

**Mechanistic and Structural Insights  
into the Specificity and Biological  
Functions of *E. coli* HAD Superfamily  
Phosphatase HAD4/YihX**



**Davide Zappalà**

Cardiff University

School of Chemistry

Thesis submitted to Cardiff University for the degree of Doctor of Philosophy

Supervisors: Dr. Yi Jin, Dr. Youcef Mehellou, Dr. Louis Luk

**June 2023**

## I. Abstract

The HAD superfamily has been extensively studied in literature, representing a major class of enzymes which are responsible mainly for phosphoryl transfer in prokaryotes and eukaryotes. The characteristic structural motifs are conserved throughout the whole superfamily and reaction mechanisms for many HAD enzymes have been validated. In this work, the spotlight is focused on YihX, encoded in the *E. coli* genome. This peculiar case is of interest since this enzyme features an uncommon DxG signature that still enables the catalysis of phosphoryl transfer while lacking key structural and catalytic features needed for the currently known mechanism. The first part will focus on the exploration of the possibility of YihX being part of the sulfoglycolytic pathway, and the D-tyrosine metabolism operon due to its collocation in the genome, located directly downstream of the sulfoglycolysis and upstream of the D-tyrosine metabolism operon. In the second part, new experimental evidence is produced in the form of crystal structures, kinetic studies and mechanistic observations. The only structure present in the PDB shows some substantial flaws. Metal fluorides as transition state analogues led to two new crystal structures, apo conformation and a transition state analogue (TSA) complex of a rarely observed octahedral  $\text{MgF}_3(\text{H}_2\text{O})^-$  entity. The formation of the TSA complex and chemical environment around it is supported by  $^{19}\text{F}$  NMR studies. A new reaction mechanism has been proposed for phosphoryl transfer catalysed by HAD phosphatases containing the DXG motif. Serendipitous discovery of a new phosphoryl transfer activity for YihX, ATP synthesis from ADP, showed that it is a dual-activity enzyme. Finally, a new expression system was created and tested for recombinant human Thymidine Kinase 2 to produce high quantities of soluble protein for crystallisation trials to hopefully obtain a long overdue crystal structure to boost the therapeutics field for mitochondrial DNA depletion syndromes.

## II. Acknowledgements

I want to acknowledge all the people and institutions that have made possible this 4-year long journey that has given me invaluable life experiences and professional skills. For this reason, huge thanks to the EPSRC for the funding allocated for my doctoral studies.

I want to thank all the members of the Jin group in Cardiff and Manchester starting with my main supervisor Dr. Yi Jin, for her relentless attention in pushing my limits and the thousands of scientific insights. With her, I learned how to correctly face the challenges in the scientific field with critical thinking and nerves of steel during periods when things are not going in the best way. Even with the bar set high, she always demonstrated that she cared about me and my career, and I will never forget it. Second from the top, my personal guide, as a brother-in-science, Dr. Patrick Baumann for his kindness, knowledge, attitude, and availability during my whole time as a PhD student. Last, but not least, Mochen Dong for the laughs and cooking in and outside the office and her positivity.

I want to thank my other supervisors, Dr. Youcef Mehellou for scientific and life coaching, and Dr. Louis Luk for creating a research group that I regarded as a second family. Even if there are too many people, I want to thank all the past and present members of the Luk-Tsai-Jin group that made my life here in Cardiff the best period of my life by a very long shot. In no particular order, thank you Nicolò Santi, Thomas Williams, Alexander Nödling, Zuyan Wu, Davide Cardella, Alexander Lander, Simon Tang, Luke Spear, Emily Mills, Heather Hayes, Victoria Barlow, Adriana Coricello, and Muge Ma. Additional thanks to Heather Hayes for introducing me to the best sport I ever practised, climbing, and to Victoria Barlow for the good times spent as flatmates. Special thanks go to the support from the Cardiff University staff over the years, especially Dr. Robert Jenkins for his knowledge and help of NMR spectroscopy.

