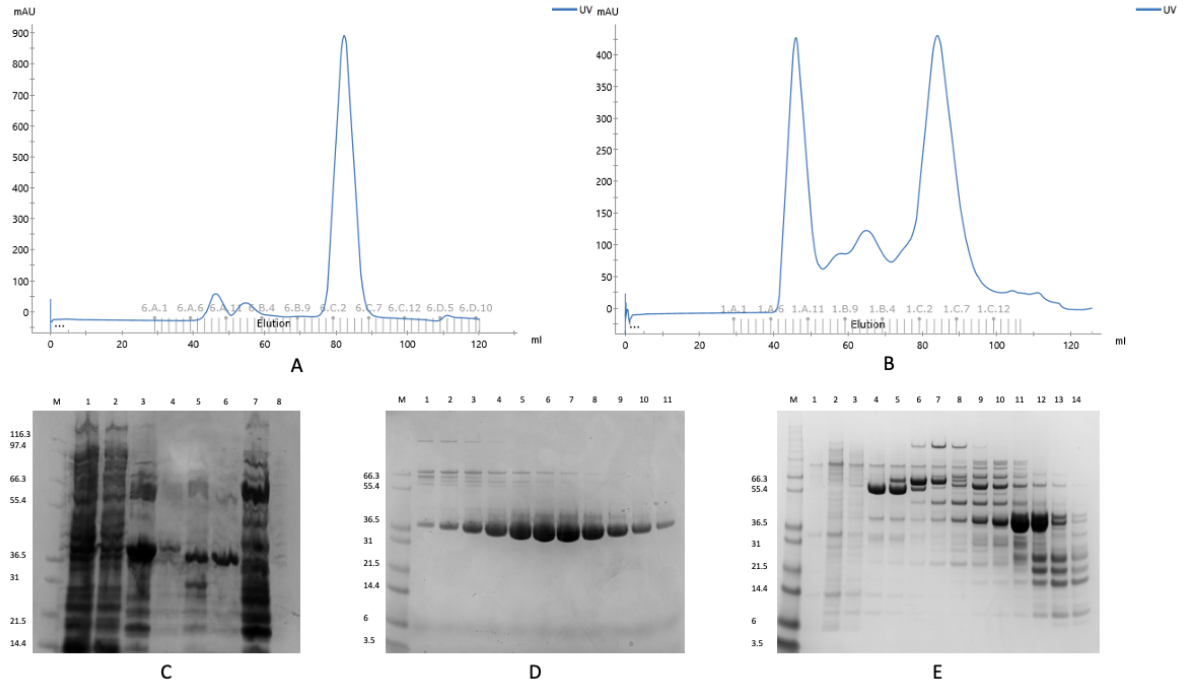



**Figure 2.2.** Figure A shows the gel filtration curve for ForT native protein purification after TEV cleavage and reverse nickel column. The correct eluted at 80 – 90 ml elution time. B shows the gel filtration curve of ForT native protein after first nickel purification. Impurities and aggregations are eluted first, followed by ForT with His-tag eluted at 80 – 90 ml elution time. C shows the SDS-PAGE analysis during ForT purification process. M is marker (Marker 12 with molecular weight lane labelled on the left), 1: Ni load, 2: Ni flow through, 3: 1$^{st}$ Ni elution, 4: Desalt elution (Pre-TEV), 5: Post TEV, 6: 2$^{nd}$ Ni flow through, 7: 2$^{nd}$ Ni elution. D and E shows the SDS-PAGE analysis of gel filtration from ForT purification corresponded with A and B, confirmed the correct size of the protein from elution.

### 2.3.2. Crystallization Optimization

Crystals of purified ForT at 43 mg/ml together with 10 mM PRPP appeared in JCSG screen containing 40 mM $KH_2PO_4$, 20% (v/v) glycerol, 16% (v/v) PEG 8000 at room temperature. The condition was further optimised with 20 mM Bis-Tris propane buffer at pH 7 and grow within 5 days of setting up crystallization plates (Figure 2.3 A-D). The crystals that resulted were relatively small and further optimisation was carried out by reducing the protein concentration to 25 mg/ml and increasing PRPP concentration to 50 mM. With reduced protein concentration, the crystals took one to two days longer to grow, but were larger (Figure 2.3 E-H). This condition was transferred to hanging drop technique and diffraction data were collected at Diamond Light Source I04-1. The resolution has been processed to 2.5Å with processing software XIA2 and DIALS [177,178].

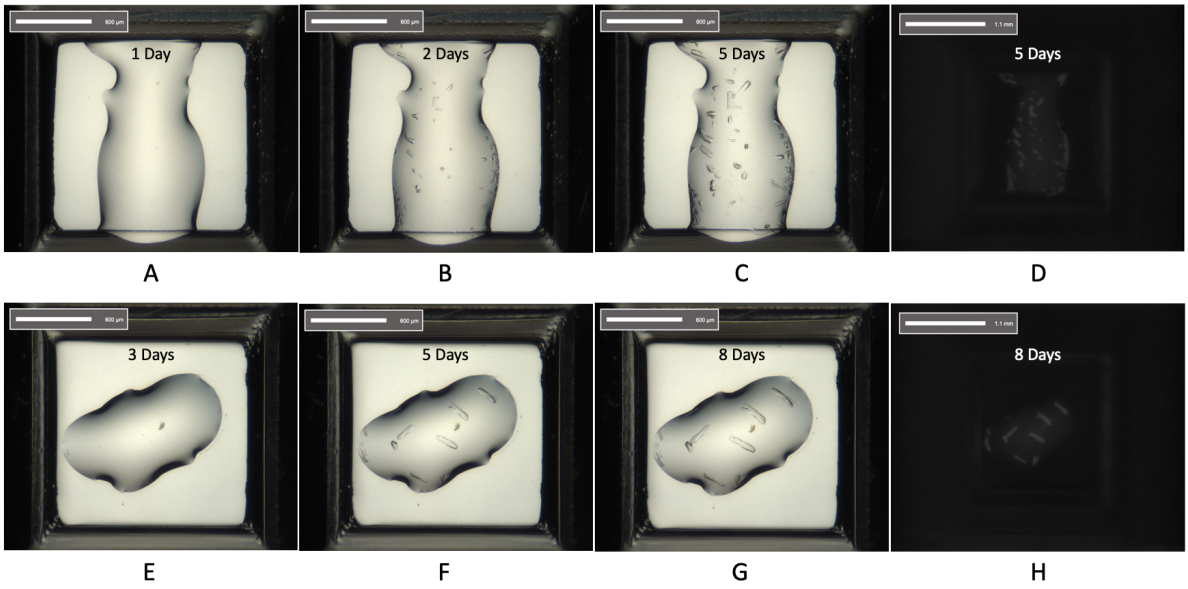



**Figure 2.3.** Shows the crystallization process of ForT-PRPP with optimisation. A-D shows the ForT-PRPP complex with 43 mg/ml protein concentration. E-H shows the ForT-PRPP complex with 25 mg/ml protein concentration. Both proteins are incubated with 50 mM PRPP on ice for an hour before crystallization plates were setup. D and H the crystals imaged with UV light, confirmed the crystals contained protein.

### 2.3.3. Overall Structural Analysis

Basic bioinformatics studies show the target protein ForT consists of 341 amino acids with a molecular weight of 36.6 KDa. Wildtype ForT recombinant gene has been transformed into *Escherichia coli* for overexpression and purification. Final purified protein was concentrated to 25 mg/mL to co-crystalize with 50 mM PRPP. The ForT/PRPP co-crystals have been soaked with 200 mM PRPP before shot with X-ray beam at Diamond Light Source for data collection. The data were processed with Diamond online processing software XIA2 and DIALS and the complex structure was solved by molecular replacement in PHASER with previous solved structure of β-RFAPS [177,178,180]. Table 2.2 displays the data collection and refinement statistics ForT-PRPP complex.

In the crystal structure, ForT has one subunit in each ASU, and crystal symmetry together with PISA calculation suggests it is a dimer [182–184]. Chromatography also suggested ForT forms dimer in the solution (Figure 2.2A). The modelled structure according to electron density map contains residues from 11-171, 180-205 and 209-341 (C-N) the missing residues are assumed to be disordered. On secondary structure level, ForT consists of 3 anti-parallel beta sheets and 9 alpha-helixes, where two of the beta-sheets are stacked on each other and perpendicular to the third one. Two alpha-helixes are geometrically parallel to the third beta-sheet and together with two more smaller alpha-helixes to form the core structure of the enzyme. The 2.5 Å electron density map suggests PRPP fits in the cavity pocket formed by these four alpha-helixes in the core structure. The other two alpha-helical bundles are suggested to contribute in dimer formation, with PISA [182–184] calculation it gives a buried surface area of 936.7 $Å^2$. The second monomer is positioned with a rotation of approximately 60 degrees from the other monomer (Figure 2.4 a,b).

Similarity studies between ForT and different homoserine kinases has been carried out by pairwise superposition despite they only shared limited sequence identity with ForT. The resulting Cα r.m.sd for over 300 structurally equivalent residues are 3.15 Å, 3.00 Å, and 3.05 Å for ForT against HSK (homoserine kinase from *Methanocalococcus jannaschii*, PDB: 1fwk), MK (mevalonate kinase from *Methanosarcina mazei*, PDB: 6mde), and MVK (mevalonate kinase from *Methanosarcina mazei*, PDB: 4hac) respectively. In contrast, the sequence

identity is only 11.4%, 17.2% and 17.3% respectively. The substrate binding site for ForT and HSK shares some level of similarity, they both contain phosphate binding loop (P-loop) for either ADP or PRPP entry and binding. The P-loop contains amino acid residues that makes intensive hydrogen bonding interactions with phosphate group from ADP or PRPP. However, the ribose moiety is positioned differently between PRPP and ADA in ForT and HSK, where superposition shows that ADP clashes with an α-helix of ForT (Figure 2.11). The corresponding α-helix on HSK is further away, leaves larger cavity for its ADP binding orientation.

PRPP exhibits as unusual stringed conformation in the enzyme active site with hydrogen bonding interactions through amino acids side chains. The pyrophosphate moiety of the PRPP is bound by the loop of protein from residues His99 to Ser104 through hydrogen-bonding with their backbone amide and the side chain of His99 with two additional water molecules. Arg19 forms salt bridges with both pyrophosphate moiety and 5'-phosphate in different binding pockets, where 5'-phosphate forms an additional salt bridge with Ser139 (Figure 2.4 d). Hydrogen bonds are made to the 5'-phosphate by the side chain of Thr106, Ser139, and Ser142, together with backbone amide of Ser142. No binding affinity was detected between the enzyme and ribose-5'-phosphate (R5P) itself, suggesting the pyrophosphate moiety is essential for PRPP recognition at the binding pocket. Initial modelling by our Cardiff collaborators in Prof. Nigel Richard's group suggested S21, T138, R19, T295, K263, Q266 and S142, may play a role in binding the second substrate APDA 16 (Figure 2.4 e).

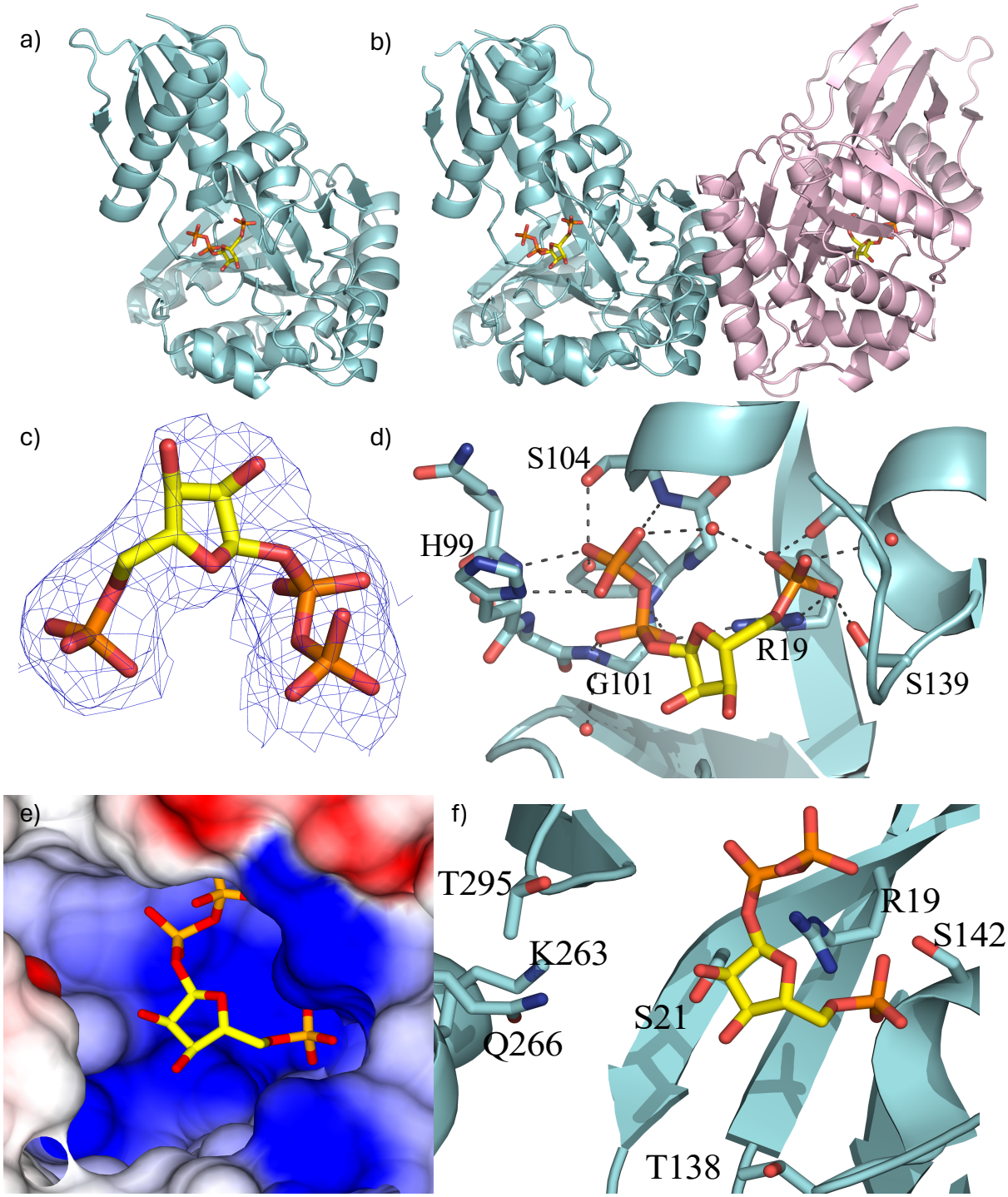



**Figure 2.4.** 2.5Å resolution X-ray crystal structure of wildtype ForT/PRPP complex. (a) The monomeric structure obtained from X-ray crystallography. ForT shown as cyan cartoon, PRPP shown in sticks (carbon yellow, phosphate orange, oxygen red). (b) Dimer structure of ForT from crystal symmetry, the second ForT unit shown in pink. (c) The electron density map of PRPP solved from ForT/PRPP complex structure. (d) PRPP binds with the enzyme by using several hydrogen bonding interactions. The pyrophosphate moiety of PRPP is bound by

interacting with residues H99, S104, and G101. The ribose moiety of PRPP interacts with residue R19 and S139. The loop from Gln98 to Ser108 is highly conserved from GHMP kinase superfamily, which plays a key role in PRPP stabilization and binding. Protein carbon atoms shown in cyan, carbon from PRPP shown in yellow, nitrogen shown in blue. (e) The surface structure of PRPP binding pocket in ForT, which consists of several polar residues in a positively charged cavity. (d) PRPP binding pocket in atomic structure. The positively charged surface is contributed by labelled residues. (Figure modified from [170])

**Table 2.2.** Data collection and refinement statistics ForT-PRPP complex

| | ForT-PRPP |
|---|---|
| Beamline | i04-1 (DLS) |
| Wavelength (Å) | 0.9159 |
| Space group | $P4_12_12$ |
| a, b, c (Å) | 81.1, 81.1, 111.0 |
| α, β, γ (°) | 90, 90, 90 |
| Resolution range (Å) | 65.5 – 2.5 (2.57 – 2.50) |
| $R_{meas}$ | 0.066 (2.15) |
| I/ σ (I) | 13.2 (1.1) |
| Completeness (%) | 100 (99) |
| Multiplicity | 12.5 (13.1) |
| No. reflections | 287431 (14481) |
| $R_{work}$ | 0.227 (0.225) |
| $R_{free}$ | 0.27 (0.42) |
| Number of atoms | |
| Total | 2435 |
| Protein | 2401 |
| Ligands / ions | 28 |
| Waters | 6 |
| B-factor | |
| Average | 62 |
| Protein only | 62 |
| Ligands/Ion | 97 |
| Water | 72.1 |
| | |
| Bond lengths (Å) | 0.005 |
| Bond angles (°) | 1.13 |
| RSCC of PRPP (occupancy 1) | 0.893 |
| Ramachandran plot favoured / outlier (%) | 94.3 / 2.6 |

### 2.3.4. Chemical Synthesis of 4-amino-1*H*-3,5-pyrazole dicarboxylic acid (APDA)

4-amino-1*H*-pyrazole-3,5-dicarboxylic acid (APDA) (**26**) has been reported to be the second substrate for ForT after PRPP. ForT catalyses the reaction between the carboxylic group on APDA (**26**) and ribose on PRPP to form the important C-C bond nucleoside intermediate in formycin biosynthesis pathway. Since the compound APDA (**26**) is not commercially available, a three-step organic synthesis pathway was employed using 3,5-dimethyl pyrazole (**31**) as the starting material [185] (Figure 2.5 a).

Because of the difficulty and low yield of oxidation reaction to 3,5-dimethyl pyrazole (**31**), we started synthesis of APDA (**26**) from commercially available compound 3,5-pyrazole dicarboxylic acid (**32**). Concentrated sulfuric acid (45 mL) was added dropwise into concentrated nitric acid (5 mL) on ice over 1 hour. 5 grams of 3,5-pyrazole dicarboxylic acid (**32**) were added slowly to the acid solution over 1 hour with stirring and the internal temperature kept below 50 ˚ C. Once the addition was complete, the reaction was heated to 65 ˚ C and held for 20 hours. The mixture was transferred into ice water (200mL) and partly neutralized with potassium hydroxide until pH reaches 2. The precipitate was discarded, and the clear solution was left at room temperature for 2 days for re-crystallization. The crystallized material was re-suspended in 50 mL water and concentrated sodium hydroxide added dropwise until no more solid dissolved. The mixture was then filtered and re-acidified with hydrochloric acid until the pH reached 2. The resulting white solid precipitation was filtered and dried with a freeze dryer overnight to give 4 grams of product 4-nitro-3,5-pyrazole dicarboxylic acid (NPDA) (**33**). The authenticity of the product was confirmed with $^{13}C$ NMR ($D_2O$) which found peaks 163.76 and 138.88 ppm (Figure 2.5).

Sodium dithionite has been reported to reduce nitro group into amino group [185]. To dissolve NPDA (**33**) (1 g, 5.3 mmol, 1 eqv.) into 50 mL of water with a stirring bar. Fresh sodium dithionite (4.6 g, 26.5 mmol, 5 eqv.) was added into mixture in 5 portion with 2 minutes interval. Once addition was complete, the reaction was heated up to 95 ˚ C for 15 minutes with stirring. After cooling to room temperature, the mixture was re-acidified with hydrochloric acid to reach pH 2. The solution was left at room temperature overnight to

recrystallize the product and the solid freeze dried. The product APDA (**26**) is found as a white solid 250 mg and confirmed by $^{13}C$ NMR ($D_2O$) found: 169.74 and 131.56 ppm (Figure 2.5).

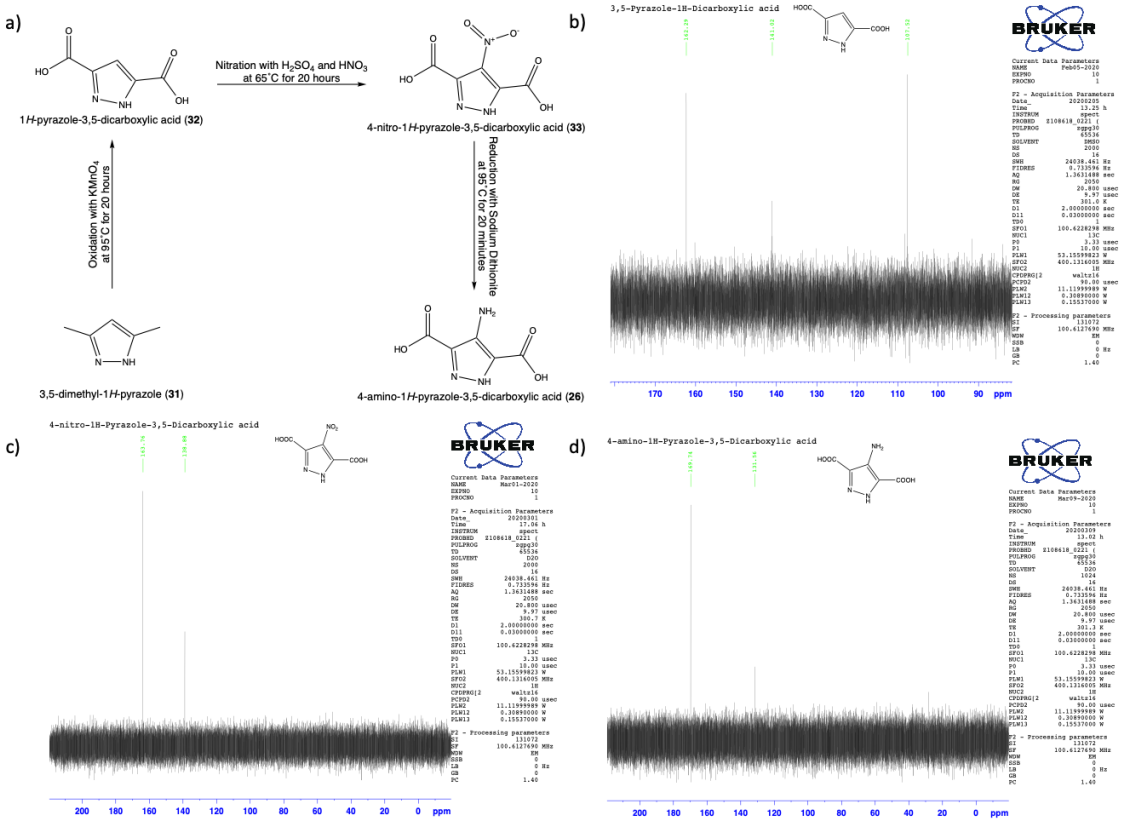



**Figure 2.5.** Organic synthesis pathway of APDA (**26**) from commercially available 3,5-dimethyl-1*H*-pyrazole (**31**). Products from each step have been confirm with $^{13}C$ NMR spectrum.

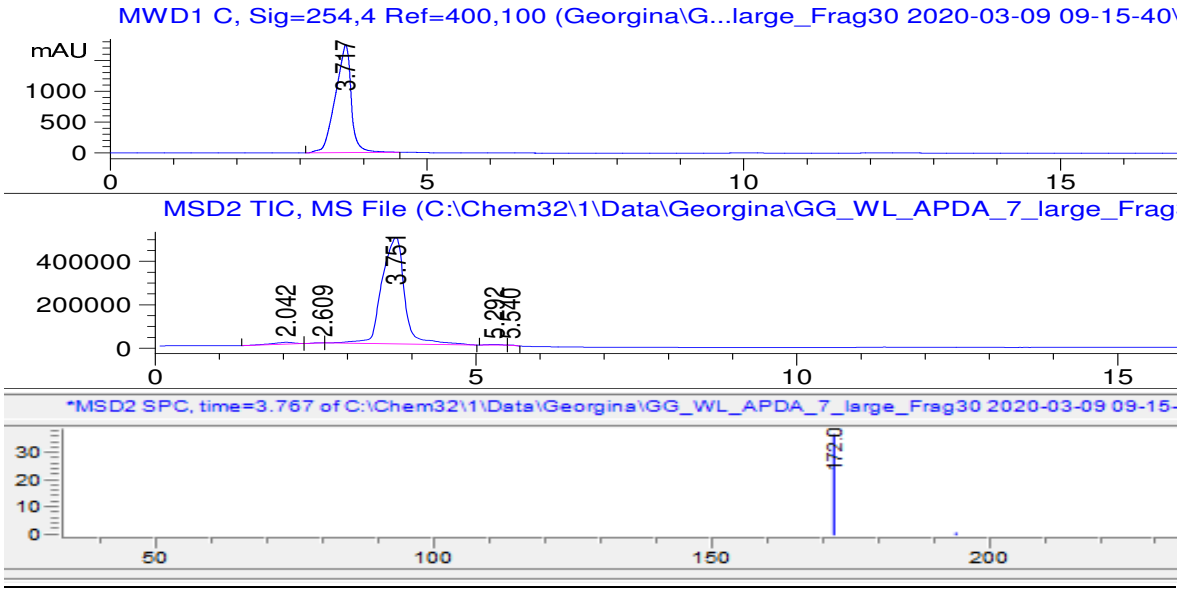



**Figure 2.6.** LCMS confirms the correct mass of APDA (**26**) gives m/z = 172. MWD1 shows the retention time of APDA (**26**) at 3.71 mins with UV detection at 254 nm. MSD2 is the mass scan for APDA (**26**) at 172 corresponds to the $[M+H]^+$.

### 2.3.5. LC-MS Activity Assay and MIMS Assay

The LC-MS assay confirmed the enzyme can catalyse the reaction using PRPP (**27**) and APDA (**26**) to form the expected carboxyaminopyrazole ribonucleotide (CARP) (**28**) product following the method discussed in section 2.2.7. LC using the same $C_{18}$ column shows that the product CAPR has elution peak at 13.75 minutes, while mass spectrometry detection has confirmed the m/z = 340. MSD3 and MSD4 has been set to search signal for APDA and PRPP at m/z = 172.2 and m/z = 390.1 respectively. (Figure 2.8)

Membrane inlet mass spectrometry (MIMS) (HDR-20, Hiden) [176] assay was carried out in University of Cardiff assisted by collaborator Prof. Nigel Richard's group. It monitors the rate of $CO_2$ formation in the ForT-catalysed reaction. A solution of 50 mM TES buffer, pH 6.8, containing 4-amino-*1H*-pyrazole-3,5-dicarboxylic acid (**26**) (1 mM), $MgCl_2$ (2 mM) and WT ForT (6.6 µM) was placed in the reaction chamber (2 mL total volume) and equilibrated for 2 min at 25 °C. Reactions were initiated by the addition of a freshly prepared solution of PRPP (**27**) (to a final concentration of 2 mM), and $CO_2$ production was determined in real time by

monitoring the ion current at 44 m/z. This procedure was also used to evaluate the $Mg^{2+}$ dependence of catalytic activity where EDTA (10 mM final concentration) was added after 15 min, resulting in the loss of activity. In the case of the ForT variants, which were found to precipitate easily, this procedure was altered so that reaction was initiated by the addition of enzyme (6.6 µM S104A or 46 µM S174A) to the substrates in 50 mM TES buffer, pH 6.8. The ion current signal was converted to the concentration of $CO_2$ using a standard curve, which was determined by adding known amounts of $K_2CO_3$ to 10 mM aq. AcOH, pH 2.5. The $CO_2$ concentration was calculated using the apparent dissociation constant for $H_2CO_3$ (pKa = 6.35) under the assumption that all of the $H_2CO_3$ dissociates into $CO_2$ [186].

The site-specific variants (Table 2.3) as well as the wild type of ForT were measured with MIMS assay. However, most of the variants were unstable under the assay conditions which didn't produce readable data. The wild-type ForT has been measured in triplicate with activity at 0.002 U/mg, and variants S174A and S104A has activity at 0.0003 U/mg and 0.0007 U/mg respectfully. Detailed data shown in Table 2.3 together with binding affinity of PRPP (**27**) which were determined by ITC. Figure 2.7 shows the triplicate runs data used to obtain the ForT WT activity (Modified from [170]).

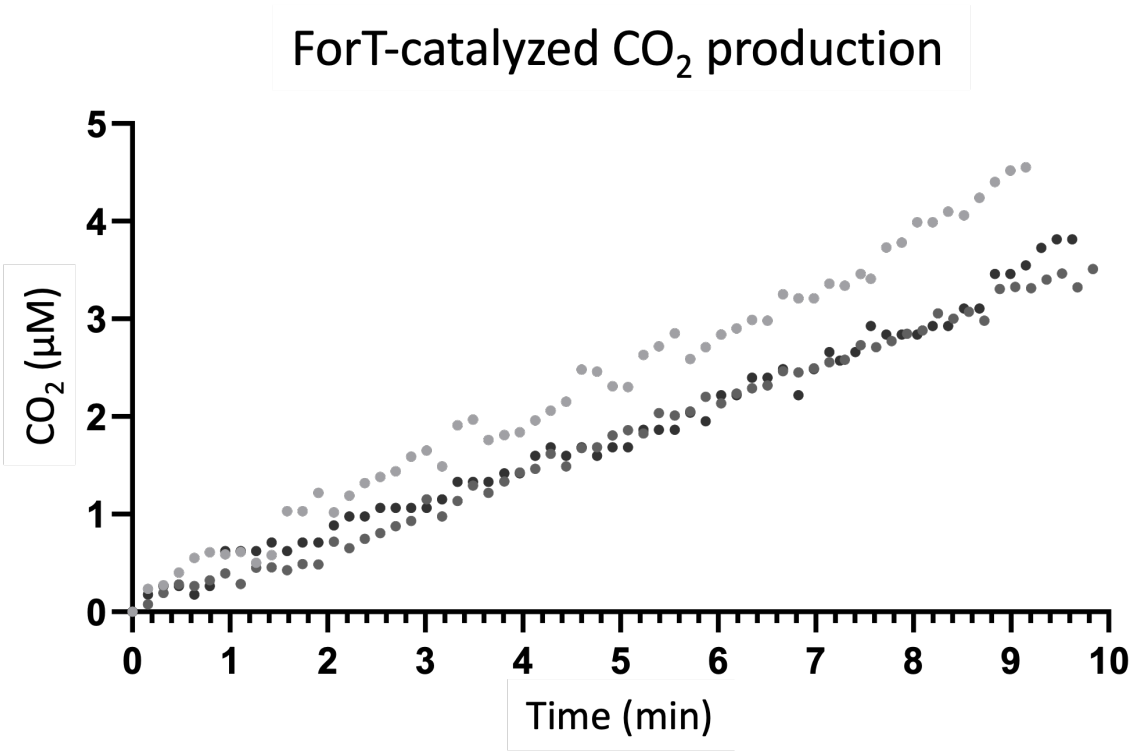



**Figure 2.7.** MIMS data with triplicate runs used to obtain the specific activity of ForT WT.

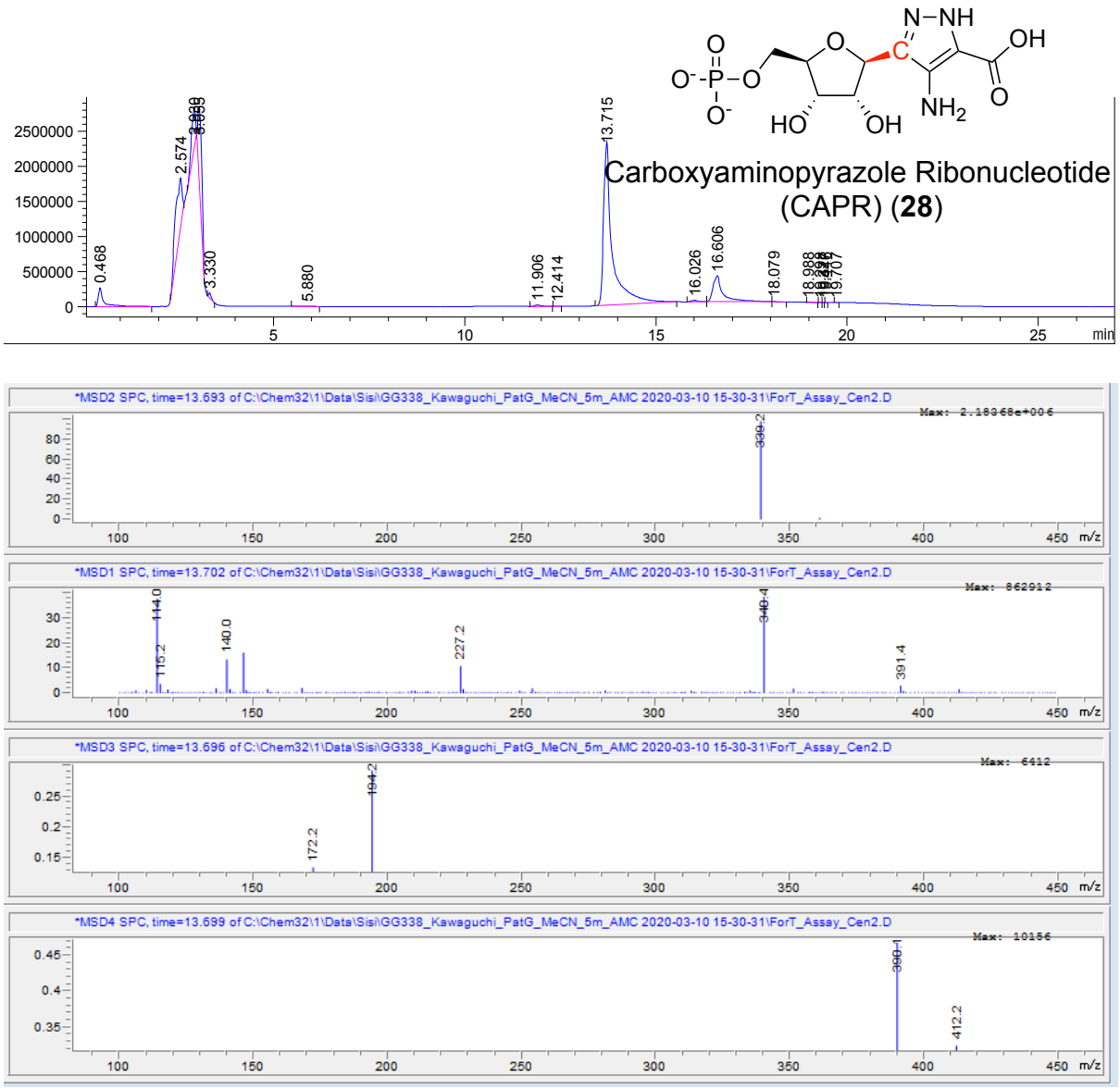



**Figure 2.8.** (a) Raw data with the mass spectrometry detection set for m/z = 340 $[M+H]^+$. (MSD2) to detect product carboxyaminopyrazole ribonucleotide (CAPR) 17 with elution peak at 13.75 minutes. (b) Mass detection with retention time around 13.7 min. MSD1: Total mass scan; MSD2: Mass detection set for product CAPR 17; MSD3: Mass detection set for substrate APDA 16 (m/z = 172.2 $[M+H]^+$); MSD4: Mass detection set for PRPP (m/z = 390.1 $[M+H]^+$).

### 2.3.6. ITC Binding Affinity with PRPP

Based on the ForT-PRPP complex structure, a series of residues around the PRPP binding pocket and catalytic site were mutated (section 2.2.3) to test if they are essential for binding to the enzyme. Mutation of H99, R19, R135, S142 (all of which participate in Hydrogen bonding interactions with PRPP) to alanine resulted in no heat change for PRPP being detected by ITC experiments (Figure 2.10). Purification and gel filtration profile for those mutants confirmed that mutant proteins were correctly folded. Figure 2.9 shows the ligand plot illustrating PRPP interactions with amino acid side chains for hydrogen bonding. This suggests these residues are key for the binding interaction with PRPP. R19K has a $K_D$ for PRPP at 10.8 ± 0.9 μM, which supports the structural data that it is the positive charge is essential for the PRPP recognition and binding. T106 and T138 have been mutated to valine to eliminate the hydroxyl group while leaving the overall side chain structures. T138V shows the $K_D$ for PRPP at 1.2 ± 0.2 μM, which is at the same level of wild type protein at 3.6 ± 0.3 μM. In contrast, T106V shows no binding affinity with PRPP, S139A and G134P have a $K_D$ with PRPP over 100 μM, indicating they are also important for PRPP binding. S104A (Ser104 which interacts with the phosphate group on PRPP) has a reduced $K_D$ for PRPP of 11.4 ± 0.8 μM. Table 2.3 shows the binding affinity data and the $K_D$ values measured by ITC and MIMS assay. Figure 2.10 shows the raw plots of ITC measurement for binding affinity between ForT variants and PRPP.

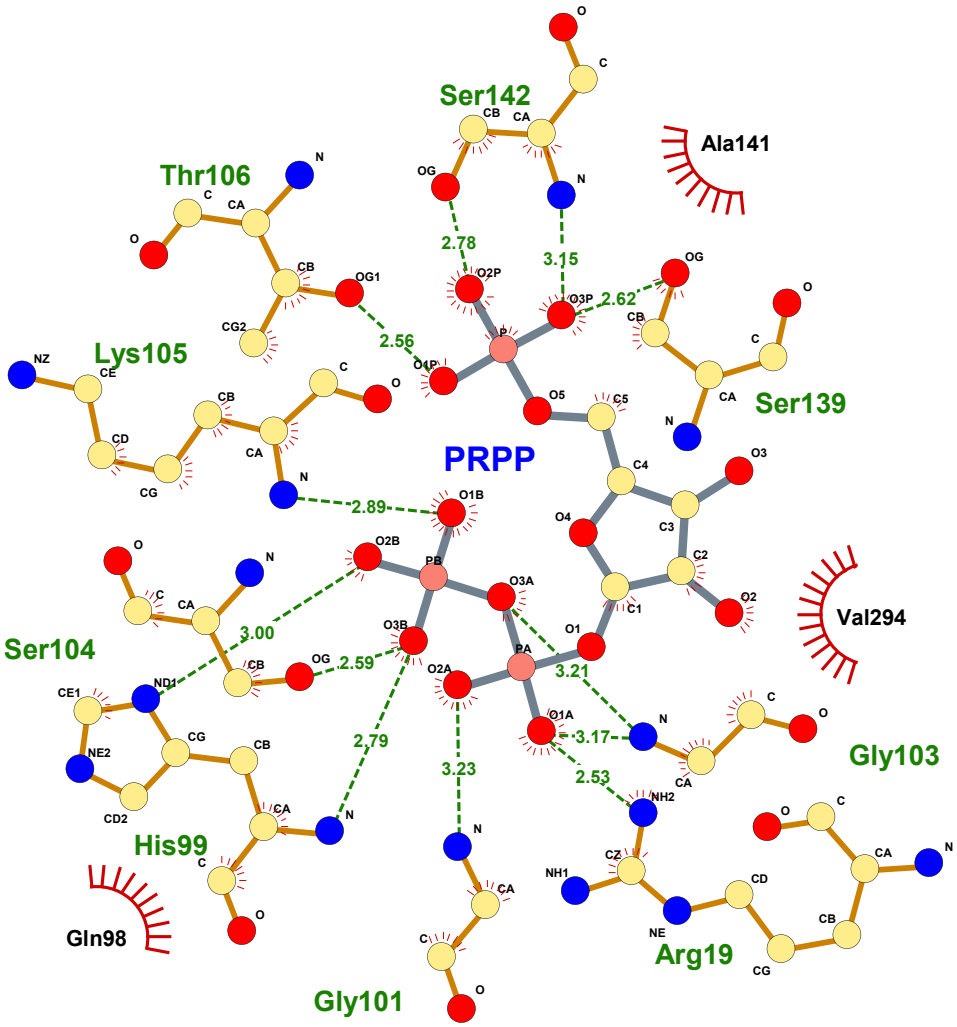



**Figure 2.9.** Ligand plot is generated by LigPlot+ suit [187] shows the interaction between PRPP and amino acids from ForT. The green dot line indicates the interactions with distance shown in angstroms. Colour of atoms: yellow – carbon; red – oxygen; blue – nitrogen; pink – phosphate.

**Table 2.3.** Binding affinity of PRPP (determined by ITC) and specific activities (determined by MIMS assay) for WT ForT and site-specific ForT variants.

| Enzyme | $K_D$ (μM) | Activity (U/mg) |
|---|---|---|
| WT ForT | 3.6 ± 0.3 | 0.002 |
| T138V | 1.2 ± 0.2 | Not Determined |
| S174A | 6.3 ± 0.3 | 0.0003 |
| S21A | 9.6 ± 2.1 | Unstable under assay condition |
| R19K | 10.8 ± 0.9 | Unstable under assay condition |
| S104A | 11.4 ± 0.8 | 0.0007 |
| S174L | 70 ± 9 | Unstable under assay condition |
| S139A | > 100 | Unstable under assay condition |
| G134P | > 200 | Not Determined |
| H99A | No Binding | Unstable under assay condition |
| R19A | No Binding | No Activity |
| R135A | No Binding | Unstable under assay condition |
| S142A | No Binding | Not Determined |
| T106V | No Binding | Not Determined |

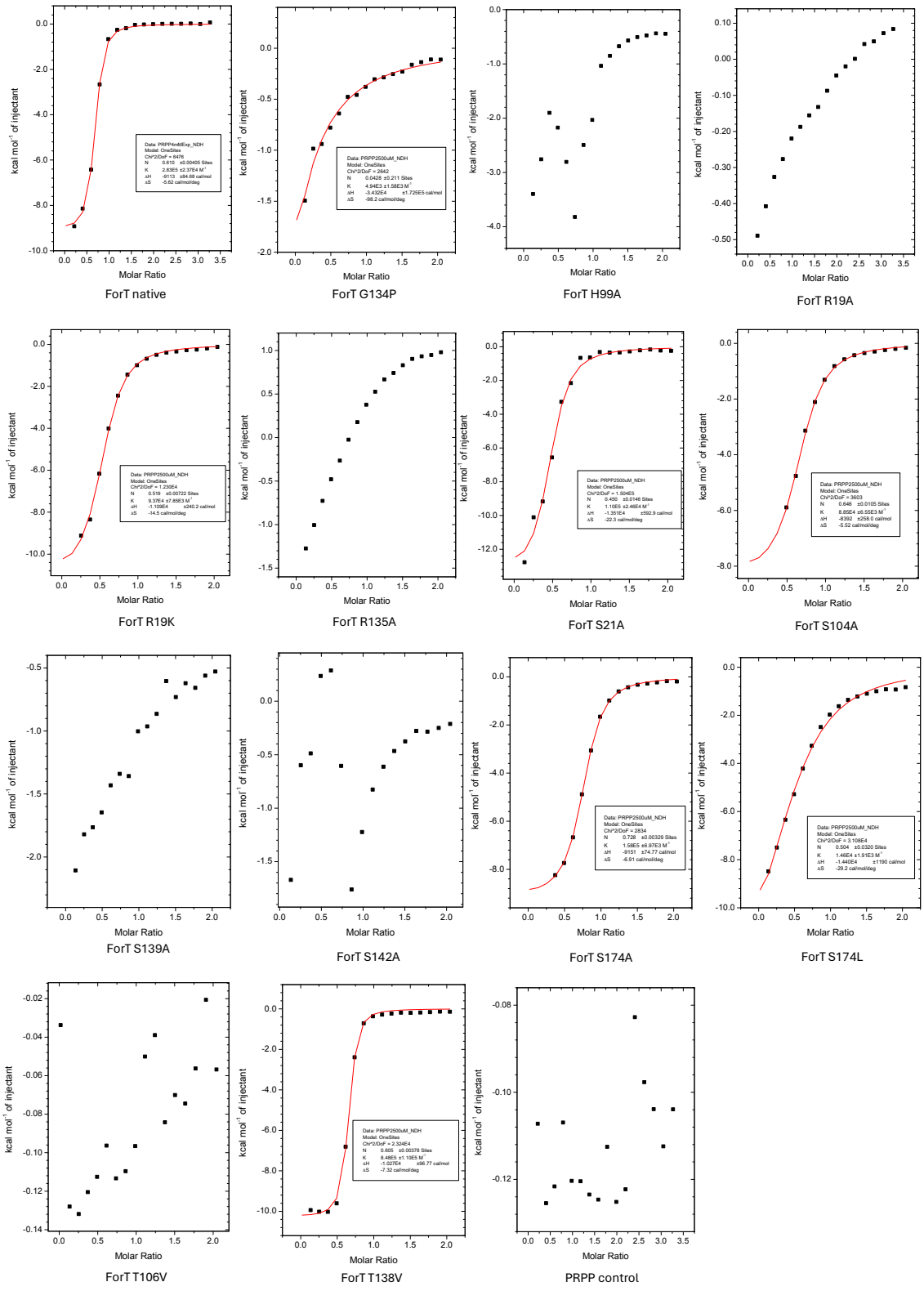



**Figure 2.10.** Isothermal titration calorimetry (ITC) measurement of binding affinity between site-directed variants ForT and PRPP.

## 2.4. Discussion

The recombinant wildtype ForT used to obtain X-ray crystal structure was tested in a series of activity assays. LC-MS assay confirmed the enzyme can catalyse the reaction using PRPP and APDA (**26**) to form the expected carboxyaminopyrazole ribonucleotide (CARP) (**28**) product (Figure 2.8) [7]. Site-directed mutagenesis was informed by the structural studies of ForT to probe different residues at the active site. Because of $CO_2$ is a product of the catalysed reaction, MIMS was used to monitor $CO_2$ production in real-time in order to determine whether it was being produced during the reaction (Figure 2.7) [175,176,188]. It can directly measure the reaction produced gas $CO_2$ to avoid errors caused by indirect measurements. However, it requires a much greater reaction volume (2 mL) hence it requires more protein and substrate to generate good quality data. The WT ForT exhibits a specific activity at 0.002 U/mg, combined with isothermal titration calorimetry (ITC) data, demonstrates a binding affinity with PRPP at 3.6 ± 0.3 µM. There was no binding affinity shown in the ITC assay between WT ForT and ribose 5-phosphate, meaning pyrophosphate is essential for PRPP binding with enzyme. For the ForT variants, there was no PRPP binding affinity observed in ITC for R19A, H99A, T106V, R135A and S142A, confirming that these residues are essential for PRPP recognition and binding (Figure 2.4D).

$Mg^{2+}$ was shown to be required for turnover, as addition of EDTA stopped the reaction (Section 2.3.5). Interestingly, ITC showed there is no binding affinity detectable between WT ForT and APDA (**26**), irrespective of the presence of $Mg^{2+}$. The much higher affinity for PRPP (3.6 ± 0.3 µM) than APDA (> 1mM) suggests it binds first and may induce a conformational change for APDA (**26**).

Data from preliminary MIMS showed that none of the ForT variants retained similar activity as WT ForT. ITC showed that the T138V (1.2 ± 0.2 µM) variant has a slightly higher binding affinity with PRPP than ForT WT (3.6 ± 0.3 µM) had (Table 2.3). There was no $CO_2$ production detected by MIMS maybe due to inactive mutant or reaction was too slow and below the limit of detection of the instrument. S174A and S104A show slightly weaker binding affinities with PRPP, and corresponding decrease in enzymatic activity. Many of the ForT variants

precipitated and exhibited poor stability during MIMS tests. The binding between C-1'-pyrophosphate and ForT binding pocket is vital for recognition and reaction initiation.