### III. Table of contents

I. Abstract	II
II. Acknowledgments	III
III. Table of contents	IV
IV. List of Figures	VII
V. List of Tables	XII
VI. List of Schemes	XIII
VII. Abbreviations	XIV

<b>1. Introduction.....</b>	<b>1</b>
1.1 The HAD superfamily: structure .....	1
1.2 Mechanism of reaction .....	3
1.3 Metal fluorides as transition state analogue complex.....	5
1.4 <sup>19</sup> F NMR spectroscopy as a tool for characterization of MF <sub>x</sub> TSA complex.....	6
1.5 Importance of HAD superfamily enzymes .....	6
1.6 Substrate specificity.....	7
1.7 The sulfoglycolytic pathway and comparison with glycolysis .....	9
1.7.1 Possible Role of YihV as a Checkpoint in pathway regulation .....	12
1.7.2 Identification of the <i>yihX</i> gene as a possible member of the sulfoglycolysis pathway.....	13
1.7.3 A closer look at neighbouring genes: <i>yihX</i> operon .....	15
1.8 Aims and Objectives .....	17
<b>2. Materials and Methods.....</b>	<b>18</b>
2.1 Molecular Cloning.....	18
2.2 List of <i>E. coli</i> Strains .....	18
2.3 Buffers.....	19
2.4 Plasmids.....	19
2.5 Primers.....	21
2.6 Growth Media .....	22
2.7 Competent Cells .....	23
2.8 Transformation Protocol .....	24
2.9 Plasmid miniprep.....	24
2.10 PCR protocol .....	24

<b>2.11 Site Directed Mutagenesis (SDM) protocol</b> .....	<b>24</b>
<b>2.12 Gibson Assembly protocol</b> .....	<b>25</b>
<b>2.13 High Performance Liquid Chromatography</b> .....	<b>25</b>
<b>2.13.1 Sample Preparation and Method for HPLC Analysis of ATP Formation and AMP Dephosphorylation</b> .....	<b>25</b>
<b>2.14 Malachite Green Phosphate Assay</b> .....	<b>26</b>
<b>2.15 Nuclear Magnetic Resonance (NMR) spectroscopy</b> .....	<b>26</b>
2.15.1 Culture growth and sample preparation for whole-cell NMR assay with M9-SQ .....	27
<b>2.16 Mass spectrometry</b> .....	<b>28</b>
<b>2.17 Fast protein liquid chromatography (FPLC)</b> .....	<b>28</b>
<b>2.18 Gel Electrophoresis</b> .....	<b>29</b>
<b>2.19 Spectrophotometric measurement of protein concentration</b> .....	<b>30</b>
<b>2.20 Thin Layer Chromatography (TLC)</b> .....	<b>30</b>
<b>2.21 Protein Expression and Purification</b> .....	<b>30</b>
2.21.1 His <sub>6</sub> -YihS recombinant protein .....	30
2.21.2 YihX-His <sub>6</sub> recombinant protein .....	31
2.21.3 YihV-His <sub>6</sub> recombinant protein.....	32
2.21.4 YihX <sub>G8D</sub> -His <sub>6</sub> recombinant protein.....	32
2.21.5 YihX <sub>D6N</sub> -His <sub>6</sub> recombinant protein.....	33
2.21.6 YihX <sub>K141L</sub> -His <sub>6</sub> recombinant protein.....	34
2.21.7 YihX <sub>N108L</sub> -His <sub>6</sub> recombinant protein .....	35
2.21.8 PG_0725-His <sub>6</sub> recombinant protein .....	35
2.21.9 His <sub>6</sub> -Usp2-cc recombinant protein .....	36
2.21.10 His <sub>6</sub> -TF-Ub-TK2 recombinant protein .....	37
<b>2.22 Chemical synthesis</b> .....	<b>38</b>
2.22.1 Synthesis of SQ.....	38
<b>3. Results and discussion</b> .....	<b>40</b>
<b>3.1 Investigating YihV as a checkpoint of sulfoglycolysis</b> .....	<b>40</b>
<b>3.2 YihX as a backup enzyme for Dtd</b> .....	<b>43</b>
<b>3.3 YihX as putative SFP phosphatase</b> .....	<b>45</b>
3.3.1 SQ synthesis .....	45
3.3.2 YihS, YihV and YihX recombinant proteins .....	45
.....	46
3.3.3 YihS reaction: isomerisation of SQ to SF .....	47
3.3.4 YihV reaction: SFP synthesis.....	48
3.3.5 YihX reaction: SFP dephosphorylation to SF .....	49