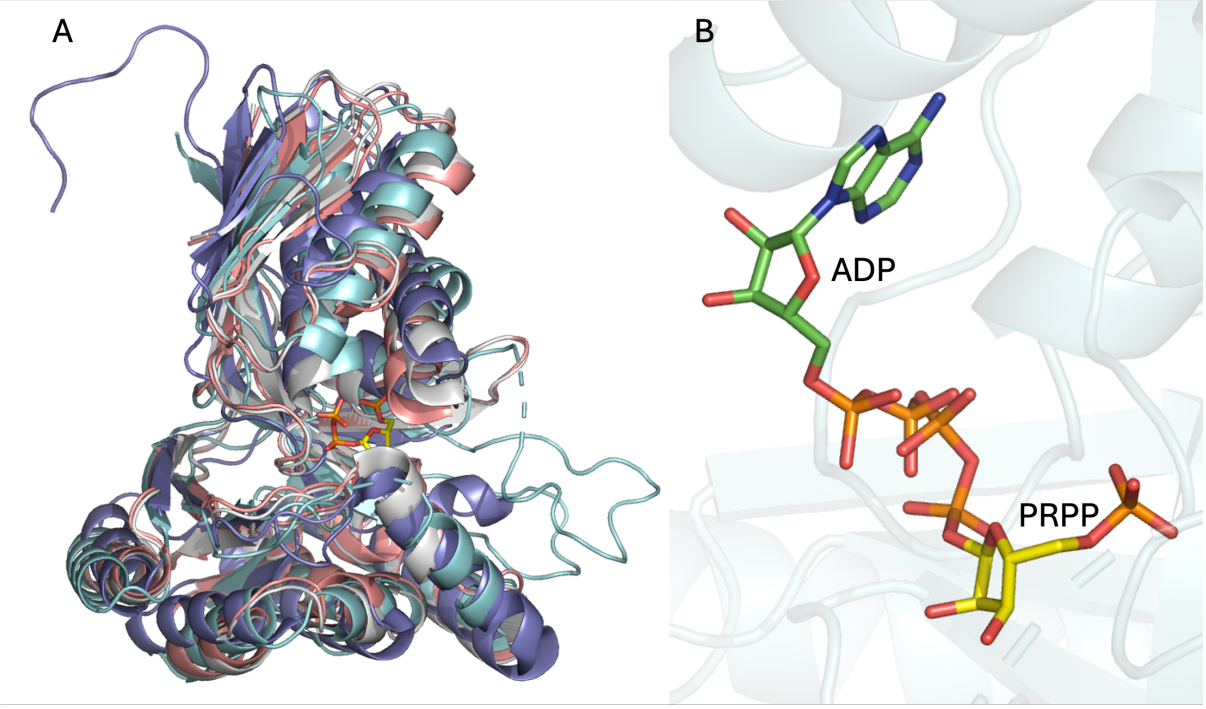



**Figure 2.11.** Figure A shows the superimposed structure of ForT (cyan) with HSK (PDB: 1FWK) (grey), MK (PDB: 6MDE) (pink), and MVK (PDB: 4HAC) (purple). B shows the ribose ring of PRPP from ForT (yellow) is 90 degrees away from ADP (Green) in HSK.

ForT shows structural homology with homoserine kinase (PDB 1h72), where the phosphate group on ADP overlapped with pyrophosphate group on PRPP by superimposed structure [189,190] (Figure 2.11). The structure of enzymes is shaped by their function, the stabilization of the transition state. Thus, structural similarity usually indicates at some level commonality with mechanism. Homoserine kinase (HSK) is an enzyme that catalyses an essential step in the threonine biosynthesis pathway. HSK belongs to the GHMP kinase superfamily that catalyses the phosphorylation of L-homoserine to O-phospho-L-homoserine. The crystal structure of HSK has provided insight into the nucleotide binding site which is structurally conserved not only in the HSK family but also a variety of enzymes in the GHMP kinase superfamily. The catalytic mechanism of HSK involves the transfer of the gamma-phosphate

of ATP to the hydroxyl group of the L-homoserine through formation of a pentacoordinate intermediate. The pentacoordinate intermediate then collapses, releasing the ADP and forming O-phospho-L-homoserine [189,190].

In the GHMP kinase superfamily, the enzyme uses the kinase loop and a $Mg^{2+}$ ion to stabilize emerging negative charge by nucleophilic attack on gamma-phosphate [191]. The high degree of structural similarity suggests the pyrophosphate moiety of PRPP undergoes dissociation to form a transition state with oxocarbenium ion on the ribose ring only in the presence of a second substrate. The breaking of the pyrophosphate group creates a highly active electrophile which the second substrate (the pyrazole ring on APDA) (**26**) attacks as the nucleophile, forming the C-C bond. Driven by re-establishing aromaticity, the intermediate undergoes decarboxylation as $S_N2$ like mechanism. (Figure 2.12) The common requirement of $Mg^{2+}$ ion in both confirms that the reaction needs to be stabilized on pyrophosphate moiety. The release of carbon dioxide from decarboxylation is the major driving force, making this step irreversible.

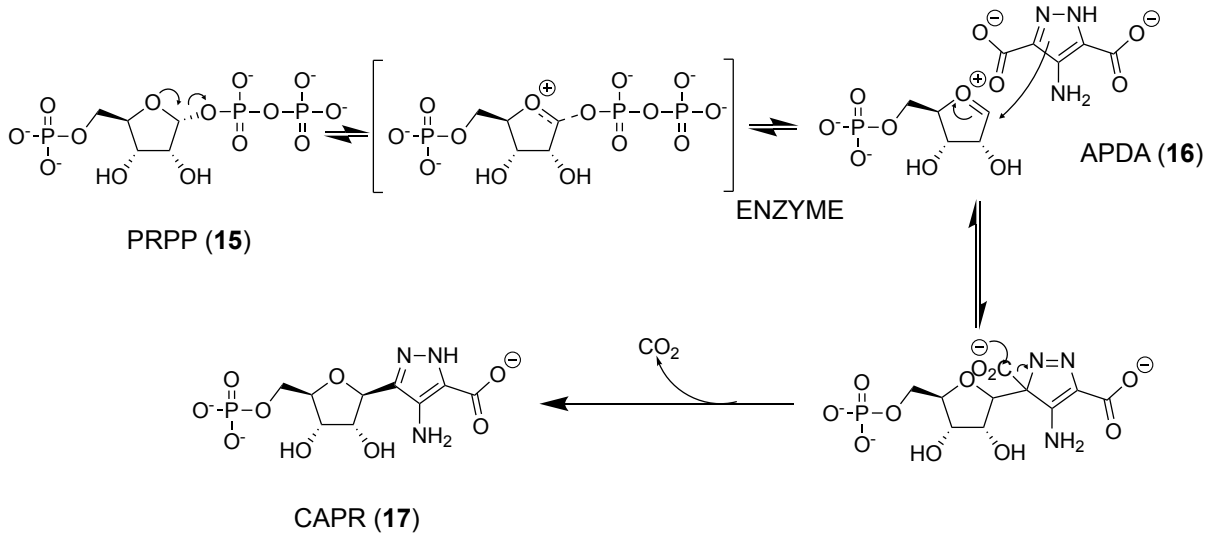



**Figure 2.12.** Proposed mechanism of ForT Catalysed C-C bond formation reaction.

# Chapter Three

# Structural and Functional Studies

## 3.1. Introduction

Driven, in part, by the success of remdesivir in the treatment of viral infections, such as SARS-CoV-2 and Ebola [192,193], and the development of elegant pro-drug strategies for increasing the bioavailability of functionalized nucleotides [194,195], there is renewed interest in C-nucleoside derivatives as anti-viral and anti-cancer drugs [93,109,196,197]. C-nucleosides and C-nucleotides are, of course, resistant to inactivation by hydrolysis of the sugar base linkage, in contrast to the more common C-N nucleotides [91]. Chemical synthesis of C-nucleosides uses two broad approaches: forming the heterocyclic nucleobase from an acyclic substituent attached to the sugar moiety or coupling a preformed heterocycle to the carbohydrate [198–203]. Both of these strategies often require the use of protecting groups, thereby increasing synthetic complexity. Biocatalysis offers an attractive alternative if we can understand the pathways that give rise to naturally occurring C-nucleosides and C-nucleotides [204–206]

Insights into the biosynthetic origins of bioactive C-nucleosides have resulted from the identification of the gene clusters encoding a number of these secondary metabolites [102,207–215]. In broad terms, the cellular pathways to these natural products utilize three distinct chemical strategies to make the key C-C bond (Figure 3.1). First, in the showdomyin and minimycin pathways, C-glycosidic bond formation is catalysed by the C-glycosynthases SdmA and MinB,[211,215,216] respectively, which are both homologous to the glycosidase YeiN.[217] As a

result, it is likely that the C-C bond is formed by nucleophilic attack of double bond in the small molecule substrate on a covalent intermediate that is formed by reaction of the ring-opened sugar with a protein lysine.[218] Second, in the biosynthesis of malayamycin A and pseudouridimycin, a pseudouridine synthase catalyses cleavage of the C-N bond between uridine and ribose-5-phosphate. Re-attachment of the nucleobase to the ribose via a C-C bond then follows rotation of the base within the active site.[213,214] In the third strategy, formycin A and pyrazofurin A arise from direct coupling of a small heteroaromatic dicarboxylic acid to 5'-phosphoribose-1'-pyrophosphate (PRPP) with concomitant decarboxylation.[102,217]

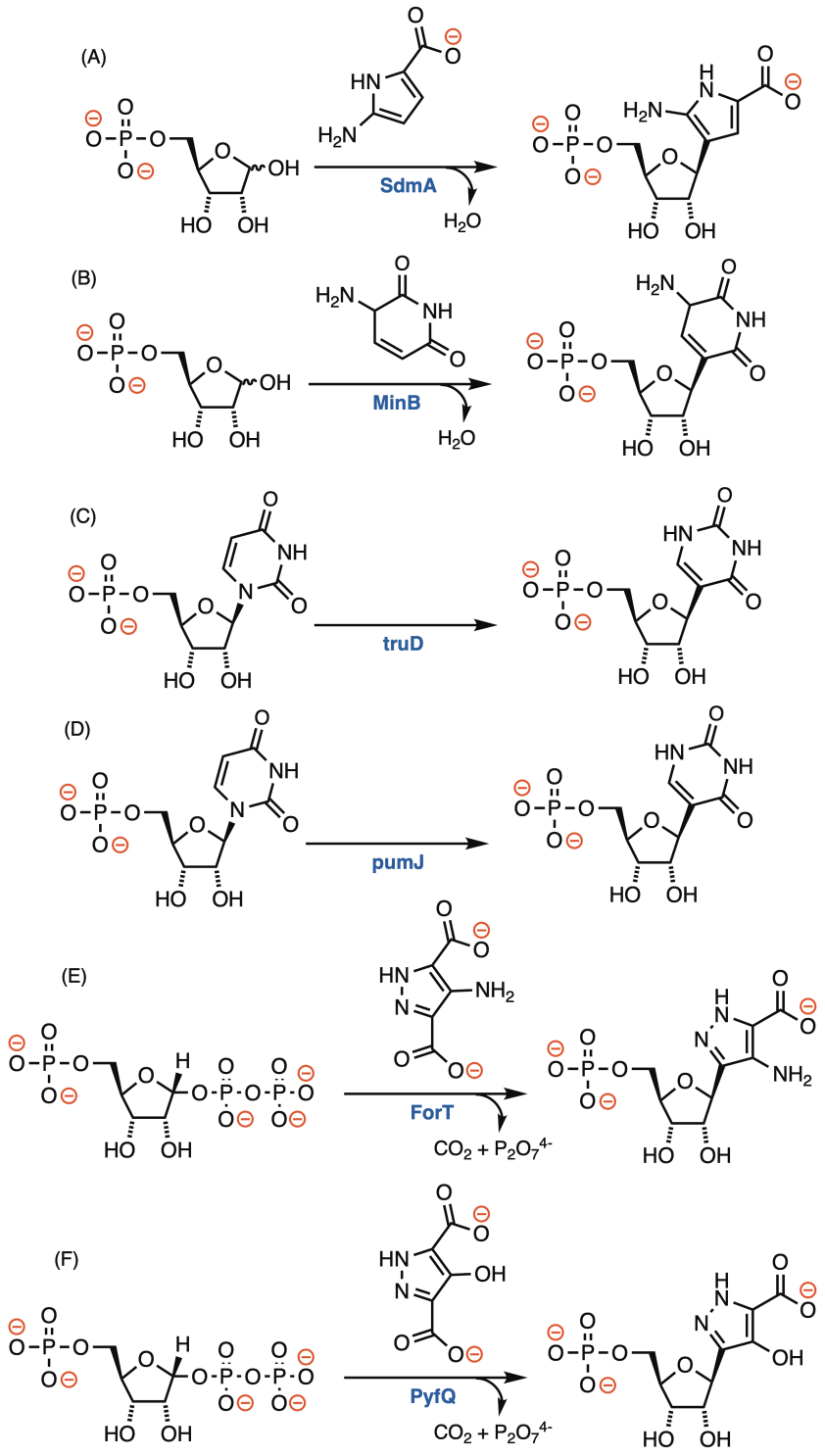



**Figure 3.1.** C-C bond forming reactions in the biosynthesis of C-nucleoside natural products. (A) Showdomycin.[216] (B) Minimycin.[211] (C) Malayamycin.[213] (D) Pseudouridimycin.[214] (E) Formycin.[217] (F) Pyrazofurin.[102] (Figure modified from [219])

In the specific case of formycin, the enzyme ForT catalyses C-C bond formation between 4-amino-1*H*-pyrazole-3,5-dicarboxylate (APDA) and PRPP, in a reaction that is driven by the elimination of $CO_2$ and inorganic pyrophosphate (PPi) (Figure 3.2).[102,217] Feeding studies suggest that glutamic acid and lysine are precursors of APDA,[112,220] although the molecular transformations in the cell that yield this highly functionalized pyrazole intermediate from hydrazinoglutamic acid[221] are, as yet, unknown. In the final steps of the biosynthetic pathway, the C-riboside product of APDA and PRPP is elaborated by four enzymes to give the pyrazolopyrimidine ring system of formycin A (Figure 3.2).[116]

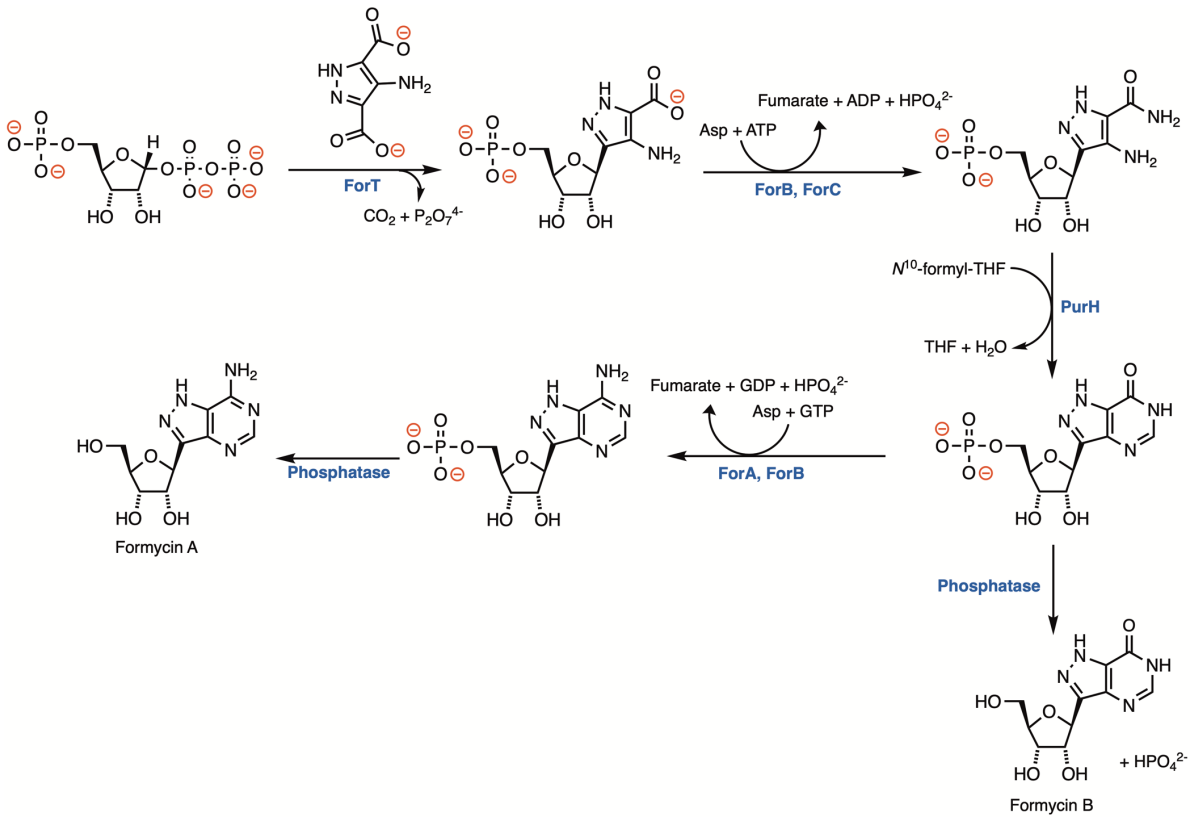



**Figure 3.2.** The later stages of formycin biosynthesis showing the role of ForT in forming the C-C bond in the C-nucleoside antibiotic [117]. (Figure modified from [219])

We have previously determined the structure of ForT, complexed with PRPP (Figure 3.3),[170] and confirmed that this enzyme is a member of the GHMP kinase superfamily.[189] We now report biophysical measurements, structural biology, and quantum chemical (QC) calculations that provide a detailed molecular picture of ForT-catalysed C-C bond formation

and the conformational changes that are coupled to catalysis. Our findings set the scene for future efforts to employ ForT in the biocatalytic production of novel, anti-viral C-nucleoside, and C-nucleotide analogues.

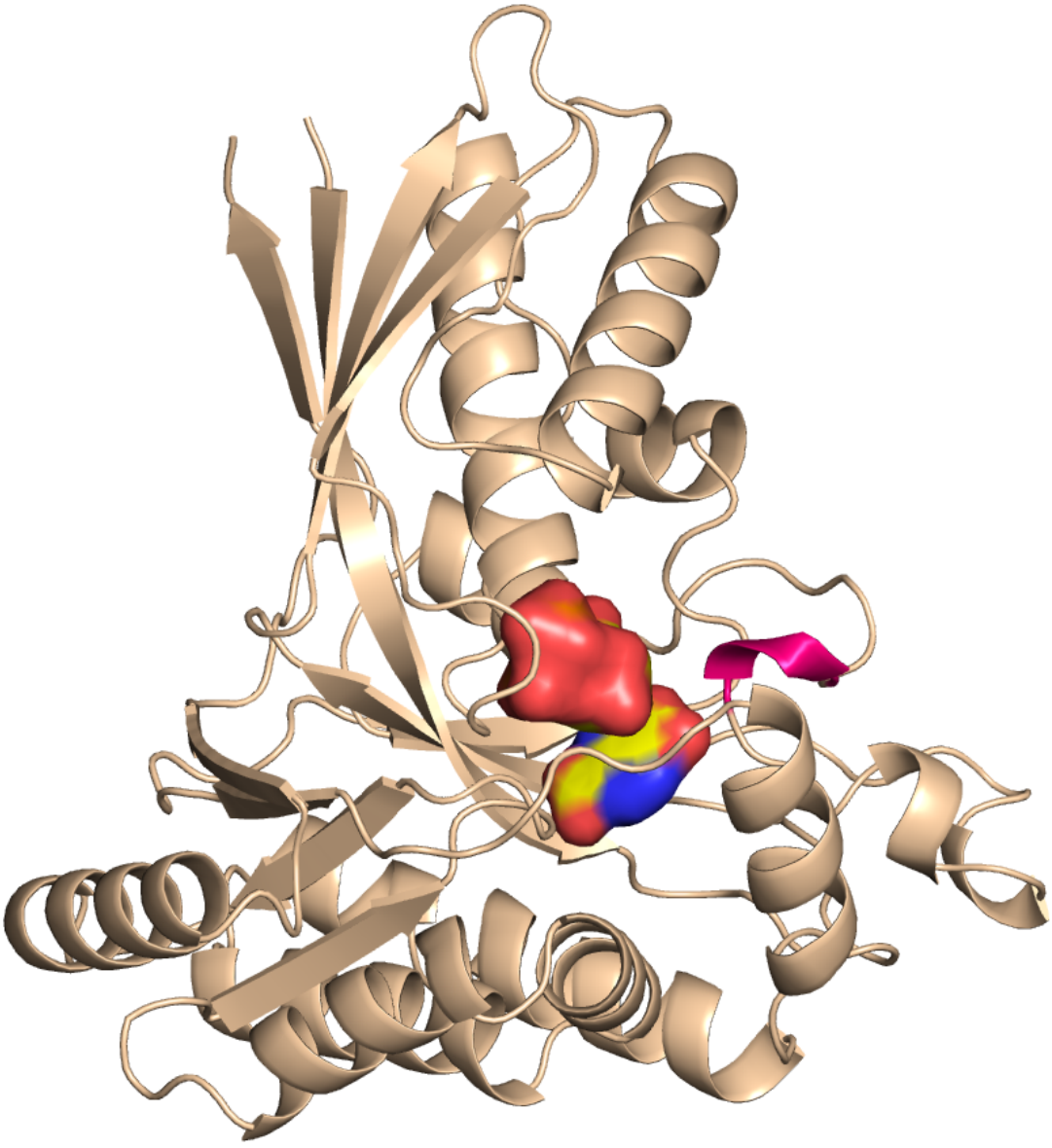


**Figure 3.3.** The structure of ForT T138V is shown in cartoon form and the substrates shown as molecular surface. The loop at Ser174 which is ordered in this structure is coloured purple. Carbon atoms are coloured by yellow, oxygen atoms coloured by red, phosphorous atoms orange and nitrogen atoms blue.

## 3.2. Material and Methods

### 3.2.1. General

Unless specified, all chemicals were purchased from Sigma-Aldrich or Fisher Scientific. Oligonucleotides were purchased from Integrated DNA Technologies. dNTP and DNA polymerase were purchased as part of the EMD Millipore Novagen KOD Hot Start DNA Polymerase kit. DNA sequencing was performed by Source Bioscience. $^{1}$H- and $^{13}$C-NMR spectra were obtained at 600 MHz and 151 MHz, respectively, and recorded using a Bruker Ascend 600 spectrometer equipped with Bruker CryoPlatform cryoprobe. NMR were referenced to residual $H_2O$. Chemical shifts are reported in parts per million (ppm) and multiplicities are indicated as: s (singlet), d (doublet), t (triplet), and m (multiplet). LC-MS and prep-LC were performed on Agilent 1260 infinity systems. ESI-MS data are reported in the form of m/z.

### 3.2.2. Site-Directed Mutagenesis

Mutations of ForT were cloned by an established site-directed mutagenesis protocol, starting from a construct coding for wild type (WT) ForT with an N-terminal His-tag/TEV cleavage site as described in chapter 2. Briefly, primer pairs (around 30 nucleotides) carrying the overlapping mutation (Table 3.1) and Q5 DNA polymerase (New England Biolabs) were employed for PCR amplification (94 °C, 7 min for initial denaturation, 12 x 95 °C, 1 min denaturation followed by $T_m$ - 5 °C 1 min annealing and 68/72 °C 5.5 min extension, 68/72 °C 10 min final extension). The PCR product was digested by DpnI for 1 hour at 37 °C, and the digested product (5 µL) was transformed into *Escherichia coli* DH5α. The mutation site was confirmed by DNA sequencing.

**Table 3.1** Primers used for the site-directed mutagenesis of ForT.

| Primers | Sequences (5'-3') |
|---|---|
| S21A-f | CAGCAAGACTTGCGTTCACGCTGATTAGCCTTGACGGCAGTTCC |
| S21A-r | GCGTGAACGCAAGTCTTGCTGGAGCACTGACCCGAACCGC |
| L24A-f | AGCTTCACGGCGATTAGCCTTGACGGCAGTTCCCTTCG |
| L24A-r | GCTAATCGCCGTGAAGCTAAGTCTTGCTGGAGCACTGAC |
| T138A-f | GGTCGTGCGTCAGGTGCTAGTACAGGCCTTTCCGCCTATGGC |
| T138A-r | CACCTGACGCACGACCCCGACCTAAGGTACGAGCAAGCTCG |
| T138S-f | GGGTCGTAGCTCAGGTGCTAGTACAGGCCTTTCCGCCTATGGC |
| T138S-r | CACCTGAGCTACGACCCCGACCTAAGGTACGAGCAAGCTCGC |
| T138V-f | GGTCGGGGTCGTGTGTCAGGTGCTAGTACAGGCCTTTCCGCCTATG |
| T138V-r | AGCACCTGACACACGACCCCGACCTAAGGTACGAGCAAGCTC |
| S174A-f | CGCCCGGCGCGTTTCGCACAGCAAGTAGCTCCGCCTAAG |
| S174A-r | GCGAAACGCGCCGGGCGTAAATACTTTTGTGGAGCCTCTGCG |
| E211A-f | CAAGAAGCGTTGGAGTGGTTTCATTCAGTCGCACCCATAC |
| E211A-r | CCACTCCAACGCTTCTTGTCCACCAAGATGACGGCCGT |
| K263A-f | GCACTTTGCGCAAGCGCAGATCGCGTTCCAGGGGGAT |
| K263A-r | CGCTTGCGCAAAGTGCCCTGTAAAGGTTATTTCGTTAACGGC |
| Q266A-f | AAGCAAGCGGCGATCGCGTTCCAGGGGGATGCAGTAG |
| Q266A-f | GCGATCGCCGCTTGCTTAAAGTGCCCTGTAAAGGTTATTTC |

### 3.2.3. ITC Binding Assay with PRPP, APDA, and Other Substrates

Equilibrium binding experiments with WT ForT and ForT variants were performed using a MicroCal PEAQ-ITC (Malvern) calorimeter in the absence of $Mg^{2+}$ or $Mn^{2+}$. Protein samples were exchanged into buffer [10 mM HEPES, pH 7.5, containing 150 mM NaCl, and 1 mM TCEP] prior carry out the experiments. ITC binding experiments were performed using a cell solution containing 100 μM protein, stirred at 750 rpm and held at 25 ˚C using a reference power of 5 μcal/s. For PRPP binding to the free enzymes, 2mM PRPP was dissolved in 10 mM HEPES, pH 7.5, containing 150 mM NaCl, and 1 mM TCEP, and, using the multiple injection method, titrated into the cell containing 100 μM protein. For APDA binding to the enzyme/PRPP binary complex, 5 mM APDA was dissolved in 10 mM HEPES, pH 7.5, containing 150 mM NaCl, and 1 mM TCEP, and titrated into the cell containing 100 μM enzyme and 5 mM PRPP. A buffer titration control in which compounds were titrated into buffer was performed for each experiment. Data processing, fitting to a nonlinear one-site binding model, and data plotting were performed using MicroCal PEAQ ITC analysis software (Malvern).

### 3.2.4. LC-MS Assay for Determining the Catalytic Activity

WT ForT (25 nM) was incubated with APDA (5-150 μM) and PRPP (500 μM) in sodium phosphate buffer (100 mM, pH 7.5) containing $MgCl_2$ (1 mM) and NaCl (100 mM) for 1 h at room temperature. Aliquots of the reaction (180 μL) were taken at time points (which varied depending on enzyme activity) and quenched by the addition of 18 μL of 10% TFA (v/v). After 5 minutes, each aliquot was neutralised by the addition of 18 μL of aq. NaOH (0.5 M) after 5 min and centrifuged at 3000 rpm for 15 mins to remove precipitated protein. The resulting supernatant was analysed by LC-MS, on a Kinetex 2.6 μm XB-C18 100A, 100 x 4.6 mm LC column with a flow rate of 1 mL/min, using a step gradient of solvent A [water containing 0.1% TFA] and solvent B [acetonitrile containing 0.1% TFA], with 5% Solvent B (0-1 min), 5-15% Solvent B (1-2 minutes), 15-100% Solvent B (2-2.2 min). UV detection was set to 254 and 280 nm. Integration of the CAPR peak at 280 nm (typically $t_R$ 0.88-1.1 min) was used to determine the amount of product, based on a standard curve (Figure 3.6), formed at each timepoint (s). Linear regression gave the velocity at each APDA concentration, and these values were used

to construct Michaelis-Menten plots (Figures 3.4 and 3.11). Non-linear curve fitting then gave estimates of the steady-state parameters, $k_{cat}$ and the apparent $K_M$ of APDA. In these experiments, APDA concentrations varied from 5-150 µM for WT ForT, 10-1000 µM for the S174A variant, 50-2000 µM for the L24A variant, 250-8000 µM for the S21A and K263A variant, 25-1000 µM for the T138A variant, and 10-250 µM for the T138V variant (Figure 3.11). Protein concentrations were 40 nM for WT ForT , 2.5 µM for S21A , L24A (575 nM), T138A (100 nM), T138V (500 nM), S174A (250 nM), and K263A (500 nM).

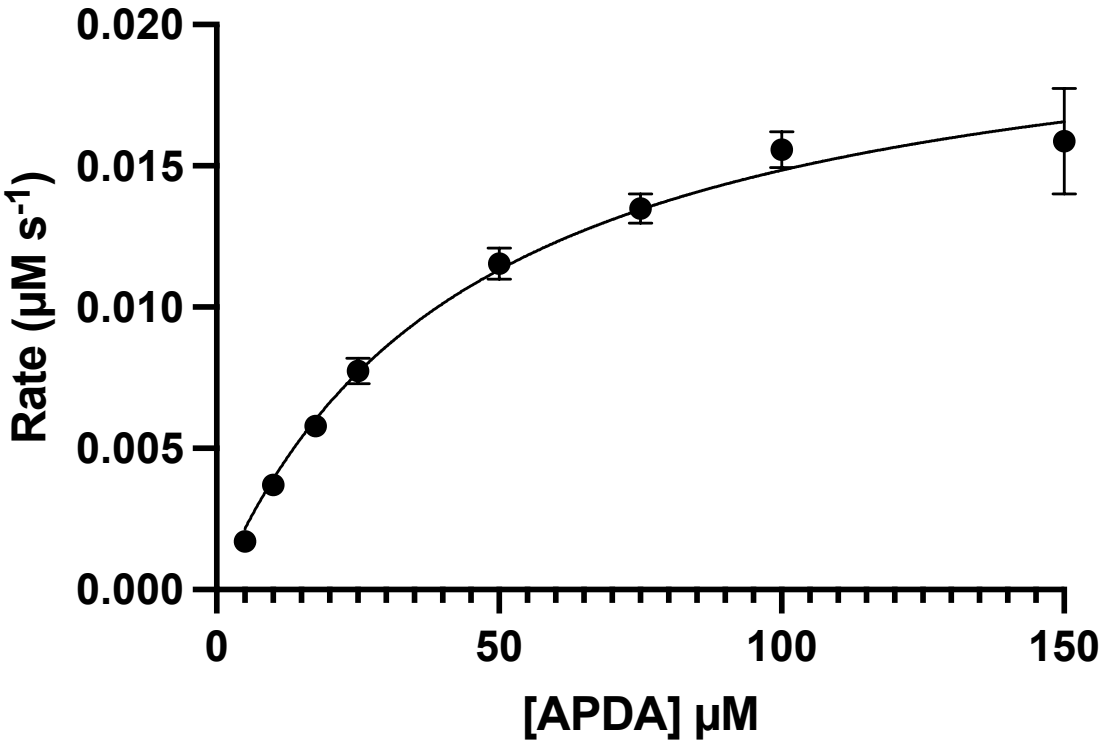



**Figure 3.4.** Michaelis-Menten curve achieved by LCMS assay for WT ForT reaction rate against varying APDA concentration. (Data from Dr. Georgina Girt)

### 3.2.5. General Screening Procedure for Establishing the Substrate Scope of ForT

ForT WT (5 μM) was incubated with substrate candidate (500 μM) in sodium phosphate buffer (pH 7.5, 100 mM) with PRPP (1 mM), $MgCl_2$ (1 mM) and NaCl (100 mM). Reaction progression was monitored using LC-MS at UV of 254 nm and 280 nm, with hits determined by new peak formation combined with substrate peak depletion. New peaks (not present in the absence of enzyme) were analysed by ESI-MS, $^{1}H$ and $^{13}C$ NMR.

### 3.2.6. ForT-catalysed synthesis of 5-amino-1-(β-d-ribofuranosyl) pyrazole-4-carboxylate 5'-monophosphate

3-amino-1*H*-pyrazole-4-carboxylic acid (6 mM) was incubated at 25 $^{o}C$ with ForT (10 μM) and PRPP (8 mM) in 100 mM sodium phosphate buffer, pH 7.5, containing and 8 mM $MgCl_2$ and NaCl (100 mM) for 3 h (10 mL total volume). The reaction was monitored by LC-MS with both m/z charge and UV at 280 nm. On completion, ForT was removed from solution by centrifugation through a 10 kDa molecular weight cut-off filter (10,000 rpm, 10 minutes). The resulting filtrate was purified by semi-preparative anion exchange HPLC (Dionex Carbopac PA1 9 x 250 mm) using a flow rate of 4 mL/ min. The pure product (iso-CAPR) was obtained using a step gradient of $H_2O$ (solvent A) and 1M aq. $NH_4OAc$ (solvent B) [10% B 0-3 min, 10-50% B 3-10 min, 50-90% B 10-15 min, 90% B 15-20 min, 90-10% B 20-21 min, 10% B 21-25 min], eluting from the column with a retention time of 16.5-17 min. Lyophilization gave iso-CAPR as a hygroscopic white powder (1.8 mg, 24 %). $^{1}H$-NMR (600 MHz, $D_2O$) $\delta$ (ppm) 7.73 (1H, s), 5.74 (1H, d, *J* = 6.3 Hz), 4.74 (1H, t, *J* = 5.9 Hz), 4.40 (1H, dd, *J* = 3.5, 5.5 Hz), 4.26 (1H, dd, *J* = 3.5, 7.0 Hz), 4.07-3.99 (2H, m); $^{13}C$ (150 MHz, $D_2O$) $\delta$ 179.7, 151.5, 142.1, 96.7, 88.7, 83.4, 71.8, 70.2, 64.9; MS (ESI) *m/z* $[M + H]^+$ 340.05; found 340.10.

This compound is known in the literature.[222] Minor differences between our spectral data and that reported previously likely result from pD differences, which we have observed can cause significant changes in $^{1}H$ chemical shifts, particularly of the pyrazole CH signal (Data from Dr. Georgina Girt).

Literature[222] $^{1}$H-NMR (300 MHz, $D_2O$) $\delta$ 7.51 ppm (1H, s), 5.56 ppm (1H, d, *J* = 6.7 Hz), 4.60 (1H, dd, *J* = 5.5, 6.7 Hz), 4.25 (1H, dd, *J* = 3.3, 5.5 Hz), 4.08 (1H, m), 3.76 (2H, m); $^{13}$C (75 MHz, $D_2O$) $\delta$ 169.3, 152.9, 143.5, 98.1, 90.1, 84.7, 73.2, 71.6, 66.3; HRMS (ESI) *m/z* $[M + H]^+$ calculated. for C9H15N3O9P 340.0546; observed 340.0545.

### 3.2.7. Structural Biology

Crystallization conditions for the two ForT variants employed 40 mg/mL protein mixed at 1:1 ratio in vapour diffusion with reservoir solution containing 40 mM $KH_2PO_4$, 20 % (v/v) glycerol, 16 % (w/v) PEG 8000 and 20 mM Bis-tris propane with the pH varied between 6 and 8. The T138V variant of ForT is easier to be purified and better crystallised comparing to the WT enzyme. At the time of the experiment, T138V was thought to be inactive initially, and so T138V ForT (20 mg/mL), PRPP (80 mM), $MgCl_2$ (5mM) and APDA (80 mM) were incubated for two hours before crystallization. Crystals, which appeared at around 10 days, were then soaked with APDA (100 mM) and PRPP (100 mM) immediately prior to be cryo-cooled in liquid nitrogen. Data were collected at the Diamond Light Source (DLS) on beamline I03 and reduced using the DIALS pipeline.[177,178] The complex structure was solved using molecular replacement. The structure of the T138V/APDA/$Mg^{2+}$/PRPP ternary complex was determined to 1.6 Å in the space group $C222_1$ by molecular replacement using PHASER.[180] WT ForT (PDB 6YQQ) without PRPP was used as the search model.[170] Refinement was performed using the CCP4 suite, and ligands were introduced into clear difference electron density during refinement when the $F_o$-$F_c$ map was judged to be unambiguous in COOT.[223] The same crystallization conditions were used to obtain the structure of the T138V/APDA/$Mn^{2+}$/PRPP ternary complex, except that 5 mM $MnCl_2$ replaced $MgCl_2$. In order to observe the $Mn^{2+}$ anomalous diffraction, data were collected at 9 keV (1.377 Å) on the DLS beamline I03. The final refined structure was obtained using restrained $Mg^{2+}$/water distances.

The L24A/CAPR/$PP_i$ complex was obtained by incubating the L24A ForT variant (25 mg/mL), with PRPP (100 mM), $MnCl_2$ (5mM) and APDA (100 mM) for two hours prior to crystallization using the conditions described above. Data were collected at the Diamond Light Source (DLS)

on beamline I03. The structure was determined to 1.75 Å resolution in space group $P4_12_12$ and refined as described for the T138V/APDA/$Mg^{2+}$/PRPP ternary complex.

## 3.3. Results

### 3.3.1. Biophysical Measurements.

Fitting ITC data yielded $K_D$ values of 0.9 µM for PRPP binding to WT ForT and of 180 µM for APDA binding to the WT ForT/PRPP complex under room temperature condition (Figure 3.5 and Table 3.2). Both of these measurements, however, were performed in the absence of $Mg^{2+}$. Typical Michaelis-Menten kinetics were observed using freshly purified WT ForT in an LC-MS assay (Figure 3.6) when APDA was varied in the presence of saturating levels of $Mg^{2+}$/PRPP at room temperature (Figure 3.4). Curve-fitting using Prism as a Michaelis-Menten[224] model gave an apparent $K_M$ for APDA of 45 µM and a turnover number ($k_{cat}$) of 0.54 $s^{-1}$. The recombinant WT enzyme can catalyse C-C bond formation in the presence of either $Mg^{2+}$ or $Mn^{2+}$ but not $Ca^{2+}$ (Figure 3.7). Little change in activity (within error) is seen over a pH range of 6-8 (Figure 3.8), although the rate of the ForT-catalysed reaction can be increased by using $NaH_2PO_4$/$Na_2HPO_4$ rather than HEPES as the assay buffer (Figure 3.9). In our experimental conditions, WT ForT loses activity when stored at 4 °C in 10 mM HEPES (pH 7.4) containing 150 mM NaCl and 1 mM TCEP. Several ForT variants including T138S and Q266A that were purified were also prone to precipitation.

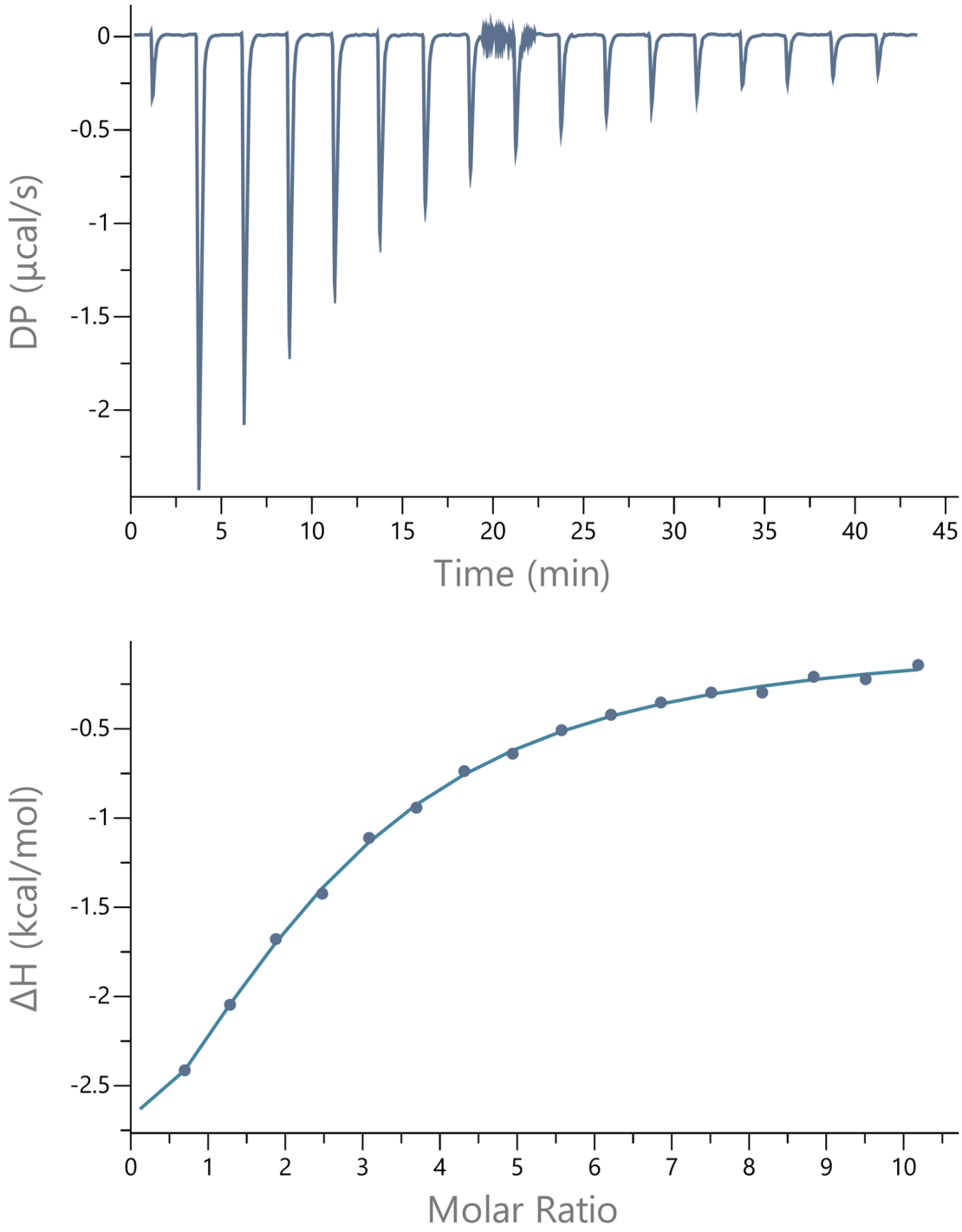



**Figure 3.5.** Single injection ITC experiment raw heat plot for injecting APDA into WT ForT/PRPP complex, gives $K_D$ for APDA is 180 µM at room temperature. Experimental data were fitted by suing one set of sites model from MicroCal PEAQ-ITC analysis software.

**Table 3.2.** ITC measurements of binding affinity, in the absence of $Mg^{2+}$, for PRPP by free ForT and APDA to the ForT/PRPP binary complex, and steady-state kinetic parameters (determined by the LC-MS assay) for WT ForT and site-specific ForT variants. Data for the ITC measurements (Figures 3.5 and 3.10) and the kinetic assays are shown in Figures 3.4 and 3.11.

| Enzyme | PRPP $K_D$ (μM) | APDA $K_D$ (μM) | APDA App. $K_M$ (μM) | $k_{cat}$ ($s^{-1}$) |
|---|---|---|---|---|
| WT ForT | 0.93 ± 0.04 | 120 ± 17 | 45 ± 10 | 0.54 ± 0.05 |
| S21A | 0.35 ± 0.02 | 108 ± 38[a] | 4300 ± 540 | 0.061 ±0.005 |
| L24A | 1.5 ± 0.10 | >1 mM[b] | 1300 ± 180 | 0.130 ± 0.008 |
| T138A | 6.7 ± 0.50 | >1 mM[b] | 220 ± 21 | 0.20 ± 0.01 |
| T138V | 0.7 ± 0.40 | 150 ± 82 | 69 ± 9 | 0.013 ± 0.007 |
| S174A | 6.0 ± 0.30 | >1 mM[b] | 210 ± 54 | 0.10 ± 0.01 |
| E211A | 0.95 ± 0.02 | 180 ± 17 | Inactive[c] | Inactive[c] |
| K263A | 20 ± 8 | >1 mM[b] | 3200 ± 1700 | 0.21 ± 0.05 |

[a] binding affinity measured at high noise and half occupancy; weaker binding confirmed with LC-MS assay.

[b] no binding measured

[c] no activity detected

Similar biophysical and kinetic studies were performed for a series of ForT variants to explore the effects of replacing conserved, active site residues on ligand binding and catalytic properties (Figures 3.10 and 3.11). All of the ForT variants exhibited reduced activity except E211A for which we confidently measured no activity; ITC measurements show that this amino acid substitution does not alter the binding of both substrates (Table 3.2). Of the remaining variants, a 10-fold reduction in $k_{cat}$ resulted from replacing Thr138 by valine (T138V). In contrast to the preliminary results $k_{cat}$ was founded to be reduced by only 4-fold in the T138A ForT variant (Table 3.2). Both of these substitutions, however, have little impact on the apparent $K_M$ of APDA (Table 3.2). In contrast, the L24A variant exhibits only a small reduction in $k_{cat}$ but a 20-fold reduction in affinity for APDA on the basis of both the $K_M$ and $K_D$ values. The T138S and Q266A ForT variants were unstable and precipitated, this prevented any useful measurements of binding or any kinetic characterization.

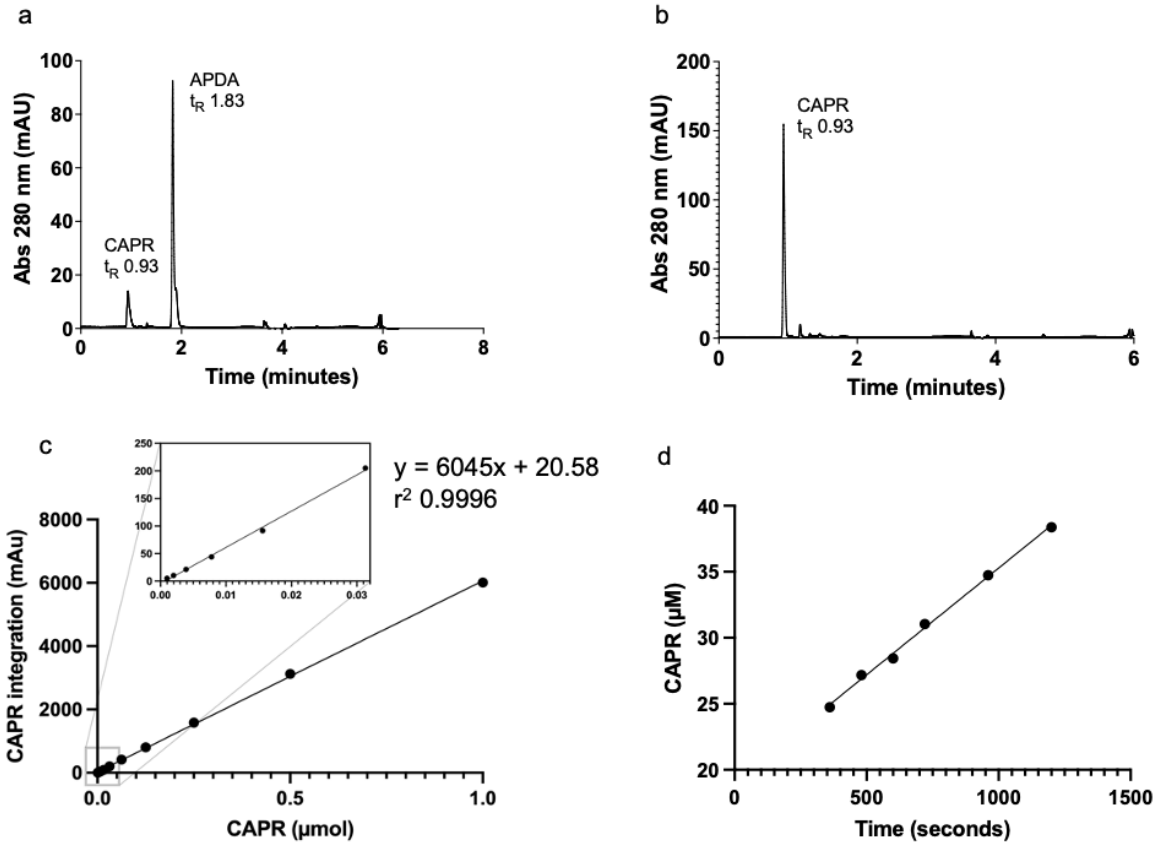



**Figure 3.6.** Integration of LC-MS UV trace used to quantify CAPR formation in samples. (a) LC trace (280 nm) of a ForT-APDA-MgPRPP reaction aliquot showing CAPR ($t_R$ 0.93 min) and APDA ($t_R$ 1.83 min). (b) LC trace (280 nm) of purified CAPR standard. c) To confirm that integration of the LC-MS CAPR peak accurately quantified product formation, we obtained a standard curve. Briefly, a known amount of CAPR (as determined by quantitative NMR) was used to prepare a 2x dilution series, and peak integration for each concentration (280 nm) was determined by LC-MS. The equation of the line of best fit was used to convert mAU to µM. (d) Plot of CAPR concentration vs time; linear regression gives rate of reaction at a given concentration of APDA for use in determining Michaelis-Menten parameters (Data from Dr. Georgina Girt).

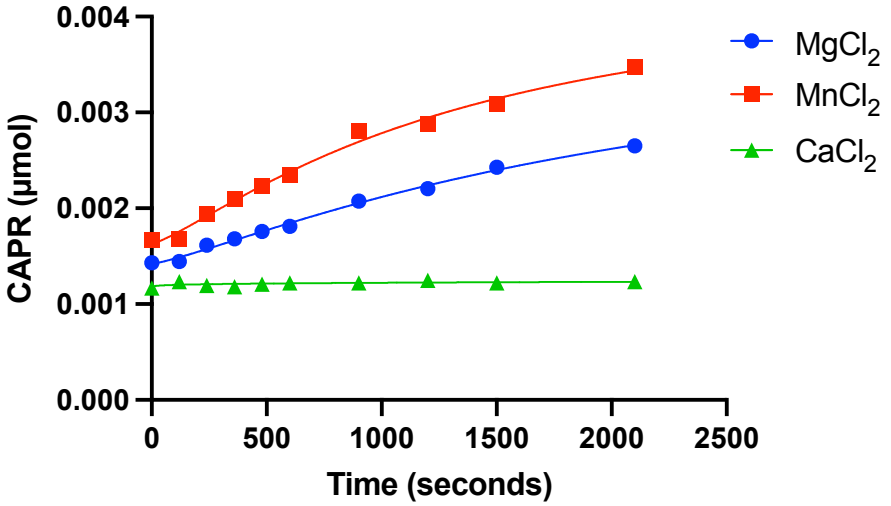



**Figure 3.7.** LC-MS analysis of reactions (800 µL total volume) in which WT ForT (50 nM) was incubated with APDA (200 µM) and PRPP (500 µM) in 100 mM sodium phosphate buffer, pH 7.5, containing $Mg^{2+}$ (1 mM), $Mn^{2+}$ (1 mM) or $Ca^{2+}$ (1 mM). Reaction progress was measured by integration of the CAPR peak at 280 nm. No product formation was observed in the presence of $CaCl_2$ (Data from Dr. Georgina Girt).

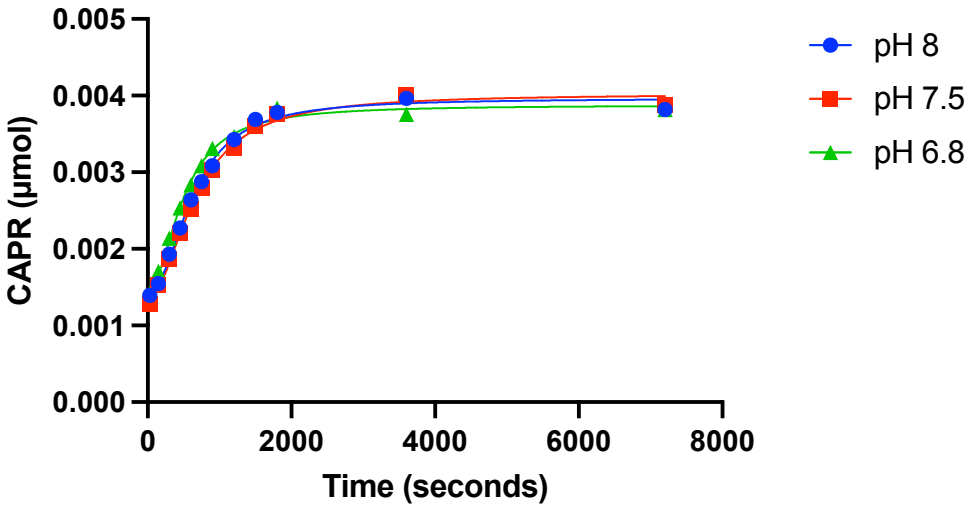



**Figure 3.8.** LC-MS analysis of reactions (800 µL total volume) in which WT ForT (200 nM) was incubated with APDA (50 µM) and PRPP (500 µM) in 100 mM sodium phosphate buffer containing NaCl (100 mm) and $MgCl_2$ (1 mM), at three pH values (8.0, 7.5 and 6.8). Reaction progress was measured by integration of the CAPR peak at 280 nm (Data from Dr. Georgina Girt).

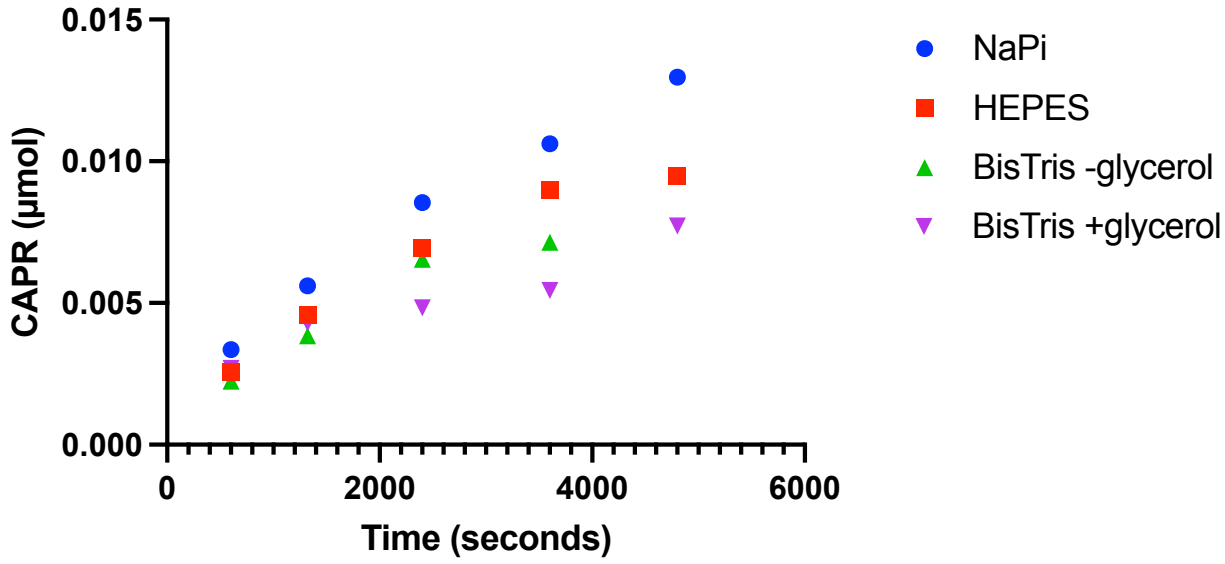



**Figure 3.9.** LC-MS analysis of reactions (200 µL total volume) in which WT ForT (100 nM) was incubated with APDA (500 µM) and PRPP (1 mM) in different buffers containing $Mg^{2+}$ (1 mM). Reaction progress was measured by integration of the CAPR peak at 280 nm. Colour scheme: blue, 100 mM sodium phosphate, pH 7.5; red, 10 mM HEPES pH 7.5; purple, 20 mM Bis-Tris, pH 7, containing 40 mM $KH_2PO_4$ and 20% glycerol; green, 20 mM Bis-Tris, pH 7, containing 40 mM $KH_2PO_4$ (Data from Dr. Georgina Girt).

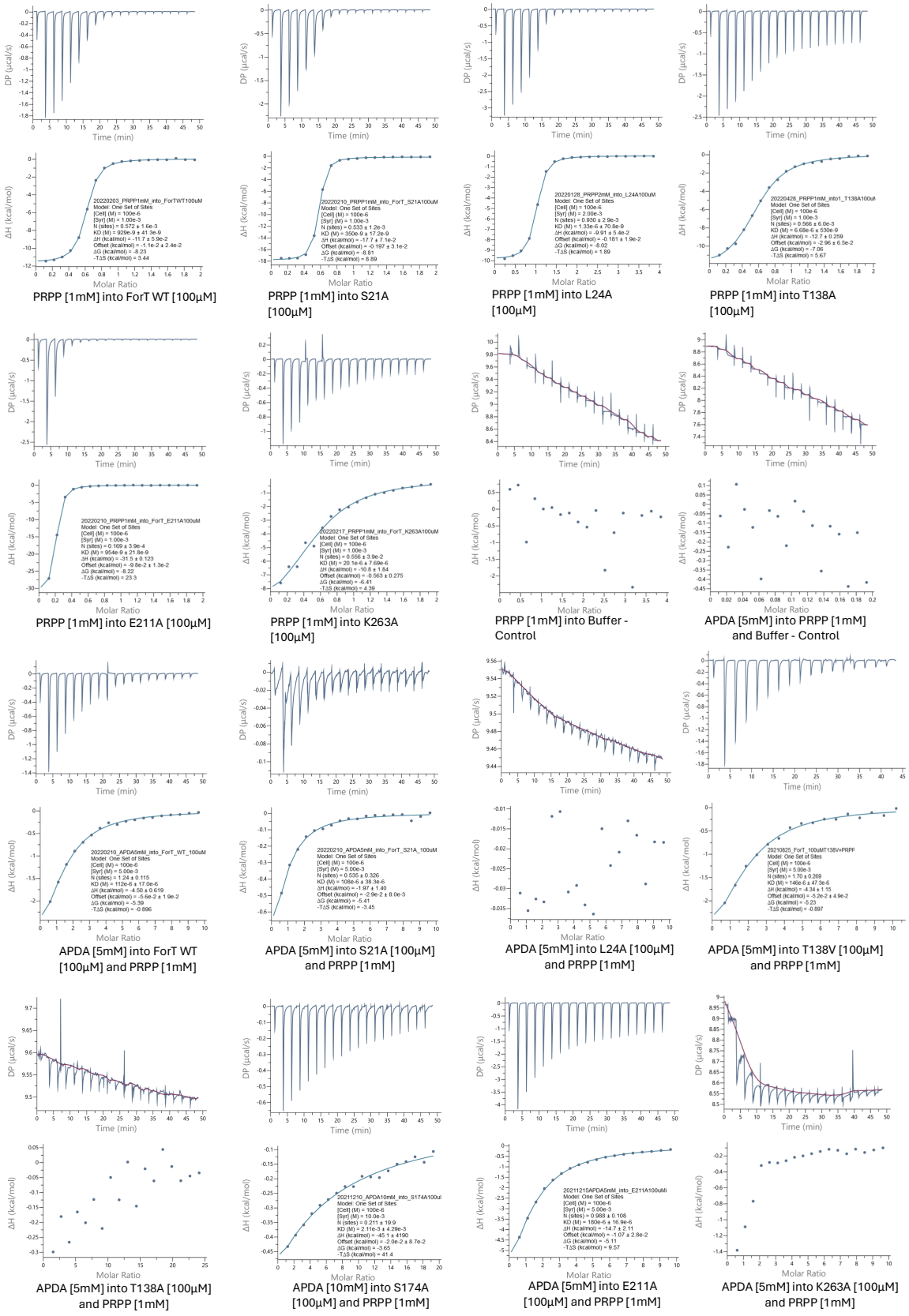



**Figure 3.10.** ITC measurements. Experimental details are provided in section 3.2.3. Experimental data were fitted by suing one set of sites model from MicroCal PEAQ-ITC analysis software.

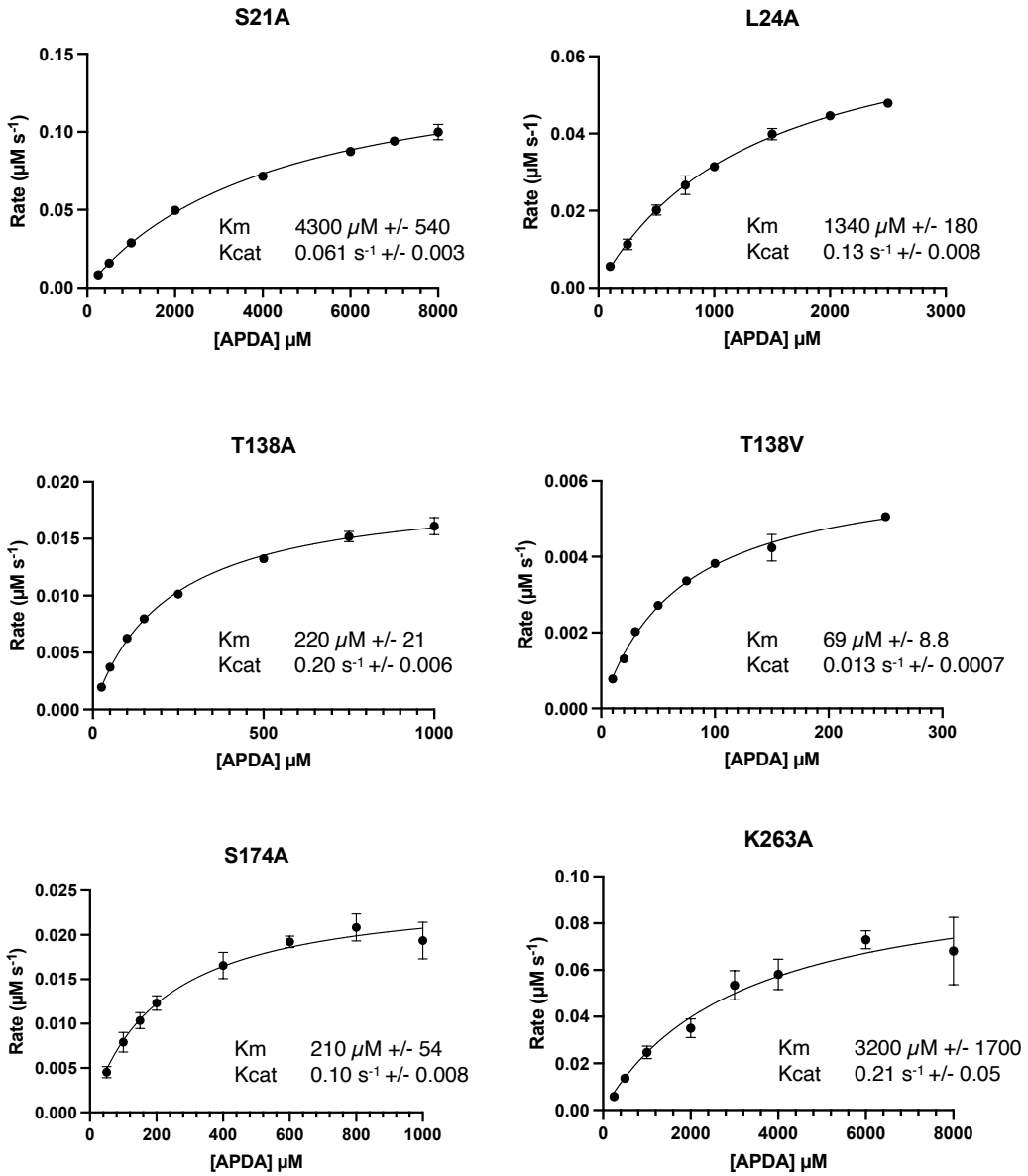



**Figure 3.11.** Michaelis-Menten kinetics of ForT WT and all active ForT variants, determined in triplicate using an LC-MS assay. Reactions containing fixed amount of enzyme, PRPP (2 mM) and $MgCl_2$ (2 mM) in 100 mM sodium phosphate buffer containing NaCl (100 mM), were incubated with eight varying concentrations of APDA, and monitored at six timepoints (see Materials and Methods for enzyme and APDA concentrations). Amount of product formation, as determined by integration of CAPR at 280 nm, was plotted as a function of time. A linear regression equation was used to fit data, and gradient of the slope was plotted vs concentration of APDA and analysed using the Michaelis-Menten equation [224] on Prism (Data from Dr. Georgina Girt).

### 3.3.2. Structural Biology

The structure of the T138V/APDA/MgPRPP ternary complex was determined to 1.6 Å in the space group $C222_1$ by molecular replacement, using WT ForT (PDB 6YQQ) without PRPP as the search model (Table 3.3).[170] Bound PRPP and APDA were easily placed in clear difference electron density in the active site (Figure 3.12a). Definitively locating a $Mg^{2+}$ ion bound to PRPP, however, proved impossible due to the appearance of the electron density is very close to a water molecule. Whilst the density suggested an ion, the configuration was unusual, and the difference density could be modelled by water. We therefore collected data at the anomalous edge of manganese on crystals of the ternary complex formed when $Mn^{2+}$ rather than $Mg^{2+}$ was present in the crystallization solution. In the resulting anomalous difference Fourier map, the highest peak appeared where we had tentatively positioned $Mg^{2+}$ in our original model (Figure 3.12b). In our structure, the $Mg^{2+}$ ion is coordinated in a distorted octahedron to both hydroxyl groups of the ribose, two oxygens in the pyrophosphate moiety, and two water molecules (O-$Mg^{2+}$ distances of 2.0-2.2 Å). This coordination in the ternary complex resembles that first seen in glutamine phosphoribosylpyrophosphate amidotransferase (GPAT), which catalyses formation of 5-phosphoribosyl-1-amine from ammonia and PRPP. [225] A similar arrangement of ligands is also observed for one of the $Mg^{2+}$ ions in the very high-resolution structure of hypoxanthine-guanine phosphoribosyltransferase (HGPRT),[226] the enzyme that catalyses C-N bond formation between PRPP and inosine or guanosine (Figure 3.13). The distorted octahedral arrangement around the $Mg^{2+}$ ion was proposed to activate PRPP for catalysis. $Mg^{2+}$ suggested to promote strong phosphate-$Mg^{2+}$ interaction to help decomposition of PRPP and cleaving pyrophosphate linkage [227]. There are two $Mg^{2+}$ ions in HGRPT that coordinate the pyrophosphate group,[226] but this is absent in our ForT/PRPP/APDA/$Mg^{2+}$ complex structure, the space being occupied by the backbone of Gly101 (Figure 3.12c). In addition, the alignment of amide dipoles in the backbone about Gly101 allows the negatively charged phosphate group to form favourable charge-dipole interactions with the enzyme (Figure 2.4).

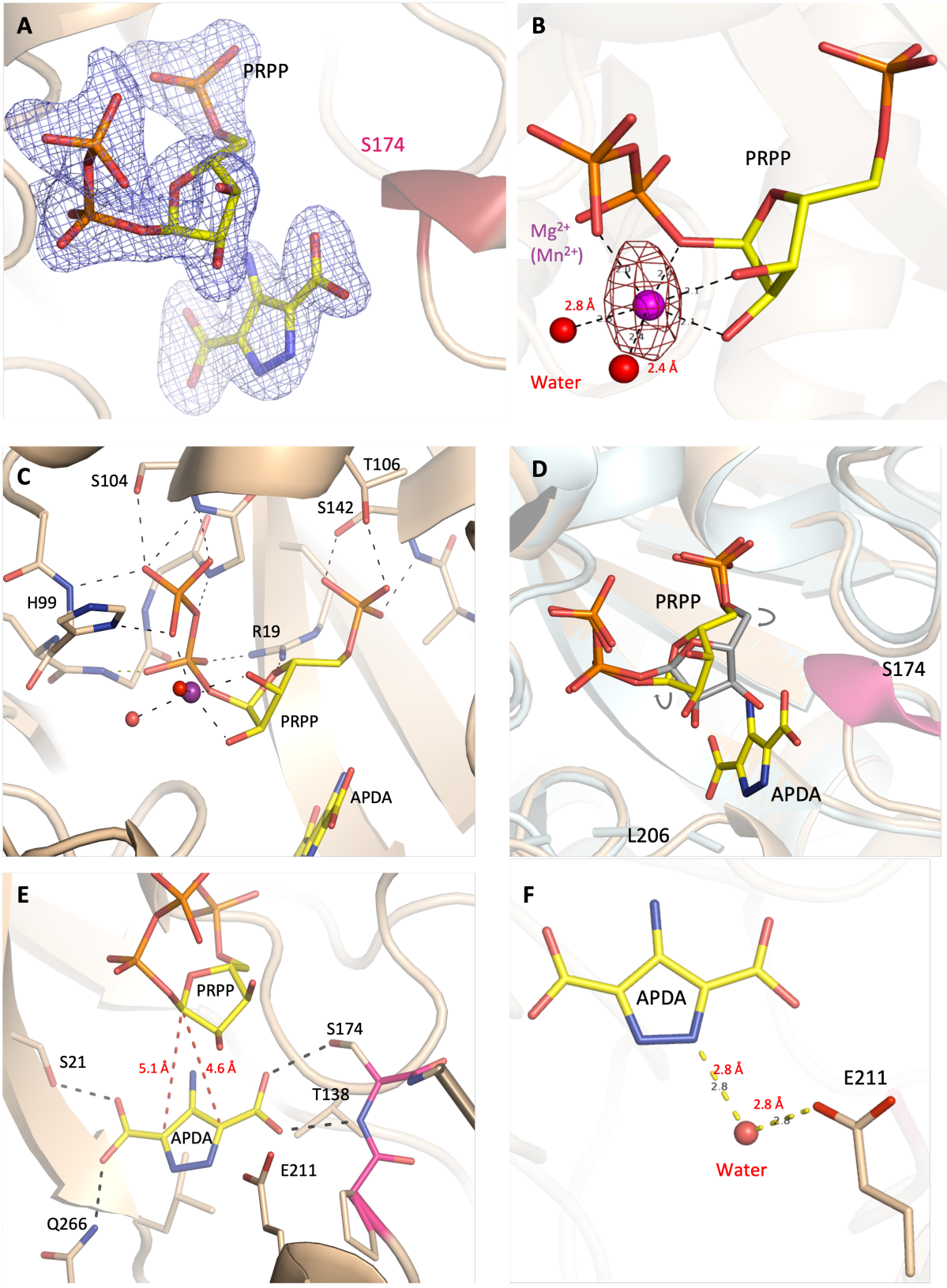



**Figure 3.12.** Structural analysis figures of ForT-PRPP-APDA complex adapted from[219]. (A) $F_o$-$F_c$ omit electron density map contoured at 3σ unambiguously locates PRPP and APDA in the

T138V protein structure. (B) The anomalous Fourier map contoured at 3σ collected at 9000 eV on $Mn^{2+}$ soaked crystals. The map confirms the location of $Mg^{2+}$ and its coordination to PRPP and two water molecules, which are distanced at 2.8 Å and 2.4 Å. (C) PRPP is anchored to the protein by the interactions of the phosphate groups. (D) Superposition with the WT ForT/PRPP binary complex (RCSB 6YQQ)[36] with carbon atoms coloured in grey showing that the ribose ring has rotated (denoted by arrows) around the two phosphate tethers. The loops at Ser 174 and Leu 206 are disordered in the binary complex but become ordered by APDA binding. (E) One carboxylate group of APDA interacts with Ser21 and Gln266, whilst the other carboxylate binds to Ser174 thereby ordering the loop. The two possible carbon atoms which could form the C-C bond are shown linked to the C1´ of ribose of PRPP. (F) The NH group on pyrazole ring is connect with Glu211 via a water molecule, distanced at 2.8 Å. (Figure modified from [219])

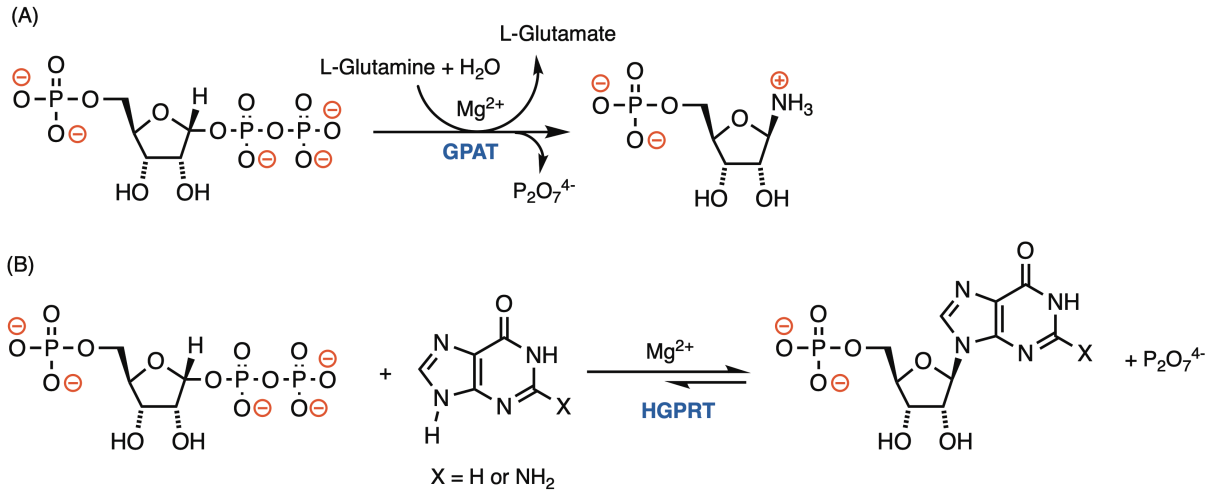



**Figure 3.13.** PRPP-dependent enzymes. (A) Glutamine PRPP amidotransferase.[225] (B) Hypoxanthine-guanine phosphoribosyltransferase.[226] (Figure modified from [219])

The overall protein structure in this ternary complex is little changed from that observed for the enzyme in the WT ForT/PRPP binary complex,[170] with an RMSD of 0.6 Å over 320 overlapping $C_\alpha$ atoms. The interactions of PRPP with the T138V variant, which are dominated by the 5'-phosphate and the pyrophosphate, are essentially unchanged from those in the WT ForT/PRPP binary complex, and the phosphate moieties are found in almost identical positions in the two structures (Figure 3.12D). The bound APDA, however, partly occupies space in which ribose was located in the WT ForT/PRPP binary complex.[170] As a result, the PRPP ribose ring in the ternary complex adopts a C2´endo conformer and rotates through an angle of 40° (Figure 3.12D). The $Mg^{2+}$ coordination environment depends on the position of the ribose ring (Figure 3.12E) and may only be fully ordered in the ternary complex. The planes of the APDA and ribose ring are approximately parallel but do not overlap, thereby positioning the $\pi$-system of the pyrazole for nucleophilic attack at C1 of PRPP, and the Val138 side chain adopts the same position and conformation as that seen for Thr138 in the WT ForT/PRPP binary complex.[170]

APDA makes a number of hydrogen bonding interactions that in agreement with the biophysical data obtained for several of our ForT variants (Table 3.2). One of the carboxylate groups of APDA hydrogen bonds to the backbone NH and side chain of Ser174, which orders residues 172-179 of the protein in the T138V/APDA/$Mg^{2+}$/PRPP ternary complex (Figure 3.12e) and positions the adjacent pyrazole carbon 4.6 Å from C1 of PRPP. This carboxylate also lies at a distance of 4.6 Å from the positively charged side chain of Lys263, which is bound to both Gln266 and a network of water molecules that engage the pyrazole carboxylate and Ser21. We note that replacing Lys263 by alanine gives a ForT variant that has low activity and reduced binding affinity for both PRPP and APDA. The other carboxylate of APDA hydrogen bonds to the side chains of Ser21 and Gln266, as well as two water molecules that form a network with the side chain of Glu211. Not only do these interactions order the loop at Leu206, which is connected to Glu211 and disordered in the WT ForT/PRPP binary complex (Figure 3.3),[170] but they also position the adjacent pyrazole carbon 5.1 Å from C1 of PRPP. We note that the S21A variant exhibits 10-fold and 100-fold reductions in $k_{cat}$ and binding affinity for APDA, respectively (Table 3.2). The Leu24 side chain is stacked against one face of the aromatic pyrazole ring. Surprisingly the extracyclic amine of APDA, which points towards Ile37 and Ser139, does not form any hydrogen bonds with the protein. The pyrazole NH, however,

appears to hydrogen bond to a water molecule that is positioned by a second hydrogen bond to the Glu211 side chain.

Despite the T138 variant displaying activity, it exhibited a 40-fold reduction in $k_{cat}$, poses the question as to why we are able to trap the T138V/APDA/$Mg^{2+}$/PRPP ternary complex. It is possible that the answer lies in the high concentration (100 mM of both substrates) in soaking solutions (Figure 3.9) immediately before freezing. This, coupled with the substantially decreased activity of T138V, then led to our ability to trap the ternary complex (Figure 3.12a).

**Table 3.3.** Data collection and refinement statistics.

| | **L24A/CAPR/PPi** | **T138V/APDA/Mg/PRPP** | **T138V/APDA/Mn/PRPP** |
|---|---|---|---|
| Beamline | I03 (DLS) | I03 (DLS) | I03 (DLS) |
| Wavelength (Å) | 1.3776 | 0.9762 | 1.3776 |
| Space group | $P4_12_12$ | $C222_1$ | $C222_1$ |
| a, b, c (Å) | 78.9, 78.9, 111.0 | 66.3, 139.0, 109.6 | 66.9, 138.7, 109.9 |
| α, β, γ (°) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| Resolution range (Å) | 64.3- 1.9 (1.94 - 1.90) | 60 - 1.60 (1.63 - 1.60) | 53 - 1.57 (1.60 - 1.57) |
| $R_{meas}$ | 0.106 (2.6) | 0.065 (3.7) | 0.119 (2.5) |
| I/σ(I) | 16 (0.6) | 15 (0.7) | 11 (0.4) |
| Completeness (%) | 100 (100) | 100(100) | 97 (77) |
| cc1/2 | 1 (0.9) | 1 (0.8) | 1 (0.4) |
| Multiplicity | 23 (25) | 13 (13) | 11(5) |
| No. reflections | 28,350 (1,798) | 63,680 (4,624) | 756,217 (26,285) |
| $R_{work}$ | 0.201 | 0.187 | |
| $R_{free}$ | 0.27 | 0.213 | |
| <u>Number of atoms</u> | | | |
| Total | 2661 | 2906 | |
| Protein | 2468 | 2551 | |
| Ligands / Ions | 31 | 34 | |
| Waters | 110 | 309 | |
| Buffer | 52 | 12 | |
| <u>B-factor</u> | | | |
| Protein | 52 | 32 | |
| Ligands / Ions | 56 | 25 | |
| Waters | 50 | 40 | |
| Buffer | 62 | 49 | |
| Bond lengths (Å) | 0.008 | 0.008 | |
| Bond angles (°) | 1.52 | 1.43 | |
| Ramachandran plot favoured / outlier (%) | 97 / 1 | 98/1 | |
| Molprobity centile | 98 | 98 | |

We were also able to observe structural changes that take place as a result of catalysis by solving the X-ray crystal structure of the L24A/CAPR/$PP_i$ product complex (Figure 3.14) to 1.75 Å resolution in space group $P4_12_12$ (Table 3.3). The complex was obtained by incubating the enzyme in the presence of the substrates prior to crystallisation and without soaking before data collection. Clear difference electron density was observed for both products in the active site (Figure 3.14a). The protein structure in the product complex is almost identical to that seen in the WT ForT/PRPP binary complex (RMSD of 0.41 Å over 319 overlapping $C_\alpha$ atoms),[170] including the disordered loop at Ser174 (Figure 3.14b). The location of the pyrophosphate molecule and its interactions with the protein are identical to those seen for the pyrophosphate group of PRPP in the both the ternary and binary complexes [170](Figure 3.14c). The 5´-phosphate group of CAPR is positioned identically to that seen in the PRPP complexes. The CAPR ribose ring in a position intermediate between those seen in the ternary and binary complexes. As a result of the change, the 2´-OH and 3´-OH groups hydrogen bond to the Thr138 side chain and a water molecule, respectively (Figure 3.14d). As a result, $Mg^{2+}$ binding site has been eliminated and no ion was observed. The single remaining carboxylate group of CAPR hydrogen bonds to the side chains of Ser21 and Gln266, in a similar manner to that seen for this group in the T138V/APDA/$Mg^{2+}$/PRPP ternary complex. However, the carboxylate atoms are shifted between 0.6 and 1.2 Å. Despite the difference electron density being unambiguous for the presence of the CAPR in the structure, the precise orientation of the heterocyclic ring was not as clear as for the T138V/APDA/$Mg^{2+}$/PRPP ternary complex. We therefore modelled the pyrazole ring in same orientation as in the ternary complex and in the alternate “flipped” orientation with L and R conformation (Figure 3.14E). Both models fit the electron density well, with similar R value of 17.75% and 17.73% for L and R conformation respectively. The resulting maps led us to fit the latter orientation in which the pyrazole nitrogen point towards Ile37 and Ser139 (but do not make hydrogen bonds) and the extracyclic amine is hydrogen bonded to water. Both orientations are plausible.

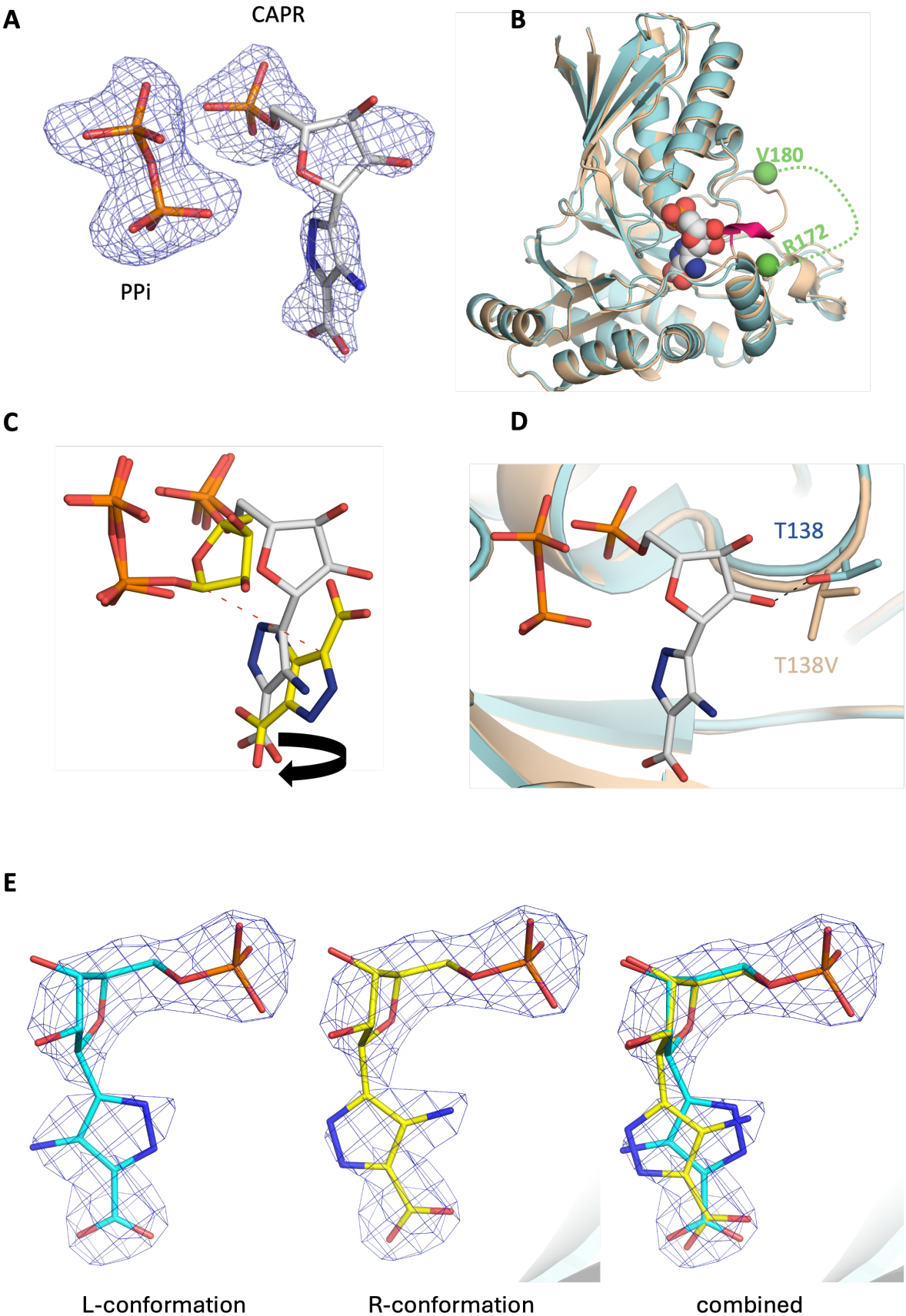

A
CAPR
PPi
B
V180
R172
C
D
T138
T138V
E
L-conformation
R-conformation
combined

**Figure 3.14.** (A) $F_o$ - $F_c$ omit electron density map contoured at 2.5σ identifies CAPR bound to the L24A protein structure. Carbon atoms are coloured white, oxygen atoms coloured by red, phosphorous atoms orange and nitrogen atoms blue. (B) The loop from R172 to V180 is disordered in the L24A/CAPR/PPi ternary complex. The CAPR molecule is shown as spheres. (C) Superposition of the protein complexes of substrates and product, reveals that the phosphate groups share identical positions. The aromatic rings of APDA and CAPR in the two X-ray crystal structures occupy a similar volume but their planes are flipped with respect to each other. It is the ribose ring which moves during catalysis. (D) Thr138 in the product complex alters its conformation from that seen in the T138V structure to make a hydrogen bond with the ribose. (E) Two different possible orientation for product CAPR in T138V product complex structure. Both models fit the electron density well, with similar R value of 17.75% and 17.73% for L and R conformation respectively. (Figure modified from [219])

### 3.3.3. Computational Modelling by Collaborators

Access to the structure of the T138V/APDA/$Mg^{2+}$/PRPP ternary complex permitted us to use quantum chemical (QC) calculations to obtain an active site "cluster" model[228] of the pre-decarboxylation intermediate proposed to link the WT ForT/$Mg^{2+}$/PRPP/APDA Michaelis complex with the WT ForT/CAPR/$PP_i$ product complex. To help resolve the ambiguity of the pyrazole orientation in the active site and the location of the pyrazole NH group, four cluster models of the Michaelis complex (**O1**-**O4**) (Figure 3.15a) were built and optimized using a localized, semi-empirical description[229] of an active site model after *in silico* reversion of Val138 to a threonine residue. All experimentally observed hydrogen bonds between the protein and both APDA carboxylates were maintained in each of the optimized cluster models of the Michaelis complex. These structures were then used to obtain cluster models of the four possible pre-decarboxylation intermediates (**P1**-**P4**) (Figure 3.15a) by systematically decreasing the distance between C3 (APDA) and C1' (PRPP) and optimizing the resulting models. As all other degrees of freedom were allowed to relax in the active site, the ribose-pyrophosphate (C1-O) bond lengthened as the pyrazole-ribose (C3-C1) bond was constrained to smaller lengths. This procedure gave optimized structures in which the C1'-O bond was fully broken (2.9 Å), corresponding to PPi release, when the C3-C1' bond reached its optimum

length. The heats of formation for all eight models were then computed using the PM7 parameterization[230] to give qualitative estimates of the activation energy barriers for this step. Assuming that each pre-decarboxylation intermediate resembles the transition state for this endothermic step, the lowest activation energy calculated for forming the pre-decarboxylation intermediate is when APDA adopts the orientation and tautomer (**O4**) seen in the T138V/APDA/$Mg^{2+}$/PRPP ternary complex (Table 3.4).

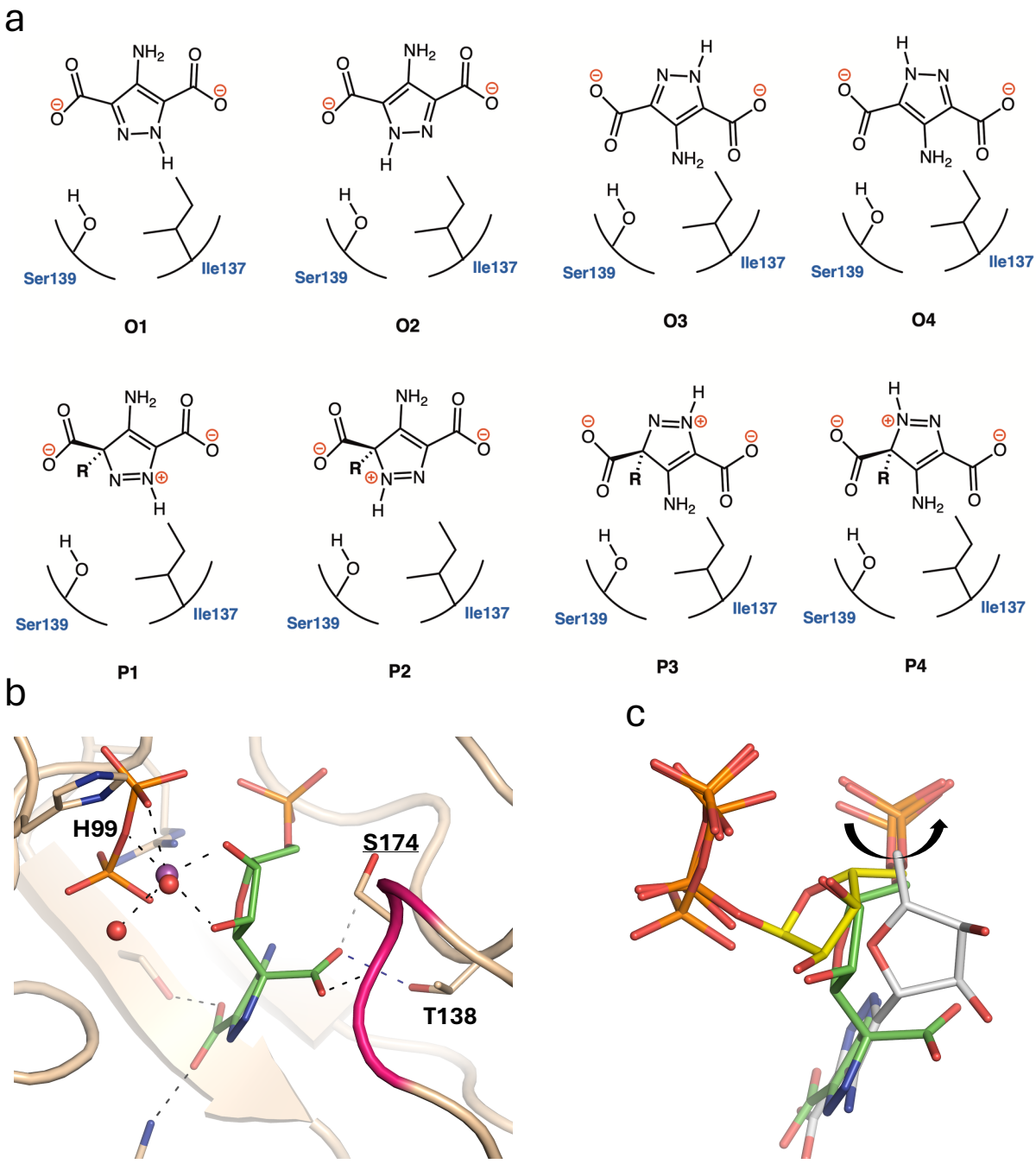



**Figure 3.15.** (a) Schematic representations of the cluster models used in the QC calculations. We considered both tautomers as well as the flipped orientation of the heterocycle. (b) The calculated intermediate retains $Mg^{2+}$ ion binding and is attached to the protein using the 5'-phosphate and retained carboxylate. Carbon atoms are coloured yellow, oxygen atoms coloured red, phosphorous atoms orange and nitrogen atoms blue. (c) Superposition using the protein atoms of the complex with substrates (yellow), calculated inter-mediate (green) and CAPR (grey). The ribose pivots (shown as black arrow) around the 5'-phosphate during the reaction.

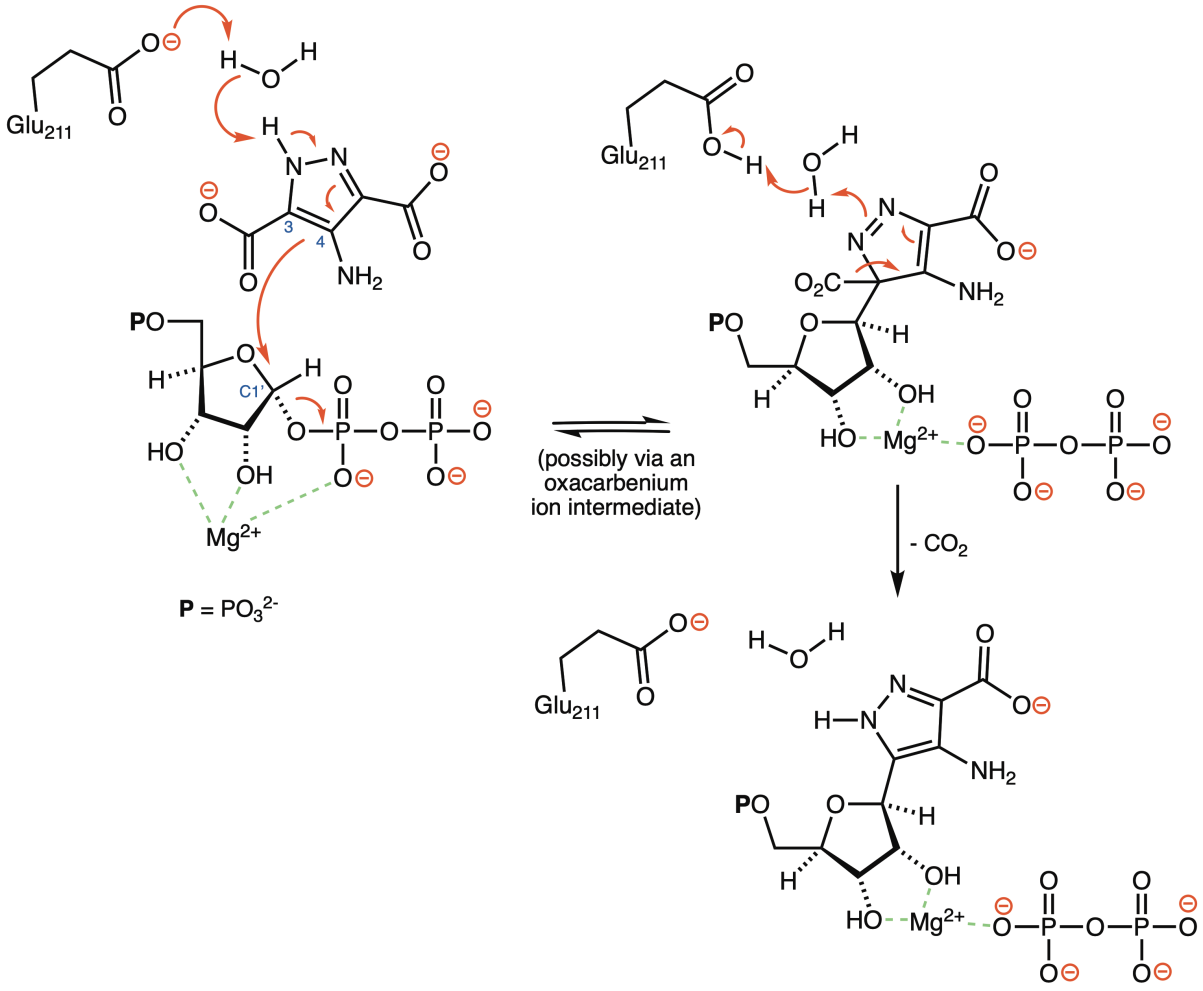



**Figure 3.16.** Proposed catalytic mechanism for ForT-catalysed C-C bond formation. (Figure modified from [219])

**Table 3.4.** Calculated heats of formation (kcal/mol) for the active site cluster models of the WT ForT/MgPRPP/APDA ternary complex (**O1**-**O4**) and their associated pre-decarboxylation intermediates (**P1**-**P4**). (Table modified from [219])

| | **O1** | **P1** | **O2** | **P2** | **O3** | **P3** | **O4** | **P4** |
|---|---|---|---|---|---|---|---|---|
| $\Delta H_f$ (kcal/mol) | -17175 | -17126 | -17173 | -17138 | -17168 | -17128 | -17171 | -17145 |
| $E_a$ (kcal/mol)[a] | | 49 | | 35 | | 40 | | 26 |

[a]Estimates of the activation energies are a lower bound and assume that the transition state for this endothermic step closely resembles the intermediate.