3.4 A new activity for YihX: ADP phosphorylation to ATP .....	50
3.4.1 Result validation: HPLC characterisation.....	51
3.4.2 Kinetic parameters determination for ATP synthesis activity.....	52
<b>3.5 Structural and kinetic characterisation of YihX.....</b>	<b>53</b>
3.5.1 Mistakes in the PDB 2B0C structure and its value .....	53
3.5.2 Refinement of YihX structure from PDB 2B0C.....	54
3.5.3 YihX catalytic mechanism: a deviation from canonical HAD phosphatase.....	55
3.5.4 Crystallisation of YihX Apo and YihX-MgF <sub>3</sub> (H <sub>2</sub> O) <sup>-</sup> -H <sub>2</sub> O transition state analogue complex.....	56
3.5.5 Structural analysis of YihX-MgF <sub>3</sub> (H <sub>2</sub> O) <sup>-</sup> -H <sub>2</sub> O transition state analogue (TSA) .....	59
3.5.6 A new mechanism for YihX-catalysed hydrolysis: solvent assisted catalysis .....	62
3.5.7 DxD knock-in mutant.....	63
3.5.8 YihX <sub>G8D</sub> TSA complex.....	64
3.5.9 Site-directed mutagenesis of catalytic site residues .....	68
3.5.10 The PG_0725 gene: another DxG motif HAD phosphatase.....	69
3.5.11 Testing PG_0725 for ADP phosphorylation .....	72
3.5.12 αG1P dephosphorylation catalysed by PG_0725 .....	73
<b>4 Introduction to Thymidine Kinase 2 .....</b>	<b>75</b>
<b>4.1 Mitochondrial DNA.....</b>	<b>75</b>
<b>4.2 Mitochondrial DNA Depletion Syndrome (MDDS) .....</b>	<b>76</b>
<b>4.3 Evolution of therapeutics for human TK2 .....</b>	<b>77</b>
<b>4.3 The lack of a TK2 crystal structure .....</b>	<b>82</b>
<b>4.4 Project aims.....</b>	<b>83</b>
<b>4.5 Results .....</b>	<b>83</b>
4.5.1 A new expression system for TK2.....	83
4.5.2 Recombinant TK2 purification.....	85
4.5.3 Cleavage of TK2 fusion protein .....	86
<b>5 Summary.....</b>	<b>89</b>
<b>6 References .....</b>	<b>91</b>
<b>7 Appendix .....</b>	<b>98</b>
<b>7.1 HPLC stop-assay for ADP phosphorylation by YihX .....</b>	<b>98</b>
<b>7.3 NMR spectra.....</b>	<b>110</b>
7.3.1 SQ synthesis spectra .....	110
7.3.2 SFP synthesis spectra .....	118
7.3.3 YihX ADP synthesis spectra .....	119
7.3.4 <sup>1</sup> H NMR whole-cell NMR assay .....	123
7.3.5 <sup>31</sup> P NMR AMP hydrolysis .....	124

<b>7.4 Extended data for YihV whole-cell NMR assay .....</b>	<b>125</b>
<b>7.5 Mass Spectrometry .....</b>	<b>126</b>
7.5.1 YihX protein .....	126
7.5.2 YihV protein .....	126
7.5.3 SFP .....	127
<b>7.6 SDS-PAGE .....</b>	<b>128</b>
.....	128
<b>7.7 Recombinant Protein Sequences .....</b>	<b>132</b>
<b>7.8 Gene sequences .....</b>	<b>133</b>
<b>7.9 Pulse sequence for F<sup>-</sup> suppression in <sup>19</sup>F NMR .....</b>	<b>135</b>