The optimized computational model of the pre-decarboxylation intermediate (**P4**) shows the reorientation of the departing carboxylate due to rehybridization of the adjacent carbon due to C-C bond formation between C1' of the ribose and the pyrazole carbon. As a result, the carboxylate lies within hydrogen bonding distance of the Thr138 side chain and the backbone NH of Ser174. The lack of interaction with the Ser174 hydroxyl is consistent with the catalytic properties of the S174A variant (Table 3.2). These QC calculations support the hypothesis that the T138 side chain can stabilize the intermediate (and, by the Hammond postulate [231], the transition state for C-C bond formation) by hydrogen bonding. Experimentally, removing this potential hydrogen bond leads to a 40-fold reduction in $k_{cat}$ for the T138V variant relative to ForT WT. On the other hand, the corresponding reduction is only 3-fold for the T138A variant (Table 3.2). We suggest that a water molecule could enter the active site and "rescue" the activity of the T138A variant by supplying a hydrogen bond to the carboxylate, an idea for which there is ample precedent.[232] A water that bridges the pyrazole NH and the Glu211 is retained in the optimized structure, and the pyrophosphate leaving group is located within the binding pocket seen in the crystal structure of the L24A/CAPR/PPi product complex (Figures 3.15b and 3.15c).

### 3.3.4. Substrate Scope of WT ForT

We assayed a series of pyrazoles as ForT substrates in which the extracyclic amine was replaced by hydrogen (PDA), nitro- (NPDA), methyl- (MPDA) and iodo- (IPDA) groups (Figure 3.17a and Table 3.5). None of these pyrazoledicarboxylates proved to be substrates for the enzyme in our LC-MS assay (data from Dr. Gerogina Girt not shown here). We also confirmed the literature finding[217] that 4-hydroxy-*1H*-pyrazole-3,5-dicarboxylic acid (HPDA) is not a substrate for the ForT-catalysed coupling reaction even though HPDA binds to the ForT/PRPP complex with a $K_D$ value of 37 μM (Figure 3.19). The monocarboxylate 4-amino-*1H*-pyrazole-3-carboxylate (APCA) was coupled to PRPP by ForT to give two products, albeit in less than 5% overall yield (Figure 3.17b).[217] 3-amino-*1H*-pyrazole-4-carboxylate (*iso*-APCA) and $Mg^{2+}$/PRPP were incubated with ForT and LC-MS analysis showed that a new product had been formed with a molecular mass of 339.1. This mass is consistent with ForT-catalysed nucleotide formation where the carboxylate substituent on the pyrazole ring. The $^{1}H$ NMR showed signals for an aromatic CH group and the hydrogen connected to the ribose C1', which was shifted downfield (5.74 ppm) relative to the signal for the cognate proton in the C-nucleotide CAPR (4.91 ppm) (Figure 3.18). We reasoned that the product was one of two possible C-N nucleotides. The $^{13}C$-$^{1}H$ HMBC NMR spectrum shows an interaction between C5 of the pyrazole ring and C1'-H, as well as the absence of interactions between C3 and C1'-H. These data are most consistent with the product being 5-amino-1-(β-D-ribofuranosyl) pyrazole-4-carboxylate 5'-monophosphate (iso-CAPR) (Figure 3.17c).[222]

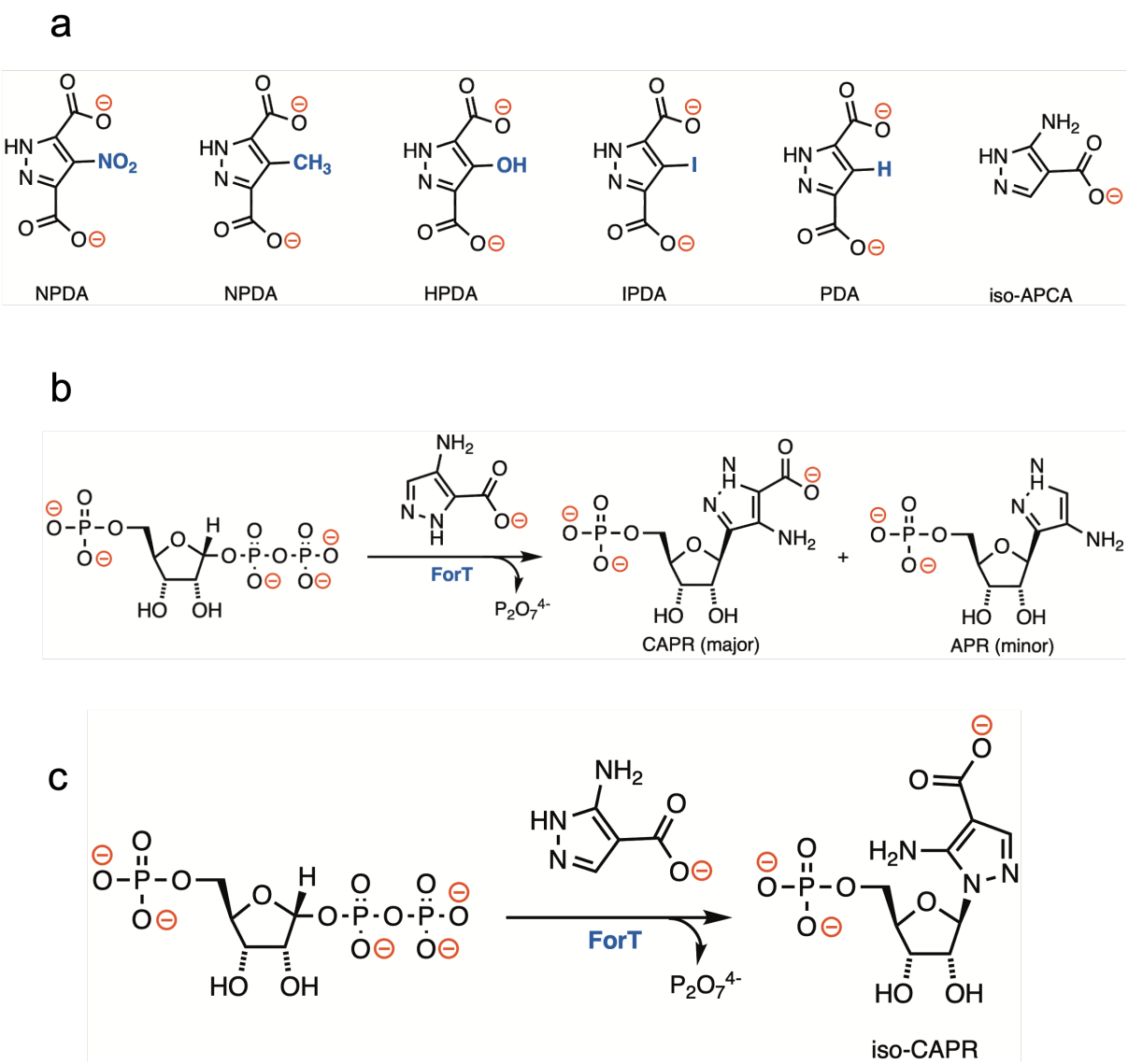



**Figure 3.17.** (a) Functionalized pyrazole-dicarboxylates used to explore the substrate scope of ForT. (b) ForT-catalysed reaction when APCA is the substrate.[217] (c) ForT-catalysed reaction when *iso*-APCA is the substrate. (Figure modified from [219])

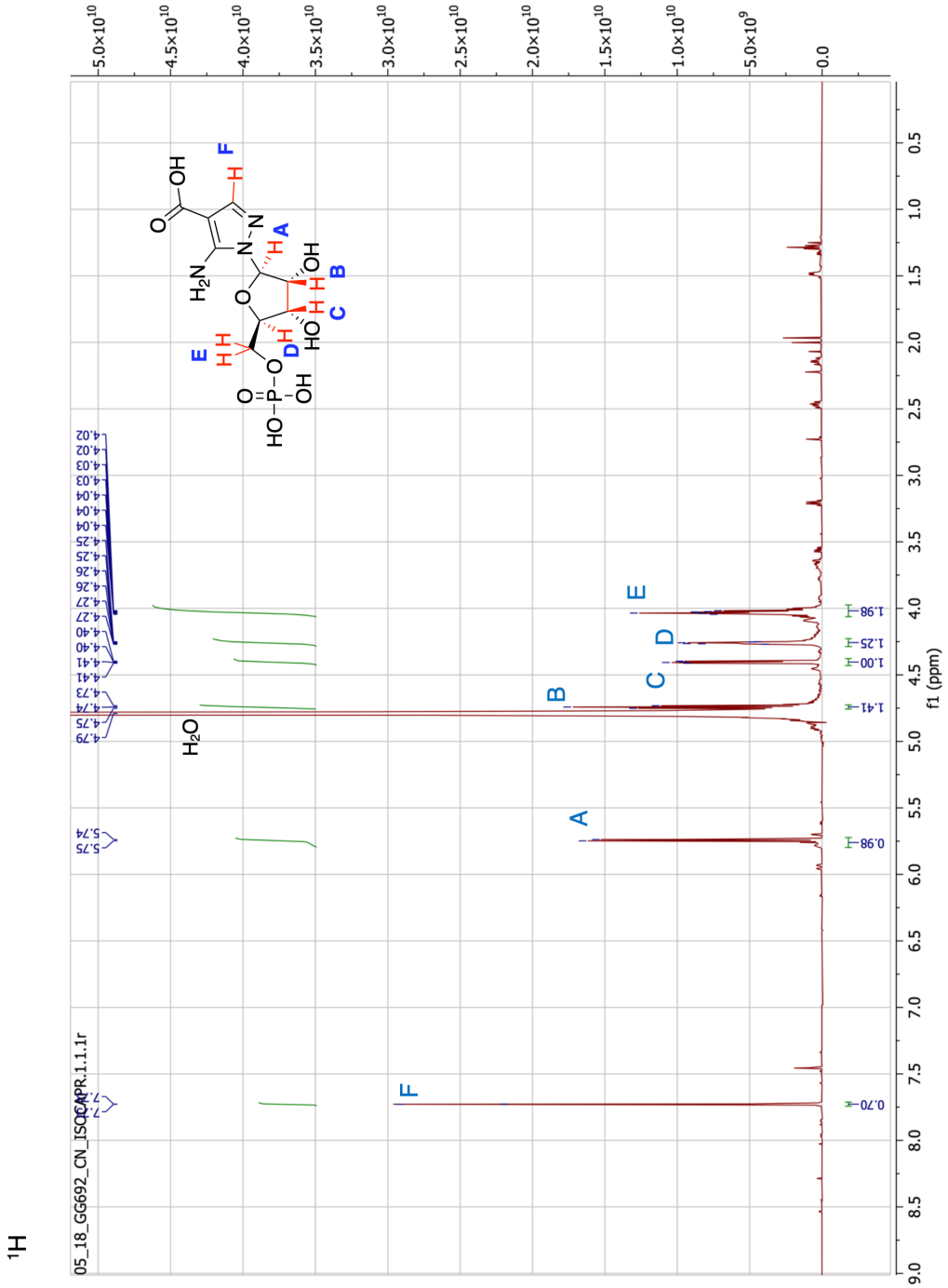



**Figure 3.18.** $^{1}$H-NMR (600 MHz) spectrum of the purified product formed (*iso*-CAPR) by incubating Mg-PRPP and *iso*-APCA in the presence of WT ForT. (Data from Dr. Georgina Girt)

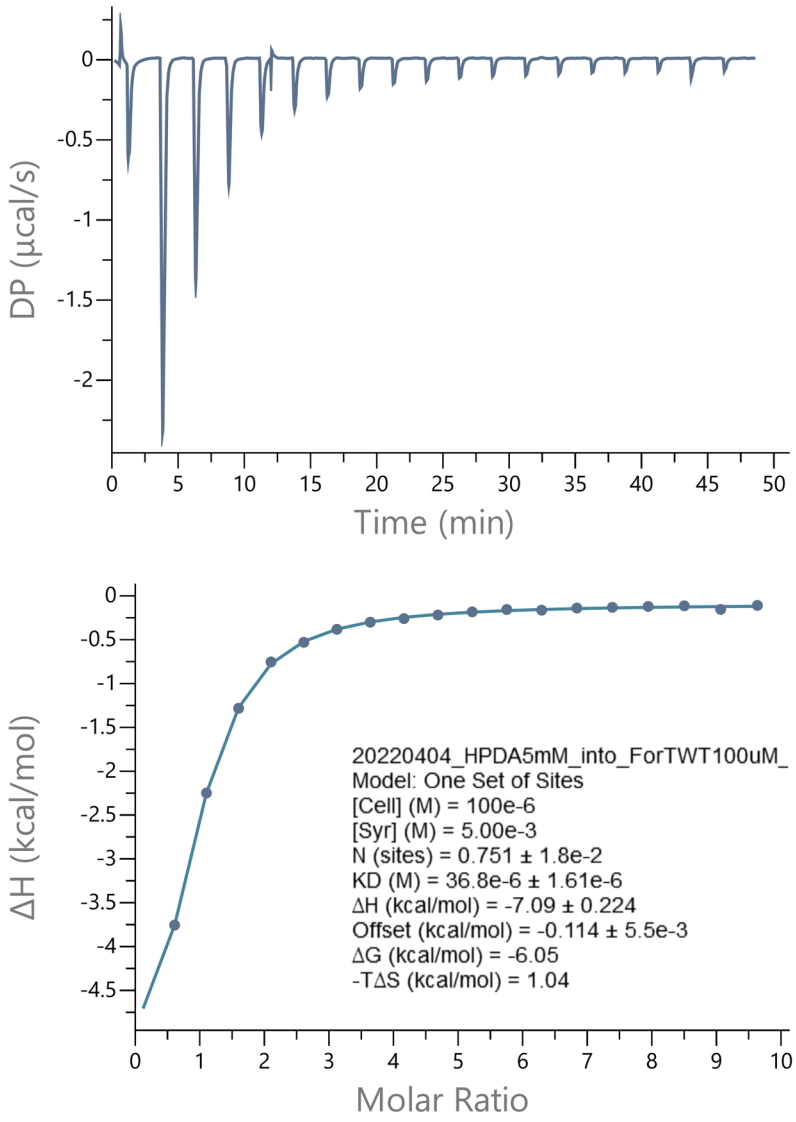
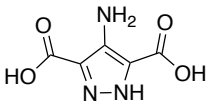
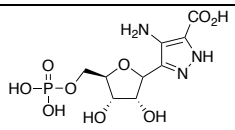
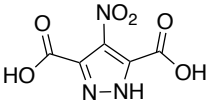
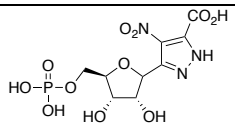
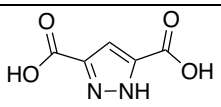
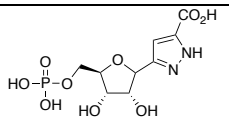
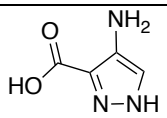
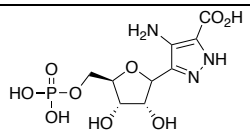
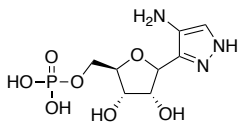
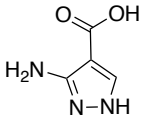
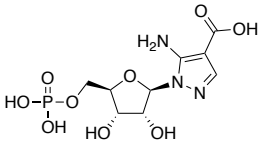
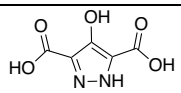
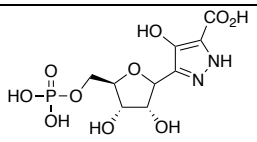
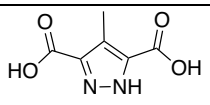
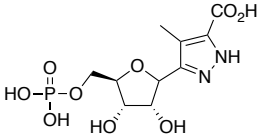
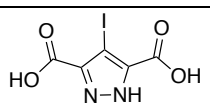
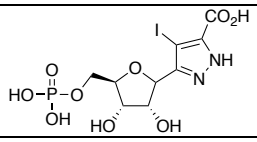



**Figure 3.19.** ITC multiple injection experiment. Injecting HPDA into WT ForT. Methods described in section 3.2.3. HPDA has $K_D$ with WT ForT at 36.8 ± 1.6 µM with the N (site) = 0.75 ± 0.01. Experimental data were fitted by suing one set of sites model from MicroCal PEAQ-ITC analysis software.

**Table 3.5.** Substrate scope of ForT. Substrates (5 mM) were typically incubated with WT ForT (500 nM) in 100 mM sodium phosphate buffer, pH 7.5, containing PRPP (1 mM), $MgCl_2$ (1 mM) and 100 mM NaCl. Reaction progression was monitored by LC-MS (m/z charge, UV 254nm and 280 nm).

| Short name | Full name | Structure | Product short name | Full name | Product structure(s) | Product formation |
|---|---|---|---|---|---|---|
| APDA | 4-amino-1*H*-pyrazole-3,5-dicarboxylic acid |  | CAPR | 4-Amino-5(3)-(β-D-ribofuranosyl)pyrazole-3(5)-carboxylate-5'-monophosphate |  | Yes.[217] |
| NPDA | 4-nitro-1*H*-pyrazole-3,5-dicarboxylic acid |  | CNPR | 4-Nitro-5(3)-(β-D-ribofuranosyl)pyrazole-3(5)-carboxylate-5'-monophosphate |  | no |
| PDA | *1H*-pyrazole-3,5-dicarboxylic acid |  | CPR | 5(3)-(β-D-ribofuranosyl)pyrazole-3(5)-carboxylate-5'-monophosphate |  | no |
| APCA | 4-amino-1*H*-pyrazole-3-carboxylic acid |  | CAPR<br><br><br><br>APR | 4-Amino-5(3)-(β-D-ribofuranosyl)pyrazole-3(5)-carboxylate-5'-monophosphate<br><br>4-Amino-5(3)-(β-D-ribofuranosyl)pyrazole-5'-monophosphate | <br> | CAPR major, APR minor, determined in previous work.[217] |
| Iso-APCA | 3-amino-1*H*-pyrazole-4-carboxylic acid |  | Iso-CAPR | 5-Amino-1-( β -D-ribofuranosyl)pyrazole-4-carboxylate 5'-monophosphate |  | Yes |
| HPDA | 4-hydroxy-1*H*-pyrazole-3,5-dicarboxylic acid |  | CHPR | 4-Hydroxy-5(3)-(β-D-ribofuranosyl)pyrazole-3(5)-carboxylate-5'-monophosphate |  | no |
| MPDA | 4-methyl-1*H*-pyrazole-3,5-dicarboxylic acid |  | CMPR | 4-Methyl-5(3)-(β-D-ribofuranosyl)pyrazole-3(5)-carboxylate-5'-monophosphate |  | no |
| IPDA | 4-iodo-1*H*-pyrazole-3,5-dicarboxylic acid |  | CIPR | 4-Iodo-5(3)-(β-D-ribofuranosyl)pyrazole-3(5)-carboxylate-5'-monophosphate |  | no |

## 3.4. Discussion

The transformation catalysed by ForT is of chemical interest for several reasons. First, it is one of only a few examples in which C-C bond formation takes place via an electrophilic substitution of a small, aromatic heterocycle. Second, the reaction is regioselective despite the heteroaromatic substrate having multiple reactive nitrogen atoms that could react with the electrophilic carbon of PRPP. Third, ForT can discriminate between the aminopyrazoledicarboxylic acid and its analogue in which the amine is replaced by a hydroxyl group; a remarkable feat given the steric and electronic similarities of the two molecules. The overall reaction is rendered essentially irreversible by coupling C-C bond formation to decarboxylation and PPi formation.

ITC measurements reveal that (experiment in absence of $Mg^{2+}$) APDA binds to ForT only in the presence of PRPP (Figure 3.5). The binding data suggest that reaction likely follows an ordered bi-bi kinetic mechanism in which PRPP binds first and creates the binding site for APDA. Structural support for this hypothesis is provided by the observation that the bound APDA molecule blocks PRPP from solvent in the T138V ternary complex (Figure 3.12e). In addition, APDA binding orders a loop due to main and side chain interactions of Ser174 with the carboxylate group of the substrate. This newly ordered loop encloses the active site. This loop is disordered in the ForT/PRPP binary complex[170] and in the crystal structure of the L24A/CAPR/PPi "product" complex (Figure 3.14b). In both these complexes, one of the APDA carboxylate groups is absent. The missing carboxylate makes multiple hydrogen bonds to this loop. We propose this loop acts a gate keeper for substrate entry into, and product release from, the active site. Deleting the side chain of Ser174 increases $K_M$ for APDA only five-fold, leading to the conclusion that is the hydrogen bond to the backbone amide of Ser174 is critical for substrate recognition and loop organization.

The structural data allow the rationalisation of the regioselectivity of the reaction where one of the two possible carboxylate groups is lost. One carbon atom in the APDA ring is positioned 4.6 Å from C1' of PRPP with the correct orbital alignment for electrophilic substitution (Figure 3.12e) and this is connected to the carboxylate which binds (and orders) the loop at Ser174.

It is this carboxylate that is lost during the reaction. Another key observation from both the structural data and calculated model of the pre-decarboxylation intermediate is that the phosphate (PRPP, CAPR) and the pyrophosphate groups (PRPP, PPi) essentially stay fixed in position throughout the entire catalytic cycle (Figure 3.14c). In a similar manner the position of the "retained" carboxylate and the plane of the heterocyclic ring of APDA remain the same in the model of the pre-decarboxylation intermediate **P4** and product (modelled) complexes (Figures 3.12c and 3.15b). The $\pi$-system of the aromatic heterocycle interacts with the methyl substituents of the L24 side chain (Figure 3.12e) which with the carboxylate interactions we propose fixes the ring in place and thus ready for the reaction. The interaction may also increase the electron density of the $\pi$-system. Deletion of this this side chain in the L24A variant, reduces both the affinity of the enzyme for APDA, and catalytic efficiency by five-fold (Table 3.2). We take this as support for the import role L24 plays in positioning the substrate.

The retained carboxylate that is present in APDA, the pre-decarboxylation intermediate and CAPR, is anchored to the protein by direct contacts to Ser21 and Gln266 (Figures 3.12e and 3.15b). Mutations of Ser21 and Lys263 (part of a wider network of water molecules) profoundly weaken binding and catalysis consistent with this role. The mutation Gln266 leads to an unstable protein that cannot be kinetically characterized (Table 3.2). Compared to the binary complex,[170] the ribose ring in the T138V ternary complex "swings" around the fixed points of the phosphate and pyrophosphate groups. The consequence is that APDA binding pocket is enlarged by the ring movement (Figure 3.12d). The ribose ring has adopted a higher energy C2´-endo conformation (Figures 3.12a and 3.12d). It is this configuration that positions the O2´ and O3´ atoms in such a way that along with the pyrophosphate group they form a binding site for the $Mg^{2+}$ ion (Figure 3.12b). Two water molecules complete the $Mg^{2+}$ coordination sphere (Figure 3.12b). We proposed that the catalytically essential $Mg^{2+}$ (or $Mn^{2+}$) ion, most likely helps to stabilizes the higher energy C2´ endo conformation. The higher energy configuration would seem likely to make PRPP more reactive. This arrangement and role of $Mg^{2+}$ was first seen in HGPRT.[226] The ligated water molecules are part of a network that involves Glu211 (Figure 3.12c).

Combining the structural data from the ForT/PRPP binary,[170] the T138V ternary complex, the model structure of the pre-decarboxylation intermediate, and the L24A product complex

provides an insight picture of the substrate and protein motions that take place during catalysis. Our analysis shows that it is the ribose and thus C1´ atom that moves with very little change in the other atoms (C1´ of APDA, oxygen of the pyrophosphate) that are involved in bond breaking or bond making. The change from the ternary complex to the proposed intermediate requires breaking of the ribose-pyrophosphate (C1´-O) bond. This results in the "untethering" of the ring from one the fix points. This permits the ribose to move in way which rotates around the 5´ phosphate towards APDA (Figure 3.14c). In the intermediate the ribose ring has become almost planar; it has lost the C2´ endo pucker. The model suggests that the intermediate retains the $Mg^{2+}$ ion binding site (Figure 3.15b). The model also suggests that the carboxylate group that ultimately becomes $CO_2$, forms a hydrogen bond to the side chain of Thr138 (Figure 3.15b). We propose that the intermediate undergoes rapid decarboxylation, this allows the ribose to continue its motion towards the APDA. The ribose having moved positions the O2´ such that it could form a hydrogen bond to Thr138 (Figure 3.14d). The trajectory of the reaction can therefore be thought of as the movement of ribose which is at all times constrained by two fixed points. At the start of the reaction these are the 5´-phosphate and the pyrophosphate (in PRPP). As the pyrophosphate bond breaks and the new C-C bond forms, the fixed points become the retained carboxylate and the 5´-phosphate (Figure 3.15c).

An important remaining mechanistic question concerns the precise order of the chemical steps; breakage of the C1'-O bond, decarboxylation, and formation of the new C-C bond. Were the C1'-O bond to break as an isolated chemical step thus yielding an extremely reactive oxocarbenium ion, we would expect some evidence of quenching by hydrolysis. This is because the crystal structure of active site of the enzyme reveals the presence of water molecules. We have bot observed any such quenched product (ribose 5´-phosphate). Further we observed well-ordered PRPP in both binary and ternary complexes, indicating that enzyme itself does not break the bond; rather the breaking of the bond requires chemistry of APDA.

APDA is itself stable in under the pH range of our assay conditions, the calculated intermediate at the active site shows no evidence for decarboxylation. Moreover, the extra cyclic hydroxyl, nitro and methyl substrate variants are inactive. This suggested to us that decarboxylation to form a full anion is unlikely to be an isolated first chemical step. The QC

calculations show the active site side chain residues to be unchanged by formation of the pre-carboxylation intermediate **P4**. This is most consistent with decarboxylation occurring after C-C bond formation (Figure 3.15b).

The sequence of events we propose is the binding of PRPP, followed by binding of APDA which orders a loop in the protein. In order to accommodate the APDA, PRPP has to move and since it is fixed by the phosphates, the ribose ring swings deeper in a pocket. As a consequence, the ring adopts a higher energy state which is stabilised by $Mg^{2+}$ (Figure 3.15c). Next step, the C-C bond formation operates in an $S_N2$ like manner where both bond breaking (C1'-O bond) and bond making (C1'-C bond) occur at the same time. The ribose rings pivots around the C5' towards the APDA ring to form the pre-decarboxylation intermediate (Figure 3.15c). The formal positive charge in the heterocyclic ring renders APDA unstable and the drive to regain aromaticity triggers the decarboxylation (release of $CO_2$). The loss of the carboxylate breaks the interactions with the loop at Ser 174 disordering it. The disordered loop permits the CAPR molecule to diffuse out of the active site. Only then would it be possible for the pyrophosphate to leave, typically it does so with $Mg^{2+}$, although we see no evidence in our structure it is possible that a $Mg^{2+}$ binds during dissociation [233]. Until PPi has left the active site apparent $K_M$ for APDA is smaller than its $K_D$ value is evidence that product release is slow relative to substrate binding [234]. The only residue which appears to absolutely essential for catalytic activity is Glu211, but ITC shows that residue is not required for binding of either substrate (Table 3.2). In the ternary complex we observed that by a bridging water molecule linked Glu211 to the one of the ring nitrogen atoms (Figure 3.13F). Taking together our experimental and computational data, we propose a mechanism where Glu211 acts as a general base via a water relay to activate the APDA molecule (Figure 3.15d). The removal of the proton would increase the electron density within the APDA molecule and thus active it as a nucleophile.

That the pyrazole ring is apparently “flipped” around the new C-C bond in the L24A/CAPR/PPi product when compared to the ternary complex has to be explained. All other pyrazole/protein interactions are largely unchanged by changing the APDA orientation. The quasi two-fold symmetry of this portion of the molecule complicates the placement in more ambiguous electron density in the L24A/CAPR/PPi product structure. As a result of this flip,

the orientation of the amino group becomes opposite to what is seen in the T138V ternary complex. We considered whether this reorientation could take place as part of the catalytic cycle. We rejected this possibility as extremely unlikely. Our reasoning was based on the structural observation of the extensive hydrogen bonds of the retained carboxylate would have to be disrupted to allow rotation (Figure 3.12E). This would seem to present a very high activation energy barrier during the reaction. We then considered whether the mutations in the respective complex could have led to an artifact. The change T138V does not create any pocket or increase steric hinderance. In the product and modelled intermediate we do observe a hydrogen bond that does indicate the hydroxyl plays a catalytic role, but there is no obvious explanation as to why it would favour the wrong orientation of the substrate. The absence of the Leu24 side chain in the L24A mutant creates a pocket, whilst the loss of the hydroxyl in the T138V creates a different hydrogen bonding environment. Superimposing the CAPR into the T138V or the native binary complex, does result in a clash with Ile37[170]. This clash is avoided in the L24A variant because Ile37 adopts a new conformation, in this new conformation it fills the pocket created by the L24A substitution. Manual modelling, however, reveals that the potential clash ForT can be avoided by slightly adjusting the position of the pyrazole ring while apparently maintaining (i) all hydrogen bonds between the enzyme and the retained carboxylate, and (ii) the distance between the reactive carbon and C1' of ribose. Thus, this analysis was not definitive. We then considered whether the alternative orientation could plausibly be catalytic. When the APDA ring is flipped, the extracyclic amine points towards the water filled pocket and lies at a distance of only 4.5 Å from the carboxylate of Glu211. However, we could place a water molecule such that it bridges Glu211 to the extracyclic amine. In the structure we do not observe a water molecule, but this could reflect the lower resolution and generally higher flexibility in this structure. Alternatively, some slight adjustments would allow direct contact. Similar to the mechanism outlined above Glu211 could by removing a proton from the extracyclic amine to promote the reaction. However, the extracyclic amine is likely to be Sp2 hybridised in order to conjugate with the aromatic ring and therefore unlikely to protonated, lessening the case for the essential nature of Glu211.

We explored QC calculations to investigate how the two different ring orientations might impact the activation energy for forming the putative pre-decarboxylation intermediate. In

addition, we examined the effect of placing the pyrazole NH in both possible positions. Assuming that the transition state resembles the pre-decarboxylation intermediate, these calculations suggest that the lowest barrier is associated with APDA adopting the orientation observed in the T138V ternary complex. More detailed QM/MM studies may show a different result and we do not regard the issue as definitively settled.

Our structural surmise was that initial binding of the substrate would be driven by binding of the retained carboxylate to Ser21 and Gln266 the pre-ordered region of the binding site. We therefore expected that pyrazoles containing a single carboxylate would use the carboxylate to bind to this loop. If so, then then the compound would not undergo decarboxylation rather regeneration of the aromatic $\pi$-system would take place via loss of a proton from the APDA. To our surprise, the reaction of 4-amino-1*H*-pyrazole-3-carboxylic acid (APCA), produced *both* the expected CAPR and 4-amino-3-(β-D-ribofuranosyl)pyrazole 5'-monophosphate, a minor product that has been completed decarboxylated (Figure 3.17b and Table 3.5)[217]. This mixture of products could mean that the compound can bind such the carboxylate binds to the loop at Ser174 ordering it and then decarboxylating. The alternative explanation is that the carbon attached to carboxylate group close to Gln266 makes a nucleophilic attack. This carbon since it is attached to the carboxylate would be more reactive than the "correctly" positioned carbon. The structure shows the carbons are relatively close, this would be plausible (Figure 3.12E). The compound 3-amino-1*H*-pyrazole-4-carboxylic acid (*iso*-APCA) (Figure 3.17c) was chosen to test this hypothesis. If the single carboxylate binds in both possible orientations, a mixture of compounds would be expected. However, if the compound only binds to the fixed site (Gln 266 and Ser21), then only the C-N bond should form because the nitrogen atom would be expected to be much more reactive. Indeed, we observed the C-N nucleotide 5-amino-1-(β-D-ribofuranosyl) pyrazole-4-carboxylate 5'-monophosphate as the only product when PRPP and iso-APCA were incubated with ForT (Figure 3.17c and 3.18). The formation of this C-N nucleotide consistent with the carboxylate of *iso*-APCA binding to Ser21 and Gln266, thus positioning the nitrogen atom for nucleophilic attack at C1' of PRPP. We therefore concluded that simplest explanation for the formation of 4-amino-3-(β-D-ribofuranosyl) pyrazole 5'-monophosphate is that the carbon adjacent to the pyrazole carboxylate comes close enough to C1' to permit reaction.

# Chapter Four

# Time Resolution Approach

## 4.1. Introduction

Structural and functional studies that were carried out for ForT were described in previous chapters. These studies led us to propose that the enzyme operates via an $S_N2$ like mechanism during catalysis of C-C nucleoside bond formation. In this model, the APDA ring acts as the nucleophile displacing PPi to form pre-decarboxylation intermediate with C1' position of the PRPP. In an $S_N2$ mechanism the bond breaking (C1' – O) and bond making (C1' – C) occur simultaneously. The liberation of PPi supplies the energy for the formation of the intermediate. The subsequent decarboxylation event converts the intermediate to the product with the release of $CO_2$ as the irreversible step and driving force. The mechanism is based on inference from experiments and calculations. The goal of the work in this chapter is to build towards a direct experimental structural study of the mechanism. Since chemical steps occur at the ns timescale, such a time resolved study will require X-ray free electron laser. The key requirement for such a study is to "freeze frame" the enzyme during turnover. Since the X-ray diffraction produces an electron density map that is an average of all the unit cells in the crystals, it is crucially important that the molecules are at the same stage of the catalytic cycle. If not, the electron density map could reflect an average of different catalytic steps. (Of course, this is dependent on the timings of the steps in the mechanism.) We therefore reasoned that for a multi-step reaction, the most successful approach would be simultaneously activate the enzyme, the most practical route being laser uncaging of a

substrate. Photo caging of substrates is well-established [235]. We considered temperature and or pH jump [236,237] type experiments but could not identify any experimental routes. As a contingency plan we contemplated using rapid mixing [238,239].

Visual inspection of both ForT-PRPP-APDA and ForT-product complex [219], identifies that there is a large cavity around the ring. In the APDA complex the space is around ring nitrogen whilst in the product complex it is adjacent to the extracyclic nitrogen. The cavity connects to the bulk solvent (Figure 4.1). Simple modelling in COOT and PyMol suggested that there might be room for a phenyl group. A useful photo-caged compound needs to be able to bind to the enzyme in the same orientation as the substrate, to be resistant to catalysis and to be easily activated by a laser pulse. The photo cage group when activated converts the caged compound into the authentic substrate. The uncaging reaction must be (and usually is) faster than the enzymatic reaction. The crystallography experiment further requires that the photo caged APDA compound can be co-crystallized with ForT, PRPP and $Mg^{2+}$ present. By setting a time between flash of the activating laser and the probing X-ray pulse (for diffraction), it should be possible to study the enzymatic catalysed C-C bond formation.

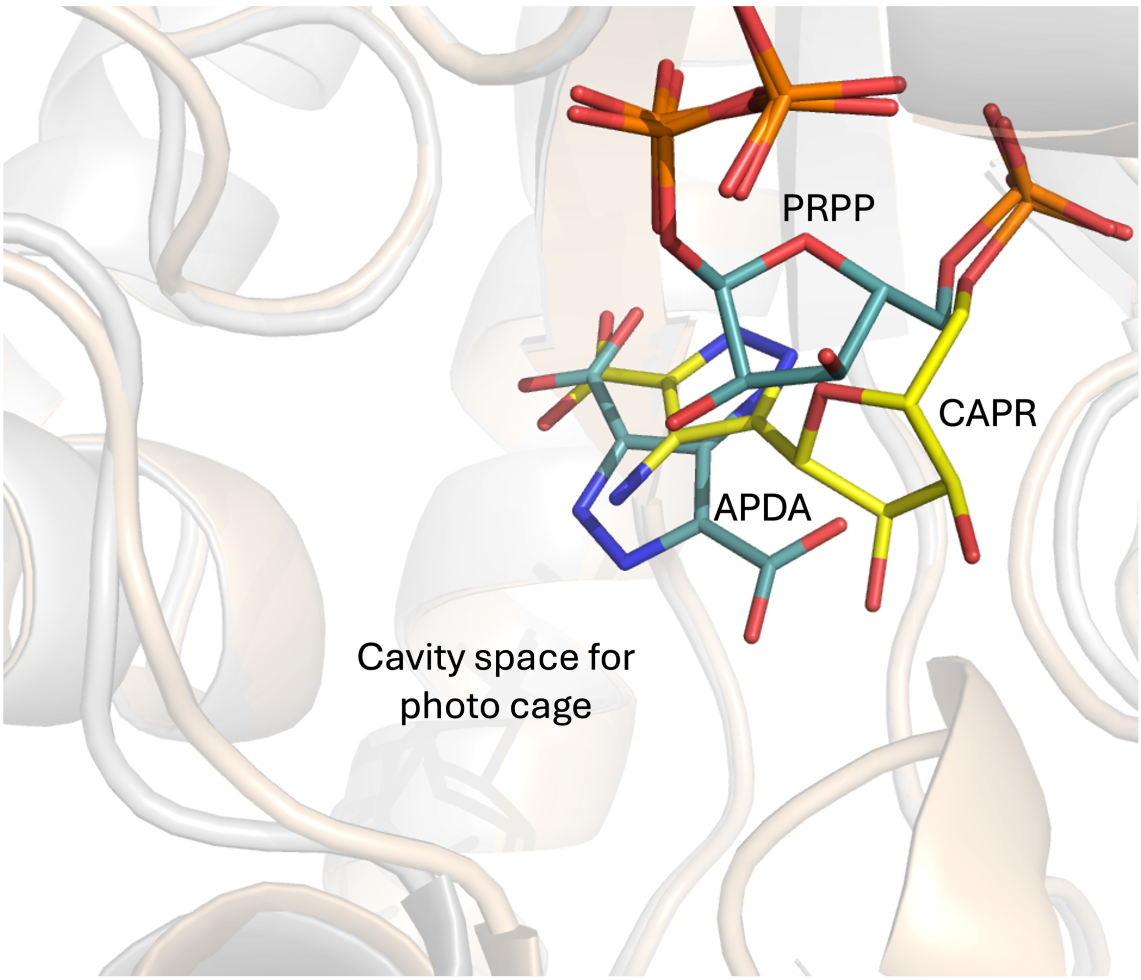



**Figure 4.1.** Superimposed structure of ForT-PRPP-APDA complex (carbon in cyan) and ForT-product complex (carbon in yellow), shows the cavity space potential for photo caging group. In contrast, the extracyclic side of the APDA shows limited space for adding caging group.

This chapter will describe and report some preliminary data as we built towards time resolved structural studies. The chapter includes the design of photo-cage compounds design, their synthesis, characterisation of their binding energy, co-crystallization with ForT, micro-crystallisation, structural analysis of caged compounds at the active site, and laser activation assay with an estimation quantum yield calculation. With the available data, this provides feasibility and further experimental path for the project.

## 4.2. Material and Methods

### 4.2.1. Synthesis of Cage Compounds

Design and synthesis of photo cage compounds was carried out by Dr. Georgina Girt.

#### 4.2.1.1. General Synthesis of Intracyclic Photocages

**2-nitrobenzyl-APDA (NB-APDA), 4,5-dimethoxy-2-nitrobenzyl-APDA (DMO-APDA), 4-hydroxyphenacyl-APDA (HPA-APDA)**
**and 6,7-dimethoxycoumarinyl-APDA (DMOC-APDA)**

APDA ethyl ester (1 eq, 1 mmol) was suspended in a mixture of acetonitrile-water (1:1, 4 mL). Potassium carbonate (100 mM, 1 mL) was added, and the suspension stirred and heated to 30C as dissolution occurred. Electrophile (2-nitrobenzyl chloride or 4,5-dimethoxy-2-nitrobenzyl bromide or 4-hydroxyphenacyl bromide or 4-bromomethyl-6,7-dimethoxycoumarin) (1.1 eq, 1.1 mmol as a suspension in or solution in MeCN/H2O, 5 mL) was then added dropwise. The reaction was then heated to 70C and stirred for 2 hours with monitoring by TLC, before cooling to room temperature. NaOH (10 M, 2mL) was then added, and the solution left stirring overnight at room temperature, after which point the sample had separated into two phases. The upper organic layer was removed, and the aqueous layer acidified with HCl (3M, 10 mL) to pH ~3. The product was then extracted with acetonitrile (3x 5 mL). Organic extracts were combined and dried in vacuo to yield the final product.

#### 4.2.1.2. General Synthesis of Extracyclic Photocages

**2-nitrobenzyl-APDA (e-NB-APDA) and 4,5-dimethoxy-2-nitrobenzyl-APDA (e-DMO-APDA)**

APDA ethyl ester (1 eq, 1.32 mmol) was dissolved in DCM (5 mL) with 2-nitrobenzaldehyde or 6-nitroveraldehyde (1.2 eq, 1.59 mmol). Trifluoroacetic acid (TFA) (3 eq, 0.3 mL) was then added dropwise, and the solution heated to 40C. Triethylsilane (TES) (2.6 eq, 0.55 mL) was then added dropwise. The solution was stirred at 40 C for 6 hours and monitored by TLC (10%

MeOH/DCM). The solution was then purified using flash column chromatography (0-10% MeOH in DCM), Fractions containing product were pooled and solvent removed under vacuum. NaOH (2M, 15 mL) was then added, and stirred at room temperature overnight. HCl (3M, 3 mL) was then added until pH ~3 reached and/or precipitation occurred. Precipitate was collected, washed with water, and lyophilised overnight.

### 4.2.2. ITC Binding Assay

Equilibrium binding experiments with WT ForT and caged compounds were performed using a MicroCal PEAQ-ITC (Malvern) calorimeter in the absence of $Mg^{2+}$ to prevent any enzymatic turnover. Protein samples were exchanged into buffer [10 mM HEPES, pH 7.5, containing 150 mM NaCl, and 1 mM TCEP] using gel filtration Cytiva® Superdex 200 16/600 column. ITC binding experiments were performed using a cell solution containing 250 μM protein, stirred at 750 rpm and held at 25 ˚ C using a reference power of 5 μcal/s. For APDA derived cage compounds binding to the enzyme/PRPP binary complex, 5 mM APDA was dissolved in 10 mM HEPES, pH 7.5, containing 150 mM NaCl, and 1 mM TCEP, and titrated into the cell containing 100 μM enzyme and 5 mM PRPP. A blank titration control where compounds were titrated into buffer was performed for each experiment. Data processing, fitting to a nonlinear one-site binding model, and data plotting were performed using MicroCal PEAQ ITC analysis software (Malvern).

### 4.2.3. LC-MS Turnover Assay with Cage Compounds

The LC-MS turnover assay was used to test if the APDA derived cage compounds were turned over by the enzyme. The previously described experimental protocol from chapter 2.2.7 was followed. Briefly different cage compounds (2.5 mM) and PRPP (5 mM) were incubated with WT ForT (25 μM) in sodium phosphate buffer [50 mM sodium phosphate (pH 7.5), 100 mM NaCl, and 50 mM $MgCl_2$] for overnight at room temperature with gentle rolling. The reaction was quenched by using 10 kDa cutting off concentrator to separate the protein out and the flow thought was analysed by LC-MS. This sample was eluted over a $C_{18}$ column using a gradient over 15 minutes of 5-100% acetonitrile and 5 mM ammonium carbonated at a flow

rate of 1 mL/min. Both LC elution UV spectrum and mass spectrum was analysed to determine if there was detectable product formation with caged compounds.

### 4.2.4. Laser Activation Assay

The laser activation assay experiment was carried out by our collaborator Dr. Ajay Jha from the Rosalind Franklin Institute. Transient UV/Visible absorption measurements were carried out at the LIFEtime facility at the Rutherford Appleton Laboratory. The experimental setup is based on a 100 kHz dual-amplifier ytterbium-based laser system (Pharos and Pharos SP, Light Conversion Ltd.). The laser beam from the amplifier Pharos SP is used to generate UV-Vis continuum by focusing the second harmonic of the 515 nm beam into a 1 mm thick calcium fluoride, $CaF_2$ plate, which is constantly moved in a raster pattern in the plane perpendicular to the direction of the laser beam to avoid colour-centre formation within $CaF_2$. A short-pass filter is placed in the probe beam path before its interaction with the sample to block the residual 1030 nm and 515 nm beams. The second Pharos amplifier output is used to pump optical parametric amplifier (Light Conversion Orpheus-HP), which in turn generate the excitation pump beam. The pump and probe beams are then focused into the sample to achieve spatial and temporal overlaps. The pump and probe beam spot sizes at sample are set to ~40 µm and ~100 µm FWHM, respectively. The timing control between the pump-probe beams is achieved using the combination of oscillator roundtrip timing and an optical delay line [240]. The pump energy used for the measurement is ~750 nJ. The amplifier's pulse picker can adjust the repetition rate of the pump pulses, which is set at 1 kHz for our measurements, while the repetition rate of the probe pulses is maintained at 100 kHz. After the sample, the probe beam is focused into a Newport 77400 spectrograph (Newport Corporation), and the spectrum is detected using a Teledyne e2v Ltd. OctoPlus line-scan camera with 2048 10*200 µm pixels.

Transient data analysis was performed using a python-based software, which is generously shared by Dr. Simon Bittmann, Max Planck Institute of the Structure and Dynamics of Matter, Hamburg [241]. The data analysis involves background subtraction and chirp correction. The chirp-corrected data is used to perform global fitting analysis.

The post laser flashed samples were analysed by LC-MS to determine the extent of production of uncaged compound APDA. The sample was eluted over a $C_{18}$ column using high pressure LC with 5 minutes gradient of 5 – 99% acetonitrile as mobile phase. The flow rate was set to 0.6 mL/min to give better separation. From 0 – 0.5 minutes it was 95% of water and 5% of acetonitrile, from 0.5 – 3 minutes, water reduced to 1% and acetonitrile raised to 99%, from 3 – 3.5 minutes, acetonitrile held as 95%, and reduced to 5% until 3.55 minutes, from 3.55 – 5 minutes, water and acetonitrile held at 95% and 5% respectively. Both LC elution UV spectrum and mass spectrum were analysed to determine the retention time of APDA and confirm the correct mass of uncaged product APDA.

Standard curve of APDA UV integration were carried using various concentration of APDA at 1 mM, 2 mM, 3 mM, 4 mM, 5 mM. The HPLC method for generating this standard curve has kept the same with previous set-ups to maintain same retention time of APDA. APDA was eluted at 0.34 min, and the UV integration for each concentration point were calculated by Waters integrated software. With the APDA standard curve, it is possible to calculate the APDA concentration in the post laser flashed sample, in order to obtain the quantum yield of laser activated uncaging reaction.

### 4.2.5. Co-Crystallization with Cage Compounds

The original condition (Chapter 3.2.7) served as the starting point for the co-crystallization with APDA caged compounds. ForT WT, ForT L24A, and ForT T138V proteins at a concentration of 25 mg/ml were incubated with both APDA caged compounds (NB-APDA and DMO-APDA) (Table 4.1) (50 mM) and PRPP (50mM) for one hour on ice before setting up crystallization plates. The crystals were grown in the hanging drop plates (EasyXtal 15-well DG-Tool X-seal) using 1 µL of co-crystallization protein solution (25 mg/ml) mixed with 1 µL of reservoir solution in the hanging drop. The crystals were mounted in a cryo-loop (Molecular Dimensions) and cryo-protected in solutions containing mother liquor plus 50 mM caged compounds and 50 mM PRPP, immediately frozen in liquid nitrogen and stored in a cryogenic Dewar. X-ray data were collected at Diamond Light Source I04-1.

### 4.2.6. Data Collection and Solve Structure with Cadge Compounds

X-ray diffraction data sets of ForT T138V – NB-APDA and ForT T138V – DMO-APDA complex structure were collected at Diamond Light Source I04 and processed through Diamond online processing software DIALS and XIA2 [177,178]. The phase was determined, and structure was solved by suing molecular replacement with ForT native structure as starting model through PHASER in CCP4i2 programme suit [179]. The caged compounds NB-APDA and DMO-APDA ligands was placed guided by the $F_o$-$F_c$ and $2F_o$-$F_c$ electron density maps in COOT, and the model was refined by Refmac5 [181] in CCP4i2 programme suit [179].

### 4.2.7. Microcrystal Formation

Microcrystals are essential for serial crystal data collection and XFEL data collection since each data set requires 1000's of individual crystals [161,242]. These cannot be individually mounted, instead the most common method is to jet a dense microcrystal slurries ( > $10^7$ crystals per millilitre) into the beam. Microcrystals between 10 – 30 µm are preferred as they match the beam and smaller crystals are desirable as they decrease the amount of protein required. Different methods of micro crystallization were investigated by altering the normal crystallization conditions and using seed crystals [242]. The crystal seedstock was prepared from conventional large crystals that were in crystal plates from previous ForT – PRPP, ForT T138V – PRPP, and ForT L24A – PRPP experiments. Individual large crystals were harvested together into a single tube with 10 – 20 glass seed beats. The crystals were vortexed for 30 seconds and rested on ice for 30 seconds, with this cycle repeated 10 times. Micro crystallization conditions were first screened in hanging drop with Hampton Research 24 well crystallization plates. 500 µL of ForT crystallization precipitant containing 20 mM Bis-Tris propane (pH 6 – 7.5), 40 mM KH2PO4, 20 % (v/v) glycerol, and 16 % (w/v) PEG 8000 was added to the reservoir. The ForT protein and its variant L24A and T138V at concentration of 30 mg/ml were incubated with 50 mM PRPP on ice for an hour before setting up crystallization plates. 1µL of seedstock was added on the cover glass with adding further protein and precipitant. The seedstock to protein to precipitant ratio combinations explored were 1:1:1; 1:1:2, 1:2:1 and 1:2:2. A second

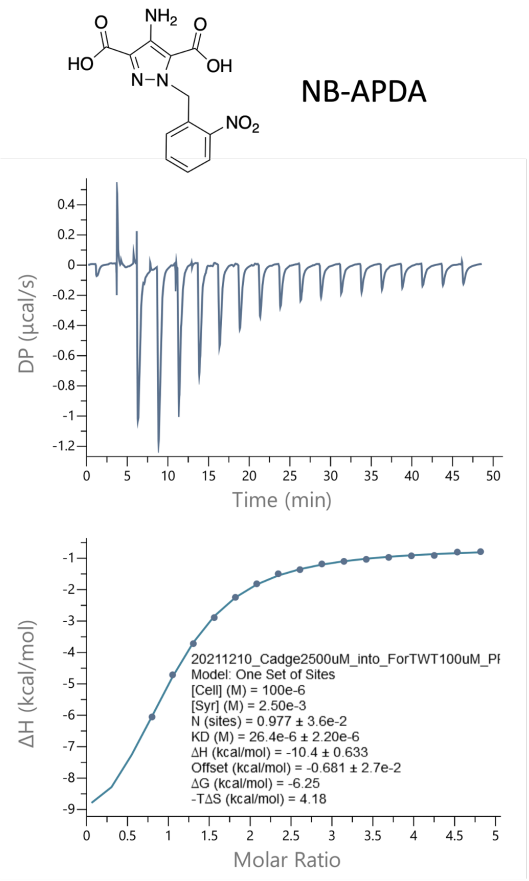
set of experiments was run in parallel with same ratios but with 10-fold diluted seedstock concentration used.

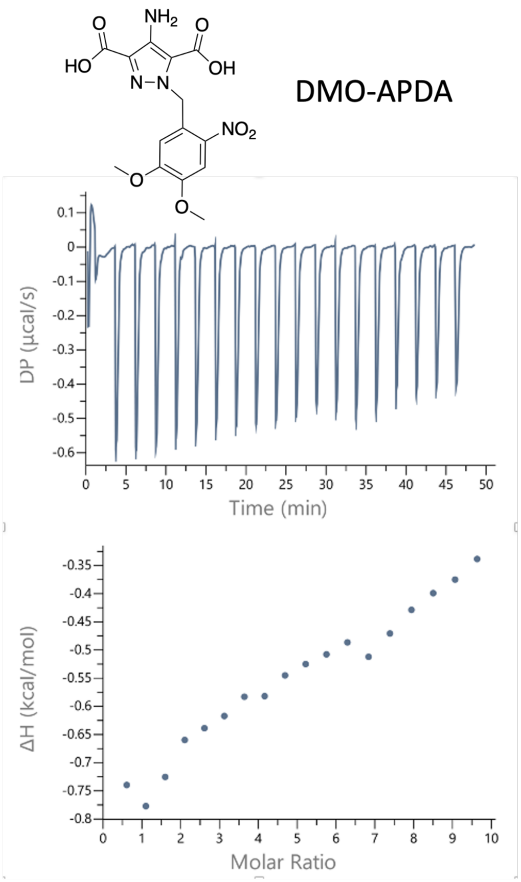
Micro crystallization condition in batch is better suited to produce sufficient quantities of micro crystals that are often need for XFEL experiments [242]. For micro-crystallization condition screening, it is important to find the balance of protein crystallization condition between nucleation and precipitation zone [243]. The final volume for batch condition is 20 µL, with addition of 1 µL seedstock, and the protein to precipitant ratio varied at 1:1, 1:2, 1:4, 1:9, 1:14, and 1:19. Micro crystals were examined under light microscope to determine the size and density.

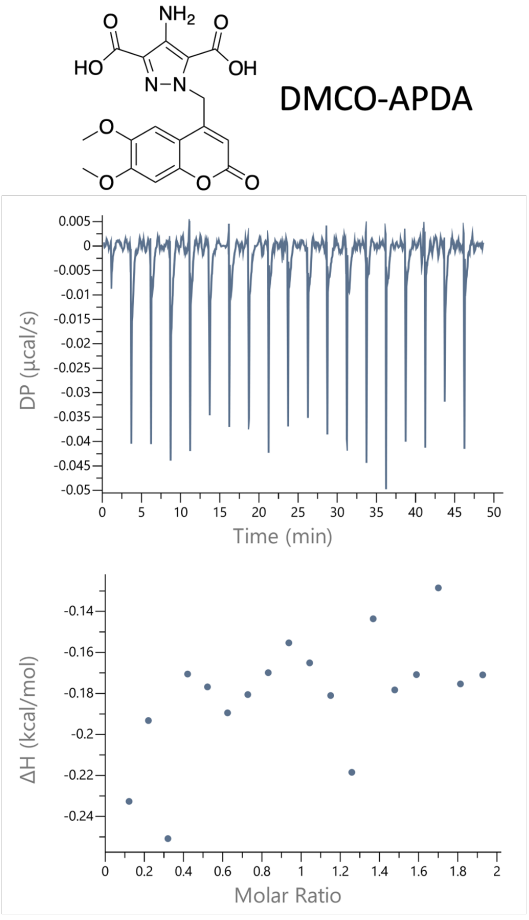
## 4.3. Results

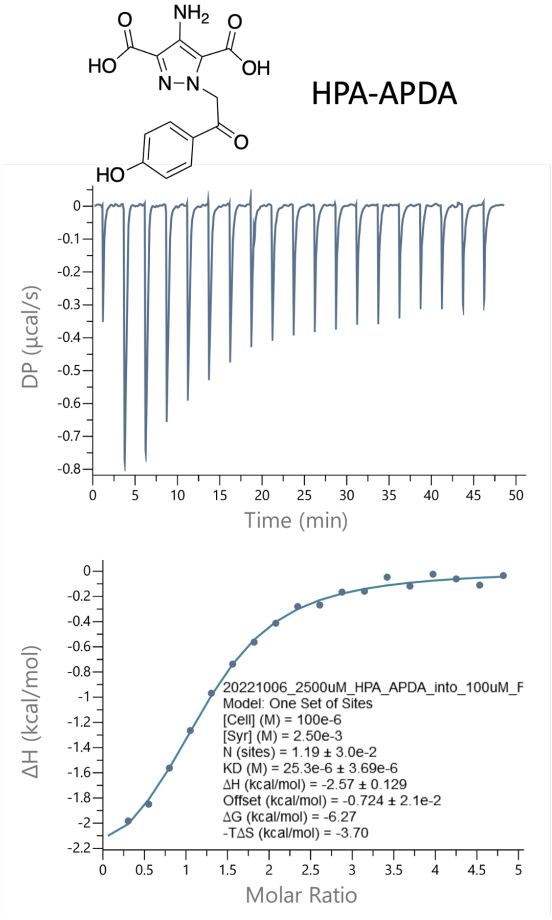
### 4.3.1. Binding Affinity of Different Cage Compounds

ITC multiple injection methods were used to initially screen the binding affinity between synthesized photo caged compounds and ForT enzyme. The photo caged compounds tested were NB-APDA, DMO-APDA, HPA-APDA, DMCO-APDA, e-NB-APDA (Figure 4.2). These compounds ranged from relatively small cage moieties to significantly larger cage. We syntheised and tested caged compounds where the extra cyclic amine group is caged and others where the ring nitrogen atoms were caged. From the ITC multiple injection plots in figure 4.2, NB-APDA and HPA-APDA clearly exhibit tight binding with ForT with measured $K_D$ 26.4 ± 2.2 µM, with N at 0.98 ± 0.04, and $K_D$ 25.3 ± 3.7 µM, with N at 1.19 ± 0.03 respectively. This compares to APDA has a $K_D$ at 120 ± 17 µM for WT ForT (chapter 3.3). In contrast, the large cage group of DMCO-APDA showed no binding affinity with ForT enzyme. The DMO-APDA is smaller than DMCO-APDA but larger than NB-APDA. The raw ITC plot showed some evidence of heat being liberated during titration of the substrate. If there is indeed binding it is below the measurement ability of the experiment. Notably, none of the extra cyclic amine addition of caged compounds showed any detectable binding. All the raw heat plots are shown in figure 4.2.


NB-APDA
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
20211210_Cadge2500uM_into_ForTWT100uM_PI
Model: One Set of Sites
[Cell] (M) = 100e-6
[Syr] (M) = 2.50e-3
N (sites) = 0.977 ± 3.6e-2
KD (M) = 26.4e-6 ± 2.20e-6
ΔH (kcal/mol) = -10.4 ± 0.633
Offset (kcal/mol) = -0.681 ± 2.7e-2
ΔG (kcal/mol) = -6.25
-TΔS (kcal/mol) = 4.18
Molar Ratio


DMO-APDA
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar Ratio


DMCO-APDA
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar Ratio


HPA-APDA
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
20221006_2500uM_HPA_APDA_into_100uM_F
Model: One Set of Sites
[Cell] (M) = 100e-6
[Syr] (M) = 2.50e-3
N (sites) = 1.19 ± 3.0e-2
KD (M) = 25.3e-6 ± 3.69e-6
ΔH (kcal/mol) = -2.57 ± 0.129
Offset (kcal/mol) = -0.724 ± 2.1e-2
ΔG (kcal/mol) = -6.27
-TΔS (kcal/mol) = -3.70
Molar Ratio

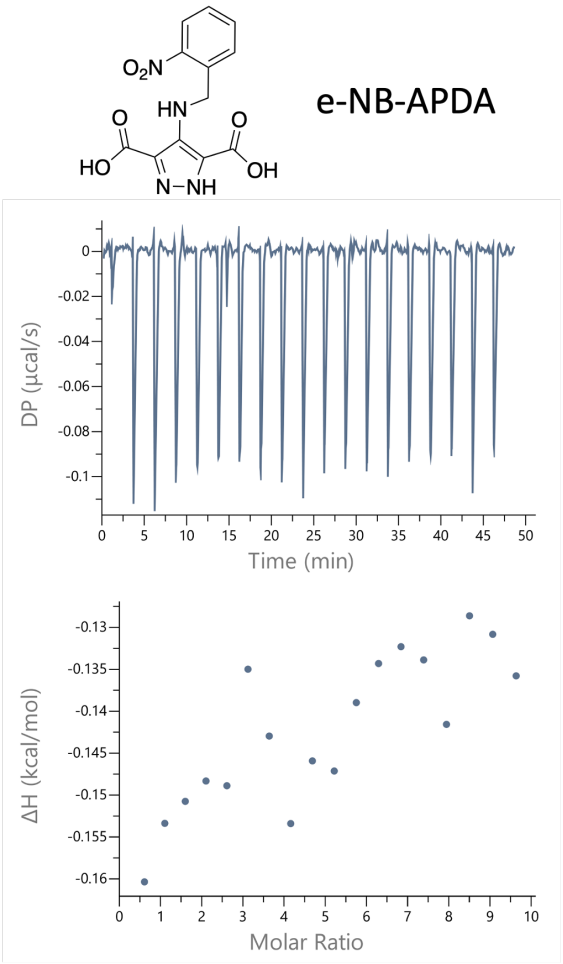



**Figure 4.2.** ITC multiple injection raw plots. Data analysis used the Malvern software suit with one site binding model. Various caged compounds titrated into ForT WT with PRPP. NB-APDA and HPA-APDA exhibits strong binding affinity with ForT with $K_D$ at 26.4 µM and 25.3 µM respectively. No binding detected under ITC experiment between DMCO-APDA, e-NB-APDA and ForT WT. DMO-APDA could not be fitted as a binder, although there is some evidence for heat liberation.

### 4.3.2. Caged Compounds are Not Substrates for ForT

This experiment was carried out by Dr. Gerogina Girt. The ForT has been incubated with photo cage compound NB-APDA overnight under LCMS activation assay condition (Chapter 2.2.7), no change in spectrum of NB-APDA photocage, with no visible depletion of substrate coupled with no detectable peak formation characteristic of APDA. MSD shows m/z 307 ($[M+H]^+$), 329 ($[M+Na]^+$) peaks for unreacted photocage. Comparable results and spectra were observed for all photocages (DMO-APDA, NB-APDA, HPA-APDA e-NB-APDA, e-DMO-APDA).

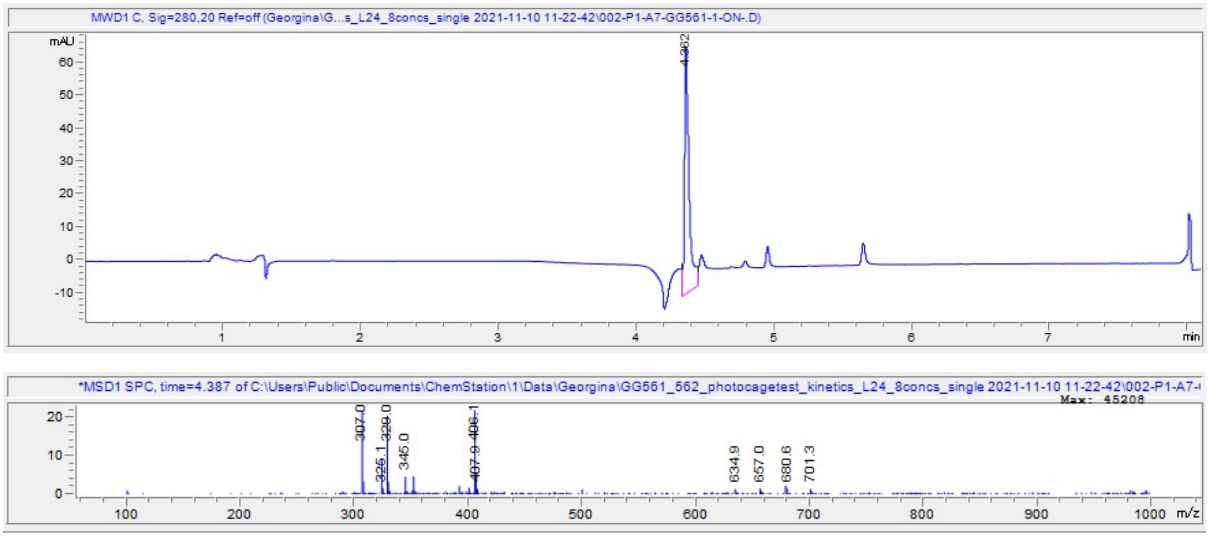



**Figure 4.3.** LCMS analysis for the overnight reaction sample of ForT and NB-APDA. No change in spectrum of NB-APDA photocage, with no visible depletion of substrate coupled with no detectable peak formation characteristic of APDA. MSD shows m/z 307 ($[M+H]^+$), 329 ($[M+Na]^+$) peaks for unreacted photocage. Top figure is the UV absorbance at 280 nm, one single peak at elution time 4.2 mins. The bottom figure is the mass detection shows MSD m/z at 307 ($[M+H]^+$), 329 ($[M+Na]^+$) peak for uncaged compound. (Data from Dr. Georgina Girt)

### 4.3.3. Laser Activation Experiments with Caged Compounds

Both NB-APDA and DMO-APDA were exposed to the laser pulse to uncage it, followed the methods described in chapter 4.2.4. (Laser uncaging experiment carried out by Dr. Ajay Jha) The resulting solution was analysed by LC-MS Figure 4.4. Figure 4.4B shows the elution curve (UV absorbance at 254 nm) together with the total mass scan in positive mode. The elution peak at 1.92 minutes is the caged compound NB-APDA. Figure 4.4A shows the mass spectrum data confirming the correct mass for NB-APDA. However, there was no APDA detected from this photo activation assay.

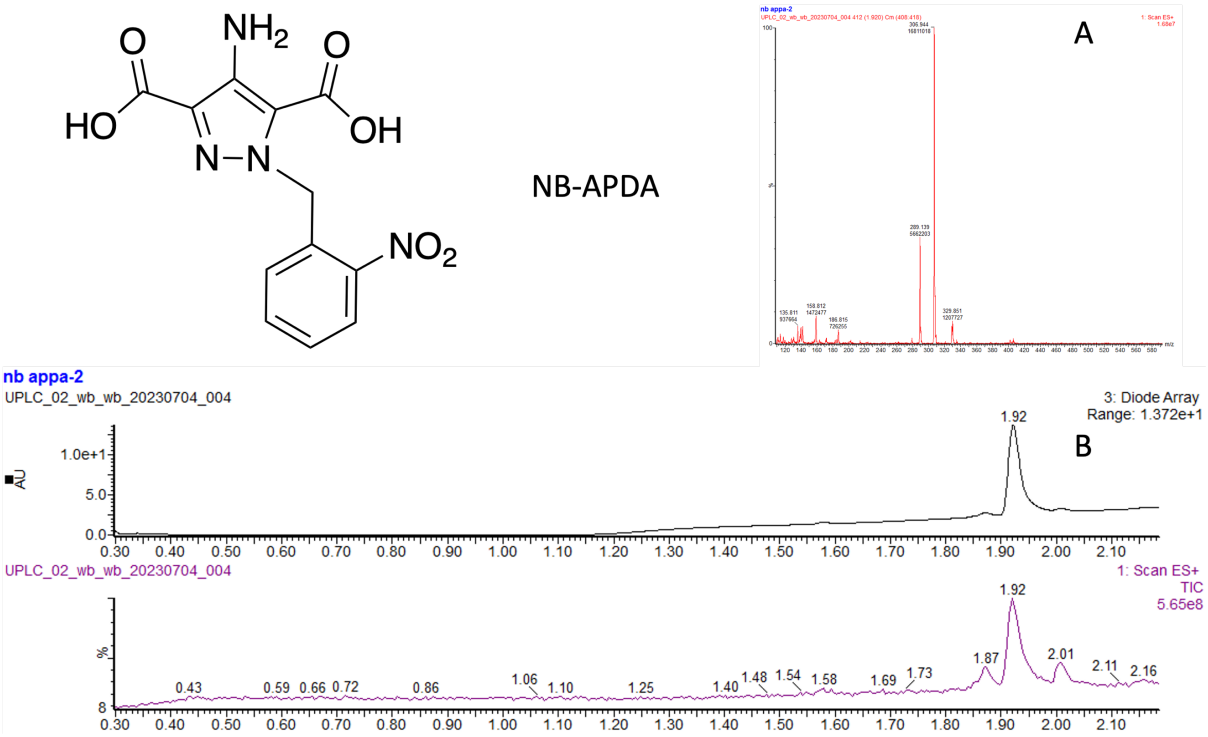



**Figure 4.4.** LC-MS assay of activation NB-APDA sample. Figure A shows the mass spectrum data confirms the correct mass for NB-APDA at elution time of 1.92 minutes. Figure B shows the elution curve measured by UV absorbance at 254 nm and total mass scan in positive mode. The 1.92 minutes peak shows the elution of NB-APDA, no uncaging was observed m/z = 306.9 $[M+H]^+$, and m/z = 329.9 $[M+Na]^+$.

In comparison, figure 4.5 shows the LC-MS analysis for DMO-APDA following uncaging. Figure 4.5A and C shows the total mass scan in both positive and negative mode, two major peaks observed at both 0.35 minutes and 1.86 minutes. APDA standard LCMS analysis experiment has been carried out to confirm the retention time of APDA in the same condition. Figure 4.5B is the UV absorbance data at 254 nm. Figure 4.5D and E confirmed the mass for the caged compound DMO-APDA in both positive and negative with m/z = 366 $[M+H]^+$, and 365 respectively. Figure 4.5F and G confirmed the mass for the released true enzyme substrate APDA in both positive and negative mode with m/z = 172 $[M+H]^+$, and 170 respectively.

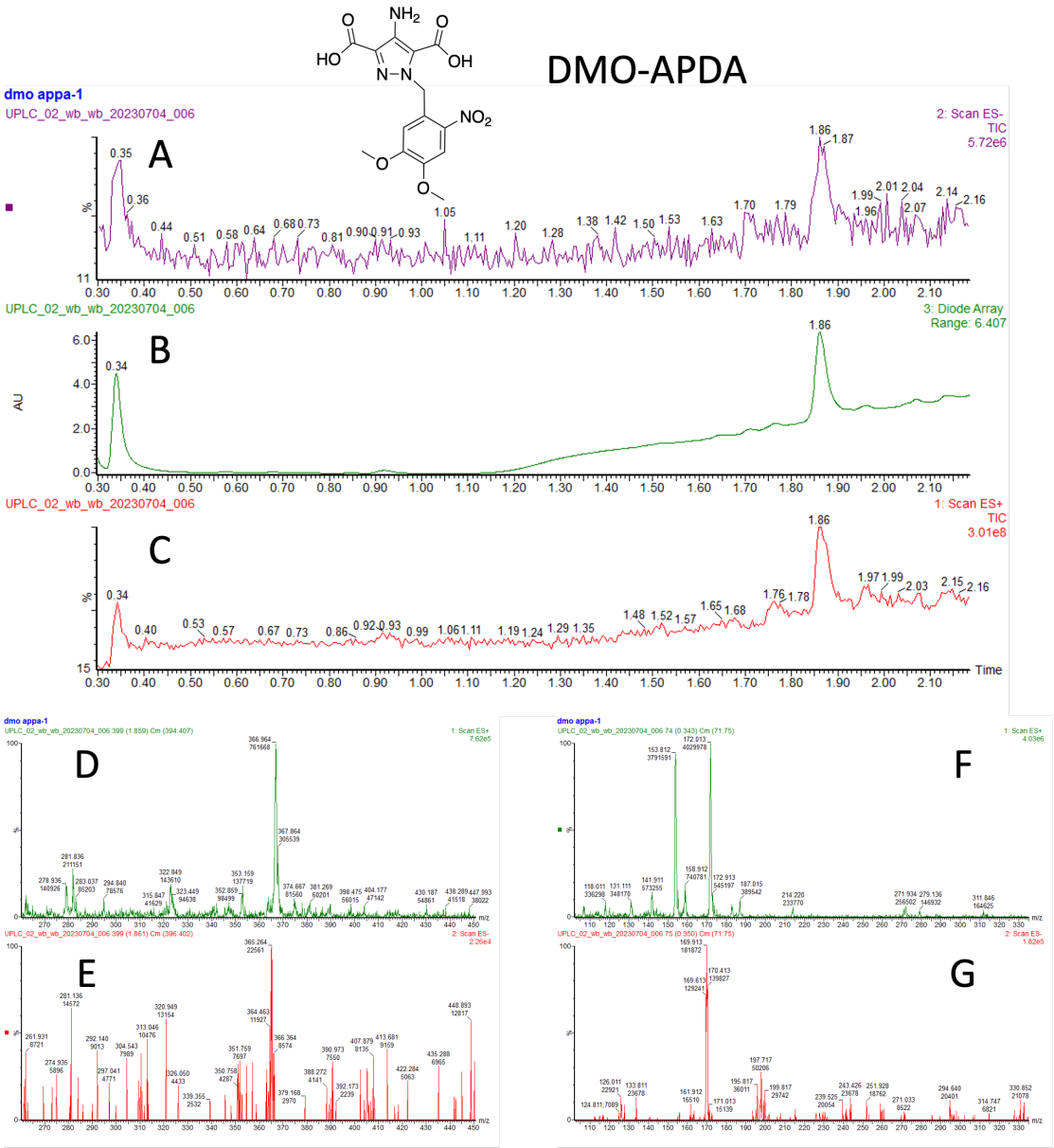


**Figure 4.5.** LC-MS analysis of uncaged DMO-APDA. Plot A, B, C shows the elution profile for both UV detection and total mass scan in both positive and negative mode. Plot D, E confirms the mass of DMO-APDA eluted at 1.86 minutes in both positive and negative mode m/z = 366 $[M+H]^+$. Plot F, G confirms the mass of photo activation released APDA eluted at 0.34 minutes in both positive and negative mode m/z = 171 $[M+H]^+$.

APDA UV absorption standard curve was generated using authentic APDA at different concentration from 1 mM to 5 mM. The UV absorbance peaks were integrated by the Waters software and the standard curve was plotted shown in figure 4.6. The standard curve was plotted by Prism with linear regression analysis, gives the equation Y = 228135*X – 8795 with R square at 0.99 confidence. The APDA peak integration from DMO-APDA photo activation sample is 106362 (Figure 4.6), together with the standard curve, the concentration of APDA presented in the post laser activation sample is 0.505 mM. The starting concentration of DMO-APDA was 5 mM, therefore, the yield of the photo cleavage reaction is calculated to be 10 %.

| [APDA] mM | UV Integration |
|---|---|
| 0.505 | 106362 |
| 1 | 167978 |
| 2 | 444749 |
| 3 | 749128 |
| 4 | 926373 |
| 5 | 1081035 |

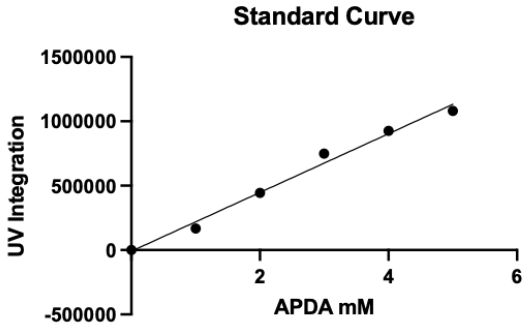



**Figure 4.6.** The standard UV absorption integration curve of APDA generated with concentrations ranging from 1 mM to 5 mM. The data were analysed in Prism with linear regression model, gives the equation Y = 228135*X – 8795 with R square at 0.99 confidence.

### 4.3.4. Quantum Yield of Cage Compounds by Laser Assay

The transient absorption (TA) measurement is performed after subjecting the DMO-APPA to a resonant excitation at 350 nm pump beam. The details of the experimental setup and measurement conditions are at chapter 4.2.4. The TA data (Figure 4.7) shows an intense excited state absorption (ESA) feature from 390 nm to beyond the probe wavelength limit of 460 nm. This feature arises immediately after excitation, and hence, was assigned to the excited singlet $S_1$ state. A clear ground state bleach signal (GSB) (~365 - 375 nm) was observed at <10 ps. The GSB feature is overlapped with the stimulated emission (SE) feature at ~380 nm. The ESA feature broadens over time and decays in intensity. At 10 ns, a significant amount of ESA signal was still observed, which points toward the generation of a distinct long-lived species after photoexcitation to $S_1$ state.

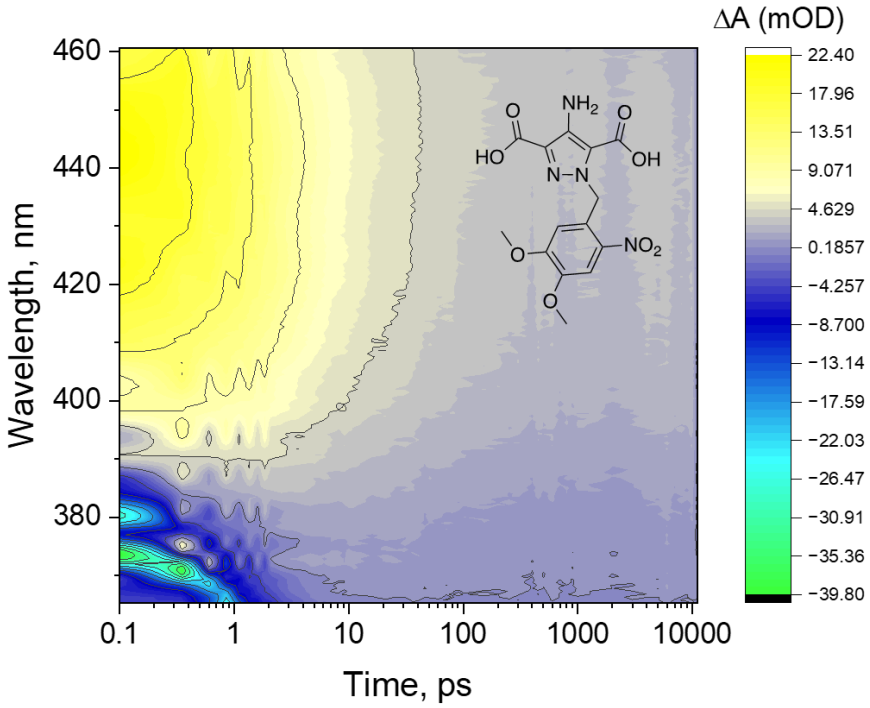



**Figure 4.7.** Transient UV/Vis absorption measurements. Pump-probe experimental data using 350 nm pump wavelength and UV/Vis probe (365-460 nm) is presented for DMO-APPA in water. Pump energy used is ~750 nJ with the beam spot size of ~100 µm FWHM to overlap with ~40 µm spot size of the probe beam. The sample cell moved in a raster pattern in the plane perpendicular to the direction of the probe beam. Yellow represents excited state

absorption (ESA) and cyan represents ground state bleach convoluted with stimulated emission. Resonant photoexcitation of DMO-APPA leads to population in excited state, which shows a broad absorption band starting from 390 nm to beyond the probe window of 460 nm. The ground state bleach (~370 nm) along with the stimulated emission band (~380 nm) seems to be masked by the evolution of ESA band after ~10 ps. (Data from Dr. Ajay Jha)

In order to better visualize the evolution of the molecule, a stack plot is shown in Figure 4.8 The differential absorption data at -6 ps, (before resonant excitation), does not show any features. After 350 nm excitation, 0.6 ps time point data clearly shows intense absorption band along with the negative SE and GSB features. The negative SE feature is rendered invisible possibly due to decay of the $S_1$ state. Additionally, the new state generated after photoexcitation seems to have broad absorption band, which masks the GSB feature as well.

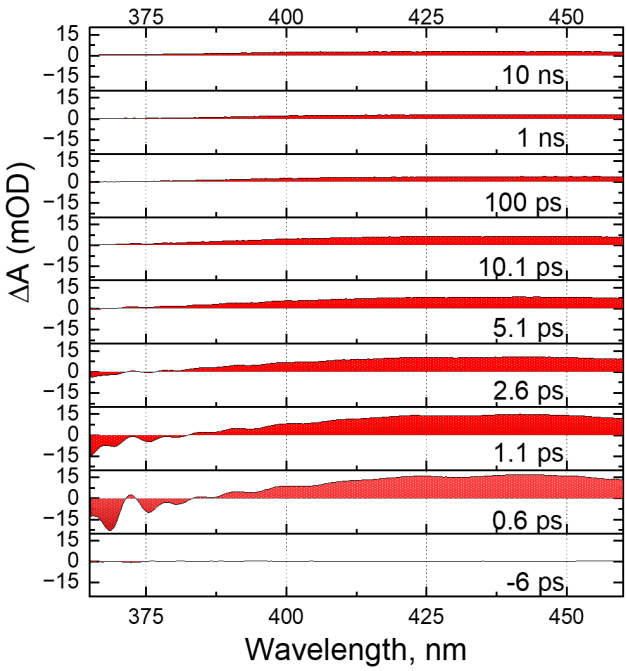



**Figure 4.8.** Transient absorption data. linecuts of the transient absorption data at different time-points after subjecting the molecules to resonant excitation with $\lambda_{ex}$ = 350 nm. (Data from Dr. Ajay Jha)

To unravel the state specific dynamics from the convoluted TA data, a global fitting analysis is performed and shown here as Figure 4.9 Three decay associated spectra (DAS) with time constants: 2.38 ps, 6.1 ns and a long-lived component (>10 ns) are obtained, as shown in Figure 3*b*. The DAS with time-constant of 2.38 ps shown a clear GSB feature at 365 nm along with a broad absorption band. This component can be assigned to singlet $S_1$ state. The second DAS with time constant 6.1 ns, shows broad and weak absorption features within the probe window. The DAS with the longest lifetime shows the GSB feature along with a broad absorption band. We assume the longest-lived component is a photoconversion product (typically singlet state decay very rapdily and triplet state formation is rare), the amplitude of GSB at 365 nm is ~8% of the initial GSB, roughly 8 % conversion.

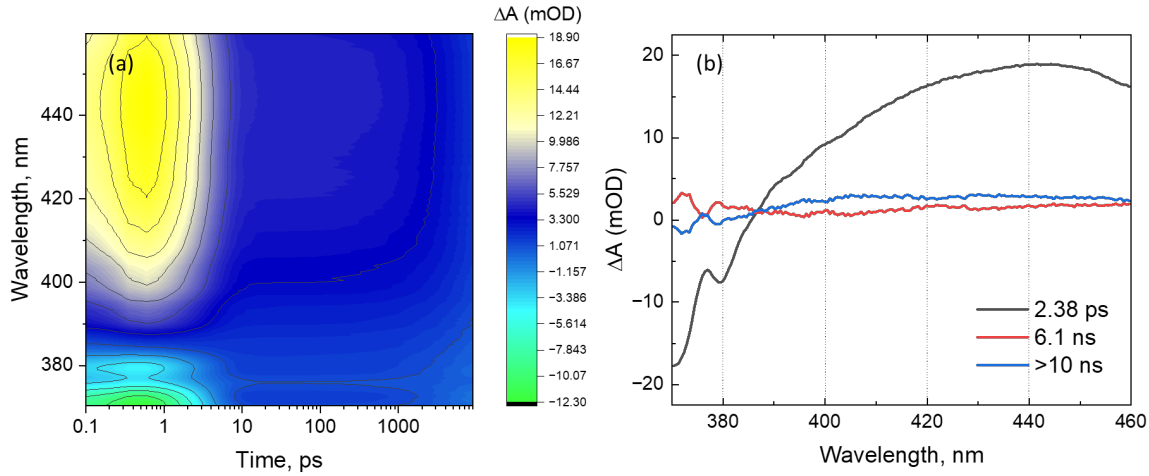



**Figure 4.9.** Global fitting analysis. The chirp corrected transient absorption data has been used to perform global fitting analysis. The global fitted data using three decay associated spectral components is shown here as (a) Yellow represents excited state absorption and cyan represents ground state bleach convoluted with stimulated emission. (b) Decay associated spectra with time-constants: 2.38 ps, 6.1 ns and >10 ns. (Data from Dr. Ajay Jha)

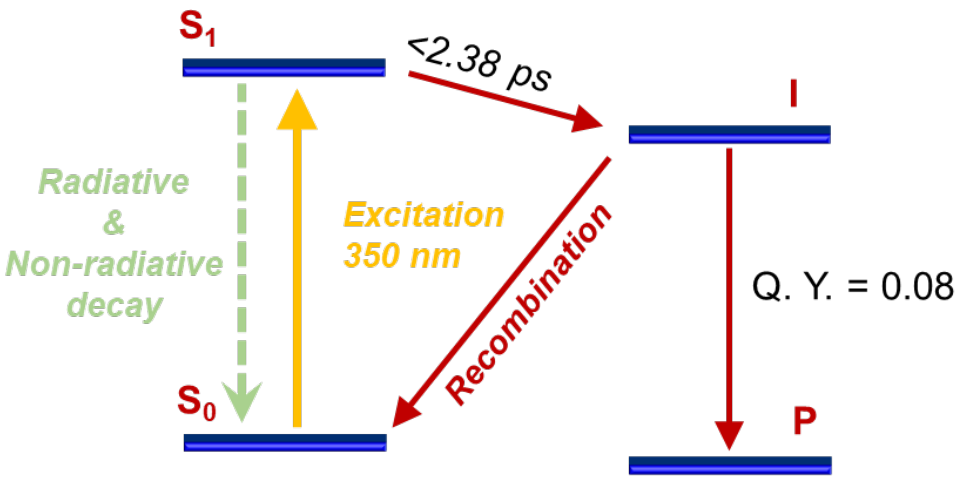



**Figure 4.10.** Proposed excited state evolution channels. Postulated excited state dynamics is depicted in energy level diagram. $S_0$ = singlet ground state, $S_1$ = singlet excited state, I = intermediate state with lifetime of 6.1 ns, and P = photoproduct state with long lifetime. The quantum yield of photoproduct formation is calculated to be 8%. (Data from Dr. Ajay Jha)

Based on the TA absorption data analysis, excited state dynamics of DMO_APPA is depicted in Figure 4.10 as postulated model. Initial photoexcitation of singlet ground state, $S_0$ populate singlet excited state, $S_1$, which decays to intermediate state, I in <2.38 ps. The intermediate state, it has a lifetime of 6.1 ns with two channels of decay: recombination to the $S_0$ state and photoproduct formation i.e., P. The quantum yield of the photoproduct formation is calculated to be 0.08. The recombination channels seem to be dominant pathway for the intermediate state to decay.

### 4.3.5. Structure of Cage Compounds in Complex with ForT

The ForT-T138V-NB-APDA-PRPP formed crystals remains the same space group with ForT-T138V substrate complex structure as $C222_1$, with one monomer in the asymmetric unit. The structure was solved to 1.34 Å resolution by molecular replacement using previous solved ForT-T138V substrate complex as a search model. The overall structure remains the same as described for the ForT-T138V PRPP APDA complex structure (Chapter 3.3.2). Figure 4.11A shows the unbiased $F_o$ - $F_c$ electron density map at 3σ that shows NB-APDA is bound at the active site replacing APDA. Figure 4.11B shows the superimposed structure with original

substrate complex structure, where yellow is the newly solved ForT-T138V-NB-APDA complex and blue is the ForT-T138V substrate complex. PRPP and $Mg^{2+}$ are identically positioned in both structures. The APDA moiety of the caged compound shares the same position and interactions with the protein as authentic APDA in the binding pocket (Chapter 3.3.2). The major difference in the binding pocket is the shift of Glu211 side chain. In the caged compound complex structure, side chain of Glu211 has shifted sideway to avoid clash with the nitrobenzoic cage moiety of NB-APDA. (Shown in Figure 4.11C) The shift of the Glu211 side chain enlarges the pocket and provides space for cage group on APDA. The disruption of the key Glu 211 side chain is consistent with the lack of turnover of the enzyme with the caged compound. The nitrobenzoic cage moiety is easily accommodated by the protein.

Another cage compound DMO-APDA has been co-crystalized with ForT T138V and the ForT T138V-DMO-APDA-PRPP complex structure has been solved to 1.77 Å resolution by molecular replacement using ForT-T138V substrate complex as a search model. It has the same space group as previous T138V complex as $C222_1$. Figure D shows the $F_o$ - $F_c$ electron density map for DMO-APDA at 1.5$\sigma$ unambiguously locates the APDA moiety of the caging compound. The caging moiety is slightly disordered due to its flexibility in the binding cavity. The electron density map shows the planner structure of the cage group to confirm orientation of the cage moiety inside of the substrate binding site.

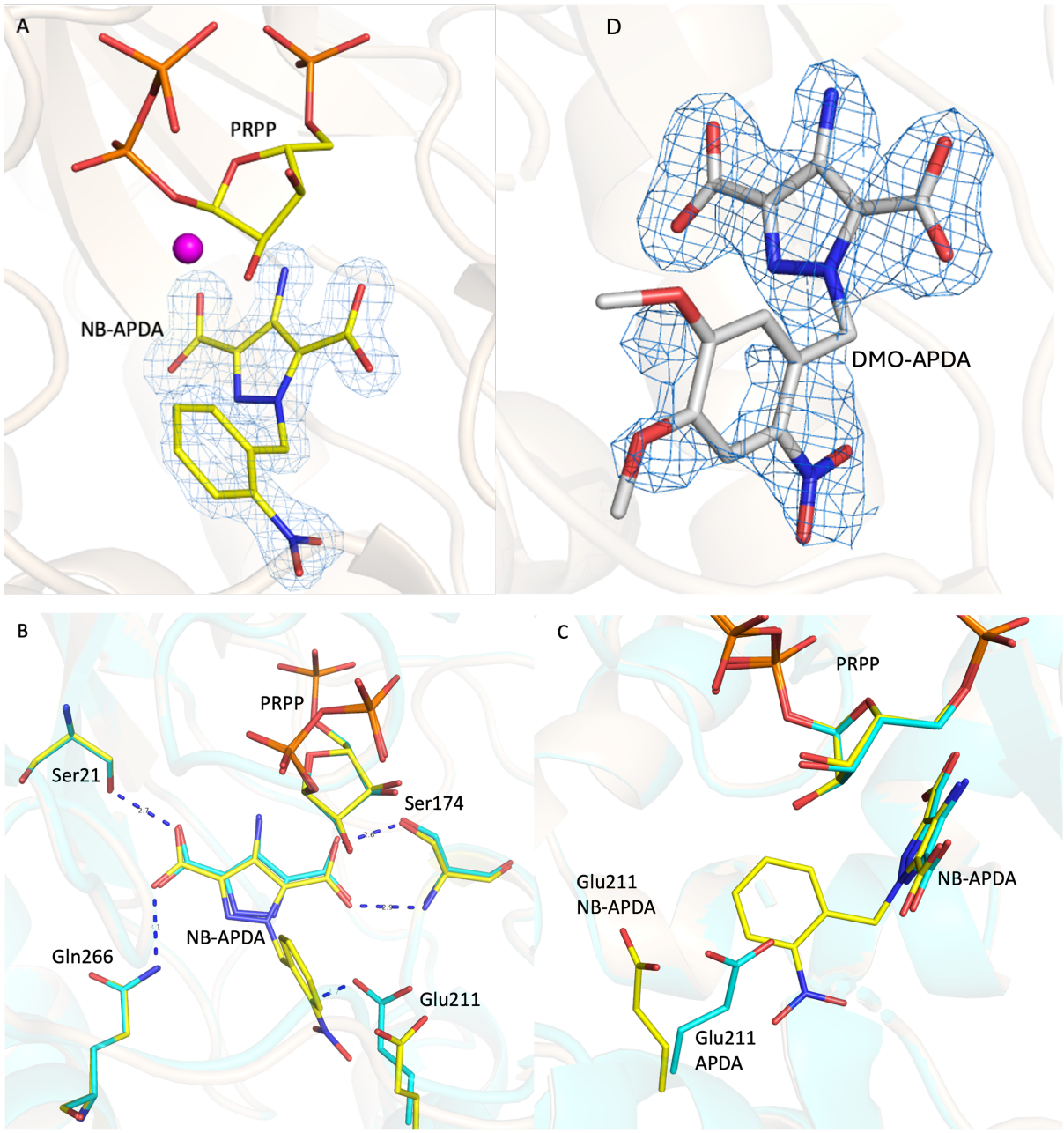



**Figure 4.11.** Structure of ForT T138V with both NB-APDA and PRPP at 1.34 Å resolution. Figure A shows the $F_o$ - $F_c$ electron density map for NB-APDA at 3σ, where the cage group extending to the far side of the PRPP binding pocket. Figure B shows the superimposed structure of T138V-APDA-PRPP complex in cyan and T138V-NB-APDA-PRPP in yellow, where confirms the PRPP and APDA moiety stays the same binding position. The caged compound NB-APDA does not alter the binding pocket side chain of Ser21, Gln266 and Ser174. Figure C shows that the side chain of Glu211 shifts away from APDA and gives space of the cage group in the enzyme active site. Figure D shows the $F_o$ - $F_c$ electron density map for DMO-APDA (carbon colour in

grey) at 1.5σ unambiguously locates the APDA moiety of the caging compound. The caging moiety is slightly disordered due to its flexibility in the binding cavity.

### 4.3.6. Micro Crystallisation Approach

Figure 4.12 shows the micro crystallization for hanging drop method with seeding technique. The two comparison groups are normal seeds stock concentration and 10 times diluted seedstock. Within those two groups, various ratios between seeds, protein and precipitant were tested as 1:1:1, 1:1:2, 1:2:1, and 1:2:2. In the normal seeds concentration, crystals were formed with under 200 µm in 1:1:1 and 1:1:2 ratio, where the 1:1:1 ratio formed slightly larger than the other. Both exhibited a slight level of precipitation in the well. For the ratio with more protein and precipitant group, larger crystals with more than 500 µm in length have grown. In comparison, with 0.1 concentration of stock seeds added, more precipitation was observed in the 1:1:1 and 1:1:2 ratio. However, the crystals formed more homogeneously in length and width than stock concentration of seeds. And the microcrystals in 1:1:2 ratio are within 100 µm in length. Similarly, in the 1:2:1 and 1:2:2 group, large single crystals have formed rather than micro crystals.

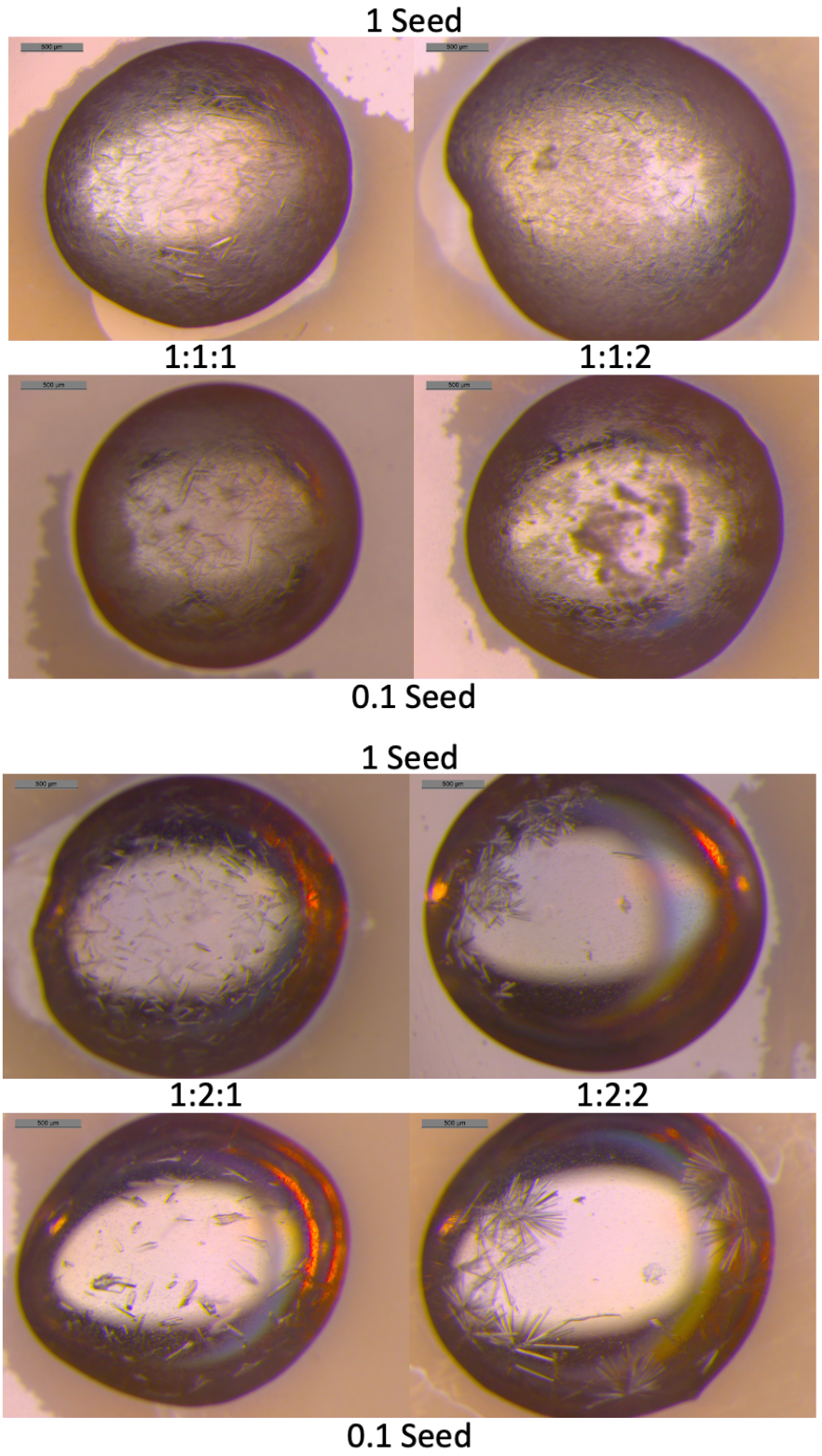



**Figure 4.12.** Micro crystallization approach of ForT WT with PRPP and $MgCl_2$ with hanging drop technique using seeds and various ratios of protein to precipitant. The scale bar is 500 µm, the micro crystals appear at undiluted seedstock concentration and the ratio between protein and precipitant is 1:1 or 1:2.

To obtain enough number of microcrystals for XFEL time resolution studies, batch condition for micro crystallization experiments was carried out (chapter 4.2.7). The final volume for each batch experiment is 20 µL and the protein to precipitant ratio was varied at 1:1, 1:2, 1:4, 1:9, 1:14, and 1:19. The original stock concentration of seeds was added to the batch at 1 µL in volume and it is 1:1 in ratio to the protein. Interestingly, no crystals observed in the 1:1 protein to precipitant ratio batch. Similarly, a few large crystals with 100 – 200 µm length crystals were observed in the 1:2 ratio batch. The rest of the batch conditions shown precipitation with undistinguishable micro crystals formation. Figure 4.13 shows the batch condition micro crystallization under light microscope where the protein form precipitations or on the edge to crystallization.

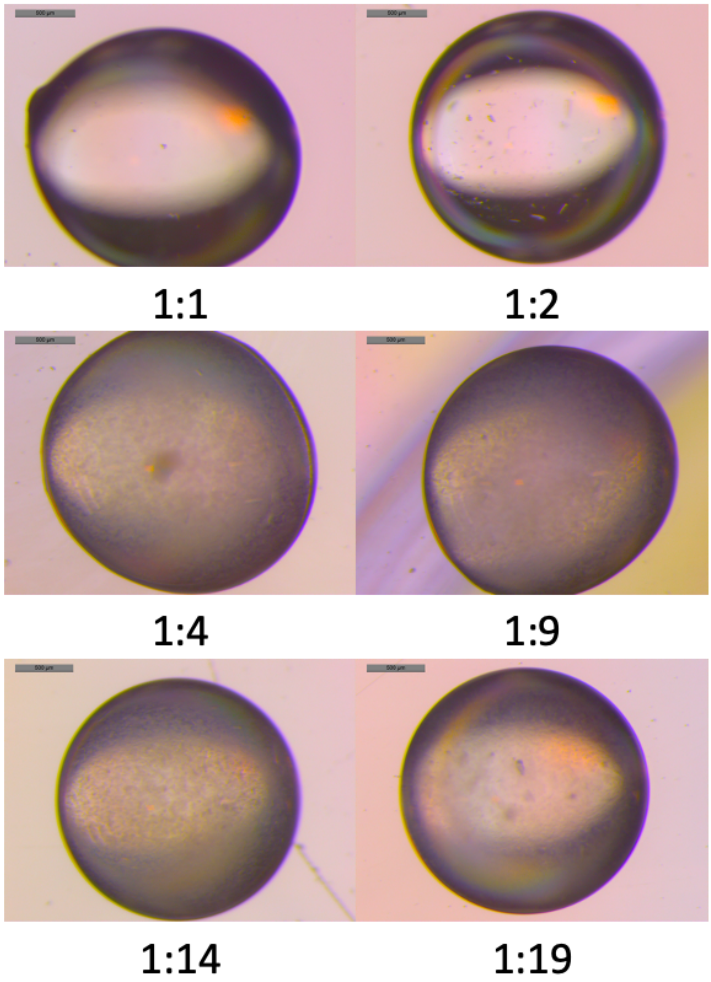



**Figure 4.13.** Micro crystallization approach of ForT WT with PRPP and $MgCl_2$ in batch condition with various protein to precipitant ratio.

For XFEL experiments micro crystals are preferred in very low or no concentration of glycerol. This because glycerol is viscous and hinders jetting. Unfortunately, our most promising experiments had glycerol. The best screening hit for ForT T138V with 50 mM PRPP co-crystallization is 10% glycerol, 17.5% PEG 8000, 40 mM $KHPO_4$, and 20 mM Bis-tris propane pH at 6.5. Figure 4.14 shows the micro crystals exhibits homogeneous morphology with the size between 50 – 100 µm in length. The density of the crystals is appeared by eye to be high in the 0.2 µL screening drop. However, it has not yet proven to scale up these conditions to produce the large quantities that XFEL experiments require.

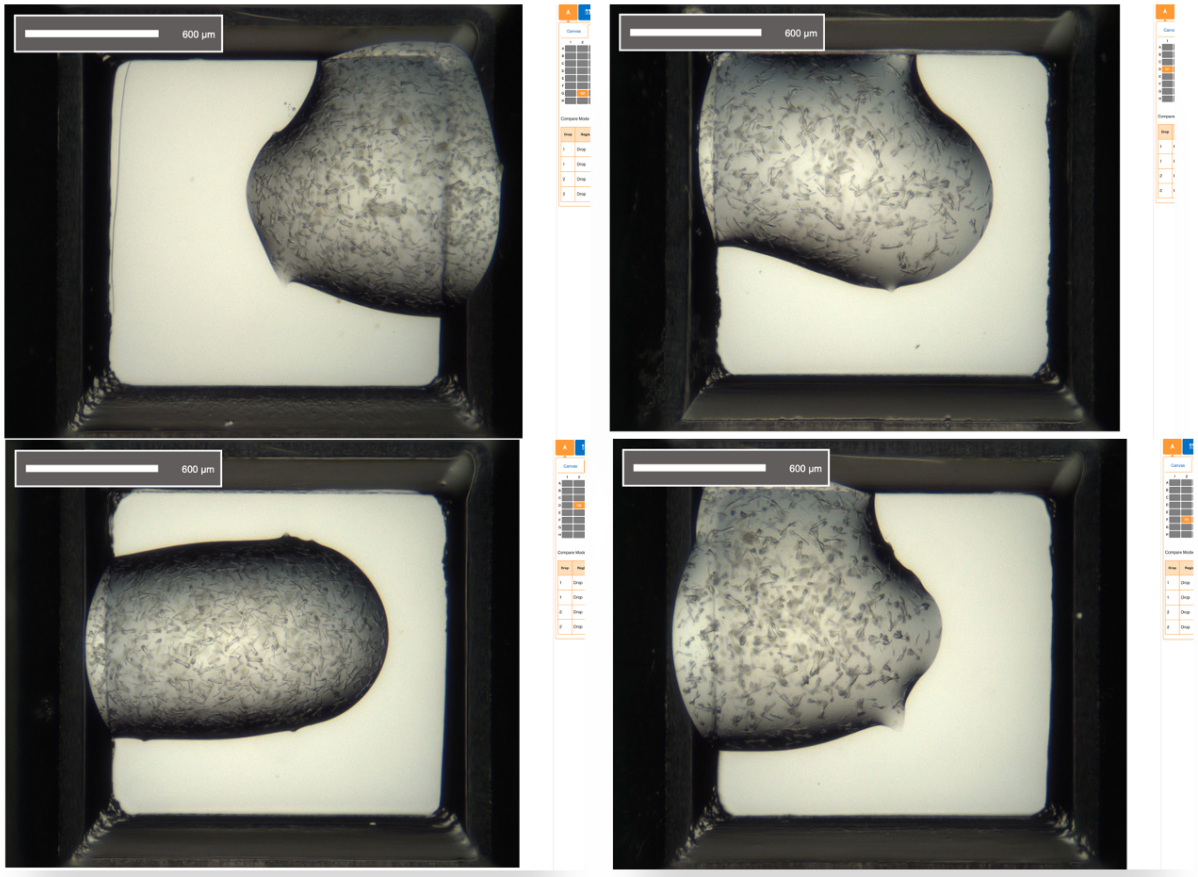



**Figure 4.14.** Micro crystallization approach of ForT T138V with PRPP and $MgCl_2$. Homogenous microcrystals formed at size between 50 – 100 µm in length. This is achieved in screening condition with 0.2 µL in volume protein and precipitant volume. The best hit condition has 10% glycerol, 17.5% PEG 8000, 40 mM $KHPO_4$, and 20 mM Bis-tris propane pH at 6.5.

## 4.4. Discussion

Structure and function are often correlated together in biological systems. However, natural system is not static, they are dynamic. The understanding of the chemical mechanism of enzymes was transformed by the knowledge of the three-dimensional structure of proteins. This allowed chemical principles to be brought to bear, the first high profile examples being the serine protease catalytic triad [244] and lysozyme [245,246]. This has breakthroughs in virtually every known biological transformation and underpins development of new medicines, biocatalysis (green chemistry) and metabolism. However, as the story of lysozyme illustrates, there can be misunderstandings, mechanism can be more complex than simply inferred [245,246]. To date, the experimental approaches to study enzyme turnover in real time have not existed. There are a number of studies where intermediates have been isolated and quasi time resolved studies carried out.

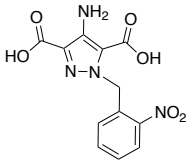
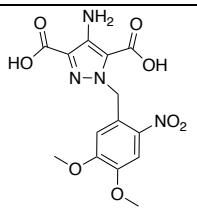
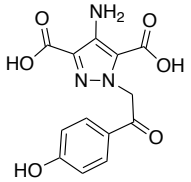
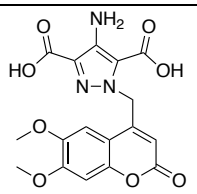
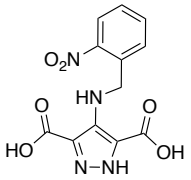
Ilme Schlichting's group used high-intensity laser pulses to induce sequential two-photon excitation in bacteriorhodopsin (bR), a light-driven proton pump, leading to isomerization of the retinal chromophore [247]. They observed ultrafast structural changes in the protein, including molecular bond length changes, vibrational modes, and differences in low-frequency vibrations between free retinal and retinal bound to the protein. Real-space refinement against extrapolated density maps was performed to analyse the structural changes. They have observed the short-lived intermediated conformational change in the retinal and its surroundings to form a vibrationally coupled network. Similarly, Valerie Panneels's group reported the conformational changes and dynamics of rhodopsin, a light-sensitive G protein-coupled receptor involved in vision [248]. The study reveals that upon photoactivation, rhodopsin undergoes anisotropic breathing motion, with changes occurring in the amino acid residues surrounding the retinal molecule. This motion is detected in the extracellular part of transmembrane helices TM3, TM5, and TM6. The researchers also observed localized intrinsic fluctuations in the dark state of rhodopsin, indicating inherent dynamics even in the absence of light stimulation.

The advent of ultra-fast diffraction (ps to fs) has opened the possibility of studying enzyme mechanism experimentally. To date, the most successful studies have involved electron movements, this is because electrons can be moved by lasers and thus synchronised in defined states. We propose to study bond forming and breaking reactions by using photoactivation of a caged substrate. There is a significant amount of preliminary work to ensure the system we have chosen is feasible.

Photo caged substrate compounds were designed in a joint manner but synthesised entirely by Dr. Georgina Girt, a postdoc researcher in the lab working on chemistry side of the project. Table 4.1 shows the summary of different photo caged compounds that were synthesized, and their binding affinity as determined by ITC. Both NB-APDA and HPA-APDA have relatively small cage group and they showed APDA like binding affinity. This is consistent with the structural analysis; the caged compound preserves all the interactions seen in the ForT T138V structure. The additional cage group makes few interactions and unsurprisingly does not add to the affinity. Equally the cage only perturbs the catalytic Glu211 but no other residues, suggesting no penalty for binding. The overlap in binding of the key atoms is important for feasibility, we can be confident that once uncaged the molecule would be positioned identically to APDA and thus react as expected.

Interestingly, the caged structure is consistent with the orientation of APDA being that seen in the original ForT APDA complex rather than the product. Further support comes from the fact the extracyclic derivatives of the amine do not bind to the protein at all. This is not definitive since the APDA like orientation may still be an artefact and there is indeed a large pocket present which would allow binding with the additional group. Secondly, the extracyclic group in the product like arrangement would position the cage group in a subtly different location, conceivably this could prevent binding. The coumarin caging group is much bigger than others we explored, and this does not bind. Visual inspection shows that there is not enough space in the substrate binding site unless the protein undergoes some conformational change (Figure 4.15). These studies give some confidence as to the size limitations placed on any likely caged molecule. The LC-MS assay confirmed that as expected the caged compounds are fully inactive to reaction by the enzyme.

**Table 4.1.** Summary of different synthesized caged compounds that have been tested on ITC for binding affinity with ForT.

| Photocage | Short Name | Structure | Binding (ITC) | Activation (UV) |
|---|---|---|---|---|
| 2-nitrobenzyl-APDA | NB-APDA |  | 26 µM | 365 nm |
| 4,5-dimethoxy-2-nitrobenzyl-APDA | DMO-APDA |  | n.d | 365 nm |
| 4-hydroxyphenacyl-APDA | HPA-APDA |  | 25 µM | X |
| 6,7-dimethoxy-coumarinyl-APDA | DMCO-APDA |  | n.d | 365 nm |
| extracyclic-2-nitrobenzyl-APDA | e-NB-APDA |  | n.d | 365 nm |

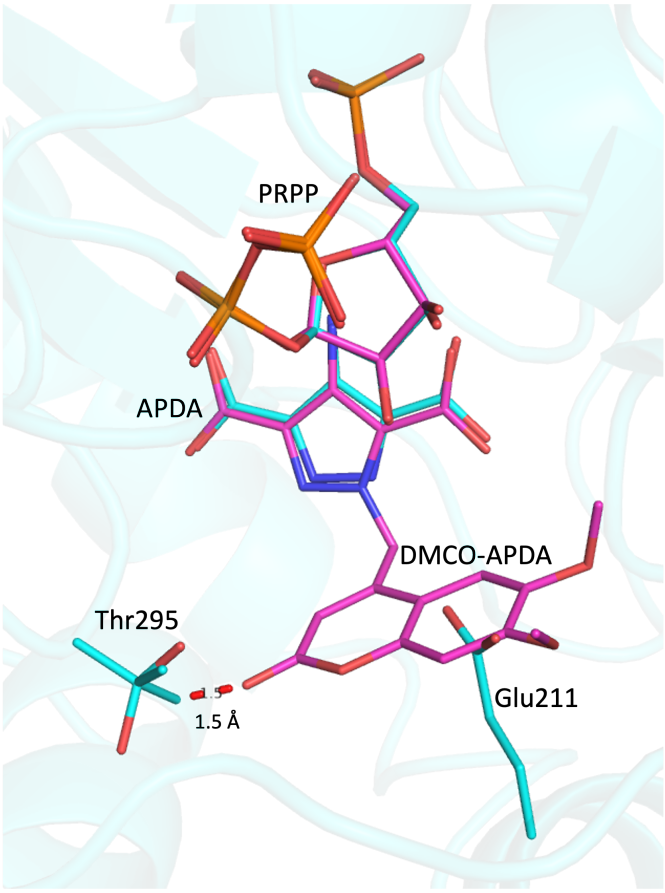



**Figure 4.15.** DMCO-APDA (carbon in pink) superimposed structure with ForT-APDA-PRPP (carbon in cyan) complex structure, shows the coumarin caging group clashes with the enzyme active site side chains Glu211 and Thr295. There is not enough space in the substrate binding site for DMCO-APDA entering unless protein undergoes major conformation change.

In this pilot time-resolved project, in collaboration with Dr. Georgina Girt, we have developed the methods for successful synthesis of photo cage compounds for the enzyme substrate APDA. (Summary of the cage compounds shown in table 4.1) The extracyclic amine cage on the pyrazole ring seem cannot entre or bind into the substrate binding site. The coumarin cage group has been proven too big for entering or binding into the binding cavity by both biophysics measurements and structural modelling. Both NB-APDA and DMO-APDA have shown similar binding affinity with its original substrate APDA with WT ForT. And the LCMS turnover assay performed by Dr. Georgina Girt proofed that the enzyme cannot turnover the caged compounds.

Laser photo activation experiment carried out by Dr. Ajay Jha demonstrated that it was possible to activate the caged compounds with pulse of laser. LC-MS analysis demonstrated only DMO-APDA was successfully un-caged with APDA release. With standard UV absorbance curved plotted against various concentrations of APDA, the yield of un-caged reaction by laser active obtained was around 10.1%. The TA data was calculated and analysed Dr. Ajay Jha, shown the quantum yield of breaking photo sensitive linkage on DMO-APDA is similar at 8%.

For future time resolved XFEL serial crystallography, micro crystallization condition has been screened for the ForT and its variants. T138V and PRPP have shown promising results in producing micro crystals in screening condition. Next step will be using current know condition to add cadged compound for co-crystallization. And eventually transfer the optimized screening condition to batch condition to produce enough ForT-PRPP-caged substrate co-crystals in micro crystals form for XFEL time resolution data collection.

# Chapter Five

# $\beta$-RFAPS Structure and Functions

## 5.1. Introduction

Folate also known as vitamin B-9, is a very important compound in human metabolism, including synthesis and repair of DNA and the formation of red blood cells. In the UK it is advised prior to conception and during early pregnancy as it reduces the risk of birth defects in the central nervous system [249–252]. Since humans and animals cannot synthesise, it is obtained by consuming other species. It is made in plants, fungi, some bacteria, and some archaea. Methanopterin is a folate analogue first identified in Archaea in early 1990s [253–255] where it plays a similar role to folate. The biosynthetic pathway for methanopterin in *Methanosarcia thermophila*, utilises 4-aminobenzoic acid (*p*ABA) and PRPP as the starting compounds [254]. The enzyme $\beta$-RFAPS, is the key enzyme in the pathway and it forms the C-glycosidic bond between *p*ABA and PRPP. $\beta$-RFAPS is a structural homologue of the ForT discussed in preceding chapters. Together with another homologue enzyme PyfQ, they constitute a class of C-glycosidic bond formation enzymes that utilize PRPP for both the ribose moiety and energy. The second (aromatic) substrate is variable reacts with the electrophilic C1' of the ribose to form the C-C glycosidic bond; to date all the natural aromatic substrates undergo a decarboxylation. Figure 5.1 compares the substrate profiles of this class of C-glycosidic bond formation enzymes [256,257].

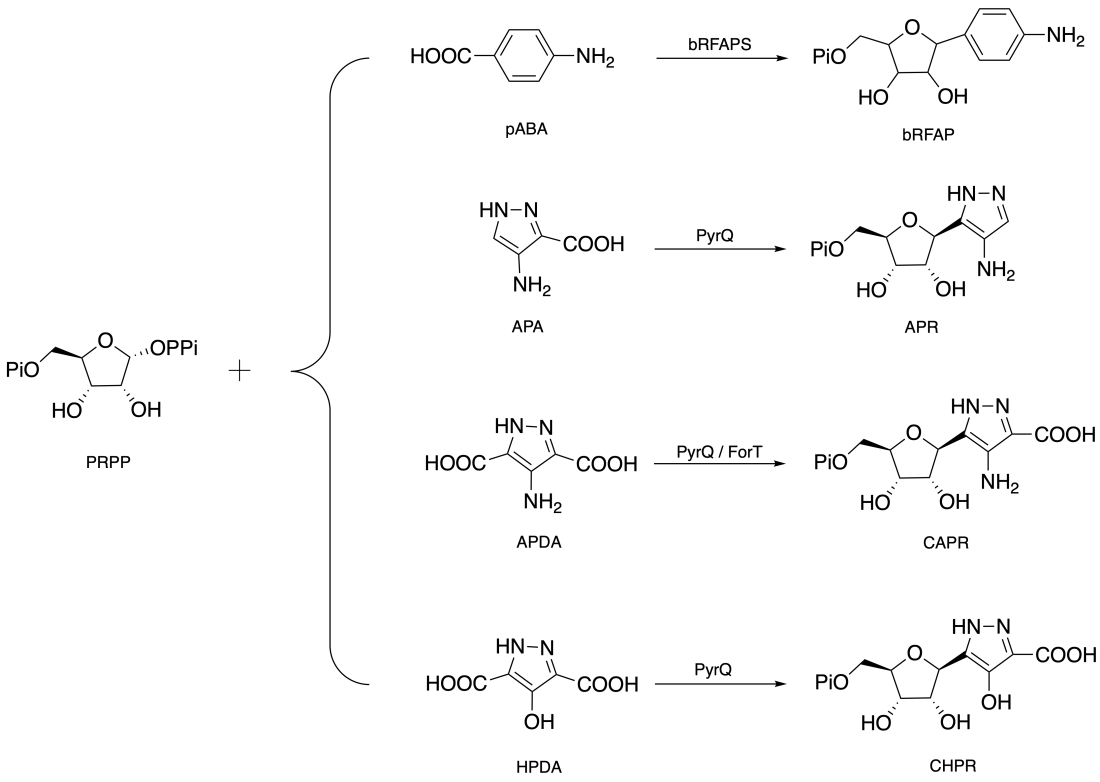



**Figure 5.1.** Aromatic substrates C-glycosides formation enzymes. $\beta$-RFAPS only utilises *p*ABA as the aromatic substrate. ForT accepts only APDA as the second substrate, whilst PyfQ has a wider substrate tolerance and tolerates APA, HPDA and APDA. Figure adapted from [256,257].

This chapter discusses the structural and functional study of this $\beta$-RFAPS enzyme, including identifying an unknown molecule bound to the enzyme active site and the development of an LCMS assay that allows the kinetic parameters of this thermophilic enzyme to be determined.

## 5.2. Material and Methods

### 5.2.1. General

All regents were purchased from Sigma-Aldrich or Fisher Scientific unless specified. Oligonucleotides were purchased from Integrated DNA Technologies. dNTP and DNA polymerase were purchased as part of the EMD Millipore Novagen KOD Hot Start DNA Polymerase kit. DNA sequencing was carried by Source Bioscience.

### 5.2.2. Cloning and Construct of $\beta$-RFAPS and N-terminal Extension

The β-*RFAPS* gene was previously isolated and cloned into pBADHISTEV vector by Dr. Huangting Liu. The construct was designed with N-terminal $His_6$-tag followed by a TEV protease cleavage site [258]. Previous studies by Dr. Sisi Gao found that insertion of a linker sequence (GAGA) between TEV cleavage site and target b-RFAPS protein sequence could improve the protein yield. The new recombinant protein has an N-terminal extension not found in the native enzyme and was termed $\beta$-RFAPS to distinguish it.

### 5.2.3. Over-Expression of Protein

2 µL solution containing 100 ng/µL of plasmid DNA encoding $\beta$-*RFAPS* gene was transformed into *E. Coli* BL21 commercial competent cells and plated onto LB agar containing 100 µg/mL ampicillin. The plate was incubated at 37 ˚ C overnight. A small-scale starter culture was carried out by picking single colony from transformation plate and inoculated into 150 mL LB media with same concentration of ampicillin. After overnight incubation of starter culture at 37 ˚ C at 200 rpm shaking speed, 10 mL of the starter culture was inoculated into 1 L large culture with LB media containing 100 µg/mL of ampicillin. The cultures were incubated at 37 ˚ C and 200 rpm until optical density at 600 nm wavelength ($OD_{600}$) reaches 0.6. L-arabinose was added into cultures at final concentration of 0.3% (v/v) to induce protein expression. Incubation remained at 37 ˚ C and 200 rpm for further 20 hours. The cells were harvested by

centrifugation at 8000 rpm for 20 minutes, and cell pellets were collected and kept at -80 ˚ C for storage and further purification.

### 5.2.4. Protein Purification

Buffer screening for protein isolation was carried out previously by Dr. Sisi Gao. The protein purification protocol that was used by me for the work in thesis is detailed here. Cell pellets was thawed and resuspended in lysis buffer [500 mM NaCl, 50 mM sodium phosphate (pH 7.5), 20 mM imidazole, and 3 mM beta-mercaptoethanol (BME)] with addition of EDTA-free protease inhibitor tablets from Roche (1 tablets per 50 mL lysis buffer), chicken egg lysosome at final concentration 1 mg/mL, and bovine DNase at 0.4 mg per gram of wet cell pellet. The cell re-suspension was lysed by being passed through constant flow cell disruptor at 30K psi at 4 ˚ C twice. The lysate was further centrifuged at 17,000 rpm, 4 ˚ C for 30 minutes to remove cell debris. The upper layer supernatant was collected and heated with water bath at 55 ˚ C for 30 minutes before centrifuge at 17,000 rpm 4 ˚ C for 1 hour to remove protein precipitations. The supernatant was then filtered before loading onto Ni Sepharose 6 FF column (GE Healthcare) equilibrated with lysis buffer. The column was washed three column volumes of washing buffer [500 mM NaCl, 50 mM sodium phosphate (pH 7.5), 30 mM imidazole, and 3 mM BME]. The target protein containing $His_6$-Tag is then eluted with elution buffer [500 mM NaCl, 50 mM sodium phosphate (pH 7.5), 250 mM imidazole, and 3mM BME]. The elution was collected by UV guidance and desalted through desalting column (16/10 Desalting, GE Healthcare) with desalting buffer [100 mM NaCl, 50 mM sodium phosphate (pH 7.5), and 3 mM BME]. Fractions containing target protein were collected and TEV protease was added at mass ratio of 10:1. EDTA was added to a concentration of 2 mM to eliminate any remaining Ni ion in the solution. The protein was digested for 3 hours at 20 ˚ C before being loaded into second nickel column which was equilibrated with desalting buffer with additional 10 mM of imidazole. The flow though was collected and concentrated to 5 mL using Vivaspin® 30 kDa concentrator. The concentrated solution was then applied to Superdex 200 (GE Healthcare) gel filtration column that was preequilibrated with gel filtration buffer [150 mM NaCl, 10 mM HEPES (pH 7.4), 10% glycerol, and 1 mM TCEP]. The gel filtration peak was

analyzed by SDS-PAGE to determine its purity, and the pure proteins were collected, concentrated, aliquoted and flash frozen with liquid nitrogen for storage in -80 ˚ C freezer.

### 5.2.5. Crystallization

The protein was incubated with 5 mM PRPP on ice for two hours before setting up crystallisation experiments. Crystallization condition for $\beta$-RFAPS was screened together with Dr. Sisi Gao using Mosquito (SPT Labtech) and commercial screens. Purified protein was tested for optimal crystallization condition with drop size at 0.1 µL or 0.2 µL using 96-well protein crystallisation plates (MiTeGen in Situ crystallisation plate). The best (judged by visual inspection) crystals were obtained in PACT screen from Molecular Dimensions at the crystallization condition [0.2 M potassium sodium tartrate tetrahydrate, 0.1 M Bis-Tris propane (pH 6.5), 20% (v/v) PEG 3350] with 30 mg/mL protein in 4 ˚ C Formulatrix incubator. Optimisation of crystallisation conditions was performed by varying concentration of PEG 3350 from 15% (v/v) to 25% (v/v). There was no significant difference in visual crystal quality observed amongst the optimisation screens. The crystals were fished and mounted onto a cryo-loop (Molecular Dimensions) at 4 ˚ C with cryo-protectant of mother liquor plus 20% (v/v) glycerol. The crystals were then flash frozen into liquid nitrogen and stored in cryogenic dewar for transporting to Diamond Light Source beamlines I04 and I24 for data collection.

### 5.2.6. Structural Biology

Structural data for $\beta$-RFAPS were collected at Diamond Light Source I04 and I24 and processed through Diamond online processing software DIALS [259] and XIA2 [260] with molecular substitution using selenomethionine labelling carried out previously by Dr. Sisi Gao. The structure was solved to 2.2 Å resolution with a space group of $P6_1$ 2 2. The structure revealed density at the active site that was not PRPP or the product. We were unable to remove the density from the structure by pre-treatment of the protein (dialysis in the presence of denaturants, heating, sonication etc). After many attempts to model ligands into the density, guanosine tetraphosphate (ppGpp) (following work detailed at chapter 5.3.3) was introduced

into electron density in COOT [261] . The final model was refined by Refmac5 [262] in CCP4i2 programme suit.

### 5.2.7. LCMS End Point Assay

Based on previous studies by Rasche and White [256], *p*ABA and PRPP were incubated with $\beta$-RFAPS (25 µM) in sodium phosphate buffer [50 mM sodium phosphate (pH 7.5), 100 mM NaCl, and 5 mM $MgCl_2$] for 1 hour at 55 ˚ C with 200 rpm shaking speed. The reaction was stopped by centrifugation through a 10kDa mini concentrator (at 4 ˚ C where the enzyme is very slow for 20 minutes), the concentrator retains the protein. The flow through was analysed by LC-MS using $C_{18}$ column with 15 minutes gradient of 5-100% acetonitrile and 5 mM ammonium carbonate at a flow rate of 1 mL/min. The mass spectrum was set for positive and negative detection where the formed C-nucleoside product has a m/z = 303.

### 5.2.8. Native Mass Identification of Ligands as Guanosine Tetraphosphate (ppGpp)

After failing to remove or identify the ligand by analysis of electron density or by denaturing the protein and mass spectrometry. Purified $\beta$-RFAPS was buffer-exchanged into 200 mM ammonium acetate buffer (pH 7.4) at a concentration of 0.5 mg/mL using Micro Bio-Spin P-6 column (Bio-Rad). 3 µL of desalted protein was loaded into a glass capillary coated with gold before MS analysis using a modified Q-Exactive mass spectrometer (Thermo Fisher Scientific). For a typical mass spectrometry experiment of $\beta$-RFAPS, the following setting was applied: capillary voltage at 1.2 kV during nanoelectrospray, injection flatapole at 5 V, inter flatapole lens at 4 V, bent flatapole at 2 V, resolution to 17500 at m/z 200, high-energy C-trap dissociation (HCD) cell at 1-10 kV. Data were analysed using Xcalibur 4.4. Errors in measurement were derived from the deviation of peak centroids of differently charged states of the same species.

To detect the dissociation of ligand on native mass spectrometry, 0.5 µL of formic acid was added into 9.5 µL of 200 mM ammonium acetate buffer containing 0.5 mg/mL $\beta$-RFAPS.

Sample was applied for native MS analysis. To detect the association of ligand to $\beta$-RFAPS, FA-treated samples were titrated back to pH 8.0 before native MS analysis.

### 5.2.9. ITC Binding Assay with PRPP

Isothermal titration calorimetry (ITC) tests equilibrium binding experiment between $\beta$-RFAPS and PRPP using a MicroCal PEAQ-ITC (Malvern) machine. Both $\beta$-RFAPS and formic acid treated $\beta$-RFAPS were buffer exchanged into gel filtration buffer [150 mM NaCl, 10 mM HEPES (pH 7.5), and 1 mM TCEP]. PRPP was dissolved into gel filtration buffer to give final concentration of 2.5 mM as a titration solution. The cell chamber of ITC experiment contains 250 µM of protein. The experiments were carried out at a stirring speed of 750 rpm at 55 ˚ C with using a reference power of 5 µcal/s. Control experiment was achieved by a buffer-to-buffer titration to establish a baseline. Data was processed by using MicroCal PEAQ ITC analysis software (Malvern) to obtain a nonlinear one-site binding model to calculate enzyme kinetics information.

### 5.2.10. LCMS Kinetic Assay

$\beta$-RFAPS at final concentration of 50 µM was incubated with *p*ABA (1 – 50 mM) and PRPP (50 mM) in phosphate buffer [100 mM sodium phosphate (pH 7.5), 100 mM NaCl, and 1 mM $MgCl_2$] for 4 hours at 55 ˚ C. Six experiments were carried out with varying *p*ABA concentration at 1 mM, 2.5 mM, 5 mM, 10 mM, 25 mM, and 50 mM. The total reaction volume was 250 µL, and 40 µL of aliquot was taken from reaction tube at each time point of 15 mins, 30 mins, 1 hour, 2 hours, and 4 hours. The reaction was stopped by centrifugation with 10 kDa concentrator at 10,000 g at 4 ˚ C to filter out the protein. The resulting flow through was then analysed by LC-MS, on a Kinetex 2.6 µm XB-C18 100A, 100 X 4.6 mm LC column with a flow rate of 1 mL/min using a step gradient of solvent A [water containing 0.1% TFA] and solvent B [acetonitrile containing 0.1% TFA], with 5% solvent B (0 – 1 min), 5 – 15% solvent B (1 – 2 minutes), 15 – 100% solvent B (2 – 2.2 min). UV detection was set to 254 and 280 nm.

## 5.3. Results

### 5.3.1. Protein Expression and Purification

$\beta$-RFAPS native protein from pEHISTEV vector purification was prone to precipitation and aggregation during purification and gel filtration process. Further studies showed that the same $\beta$-RFAPS protein sequence from a pBADHISTEV vector gave higher yield during purification process. However, the TEV protease was unable to cleave the $His_6$-Tag. We reasoned this was because the $His_6$-Tag linkage may be partially buried by the protein, preventing the protease from accessing the recognition site. A GAGA linker was introduced to the $\beta$-RFAPS protein between His-Tag and TEV protease cleavage site to investigate if this allowed cleavage to occur. The newly constructed protein $\beta$-RFAPS was rapidly cleaved by incubating with TEV protease for 3 hours at room temperature. Since the protein is a thermophile, heat treatment after cell lysis was explored and this did indeed improve the purity of protein. Figure 5.2 A shows the gel filtration profile of $\beta$-RFAPS, and figure 5.2 B shows the SDS-PAGE analysis.

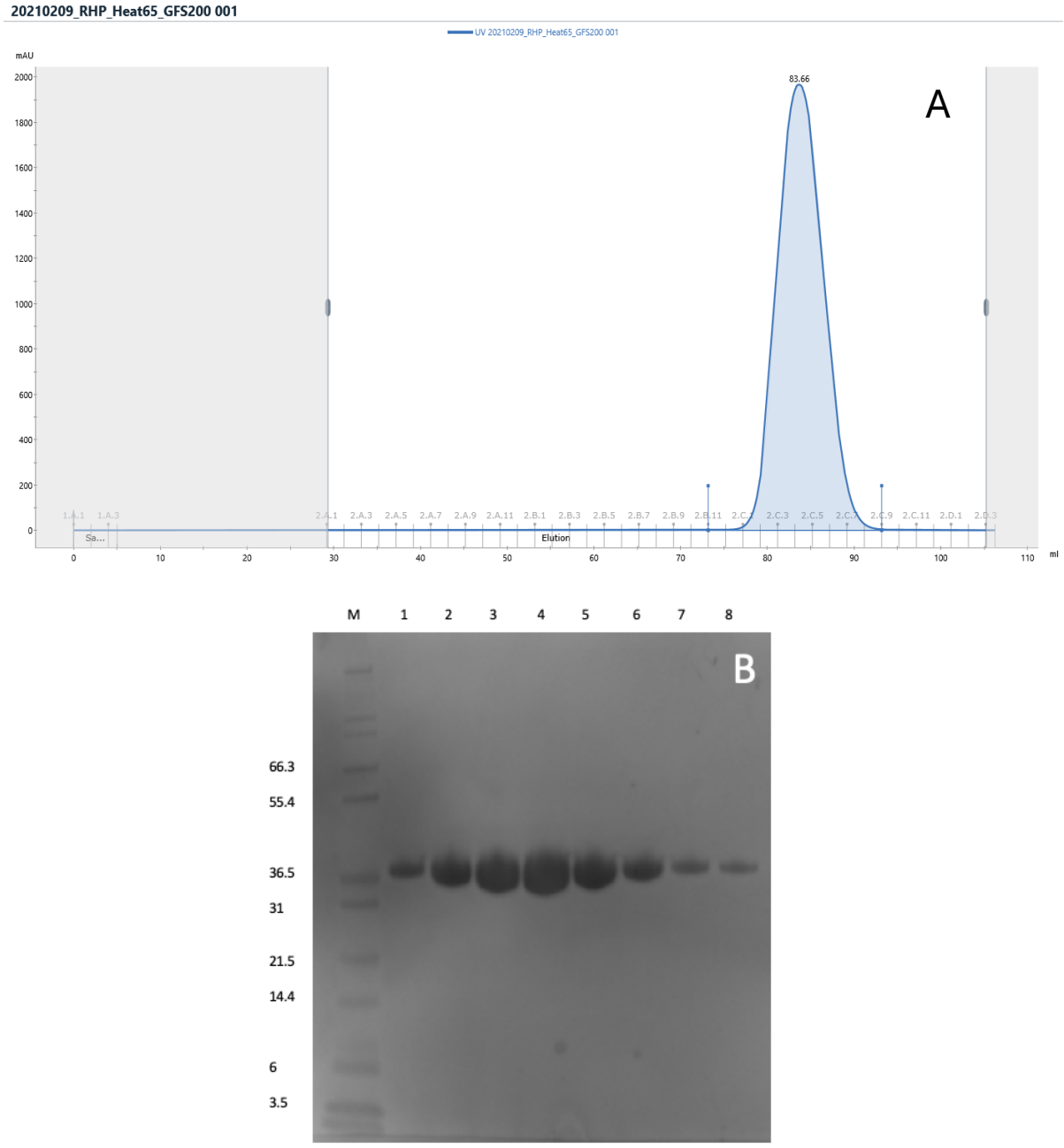



**Figure 5.2.** $\beta$-RFAPS gel filtration (A) shows the corrected elution time for $\beta$-RFAPS is around 80 – 90 ml using Superdex 200 column, shown as a dimer in solution with molecular weight at 72 kDa. (B) SDS PAGE analysis for gel filtration curve samples confirmed the purity for $\beta$-RFAPS, where M is the maker, $\beta$-RFAPS monomer has a molecular weight around 36 kDa.

### 5.3.2. Crystallization

Structural determination for ForT and $\beta$-RFAPS was difficult due to lack of structures with sufficient homology for molecular replacement ($\beta$-RFAPS was crystallised before ForT). SetMet substitution was used by Dr. Sisi Gao to solve the first structure at 2.1 Å resolution. This allowed building an atomic model of $\beta$-RFAPS which was used as molecular replacement model to solve the structure of ForT and subsequent $\beta$-RFAPS experiments. Crystal quality improved by using heat treated (higher purity) $\beta$-RFAPS protein. This resulted in a single morphology of crystals for fishing and data collection. Figure 5.3 shows the process of $\beta$-RFAPS crystal formation. The crystal plates were stored in the 4 ˚ C Formulatrix incubator, and crystals appeared after eight days incubation periods. UV fast images confirmed the observed crystals are protein crystals.

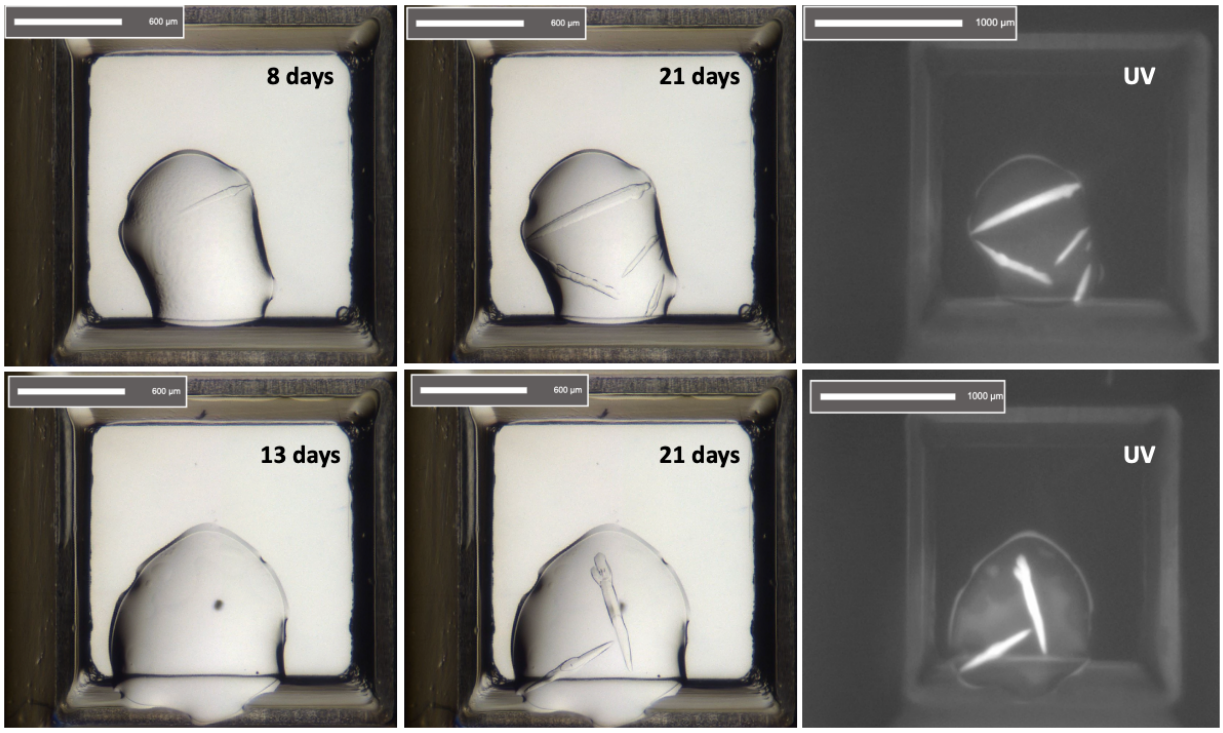



**Figure 5.3.** Crystal for $\beta$-RFAPS with 1 mM PRPP. Single crystals forms between 8 – 21 days.

### 5.3.3. Ligand Identification

Intact mass spectrometry analysis for the denatured $\beta$-RFAPS whole protein was carried out and shown in figure 5.4. The mass of the whole protein is 36568 Da. This is consistent with the theoretical calculated protein mass based on its sequence. This defined the apo mass for the $\beta$-RFAPS protein for native mass analysis experiments.

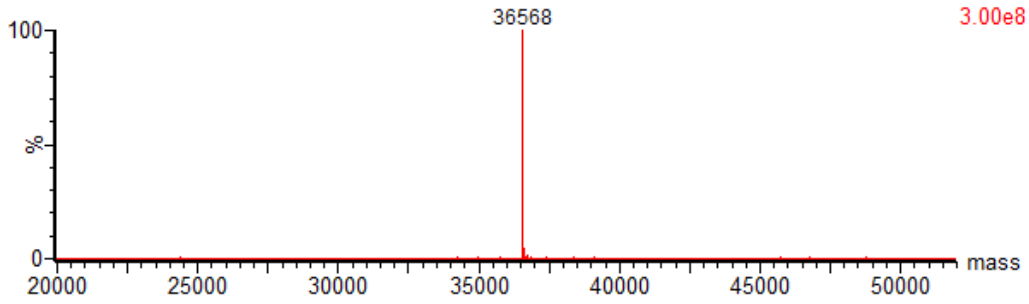



**Figure 5.4.** Intact mass spectrometry analysis for denatured whole protein of $\beta$-RFAPS gives 36568 Da.

Native mass spectrometry experiments were conducted to identify the non-covalent binding molecule. Figure 5.5A shows the deconvolved native mass spectrum using lower energy ionization mode. The protein exists as a homodimer, consistent with crystal packing and gel filtration, with the mass of 74344 Da. Subtraction of the apo mass for the dimer, leaves a mass of 1208 Da. The analysis was carried out in high energy mode (Figure 5.5B), and it can be seen to break down. The monomer mass peak was observed at 37170.7 Da, giving the mass of the ligands of 603 Da. Figure 5.5C shows the native mass data where the protein was treated with 4% formic acid, in order to break dissociate the complex. We also observed the complex which remarkably survived this treatment. This explained our previous failures to remove the ligand, the complex is very tight. More interestingly, figure 5.5D is the native mass analysis for the formic acid treated protein that been pH raised by ammonium. The complex was able to reform, indicating it was indeed non-covalent. With experimental data from the native mass analysis, we concluded the ligand was 603 Da. Together with searching in the small molecule bank with nucleotide and phosphate feature, we concluded that the 603 Da binding molecule in $\beta$-RFAPS was guanosine tetraphosphate (ppGpp).

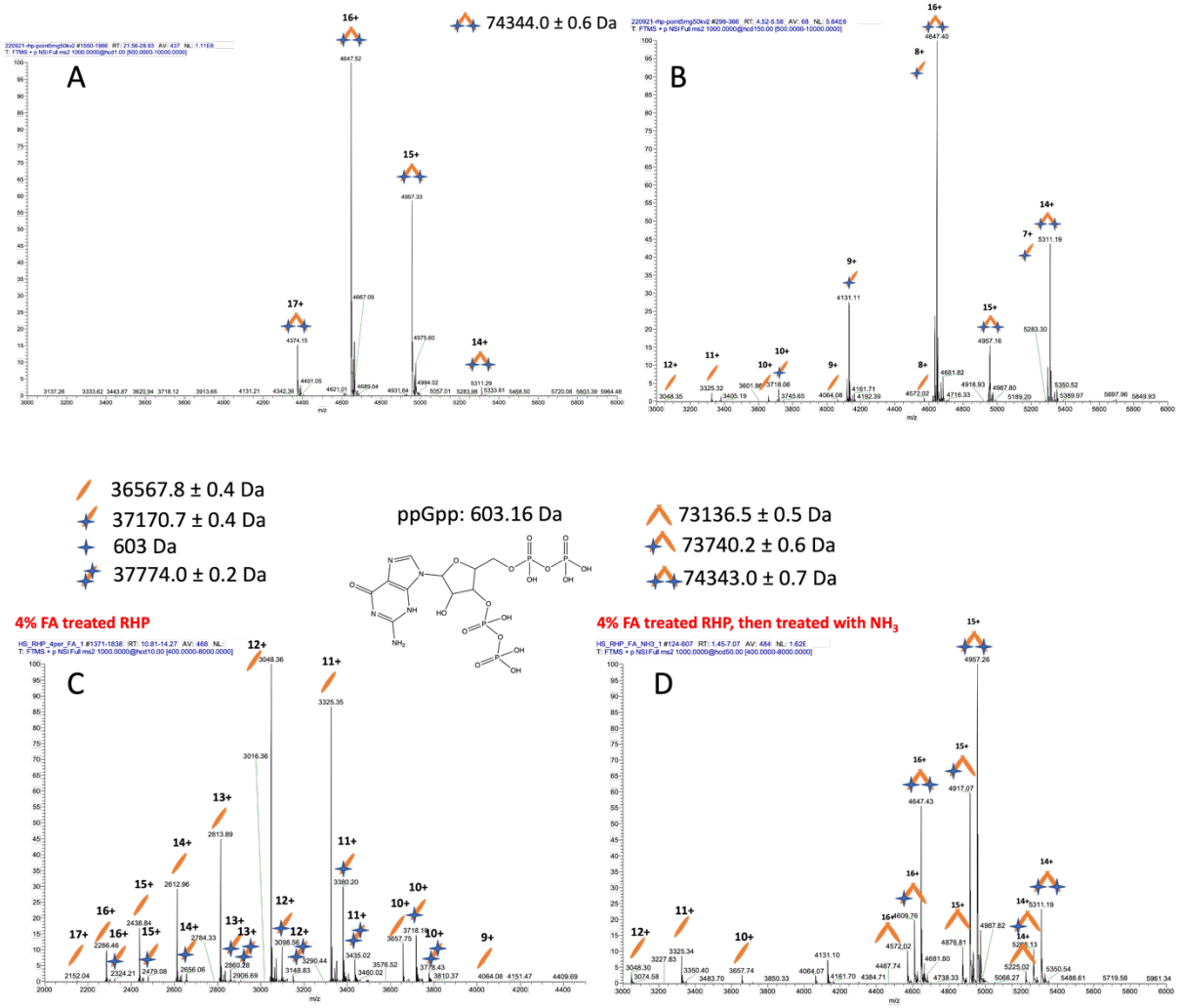



**Figure 5.5.** Native mass deconvolved spectrum analysis for $\beta$-RFAPS to determine its binding partner. Figure A shows the lower energy mass detection, $\beta$-RFAPS protein is a dimer and two ppGpp ligands. Figure B was generated with higher energy, where the dimer of the protein started break down. Figure C reports the protein pre-treated with 4% formic acid. Monomer without the ppGpp can be seen. Figure D shows the ppGpp re-binds to $\beta$-RFAPS when pH is raised. (Data from Dr. Haigang Song)

### 5.3.4. Structural Biology

The data set for $\beta$-RFAPS was collected at Diamond Light Source I03 and processed with Diamond online processing pipeline XIA2 [260] and DIALS [259]. The structure was solved using molecular replacement in PHASER [263] using the previously solved selenomethionine labelled $\beta$-RFAPS structure determined by Dr. Sisi Gao. Crystals of the $\beta$-RFAPS protein belongs to $P6_1$ 2 2, the same as the selenmethionine variant. The data collection and refinement statistics is shown in table 5.1. The structure was refined in CCP4i2 programme suit with COOT [261] and REFMAC5 [262]. $\beta$-RFAPS consists of 333 amino acids with molecular weight at 36.5 kDa. $\beta$-RFAPS consists of 10 alpha helixes and 3 beta sheets. And based on PISA calculation it forms a dimer in the crystal. Sequence alignment between $\beta$-RFAPS and ForT shows 29 % identity with key residues discovered in ForT conserved in $\beta$-RFAPS such as L24, T138 and E211. (Figure 5.7) Figure 5.6 shows the superimposed structure of $\beta$-RFAPS (yellow) and ForT T138V-APDA complex (grey).

**Table 5.1.** Data collection and refinement statistics.

| | **$\beta$-RFAPS** |
|---|---|
| Beamline | I24 (DLS) |
| Wavelength (Å) | 0.9789 |
| Space group | $P6_122$ |
| a, b, c (Å) | 74.0 74.0 253.8 |
| α, β, γ (°) | 90, 90, 120 |
| Resolution range (Å) | 64.11 - 2.20 (2.26 -2.20) |
| $R_{meas}$ | 0.17 (2.32) |
| I/σ(I) | 15.2 (1.5) |
| Completeness (%) | 99.6 (99.1) |
| cc1/2 | 1 (0.56) |
| Multiplicity | 37.1 (39.2) |
| Refinement | |
| No. reflections | 696728 (34727) |
| $R_{work}$ | 0.22 (0.34) |
| $R_{free}$ | 0.28 (0.36) |
| Number of atoms | |
| Total | 2613 |
| Protein | 2529 |
| Ligands / Ions | 33 |
| Waters | 51 |
| B-factor | |
| Protein | 52.5 |
| Ligands / Ions | 70.6 |
| Waters | 45.9 |
| Average | 52.5 |
| R.m.s deviations | |
| Bond lengths (Å) | 0.008 |
| Bond angles (°) | 1.61 |

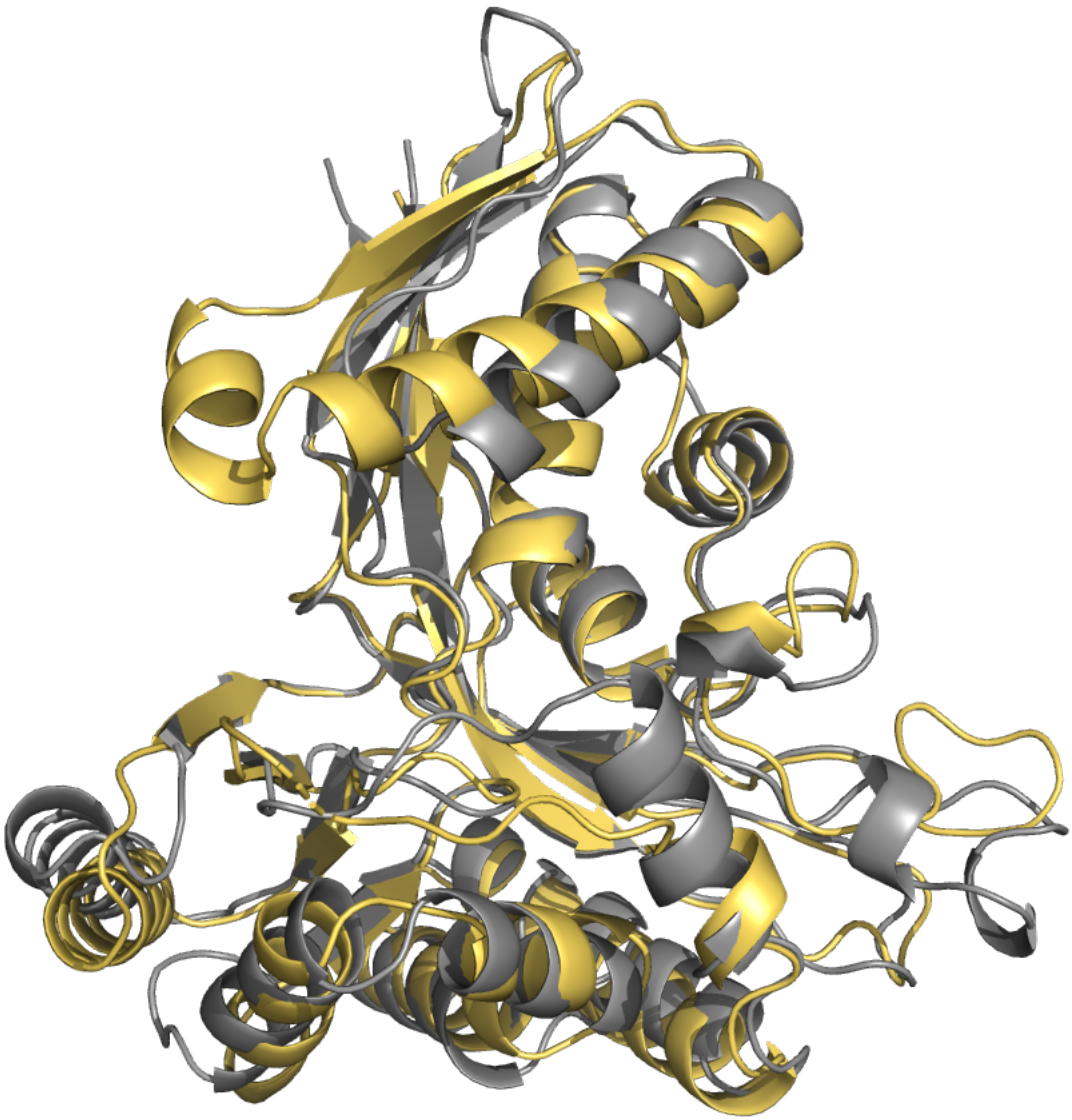


**Figure 5.6.** 3-D superimposed structure between ForT-T138V-APDA complex and $\beta$-RFAPS. $\beta$-RFAPS shows in yellow and ForT-T138V-APDA complex shows in grey.

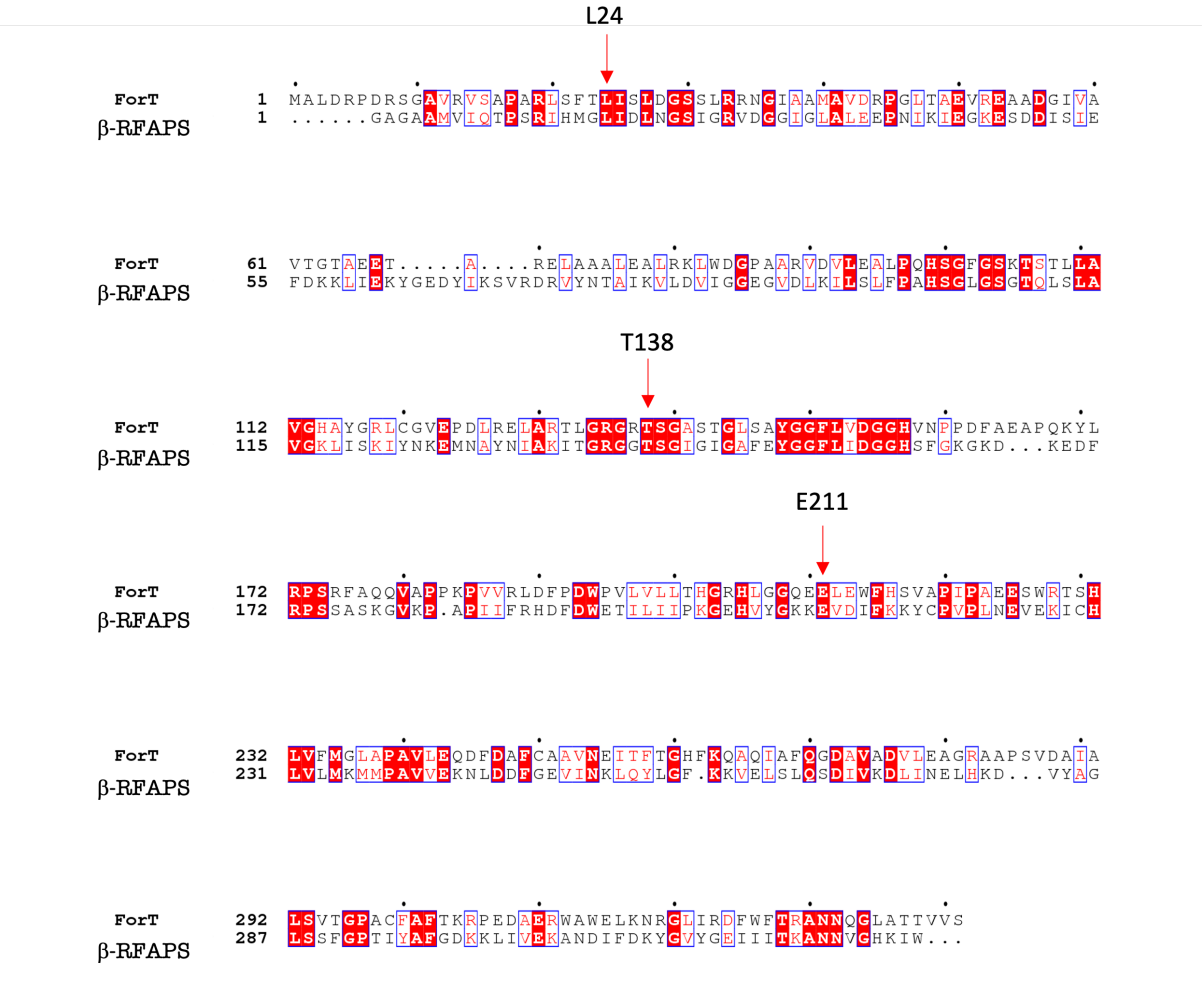



**Figure 5.7.** Sequence alignment between $\beta$-RFAPS and its analogue ForT shows 29.3% sequence identity. Important residues like L24, T138, and E211 from ForT are conserved in $\beta$-RFAPS.

Although $\beta$-RFAPS was incubated with 10 mM PRPP for two hours before setting up crystallization plates, the density was confirmed to ppGpp.

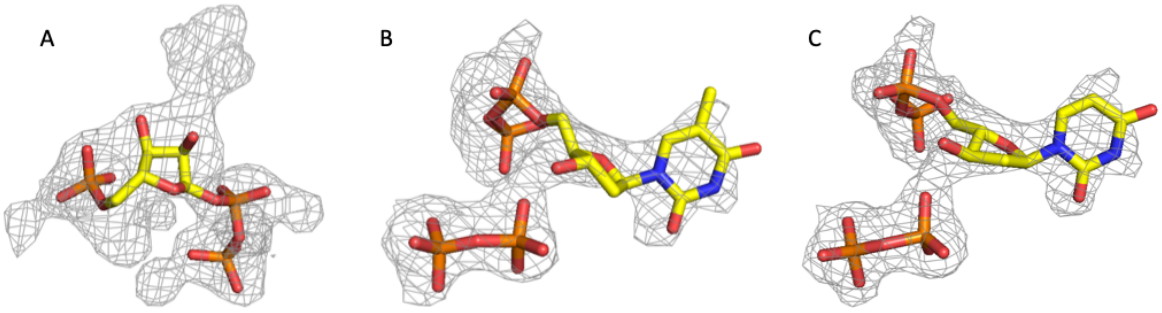



**Figure 5.8.** Attempts to fit different possible ligands to the electron density map. Fitting did not match the density. A: PRPP. B: TDP and PPi. C: UDP and PPi.

Figure 5.9 shows the ppGpp molecule fitted into the $F_o - F_c$ map. The two pyrophosphate binding sites are occupied by ppGpp, and the aromatic ring structure fits in the flat ring like density map. The C3' pyrophosphate group interacts with both enzyme side chains and back bone amide for formation of hydrogen bond and salt bridge. Figure 5.10 A shows Arg9, Ser103 forms salt bridge and hydrogen bond with ppGpp, where His98 and Gly100 interact with ppGpp through its backbone amide. On the other side of the C5' pyrophosphate, ppGpp binds enzyme side chains of Arg9, Thr105, Arg134, Ser138 and Ser170 via hydrogen bonds or salt bridge, and Gly141 uses its amide to form hydrogen bond with ppGpp.

We superimposed the $\beta$-RFAPS structure with ForT-PRPP complex structure. The ppGpp shares the same binding pocket with PRPP in ForT, where C3' pyrophosphate on both ppGpp and PRPP have almost identical position and C5' pyrophosphate from ppGpp shares the same binding pocket with C5' phosphate group from PRPP (Figure 5.11). Figure 5.12 shows the ligand plot of ppGpp inside of the $\beta$-RFAPS active site, the two pyrophosphate groups show extensive interactions with different residues to form strong hydrogen bond network.

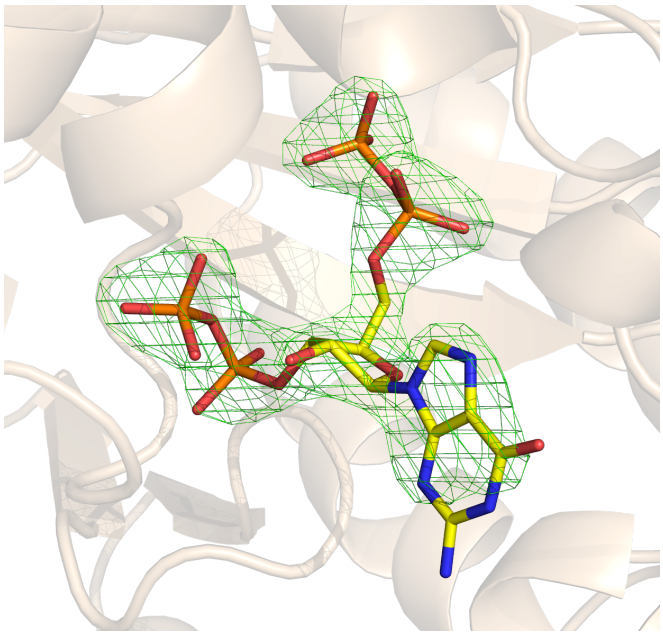


**Figure 5.9.** ppGpp fitment in the previously unexplained density.

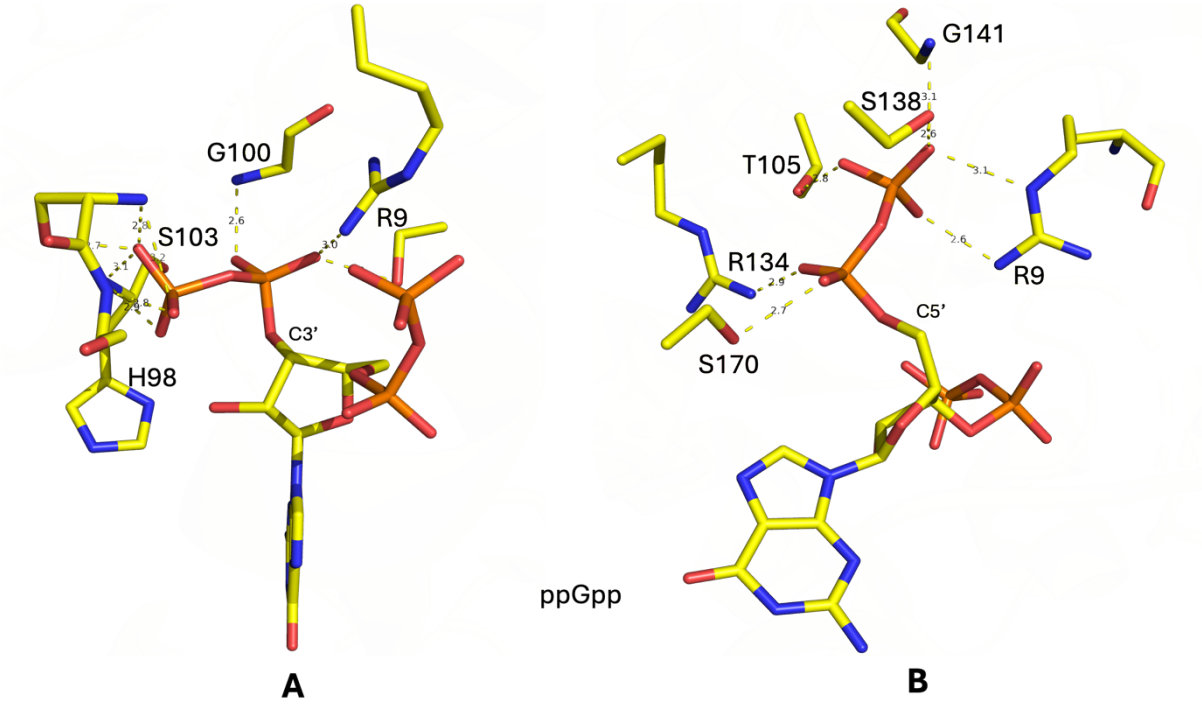



**Figure 5.10.** A: The C3' pyrophosphate interacts with Gly100, Arg9, Ser103 and His98 by hydrogen bond and salt bridges. B: The C5' pyrophosphate binds with Ser170, Arg134, Thr105 Ser138, and Arg9.

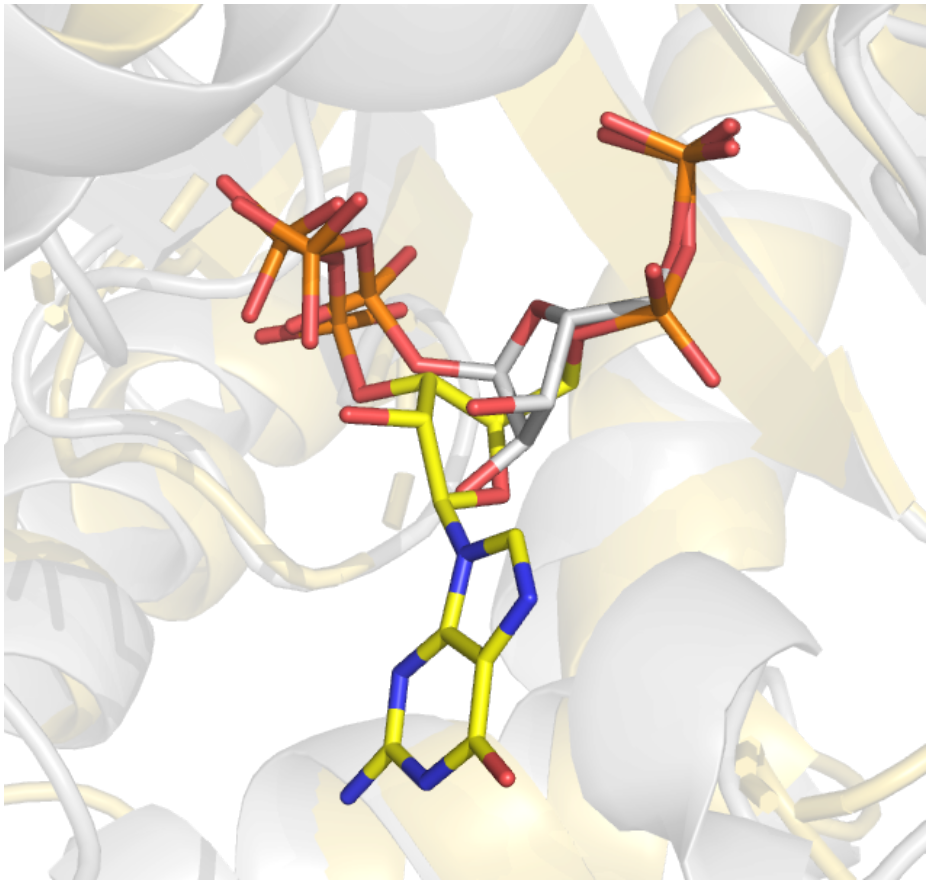


**Figure 5.11.** Superimposed structures of the ForT T138V PRPP complex and $\beta$-RFAPS ppGpp complexes. ppGpp shares same C3' pyrophosphate binding pocket used by PRPP in ForT structure. Part of the C5' pyrophosphate binding pocket in $\beta$-RFAPS occupies the same position as C5' phosphate group from ForT-PRPP complex structure.

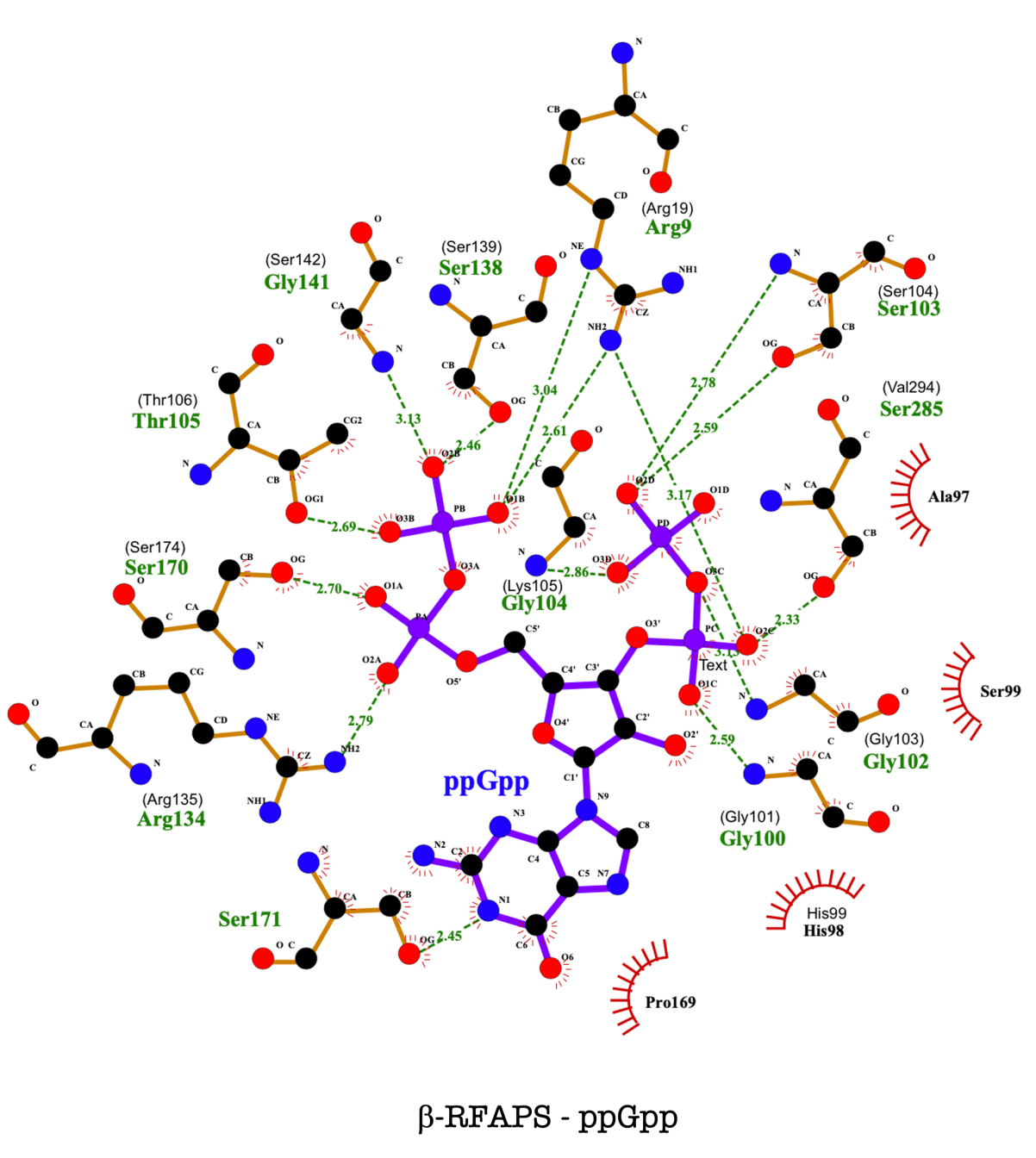
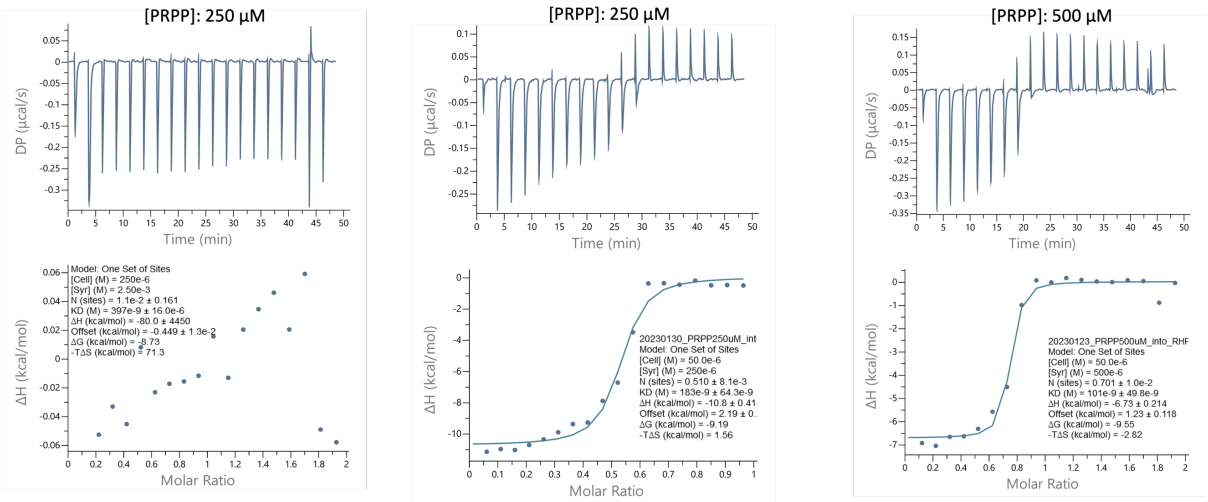



**Figure 5.12.** Ligand plot is generated by LigPlot+ suit [187] shows the interaction between ppGpp and amino acids from $\beta$-RFAPS for interactions. The green dot line indicates the interaction with distance shown in angstroms. In the bracket are the residues equivalent from ForT. The two pyrophosphate groups on ppGpp forms extensive interactions with the binding pocket in $\beta$-RFAPS.

## 5.3.5. ITC Binding Assay with PRPP

PRPP was confirmed as one of the substrates for $\beta$-RFAPS [256]. In the native untreated $\beta$-RFAPS experiment, PRPP with 2.5 mM concentration was injected into cell chamber with $\beta$-RFAPS at 250 µM concentration. There was no heat detected Figure 5.5A. We concluded that there was no binding between PRPP and the as purified protein. This was consistent with ppGpp binding in very tightly at the active site and effectively preventing PRPP binding. It is also consistent with our failure to displace the ligand during purification or crystallisation trials. $\beta$-RFAPS was incubated with 5% formic acid dialysed and then buffer exchanged back to the standard gel filtration buffer. The protein concentration in the chamber was again 50 µM and PRPP in syringe varied between 250 µM and 500 µM. Titration showed clear binding. Data was processed by MicroCal PEAQ ITC analysis software suit with nonlinear one-site binding model and both experiments were in agreement with $K_D$s of PRPP at 183 nM and 101 nM respectively. Figure 5.13 shows the raw ITC for untreated protein and treated protein.




Untreated β-RFAPS

5% FA Treated β-RFAPS

**Figure 5.13.** ITC multiple injection assay testing $K_D$ for PRPP with 5% formic acid treatment and untreated (as purified) $\beta$-RFAPS.

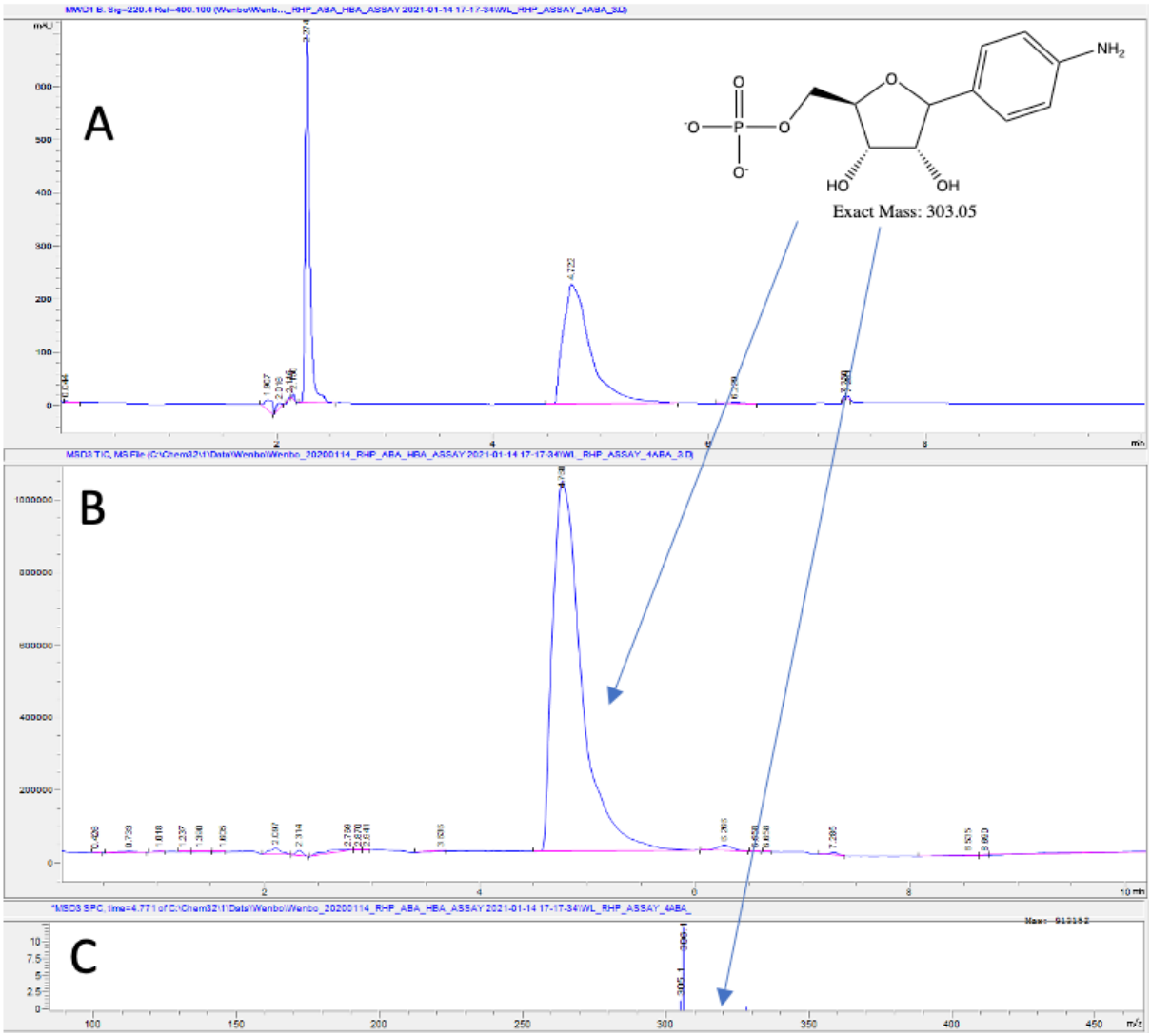



**Figure 5.14.** LC-MS assay confirms the activity of recombinant $\beta$-RFAPS enzyme that catalyse the formation of C-nucleoside from *p*ABA and PRPP in 55˚ C overnight. A. The UV detection at 220 nm wavelength, the resulting C-nucleoside product eluted at 4.7 min. B. The positive mode mass scan for 305, with elution time at 4.7 min. C. Mass detection at 4.7 min, show the mass for product for m/z = 305.1 for $[M + H]^+$.

### 5.3.6. LC-MS Assay

Figure 5.14 shows the resulting product was eluted at 4.7 minutes with confirmed correct mass at m/z = 305.1 $[M + H]^+$. as expected for the product. This confirms that the $\beta$-RFAPS construct is active.

Initial kinetics were assessed by LC-MS analysis with formic acid treated $\beta$-RFAPS (to remove ppGpp) and various *p*ABA concentration from 2.5 mM to 50 mM. With ppGpp eliminated from the enzyme active site, it was possible to observe the enzyme kinetic activity by LC-MS assay. Figure 5.16 shows the plot of product integration from 254 nm against different time points with various *p*ABA concentration. The result shows there is an increase in product with increasing the substrate concentration. At both 2.5 mM and 5 mM concentration of *p*ABA, the product formation has halted, suggesting the reaction is complete under these conditions within 15 minutes. For the reaction with *p*ABA concentration at 10 mM, 25 mM, and 50 mM, product formation continued between 15 mins to 240 mins. Fitting the linear portion of the slope showed that the gradient increased 6.284 to 11.58 and 22.06 respectively. As would be expected within the range where kinetic analysis is meaningful the rate of product formation doubled with doubling the substrate concentration. Figure 5.16 reports product formation (judged by UV) at each of the *p*ABA concentrations after 15 minutes. At *p*ABA concentration below 10 mM, the product formation is linear.

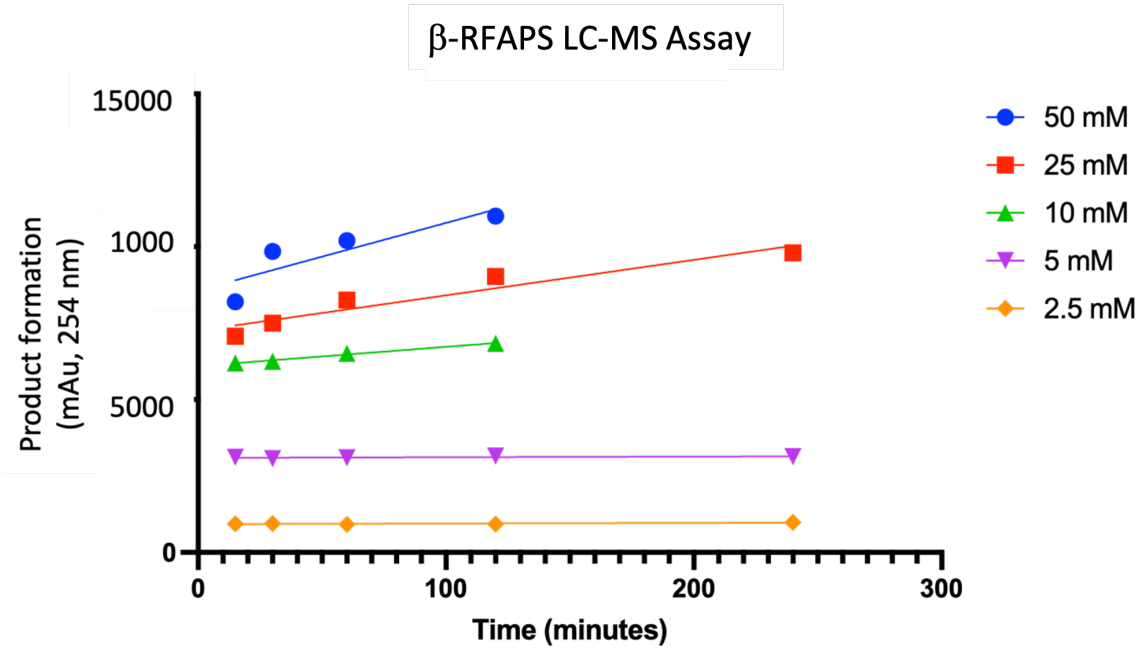



**Figure 5.15.** Initial LC-MS kinetics experiment with various *p*ABA concentration and time points. $\beta$-RFAPS concentration fixed at 50 µM, PRPP set at 50 mM, $MgCl_2$ at 10 mM, and *p*ABA varied at 2.5 mM, 5 mM, 10 mM, 25 mM, and 50 mM respectively. Reaction aliquots were taken out at 15 min, 30 mins, 60 mins, and 120 mins.

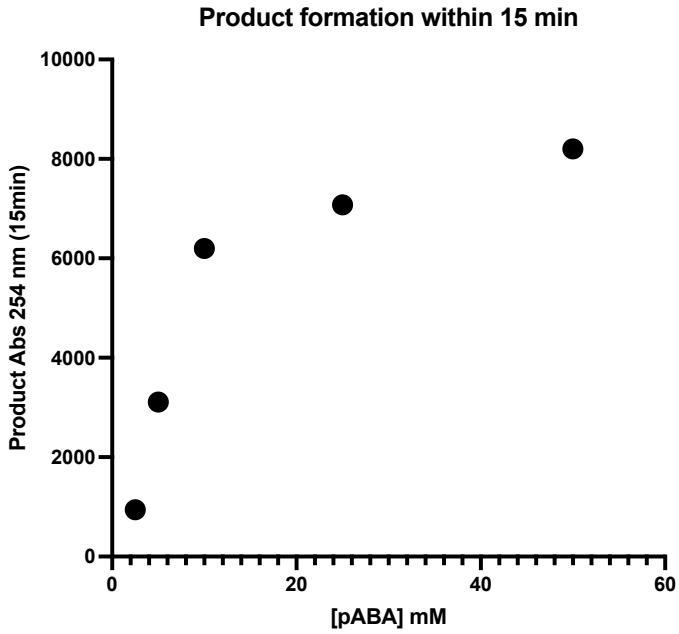



**Figure 5.16.** $\beta$-RFAPS catalysed reaction for 15 mins. Product formation measured at 254 nm UV absorbance. The *p*ABA concentration used are 2.5 mM, 5 mM, 10 mM, 25 mM, and 50 mM.

## 5.4. Discussion

β-RFAPS is the enzyme identified from thermophilic species that makes the C-glycosidic linkage between PRPP and *p*ABA. The process belongs to the methanopterin pathway. The thermal stability of the enzyme is an attractive property that could help in the deployment of the enzyme technologies in the manufacture of novel medicinal C-nucleotides. In order to study the enzyme, we made a variant that was christened $\beta$-RFAPS, the variant has an additional GAGA residue at the N-terminus.

For a significant amount of time, with the unknown molecule inside of the enzyme active site, it is difficult to characterise the enzyme properly. The ITC showed unusual (little binding) results. Although the enzyme had crystallised and its structure determined, there was unmodelled density at the active site. This was assumed to be the cause of the unusual behaviour. We tried extensive attempts to remove the molecule or displace it. None were successful, when the structures of treated proteins were solved, the density was there, with differing clarity. We tried several different ways to unfold the protein and identify any bound ligand by mass spectrometry; none were successful. We concluded that the elusive molecule was bound exceptionally tightly. Nor where we able to fit the density to our satisfaction with the molecules that we thought plausible. At this point, we switched focus to the ForT enzyme.

We revisited the issue through access to native mass spectrometry where both covalent and non-covalent interactions can be observed. Native mass established that the binding molecule in $\beta$-RFAPS has a mass at 603 Da. With cross reference in chemical structures and properties, ppGpp has the exact match mass and did fit well into the electron density map. The native mass spectrometry analysis confirmed the ligand was tightly bound. Under conditions where the intact dimer was seen, there was no dissociation of the molecule. Even when the dimeric structure of the protein began to break down under the conditions in the native mass instrument, the ppGpp molecule remains bound.

The two pyrophosphate groups on ppGpp bind to the pockets intended for the phosphates of PRPP. The extensive interactions (Figure 5.12) between the phosphates and the protein we

suggest accounts for the tightness of the binding. There are relatively few interactions with the ribose or guanidine moieties.

ppGpp is synthesized by bacteria in response to stress[264], normally the molecule is at a very low concentration. It is possible that the over expression of the thermophilic $\beta$-RFAPS does introduce stress. The expression condition is not particularly unusual to cause significantly more stress than ForT (Chapter 2.2.4). Rather we conclude it is the exceptionally high affinity of $\beta$-RFAPS for the ppGpp that is the explanation. The extensive network of interactions between the phosphates and protein (Figure 5.12) supports this. It can be crudely estimated in the low nM to pM regime based on the inability to displace it with PRPP and behaviour in mass spectrometry. The additional N-terminal residues of $\beta$-RFAPS are away from the active site, it is therefore not the reason for tight binding affinity between $\beta$-RFAPS and ppGpp. The binding of ppGpp was so tight that PRPP was unable to displace it. This explains our inability to measure the interaction of as purified $\beta$-RFAPS with PRPP.

However, during the native mass experiments, we discovered that incubating the enzyme with 5% formic acid breaks the complex. Using this insight, we saw an opportunity to eliminate the ppGpp from $\beta$-RFAPS active site for further study. This proved successful and we were able to use ITC to measure the $K_D$ of $\beta$-RFAPS for PRPP as 100 nM, where ForT as 3.6 µM.

With superimposed structure with ForT-PRPP-APDA tenary complex structure, the $\beta$-RFAPS ppGpp complex reveals the phosphate binding sites overlap. The ppGpp has an additional phosphate group and we presume this interaction (absent in PRPP) drive the additional affinity. On the second substrate position, we superimposed the APDA from ForT model and replaced it with *p*ABA. There are two possible binding conformations for the *p*ABA, as there is only one carboxylic group on the molecule. Figure 5.17 A shows the model that *p*ABA's carboxylic pointing towards Glu260, Ser285 and His11 residues, with amino group towards Ser170. The distance between carboxylic group and those residues are at 3.5 – 4 Å which is too far away for hydrogen bond interactions range. Figure 5.17 B shows the other model that *p*ABA's carboxylic group pointing towards Ser170 (Ser174 from ForT) and Glu260 making

contact with the amino group of *p*ABA. The distance is within the hydrogen bond interaction range, and hydroxyl group and backbone amide on Ser170 can form hydrogen bond and salt bridge with p-ABA for binding. Glu260 may acting similarly with Glu211 from ForT that taking a proton away from the amino group and activate the *p*ABA ring to make it as a nucleophile to attack the C1' carbon on the ribose and drive C-C bond formation. The Ser170 loop is ordered in this structure with absence of *p*ABA may because the addition phosphate on ppGpp forms hydrogen bond with the hydroxyl group on Ser170.

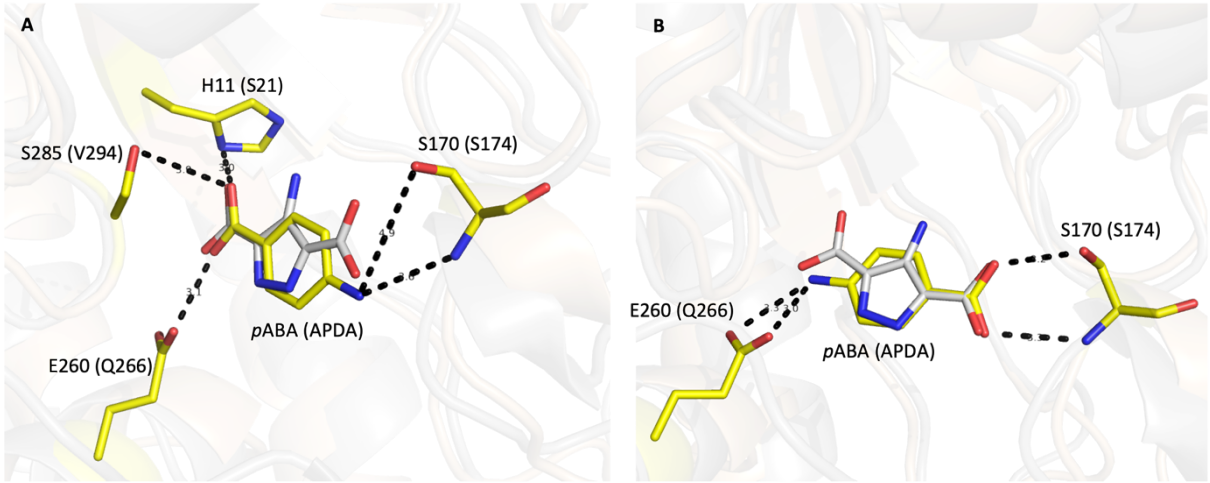



**Figure 5.17.** Superimposed structure of ForT-PRPP-APDA model to find out *p*ABA potential binding conformation. Two possible binding conformation, A shows the carboxylic group pointing toward Glu260, Ser285, and His211. B Shows the carboxylic group making contact with Ser170. The residues in bracket are the equivalent residues in ForT. APDA shown in grey, *p*ABA and $\beta$-RFAPS model shown in yellow.

The catalytic activity of $\beta$-RFAPS was confirmed by LC-MS where we observed the expected C-nucleoside product when the enzyme was incubated with both *p*ABA and PRPP at 55 ˚ C overnight. Initial kinetics assay with $\beta$-RFAPS concentration at 50 µM, PRPP held at 50 mM and *p*ABA varied from 2.5 mM to 50 mM demonstrated Michaelis-Menten like behaviour. The reaction rate increases proportionally with *p*ABA concentration between 10 mM to 50 mM at 15 minutes. Precise kinetics analysis will be difficult because the protocol is technically complex and cumbersome.

Further on this project, more in-depth studies of structural and functional information of this enzyme need to be carried out. The kinetics assay needs to be carried out in further details to obtain the $K_m$ and $k_{cat}$ for the enzyme. $\beta$-RFAPS-substrate and product complexes structure will help to understand the binding and recognition of substrates inside of the active site. We can also design the photo cage substrate with the structural information to make sure that there is enough space for entry and binding.

# Chapter Six

# Summary and future research

## 6.1. Summary

Throughout the entirety of human history, mankind has grappled with metabolic disease and pathogenic infection. Viral, parasitic, yeast and bacterial diseases comprise a significant burden for people across the world. The year 2020 presented the world with a pandemic unprecedented in the modern era, Covid-19 outbreak. The lack of the medicines during the early waves of the virus was starkly apparent. There is currently an urgent unmet need for effective antivirals. Recently, C-nucleoside derivatives, BCX4430 and GS-6620, were shown to possess antiviral activity against Ebola and Hepatitis C. Remdesivir, another C-nucleoside drug, has seen successful utilization in the treatment of Covid-19 in the United States. Biosynthetic routes to such C-nucleoside molecules would be highly valuable. The C-nucleoside, formycin, drew our interest due to its antimicrobial activity. The biosynthetic pathway contains an enzyme, ForT that makes the key C-C glycosidic bond, this bond differentiates the molecule from normal C-N glycosidic bond. A deeper understanding of the enzyme structure and mechanisms holds the potential to facilitate the construction of biosynthetic toolsets through the engineering and modification of the enzymes.

The second overarching objective of the work was to obtain “a molecular movie of enzyme catalysis”, by mapping the time domain through the application of ultrafast X-ray free electron laser X-ray diffraction. From an academic standpoint, the enzyme is appealing target

as it features several bond making and breaking steps (C-O break as PPi leaves, C-C glycoside bond formation, C-O bond formation and C-C bond break the decarboxylation). Moreover, our work has shown that there are multiple large conformational changes (ribose ring).

Our work puts forward a mechanistic proposal and established the groundwork for future XFEL work. ForT has conserved (with the mechanistically related $\beta$-RFAPS, PyrQ enzymes) binding pocket for phosphoribosyl pyrophosphate (PRPP). The binding arrangement of PRPP is also found in the GHMP kinase superfamily. In the presence of a second substrate and $Mg^{2+}$ ion, PRPP undergoes pyrophosphate moiety dissociation, forming a transition state with an oxocarbenium ion on the ribose C1' carbon. The catalytically significant $Mg^{2+}$ ion appears at a distorted octahedral coordination, engaging with the ribose's hydroxyl groups, two oxygen atoms within the pyrophosphate moiety, and two water molecules. This coordination, seen in glutamine phosphoribosylpyrophosphate amidotransferase (GPAT), stabilizes a constrained conformation of PRPP, activating it. There is one major uncertainty in our work that does pose a challenge for our XFEL work, the orientation of the APDA molecule. This will have to resolved. We favour a mechanistic model in which Glu211 removes a proton from the pyrazole via a water molecule thereby activating the pyrazole ring as a nucleophile for the attack the C1' carbon on PRPP (Figure 3.16). A product complex with a different mutant shows a different orientation of the APDA (a flip) in which Glu211 could still abstract a proton but from the extracyclic nitrogen (Figure 3.14).

The orientation of the APDA is critical in correctly designing the caged compounds. Consistent with our favoured mechanism, we were able to design and synthesis caged compounds on the ring nitrogen. These molecules bound but did not turnover and were activated by a laser pulse although the quantum yield was below 10%.

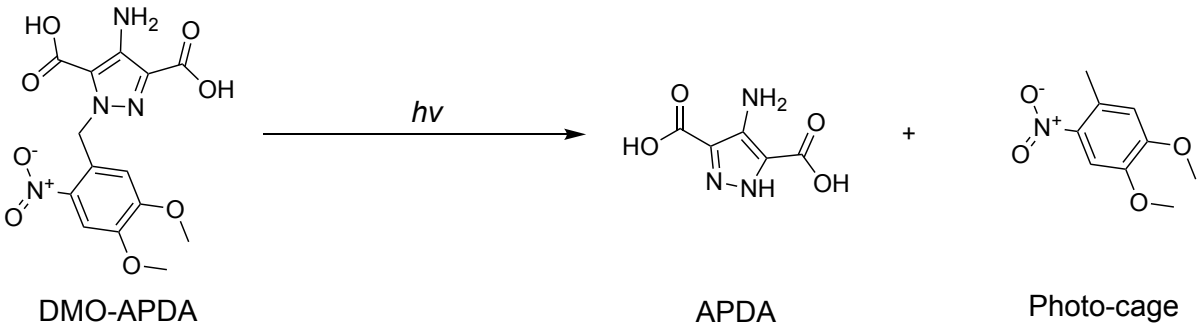



**Figure 6.2.** Proposed laser activated substrate release of APDA inside of the enzyme active site.

$\beta$-RFAPS, related to ForT, found in the methanopterin biosynthesis pathway that we originally worked on but struggled because of an unidentified density at the active site. We finally identified the molecule guanosine tetraphosphate (ppGpp) and found a means by which to remove this molecule from the active site. Removing it allowed us to determine binding and to demonstrate enzyme turnover.

## 6.2. Future research plans

It is clear that definitively answering the question as to the orientation of the APDA at the active site is the key next step. To date, our attempts to answer this have tended to favour the orientation observed for APDA at the active site. My suggestion would be repeat both sets crystals with the two caged molecules and determine the structures using the synchrotron after a series of activating pulses. This is different from time resolved structures, rather the plan is to convert enough caged compound into substrate to observe the result. If the caged compounds are bound in a non-active conformation, then we would expect to see substrate bound to the enzyme. If the compound is bound correctly, we would expect to see evidence for the product. The hypothetical risks with this approach are that it turns out be impossible to convert enough substrate without damaging the protein. Secondly, even if enough product can be converted, there is a risk the electron density is unclear to be unambiguous. Finally, there is a risk that the protein crystal packing prevents turnover, misleading us. If the compound is bound incorrectly, then a major redesign of the caged compounds would be necessary. All of the compounds synthesized which were caged on the extracyclic nitrogen (necessary if the less favoured orientation is correct) did not bind to the protein. This might require some further protein mutagenesis to enlarge the pocket since all extracyclic nitrogen substituents to date have not bound.

If the favoured orientation is correct, then there are two principal challenges for ForT. The most important hurdle is that the caged compound will need to be redesigned to improve quantum yield. Ideally, we would desire the quantum yield to be 40 % or above, otherwise density maps may be challenging to interpret. Any new compound would need to be designed, synthesized tested for binding, turnover, and quantum yield. The next challenge is to prepare sufficient quantities of micro crystals for the XFEL experiments. There was promising data suggesting this would be possible, but of course this depends on the new cage compound. Two photon uncaging mechanism can be further explored, as it has the opportunity to shift the activation wavelength away from the protein absorbance wavelength to avoid energy waste or protein damage. This could allow higher fluxes of the laser, however, there the quantum yield can be lower for such two photon approaches.

If the ForT route is blocked, the $\beta$-RFAPS enzyme is an attractive replacement in some ways. As a thermophile, the reaction may be considerably slower at room temperature, thus allowing building of intermediates. The amino benzoic acid substrate may be more synthetically tractable, and this offers the opportunity to create caged compounds. A first step in the work would be obtain more detailed understanding of the enzyme kinetics, of $K_m$ and $k_{cat}$. From a structural viewpoint, substrate complex and site directed mutagenesis will be key in exploring the tolerance for caging groups. A simple structural model does not predict a role for Glu211 equivalent in $\beta$-RFAPS.

In a similar vein the more promiscuous PyrQ enzyme is worth investigating. The structural basis for this tolerance is unclear and hints at a lack of insight from our studies. By investigating this system more thoroughly, we will shed light on both ForT and $\beta$-RFAPS. It may also clarify the role of Glu211 and the orientation of the extracyclic nitrogen. Because PyrQ is more promiscuous it may allow a bigger but more reactive caging group (such as coumarin) to be bound. The synthetic scope of this enzyme has not been explored and the ability to make C-C glycosidic bonds is valuable. A details SAR for the enzyme could reveal important biotechnological applications.

# Bibliography


1. Atanasov, A. G. *et al.* Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnol Adv* **33**, 1582–1614 (2015).
2. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* **14**, 111–129 (2015).
3. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* **79**, 629–661 (2016).
4. Waltenberger, B., Mocan, A., Šmejkal, K., Heiss, E. H. & Atanasov, A. G. Natural products to counteract the epidemic of cardiovascular and metabolic disorders. *Molecules* **21**, (2016).
5. Tintore, M., Vidal-Jordana, A. & Sastre-Garriga, J. Treatment of multiple sclerosis - success from bench to bedside. *Nat Rev Neurol* **15**, 53–58 (2019).
6. Mishra, B. B. & Tiwari, V. K. Natural products: an evolving role in future drug discovery. *Eur J Med Chem* **46**, 4769–4807 (2011).
7. Rey-Ladino, J., Ross, A. G., Cripps, A. W., McManus, D. P. & Quinn, R. Natural products and the search for novel vaccine adjuvants. *Vaccine* **29**, 6464–6471 (2011).
8. Cragg, G. M. & Newman, D. J. Biodiversity: A continuing source of novel drug leads. *Pure and Applied Chemistry* **77**, 7–24 (2005).
9. Haefner, B. Drugs from the deep: marine natural products as drug candidates. *Drug Discov Today* **8**, 536–544 (2003).
10. Butler, M. S. The role of natural product chemistry in drug discovery. *J Nat Prod* **67**, 2141–2153 (2004).
11. Feher, M. & Schmidt, J. M. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. *J Chem Inf Comput Sci* **43**, 218–227 (2003).
12. Barnes, E. C., Kumar, R. & Davis, R. A. The use of isolated natural products as scaffolds for the generation of chemically diverse screening libraries for drug discovery. *Nat. Prod. Rep.* **33**, 372–381 (2016).
13. Li, J. W.-H. & Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**, 161–165 (2009).
14. Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829–837 (2004).
15. Atanasov, A. G. *et al.* Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* **20**, 200–216 (2021).

16. Dias, D. A., Urban, S. & Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2**, 303–336 (2012).
17. Kinghorn, A. D., Pan, L., Fletcher, J. N. & Chai, H. The relevance of higher plants in lead compound discovery programs. *J Nat Prod* **74**, 1539–1555 (2011).
18. Maplestone, R. A., Stone, M. J. & Williams, D. H. The evolutionary role of secondary metabolites--a review. *Gene* **115**, 151–157 (1992).
19. McRae, J., Yang, Q., Crawford, R. & Palombo, E. Review of the methods used for isolating pharmaceutical lead compounds from traditional medicinal plants. *Environmentalist* **27**, 165–174 (2007).
20. Fabricant, D. S. & Farnsworth, N. R. The value of plants used in traditional medicine for drug discovery. *Environ Health Perspect* **109 Suppl**, 69–75 (2001).
21. Farnsworth, N. R., Akerele, O., Bingel, A. S., Soejarto, D. D. & Guo, Z. Medicinal plants in therapy. *Bull World Health Organ* **63**, 965–981 (1985).
22. Su, X.-Z. & Miller, L. H. The discovery of artemisinin and the Nobel Prize in Physiology or Medicine. *Science China. Life sciences* vol. 58 1175–1179 Preprint at https://doi.org/10.1007/s11427-015-4948-7 (2015).
23. Cragg, G. M. Paclitaxel (Taxol): a success story with valuable lessons for natural product drug discovery and development. *Med Res Rev* **18**, 315–331 (1998).
24. Gallego-Jara, J., Lozano-Terol, G., Sola-Martínez, R. A., Cánovas-Díaz, M. & de Diego Puente, T. A Compressive Review about Taxol(®): History and Future Challenges. *Molecules* **25**, (2020).
25. Nicolaou, K. C. *et al.* Total synthesis of taxol. *Nature* **367**, 630–634 (1994).
26. Cseke, L. J. *et al. Natural Products from Plants*. (CRC Press, 2016). doi:10.1201/9781420004472.
27. Nahar, L. *et al.* Naturally Occurring Calanolides: Occurrence, Biosynthesis, and Pharmacological Properties Including Therapeutic Potential. *Molecules* **25**, (2020).
28. Zhang, L. *et al.* [HPLC enantioseparation, absolute configuration determination and anti-HIV-1 activity of (±)-F18 enantiomers]. *Yao Xue Xue Bao* **50**, 733–737 (2015).
29. Wu, X. *et al.* Metabolism of F18, a Derivative of Calanolide A, in Human Liver Microsomes and Cytosol. *Front Pharmacol* **8**, 479 (2017).
30. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **115**, 6506–6511 (2018).
31. Delgado-Baquerizo, M. *et al.* A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).
32. Hoffmann, T. *et al.* Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat Commun* **9**, 803 (2018).
33. Hug, J. J., Krug, D. & Müller, R. Bacteria as genetically programmable producers of bioactive natural products. *Nat Rev Chem* **4**, 172–193 (2020).
34. Gray, A. M. & Flatt, P. R. Insulin-releasing and insulin-like activity of Agaricus campestris (mushroom). *J Endocrinol* **157**, 259–266 (1998).
35. Müller, K. Pharmaceutically relevant metabolites from lichens. *Appl Microbiol Biotechnol* **56**, 9–16 (2001).
36. Mann, J. *Murder, Magic, and Medicine*. (Oxford University Press, 1994).
37. Abraham, E. P. *et al.* Further observations on penicillin. 1941. *Eur J Clin Pharmacol* **42**, 3–9 (1992).

38. Lax, E. *The Mold in Dr. Florey's Coat: The Story of the Penicillin Miracle*. (Henry Holt and Company, 2005).
39. Wainwright, M. *Miracle Cure: The Story of Penicillin and the Golden Age of Antibiotics*. (Blackwell, 1990).
40. Mann, J. *The Elusive Magic Bullet: The Search for the Perfect Drug*. (Oxford University Press, 1999).
41. Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *J Anaesthesiol Clin Pharmacol* **33**, 300–305 (2017).
42. Tooke, C. L. *et al.* β-Lactamases and β-Lactamase Inhibitors in the 21st Century. *J Mol Biol* **431**, 3472–3500 (2019).
43. Williams, J. D. Beta-lactamases and beta-lactamase inhibitors. *Int J Antimicrob Agents* **12 Suppl 1**, S3-7; discussion S26-7 (1999).
44. Fabbretti, A., Gualerzi, C. O. & Brandi, L. How to cope with the quest for new antibiotics. *FEBS Lett* **585**, 1673–1681 (2011).
45. Stogios, P. J. & Savchenko, A. Molecular mechanisms of vancomycin resistance. *Protein Science* **29**, 654–669 (2020).
46. Choroba, O. W., Williams, D. H. & Spencer, J. B. Biosynthesis of the Vancomycin Group of Antibiotics: Involvement of an Unusual Dioxygenase in the Pathway to (S)-4-Hydroxyphenylglycine. *J Am Chem Soc* **122**, 5389–5390 (2000).
47. Moore, M. J. *et al.* Next-Generation Total Synthesis of Vancomycin. *J Am Chem Soc* **142**, 16039–16050 (2020).
48. Nambiar, S. *et al.* Product quality of parenteral vancomycin products in the United States. *Antimicrob Agents Chemother* **56**, 2819–2823 (2012).
49. Taymaz-Nikerel, H., Karabekmez, M. E., Eraslan, S. & Kırdar, B. Doxorubicin induces an extensive transcriptional and metabolic rewiring in yeast cells. *Sci Rep* **8**, 13672 (2018).
50. Cameron of Sunderland., J. The Gaelic names of plants (Scottish, Irish, and Manx), collected and arranged in scientific order, with notes on their etymology, uses, plant superstitions, etc., among the Celts, with copious Gaelic, English, and scientific indices. Preprint at (1900).
51. Newman, D. Screening and identification of novel biologically active natural compounds. *F1000Res* **6**, 783 (2017).
52. Zähner, H. Some aspects of antibiotics research. *Angew Chem Int Ed Engl* **16**, 687–694 (1977).
53. Schiewe, H. J. & Zeeck, A. Cineromycins, gamma-butyrolactones and ansamycins by analysis of the secondary metabolite pattern created by a single strain of Streptomyces. *J Antibiot (Tokyo)* **52**, 635–642 (1999).
54. Hemphill, C. F. P. *et al.* OSMAC approach leads to new fusarielin metabolites from Fusarium tricinctum. *J Antibiot (Tokyo)* **70**, 726–732 (2017).
55. Hussain, A. *et al.* Novel bioactive molecules from Lentzea violacea strain AS 08 using one strain-many compounds (OSMAC) approach. *Bioorg Med Chem Lett* **27**, 2579–2582 (2017).
56. Vartoukian, S. R., Palmer, R. M. & Wade, W. G. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol Lett* **309**, 1–7 (2010).

57. Abdel-Razek, A. S., Hamed, A., Frese, M., Sewald, N. & Shaaban, M. Penicisteroid C: New polyoxygenated steroid produced by co-culturing of Streptomyces piomogenus with Aspergillus niger. *Steroids* **138**, 21–25 (2018).
58. Moussa, M. *et al.* Co-culture of the fungus Fusarium tricinctum with Streptomyces lividans induces production of cryptic naphthoquinone dimers. *RSC Adv.* **9**, 1491–1500 (2019).
59. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).
60. Cross, K. L. *et al.* Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol* **37**, 1314–1321 (2019).
61. Nichols, D. *et al.* Use of ichip for high-throughput in situ cultivation of 'uncultivable' microbial species. *Appl Environ Microbiol* **76**, 2445–2450 (2010).
62. Wolfender, J.-L., Nuzillard, J.-M., van der Hooft, J. J. J., Renault, J.-H. & Bertrand, S. Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling, in Silico Databases, and Chemometrics. *Anal Chem* **91**, 704–742 (2019).
63. Stuart, K. A., Welsh, K., Walker, M. C. & Edrada-Ebel, R. Metabolomic tools used in marine natural product drug discovery. *Expert Opin Drug Discov* **15**, 499–522 (2020).
64. Allard, P.-M., Genta-Jouve, G. & Wolfender, J.-L. Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification. *Curr Opin Chem Biol* **36**, 40–49 (2017).
65. Eugster, P. J. *et al.* Ultra high pressure liquid chromatography for crude plant extract profiling. *J AOAC Int* **94**, 51–70 (2011).
66. Stavrianidi, A. A classification of liquid chromatography mass spectrometry techniques for evaluation of chemical composition and quality control of traditional medicines. *J Chromatogr A* **1609**, 460501 (2020).
67. Tahtah, Y. *et al.* High-resolution PTP1B inhibition profiling combined with high-performance liquid chromatography-high-resolution mass spectrometry-solid-phase extraction-nuclear magnetic resonance spectroscopy: Proof-of-concept and antidiabetic constituents in crude extr. *Fitoterapia* **110**, 52–58 (2016).
68. Liu, B. *et al.* Dual High-Resolution α-Glucosidase and Radical Scavenging Profiling Combined with HPLC-HRMS-SPE-NMR for Identification of Minor and Major Constituents Directly from the Crude Extract of Pueraria lobata. *J Nat Prod* **78**, 294–300 (2015).
69. Garcia-Perez, I. *et al.* Identifying unknown metabolites using NMR-based metabolic profiling techniques. *Nat Protoc* **15**, 2538–2567 (2020).
70. Wolfender, J.-L., Marti, G., Thomas, A. & Bertrand, S. Current approaches and challenges for the metabolite profiling of complex natural extracts. *J Chromatogr A* **1382**, 136–164 (2015).
71. Wishart, D. S. NMR metabolomics: A look ahead. *J Magn Reson* **306**, 155–161 (2019).
72. Berlinck, R. G. S. *et al.* Approaches for the isolation and identification of hydrophilic, light-sensitive, volatile and minor natural products. *Nat Prod Rep* **36**, 981–1004 (2019).

73. Jones, C. G. *et al.* The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent Sci* **4**, 1587–1592 (2018).
74. Ting, C. P. *et al.* Use of a scaffold peptide in the biosynthesis of amino acid-derived natural products. *Science* **365**, 280–284 (2019).
75. Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).
76. Viehrig, K. *et al.* Structure and Biosynthesis of Crocagins: Polycyclic Posttranslationally Modified Ribosomal Peptides from Chondromyces crocatus. *Angew Chem Int Ed Engl* **56**, 7407–7410 (2017).
77. Surup, F. *et al.* Crocadepsins-Depsipeptides from the Myxobacterium Chondromyces crocatus Found by a Genome Mining Approach. *ACS Chem Biol* **13**, 267–272 (2018).
78. Kayrouz, C. M., Zhang, Y., Pham, T. M. & Ju, K.-S. Genome Mining Reveals the Phosphonoalamide Natural Products and a New Route in Phosphonic Acid Biosynthesis. *ACS Chem Biol* **15**, 1921–1929 (2020).
79. Weber, T. & Kim, H. U. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol* **1**, 69–79 (2016).
80. Laureti, L. *et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens. *Proc Natl Acad Sci U S A* **108**, 6258–6263 (2011).
81. Ganesh, T. *et al.* Evaluation of the tubulin-bound paclitaxel conformation: synthesis, biology, and SAR studies of C-4 to C-3' bridged paclitaxel analogues. *J Med Chem* **50**, 713–725 (2007).
82. Durand, G. A., Raoult, D. & Dubourg, G. Antibiotic discovery: history, methods and perspectives. *Int J Antimicrob Agents* **53**, 371–382 (2019).
83. Ekstein, D. & Schachter, S. C. Natural products in epilepsy-the present situation and perspectives for the future. *Pharmaceuticals* **3**, 1426–1445 (2010).
84. Ribnicky, D. M. *et al.* Plants and Human Health in the Twenty-first Century. *Trends Biotechnol* **20**, 522–531 (2002).
85. Seley-Radtke, K. L. & Yates, M. K. The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Res* **154**, 66–86 (2018).
86. De Clercq, E. Antiviral drug discovery and development: Where chemistry meets with biomedicine. *Antiviral Res* **67**, 56–75 (2005).
87. De Clercq, E. Recent highlights in the development of new antiviral drugs. *Curr Opin Microbiol* **8**, 552–560 (2005).
88. Galmarini, C. M., Mackey, J. R. & Dumontet, C. Nucleoside analogues: Mechanisms of drug resistance and reversal strategies. *Leukemia* **15**, 875–890 (2001).
89. Tamborini, L. *et al.* An enzymatic flow-based preparative route to vidarabine. *Molecules* **25**, (2020).
90. Eyer, L., Nencka, R., de Clercq, E., Seley-Radtke, K. & Růžek, D. Nucleoside analogs as a rich source of antiviral agents active against arthropod-borne flaviviruses. *Antivir Chem Chemother* **26**, 1–28 (2018).
91. Drohat, A. C. & Maiti, A. Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA. *Org Biomol Chem* **12**, 8367–8378 (2014).

92. Stambaský, J., Hocek, M. & Kočovský, P. C-Nucleosides: Synthetic Strategies and Biological Applications. *Chem Rev* **109**, 6729–6764 (2009).
93. De Clercq, E. C-Nucleosides to Be Revisited. *J Med Chem* **59**, 2301–2311 (2016).
94. Temburnikar, K. & Seley-Radtke, K. L. Recent advances in synthetic approaches for medicinal chemistry of C-nucleosides. *Beilstein Journal of Organic Chemistry* **14**, 772–785 (2018).
95. NISHIMURA, H. *et al.* SHOWDOMYCIN, A NEW ANTIBIOTIC FROM A STREPTOMYCES SP. *J Antibiot (Tokyo)* **17**, 148–155 (1964).
96. Roy-Burman, S., Roy-Burman, P. & Visser, D. W. Showdomycin, A New Nucleoside Antibiotic. *Cancer Res* **28**, 1605 LP – 1610 (1968).
97. Darnall, K. R., Townsend, L. B. & Robins, R. K. THE STRUCTURE OF SHOWDOMYCIN, A NOVEL CARBON-LINKED NUCLEOSIDE ANTIBIOTIC RELATED TO URIDINE. *Proceedings of the National Academy of Sciences* **57**, 548 LP – 553 (1967).
98. HORI, M., TAKITA, T., KOYAMA, G., TADEUCHI, T. & UMEZAWA, H. A NEW ANTIBIOTIC, FORMYCIN. *J Antibiot (Tokyo)* **17**, 96–99 (1964).
99. Ishida, N., Homma, M., Kumagai, K., Shimizu, Y. & Matsumoto, S. Studies on the antiviral activity of formycin. *J Antibiot (Tokyo)* **20**, 49–52 (1967).
100. ISHIZUKA, M. *et al.* ANTITUMOR ACTIVITIES OF FORMYCIN AND LABILOMYCIN. *J Antibiot (Tokyo)* **17**, 124–126 (1964).
101. Dapp, M. J., Bonnac, L., Patterson, S. E. & Mansky, L. M. Discovery of novel ribonucleoside analogs with activity against human immunodeficiency virus type 1. *J Virol* **88**, 354–363 (2014).
102. Zhang, M. *et al.* Comparative investigation into formycin a and pyrazofurin a biosynthesis reveals branch pathways for the construction of C-Nucleoside scaffolds. *Appl Environ Microbiol* **86**, 1–14 (2020).
103. Warren, T. K. *et al.* Protection against filovirus diseases by a novel broad-spectrum nucleoside analogue BCX4430. *Nature* **508**, 402–405 (2014).
104. Bergeron-Brlek, M., Meanwell, M. & Britton, R. Direct synthesis of imino-C-nucleoside analogues and other biologically active iminosugars. *Nat Commun* **6**, 1–6 (2015).
105. Korycka, A. & Robak, J. Z. B. and T. Forodesine (BCX-1777, Immucillin H) - A New Purine Nucleoside Analogue: Mechanism of Action and Potential Clinical Application. *Mini-Reviews in Medicinal Chemistry* vol. 7 976–983 Preprint at https://doi.org/http://dx.doi.org/10.2174/138955707781662636 (2007).
106. Tyler, V. L. S. and P. C. Imino-Sugar-Based Nucleosides. *Current Topics in Medicinal Chemistry* vol. 3 525–540 Preprint at https://doi.org/http://dx.doi.org/10.2174/1568026033452465 (2003).
107. Julander, J. G. *et al.* BCX4430, a novel nucleoside analog, effectively treats yellow fever in a hamster model. *Antimicrob Agents Chemother* **58**, 6607–6614 (2014).
108. Taylor, R. *et al.* BCX4430 - A broad-spectrum antiviral adenosine nucleoside analog under development for the treatment of Ebola virus disease. *J Infect Public Health* **9**, 220–226 (2016).
109. Eyer, L. *et al.* Antiviral activity of the adenosine analogue BCX4430 against West Nile virus and tick-borne flaviviruses. *Antiviral Res* **142**, 63–67 (2017).

110. Wang, Y. *et al.* Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet* **0**, 1569–1578 (2020).
111. Buchanan, J. G. Studies in the Synthesis and Biosynthesis of C-Nucleosides. *Nucleosides Nucleotides* **4**, 13–19 (1985).
112. Buchanan, J. G., Hamblin, M. R., Sood, G. R. & Wightman, R. H. The biosynthesis of pyrazofurin and formycin. *J Chem Soc Chem Commun* 917–918 (1980).
113. Galanie, S., Entwistle, D. & Lalonde, J. Engineering biosynthetic enzymes for industrial natural product synthesis. *Nat Prod Rep* (2020) doi:10.1039/C9NP00071B.
114. Kirschning, A. & Hahn, F. Merging Chemical Synthesis and Biosynthesis: A New Chapter in the Total Synthesis of Natural Products and Natural Product Libraries. *Angewandte Chemie International Edition* **51**, 4012–4022 (2012).
115. Wang, S. A. *et al.* Identification of the Formycin A Biosynthetic Gene Cluster from Streptomyces kaniharaensis Illustrates the Interplay between Biological Pyrazolopyrimidine Formation and de Novo Purine Biosynthesis. *J Am Chem Soc* **141**, 6127–6131 (2019).
116. Ko, Y., Wang, S.-A., Ogasawara, Y., Ruszczycky, M. W. & Liu, H.-W. Identification and Characterization of Enzymes Catalyzing Pyrazolopyrimidine Formation in the Biosynthesis of Formycin A. *Org Lett* **19**, 1426–1429 (2017).
117. Wang, S.-A. *et al.* Identification of the Formycin A Biosynthetic Gene Cluster from Streptomyces kaniharaensis Illustrates the Interplay between Biological Pyrazolopyrimidine Formation and de Novo Purine Biosynthesis. *J Am Chem Soc* **141**, 6127–6131 (2019).
118. Matsuda, K. *et al.* Discovery of Unprecedented Hydrazine-Forming Machinery in Bacteria. *J Am Chem Soc* **140**, 9083–9086 (2018).
119. Gao, S. *et al.* PMP-diketopiperazine adducts form at the active site of a PLP dependent enzyme involved in formycin biosynthesis. *Chemical Communications* **55**, 14502–14505 (2019).
120. Hove-Jensen, B. *et al.* Phosphoribosyl Diphosphate (PRPP): Biosynthesis, Enzymology, Utilization, and Metabolic Significance. *Microbiol Mol Biol Rev* **81**, (2017).
121. Ren, D. *et al.* Identification of the C-Glycoside Synthases during Biosynthesis of the Pyrazole-C-Nucleosides Formycin and Pyrazofurin. *Angewandte Chemie - International Edition* **58**, 16512–16516 (2019).
122. Cooper, G. M. *The Cell: A Molecular Approach. 2nd Edition*. (Sinauer Associates 2000, 2000).
123. Curry, S. Structural Biology: A Century-long Journey into an Unseen World. *Interdiscip Sci Rev* **40**, 308–328 (2015).
124. Schneider, G. X-ray microscopy: methods and perspectives. *Anal Bioanal Chem* **376**, 558–561 (2003).
125. Cowtan, K. Phase Problem in X-ray Crystallography, and Its Solution. in *Encyclopedia of Life Sciences* (2003). doi:https://doi.org/10.1038/npg.els.0002722.
126. HODGKIN, D. C. The X-ray analysis of the structure of penicillin. *Adv Sci* **6**, 85–89 (1949).
127. HODGKIN, D. C. *et al.* Structure of Vitamin B12. *Nature* **178**, 64–66 (1956).

128. BRINK, C. *et al.* Structure of Vitamin B12: X-ray Crystallographic Evidence on the Structure of Vitamin B12. *Nature* **174**, 1169–1171 (1954).
129. Hendrickson, W. A. Jerome Karle (1918–2013). *Nature* **499**, 410 (2013).
130. Giacovazzo, C. Herbert Hauptman (1917–2011). *Nature* **479**, 300 (2011).
131. Ruprecht, J. & Nield, J. Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction. *Prog Biophys Mol Biol* **75**, 121–164 (2001).
132. O'Donoghue, S. I. *et al.* Visualization of macromolecular structures. *Nat Methods* **7**, S42–S55 (2010).
133. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
134. Thompson, M. C., Yeates, T. O. & Rodriguez, J. A. Advances in methods for atomic resolution macromolecular structure determination. *F1000Res* **9**, (2020).
135. Lee, C., Su, B.-H. & Tseng, Y. J. Comparative studies of AlphaFold, RoseTTAFold and Modeller: a case study involving the use of G-protein-coupled receptors. *Brief Bioinform* **23**, bbac308 (2022).
136. Read, R. J., Baker, E. N., Bond, C. S., Garman, E. F. & van Raaij, M. J. AlphaFold and the future of structural biology. *IUCrJ* **10**, 377–379 (2023).
137. Subramaniam, S. & Kleywegt, G. J. A paradigm shift in structural biology. *Nat Methods* **19**, 20–23 (2022).
138. Akdel, M. *et al.* A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol* **29**, 1056–1067 (2022).
139. Coquelle, N. *et al.* Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat Chem* **10**, 31–37 (2018).
140. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–78 (2011).
141. Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science (1979)* **337**, 362 LP – 364 (2012).
142. Doerr, A. Diffraction before destruction. *Nat Methods* **8**, 283 (2011).
143. Chapman, H. N., Caleman, C. & Timneanu, N. Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130313 (2014).
144. Sierra, R. G. *et al.* Nanoflow electrospinning serial femtosecond crystallography. *Acta Crystallogr D Biol Crystallogr* **68**, 1584–1587 (2012).
145. Schmidt, M. Time-Resolved Macromolecular Crystallography at Modern X-Ray Sources. *Methods Mol Biol* **1607**, 273–294 (2017).
146. Aquila, A. *et al.* Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Opt Express* **20**, 2706–2716 (2012).
147. Moffat, K. Time-Resolved Biochemical Crystallography: A Mechanistic Perspective. *Chem Rev* **101**, 1569–1582 (2001).
148. Tenboer, J. *et al.* Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science (1979)* **346**, 1242 LP – 1246 (2014).
149. Suga, M. *et al.* Time-resolved studies of metalloproteins using X-ray free electron laser radiation at SACLA. *Biochim Biophys Acta Gen Subj* **1864**, 129466 (2020).

150. Liu, H. & Spence, J. C. H. XFEL data analysis for structural biology. *Quantitative Biology* **4**, 159–176 (2016).
151. Liu, H. & Lee, W. The XFEL Protein Crystallography: Developments and Perspectives. *International journal of molecular sciences* vol. 20 Preprint at https://doi.org/10.3390/ijms20143421 (2019).
152. Kirkwood, H. J. *et al.* A multi-million image Serial Femtosecond Crystallography dataset collected at the European XFEL. *Sci Data* **9**, 161 (2022).
153. Nogly, P. *et al.* Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ* **2**, 168–176 (2015).
154. Ishchenko, A., Gati, C. & Cherezov, V. Structural biology of G protein-coupled receptors: new opportunities from XFELs and cryoEM. *Curr Opin Struct Biol* **51**, 44–52 (2018).
155. Zhang, H. *et al.* Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. *Cell* **161**, 833–844 (2015).
156. Sugahara, M. *et al.* Hydroxyethyl cellulose matrix applied to serial crystallography. *Sci Rep* **7**, 703 (2017).
157. Sugahara, M. *et al.* Grease matrix as a versatile carrier of proteins for serial crystallography. *Nat Methods* **12**, 61–63 (2015).
158. Edlund, P. *et al.* The room temperature crystal structure of a bacterial phytochrome determined by serial femtosecond crystallography. *Sci Rep* **6**, 35279 (2016).
159. Woitowich, N. C. *et al.* Structural basis for light control of cell development revealed by crystal structures of a myxobacterial phytochrome. *IUCrJ* **5**, 619–634 (2018).
160. Zhao, F.-Z. *et al.* A guide to sample delivery systems for serial crystallography. *FEBS J* **286**, 4402–4417 (2019).
161. Cheng, R. K. Y. Towards an Optimal Sample Delivery Method for Serial Crystallography at XFEL. *Crystals (Basel)* **10**, (2020).
162. Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat Methods* **14**, 443–449 (2017).
163. Fromme, P., Graves, W. S. & Martin-Garcia, J. M. Serial Femtosecond Crystallography: A Decade at the Forefront in Structural Biology. in *Encyclopedia of Life Sciences* 1–17 (2020). doi:https://doi.org/10.1002/9780470015902.a0028964.
164. Pandey, S. *et al.* Time-resolved serial femtosecond crystallography at the European XFEL. *Nat Methods* **17**, 73–78 (2020).
165. Monteiro, D. C. F. *et al.* 3D-MiXD: 3D-printed X-ray-compatible microfluidic devices for rapid, low-consumption serial synchrotron crystallography data collection in flow. *IUCrJ* **7**, 207–219 (2020).
166. Monteiro, D. C. F. *et al.* A microfluidic flow-focusing device for low sample consumption serial synchrotron crystallography experiments in liquid flow. *J Synchrotron Radiat* **26**, 406–412 (2019).
167. Levantino, M., Yorke, B. A., Monteiro, D. C. F., Cammarata, M. & Pearson, A. R. Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. *Curr Opin Struct Biol* **35**, 41–48 (2015).

168. Šrajer, V. & Schmidt, M. Watching Proteins Function with Time-resolved X-ray Crystallography. *J Phys D Appl Phys* **50**, (2017).
169. Davison, E. K. & Brimble, M. A. Natural product derived privileged scaffolds in drug discovery. *Curr Opin Chem Biol* **52**, 1–8 (2019).
170. Gao, S. *et al.* Uncovering the chemistry of C–C bond formation in C-nucleoside biosynthesis: crystal structure of a C-glycoside synthase/PRPP complex. *Chem. Commun.* **56**, 7617–7620 (2020).
171. Dumitru, R. V & Ragsdale, S. W. Mechanism of 4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate synthase, a key enzyme in the methanopterin biosynthetic pathway. *J Biol Chem* **279**, 39389–39395 (2004).
172. Rasche, M. E. & White, R. H. Mechanism for the enzymatic formation of 4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate during the biosynthesis of methanopterin. *Biochemistry* **37**, 11343–11351 (1998).
173. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63**, 102–111 (2009).
174. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91 (2008).
175. Sheng, X. *et al.* A Combined Experimental-Theoretical Study of the LigW-Catalyzed Decarboxylation of 5-Carboxyvanillate in the Metabolic Pathway for Lignin Degradation. *ACS Catal* **7**, 4968–4974 (2017).
176. Moral, M. E. G., Tu, C., Richards, N. G. J. & Silverman, D. N. Membrane inlet for mass spectrometric measurement of catalysis by enzymatic decarboxylases. *Anal Biochem* **418**, 73–77 (2011).
177. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186–190 (2010).
178. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D* **74**, 85–97 (2018).
179. Agirre, J. *et al.* The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallographica Section D* **79**, 449–461 (2023).
180. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
181. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255 (1997).
182. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797 (2007).
183. Krissinel, E. & Henrick, K. Detection of Protein Assemblies in Crystals. in *Computational Life Sciences* (eds. R. Berthold, M., Glen, R. C., Diederichs, K., Kohlbacher, O. & Fischer, I.) 163–174 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005).
184. Krissinel, E. Crystal contacts as nature's docking solutions. *J Comput Chem* **31**, 133–143 (2010).
185. Ek, S. & Latypov, N. V. Four Syntheses of 4-Amino-3,5-dinitropyrazole. *J Heterocycl Chem* **51**, 1621–1627 (2014).
186. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).

187. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *'Protein Engineering, Design and Selection'* **8**, 127–134 (1995).
188. Zhu, W. *et al.* Substrate Binding Mode and Molecular Basis of a Specificity Switch in Oxalate Decarboxylase. *Biochemistry* **55**, 2163–2173 (2016).
189. Zhou, T., Daugherty, M., Grishin, N. V, Osterman, A. L. & Zhang, H. Structure and mechanism of homoserine kinase: prototype for the GHMP kinase superfamily. *Structure* **8**, 1247–1257 (2000).
190. Krishna, S. S., Zhou, T., Daugherty, M., Osterman, A. & Zhang, H. Structural basis for the catalysis and substrate specificity of homoserine kinase. *Biochemistry* **40**, 10810—10818 (2001).
191. Schmelz, S. *et al.* AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nat Chem Biol* **5**, 174–182 (2009).
192. Williamson, B. N. *et al.* Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. *Nature* **585**, 273–276 (2020).
193. Warren, T. K. *et al.* Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. *Nature* **531**, 381–385 (2016).
194. Mehellou, Y., Rattan, H. S. & Balzarini, J. The ProTide Prodrug Technology: From the Concept to the Clinic. *J Med Chem* **61**, 2211–2226 (2018).
195. Pradere, U., Garnier-Amblard, E. C., Coats, S. J., Amblard, F. & Schinazi, R. F. Synthesis of nucleoside phosphate and phosphonate prodrugs. *Chem Rev* **114**, 9154–9218 (2014).
196. Ramesh, D., Vijayakumar, B. G. & Kannan, T. Advances in Nucleoside and Nucleotide Analogues in Tackling Human Immunodeficiency Virus and Hepatitis Virus Infections. *ChemMedChem* **16**, 1403–1419 (2021).
197. Serpi, M., Ferrari, V. & Pertusati, F. Nucleoside Derived Antibiotics to Fight Microbial Drug Resistance: New Utilities for an Established Class of Drugs? *J Med Chem* **59**, 10343–10382 (2016).
198. Panda, S., Poudel, T. N., Hegde, P. & Aldrich, C. C. Innovative Strategies for the Construction of Diverse 1'-Modified C-Nucleoside Derivatives. *J Org Chem* **86**, 16625–16640 (2021).
199. Bouton, J., Van Calenbergh, S. & Hullaert, J. Sydnone Ribosides as a Platform for the Synthesis of Pyrazole C-Nucleosides: A Unified Synthesis of Formycin B and Pyrazofurin. *Org Lett* **22**, 9287–9291 (2020).
200. Meanwell, M. *et al.* A short de novo synthesis of nucleoside analogs. *Science (1979)* **369**, 725–730 (2020).
201. Xue, F. *et al.* Improvement of the C-glycosylation step for the synthesis of remdesivir. *Org Process Res Dev* **24**, 1772–1777 (2020).
202. Temburnikar, K. & Seley-Radtke, K. L. Recent advances in synthetic approaches for medicinal chemistry of C-nucleosides. *Beilstein journal of organic chemistry* **14**, 772–785 (2018).
203. Stambaský, J., Hocek, M. & Kocovský, P. C-nucleosides: synthetic strategies and biological applications. *Chem Rev* **109**, 6729–6764 (2009).
204. Gong, R. *et al.* Harnessing synthetic biology-based strategies for engineered biosynthesis of nucleoside natural products in actinobacteria. *Biotechnol Adv* **46**, 107673 (2021).

205. Slagman, S. & Fessner, W.-D. Biocatalytic routes to anti-viral agents and their synthetic intermediates. *Chem Soc Rev* **50**, 1968–2009 (2021).
206. Zhang, M. *et al.* Biosynthesis of C-nucleoside antibiotics in actinobacteria: recent advances and future developments. *Microb Cell Fact* **21**, 2 (2022).
207. Böhringer, N., Patras, M. A. & Schäberle, T. F. Heterologous Expression of Pseudouridimycin and Description of the Corresponding Minimal Biosynthetic Gene Cluster. *Molecules* **26**, (2021).
208. Zhu, W., Liu, X., Hughes, M., de Crécy-Lagard, V. & Richards, N. G. J. Whole-Genome Sequence of Streptomyces kaniharaensis Shomura and Niida SF-557. *Microbiol Resour Announc* **9**, (2020).
209. Zhao, G. *et al.* The Biosynthetic Gene Cluster of Pyrazomycin-A C-Nucleoside Antibiotic with a Rare Pyrazole Moiety. *Chembiochem* **21**, 644–649 (2020).
210. Wang, S. A. *et al.* Identification of the Formycin A Biosynthetic Gene Cluster from Streptomyces kaniharaensis Illustrates the Interplay between Biological Pyrazolopyrimidine Formation and de Novo Purine Biosynthesis. *J Am Chem Soc* **141**, 6127–6131 (2019).
211. Kong, L. *et al.* Divergent biosynthesis of C-nucleoside minimycin and indigoidine in bacteria. *iScience* **22**, 430–440 (2019).
212. Maffioli, S. I., Sosio, M., Ebright, R. H. & Donadio, S. Discovery, properties, and biosynthesis of pseudouridimycin, an antibacterial nucleoside-analog inhibitor of bacterial RNA polymerase. *J Ind Microbiol Biotechnol* **46**, 335–343 (2019).
213. Hong, H., Samborskyy, M., Zhou, Y. & Leadlay, P. F. C-nucleoside formation in the biosynthesis of the antifungal malayamycin A. *Cell Chem Biol* **26**, 493–501 (2019).
214. Sosio, M. *et al.* Analysis of the pseudouridimycin biosynthetic pathway provides insights into the formation of C-nucleoside antibiotics. *Cell Chem Biol* **25**, 540–549 (2018).
215. Palmu, K. *et al.* Discovery of the showdomycin gene cluster from Streptomyces showdoensis ATCC 15227 yields insight into the biosynthetic logic of C-nucleoside antibiotics. *ACS Chem Biol* **12**, 1472–1477 (2017).
216. Ren, D., Kim, M., Wang, S.-A. & Liu, H.-W. Identification of a Pyrrole Intermediate Which Undergoes C-Glycosidation and Autoxidation to Yield the Final Product in Showdomycin Biosynthesis. *Angew Chem Int Ed Engl* **60**, 17148–17154 (2021).
217. Ren, D. *et al.* Identification of the C-Glycoside Synthases during Biosynthesis of the Pyrazole-C-Nucleosides Formycin and Pyrazofurin. *Angewandte Chemie International Edition* (2019) doi:10.1002/anie.201910356.
218. Huang, S., Mahanta, N., Begley, T. P. & Ealick, S. E. Pseudouridine monophosphate glycosidase: a new glycosidase mechanism. *Biochemistry* **51**, 9245–9255 (2012).
219. Li, W. *et al.* Experimental and computational snapshots of C-C bond formation in a C-nucleoside synthase. *Open Biol* **13**, 220287 (2023).
220. Ochi, K., Yashima, S., Eguchi, Y. & Matsushita, K. Biosynthesis of formycin. Incorporation and distribution of 13C-, 14C-, and 15N-labeled compounds into formycin. *J Biol Chem* **254**, 8819–8824 (1979).
221. Zhao, G. *et al.* Molecular basis of enzymatic nitrogen-nitrogen formation by a family of zinc-binding cupin enzymes. *Nat Commun* **12**, 7205 (2021).

222. Firestine, S. M., Wu, W., Youn, H. & Davisson, V. J. Interrogating the mechanism of a tight binding inhibitor of AIR carboxylase. *Bioorg Med Chem* **17**, 794–803 (2009).
223. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
224. Johnson, K. A. & Goody, R. S. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**, 8264–8269 (2011).
225. Krahn, J. M. *et al.* Coupled formation of an amidotransferase interdomain ammonia channel and a phosphoribosyltransferase active site. *Biochemistry* **36**, 11061–11068 (1997).
226. Héroux, A., White, E. L., Ross, L. J., Kuzin, A. P. & Borhani, D. W. Substrate deformation in a hypoxanthine-guanine phosphoribosyltransferase ternary complex: the structural basis for catalysis. *Structure* **8**, 1309–1318 (2000).
227. Meola, M., Yamen, B., Weaver, K. & Sandwick, R. K. The catalytic effect of Mg2+ and imidazole on the decomposition of 5-phosphoribosyl-alpha-1-pyrophosphate in aqueous solution. *J Inorg Biochem* **93**, 235–242 (2003).
228. Siegbahn, P. E. M. & Himo, F. The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdiscip Rev Comput Mol Sci* **1**, 323–336 (2011).
229. Stewart, J. J. P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *Int J Quantum Chem* **58**, 133–146 (1996).
230. Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J Mol Model* **19**, 1–32 (2013).
231. Hammond, G. S. A Correlation of Reaction Rates. *J Am Chem Soc* **77**, 334–338 (1955).
232. Mowat, C. G. *et al.* Engineering water to act as an active site acid catalyst in a soluble fumarate reductase. *Biochemistry* **41**, 11990–11996 (2002).
233. Gopinath, P., Ramalingam, V. & Breslow, R. Magnesium pyrophosphates in enzyme mimics of nucleotide synthases and kinases and in their prebiotic chemistry. *Proc Natl Acad Sci U S A* **112**, 12011–12014 (2015).
234. Klinman, J. P. & Matthews, R. G. Calculation of substrate dissociation constants from steady-state isotope effects in enzyme-catalyzed reactions. *J Am Chem Soc* **107**, 1058–1060 (1985).
235. Tosha, T. *et al.* Capturing an initial intermediate during the P450nor enzymatic reaction using time-resolved XFEL crystallography and caged-substrate. *Nat Commun* **8**, 1585 (2017).
236. Nunes, R. M. D., Pineiro, M. & Arnaut, L. G. Photoacid for Extremely Long-Lived and Reversible pH-Jumps. *J Am Chem Soc* **131**, 9456–9462 (2009).
237. HIROMI, K., OHNISHI, M., KANAYA, K. & MATSUMOTO, T. The pH Jump Study of Enzyme Proteins: I. Liquefying α-Amylase from Bacillus subtilis12. *The Journal of Biochemistry* **77**, 957–963 (1975).
238. Calvey, G. D., Katz, A. M., Schaffer, C. B. & Pollack, L. Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers. *Structural Dynamics* **3**, 54301 (2016).

239. Wang, D., Weierstall, U., Pollack, L. & Spence, J. Double-focusing mixing jet for XFEL study of chemical kinetics. *J Synchrotron Radiat* **21**, 1364–1366 (2014).
240. Greetham, G. M. *et al.* A 100 kHz Time-Resolved Multiple-Probe Femtosecond to Second Infrared Absorption Spectrometer. *Appl Spectrosc* **70**, 645–653 (2016).
241. Bittmann, S. F. *et al.* Ultrafast ring-opening and solvent-dependent product relaxation of photochromic spironaphthopyran. *Phys Chem Chem Phys* **21**, 18119–18127 (2019).
242. Beale, J. H. *et al.* Successful sample preparation for serial crystallography experiments. *J Appl Crystallogr* **52**, 1385–1396 (2019).
243. Stohrer, C. *et al.* Homogeneous batch micro-crystallization of proteins from ammonium sulfate. *Acta Crystallogr D Struct Biol* **77**, 194–204 (2021).
244. Blow, D. M., Birktoft, J. J. & Hartley, B. S. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature* **221**, 337–340 (1969).
245. Kirby, A. J. The lysozyme mechanism sorted — after 50 years. *Nat Struct Biol* **8**, 737–739 (2001).
246. BLAKE, C. C. F. *et al.* Structure of Hen Egg-White Lysozyme: A Three-dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **206**, 757–761 (1965).
247. Nass Kovacs, G. *et al.* Three-dimensional view of ultrafast dynamics in photoexcited bacteriorhodopsin. *Nat Commun* **10**, 3177 (2019).
248. Gruhl, T. *et al.* Ultrafast structural changes direct the first molecular events of vision. *Nature* **615**, 939–944 (2023).
249. Fekete, K. *et al.* Effect of folate intake on health outcomes in pregnancy: a systematic review and meta-analysis on birth weight, placental weight and length of gestation. *Nutr J* **11**, 75 (2012).
250. Green, R. & Datta Mitra, A. Megaloblastic Anemias: Nutritional and Other Causes. *Med Clin North Am* **101**, 297–317 (2017).
251. Wilson, R. D. *et al.* Pre-conception Folic Acid and Multivitamin Supplementation for the Primary and Secondary Prevention of Neural Tube Defects and Other Folic Acid-Sensitive Congenital Anomalies. *J Obstet Gynaecol Can* **37**, 534–552 (2015).
252. Reynolds, E. H. The neurology of folic acid deficiency. *Handb Clin Neurol* **120**, 927–943 (2014).
253. van Beelen, P. *et al.* Elucidation of the structure of methanopterin, a coenzyme from Methanobacterium thermoautotrophicum, using two-dimensional nuclear-magnetic-resonance techniques. *Eur J Biochem* **138**, 563–571 (1984).
254. Zhou, D. & White, R. H. 5-(p-aminophenyl)-1,2,3,4-tetrahydroxypentane, a structural component of the modified folate in Sulfolobus solfataricus. *J Bacteriol* **174**, 4576–4582 (1992).
255. White, R. H. Biosynthesis of methanopterin. *Biochemistry* **29**, 5397–5404 (1990).
256. Rasche, M. E. & White, R. H. Mechanism for the enzymatic formation of 4-(beta-D-ribofuranosyl)aminobenzene 5'-phosphate during the biosynthesis of methanopterin. *Biochemistry* **37**, 11343–11351 (1998).
257. Ren, D. *et al.* Identification of the C-Glycoside Synthases during Biosynthesis of the Pyrazole-C-Nucleosides Formycin and Pyrazofurin. *Angewandte Chemie - International Edition* **58**, 16512–16516 (2019).
258. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63**, 102–111 (2009).

259. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D* **74**, 85–97 (2018).
260. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186–190 (2010).
261. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
262. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255 (1997).
263. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).
264. Yang, J. *et al.* The nucleotide pGpp acts as a third alarmone in Bacillus, with functions distinct from those of (p)ppGpp. *Nat Commun* **11**, 5388 (2020).