#### IV. List of Figures

<b>Figure 1</b> A scheme representing the core structure of the typical HAD enzyme. Strands with dashed borders are not always present.....	1
<b>Figure 2</b> The two most common cap types in HAD superfamily. <b>a)</b> Crystal structure of Phosphoserine phosphatase (PDB: 1L8L) as a representative of C1 cap class and <b>b)</b> Crystal structure of Phosphomannomutase 1 (PDB: 2FUE) as a representative of C2 cap class. Cap domains are shown in green. ....	2
<b>Figure 3</b> Comparison between the natural transition state of phosphoryl transfer and the TSA MgF <sub>3</sub> <sup>-</sup> . ....	5
<b>Figure 4 a)</b> Crystal structure of KDN-9-P co-crystallised with N-acetyl-β-neuraminic acid, VO <sup>3-</sup> , Mg <sup>2+</sup> , <b>b)</b> a zoomed-in view of the interface between two chains, highlighting the “pseudocapping” formed by the oligomerisation of the enzyme. ....	7
<b>Figure 5</b> The sulfoglycolytic pathway (left) and the glycolytic pathway (right). Both pathways provide the key metabolite DHAP which can be interconverted into GAP by the activity of the Tpi enzyme. Further steps will yield the pyruvate which can be funnelled into the Krebs cycle, NADH and ATP. ....	11
<b>Figure 6</b> Characteristic <sup>1</sup> H NMR signals of Meropenem and the inactivated product formed after reaction with NDM-1. The analysis is performed with a live cell culture. ....	13
<b>Figure 7</b> The sulfoglycolysis operon and the position of the yihX gene relative to the pathway. ....	13

<b>Figure 8</b> The high throughput assay performed by K. Allen et al. on the YihX enzyme. The top table shows the different substrates tested and the bottom table shows the activity towards the specific phosphorylated substrate <sup>47</sup> .....	14
<b>Figure 9</b> The YihX operon with all the genes encoded and the position of the promoter sequence <sup>48</sup> .....	15
<b>Figure 10</b> Workflow for sample preparation for <sup>1</sup> H NMR analysis of upregulated YihV cultures in M9-SQ medium.....	27
<b>Figure 11</b> Characteristic <sup>1</sup> H NMR signals of SQ and DHPS. YihV is upregulated by induction through IPTG addition.....	40
<b>Figure 12</b> Depletion of SQ displayed using the integration of SQ signal compared to the internal standard.....	41
<b>Figure 13</b> Sodium trimethylsilylpropanesulfonate (DSS). .....	41
<b>Figure 14</b> Sulfoglycolysis, gluconeogenesis, glycolysis, and other major metabolic pathways. Pathway crosstalk points are highlighted by arrows connecting metabolites to multiple pathways. Adapted from Janice W.-Y. Mui et al(49).....	42
<b>Figure 15</b> Growth curve of parent strain and single-knockout mutants in M9 minimal medium supplied with D-tyrosine.....	44
<b>Figure 16 (Left)</b> YihX purification by Ni-NTA chromatography. L: protein ladder. Lane 1: flowthrough. Lanes 2 to 12: elution with imidazole gradient. <b>Right)</b> Plasmid map	45
<b>Figure 17 (Left)</b> YihS purification by Ni-NTA chromatography. L: protein ladder. Lane 1 to 8: elution with imidazole gradient. <b>Right)</b> Plasmid map. ....	46
<b>Figure 18 (Left)</b> YihV purification by Ni-NTA chromatography. L: protein ladder. Lanes 1 to 11: elution with imidazole gradient. <b>Right)</b> Plasmid map.....	46
<b>Figure 19</b> YihS isomerisation reaction and <sup>1</sup> H NMR spectrum of the reaction after reaching equilibrium with the presence of characteristic signals of SQ, SR and SF. 47	
<b>Figure 20</b> YihV reaction scheme (top) and <sup>31</sup> P NMR spectra of the mixture after reaction completion. ....	48
<b>Figure 21</b> <sup>31</sup> P NMR spectra of reaction mixture after addition of YihX phosphatase. ATP peaks are clearly produced from the reaction, since there was no unreacted ATP left from the previous reaction. ....	49
<b>Figure 22</b> <sup>31</sup> P NMR time-course experiment highlighting ATP formation over an extended period. The reaction was carried in 50 mM Tris-HCl buffer pH 7.00, 200 mM NaCl, 10 mM ADP, 15 mM MgCl <sub>2</sub> , 200 μM YihX, 37 °C.....	50

<b>Figure 23</b> HPLC trace of ADP phosphorylation reaction mixture. The retention time of all the species is validated with commercially available standards. Run conditions as explained in section 2.13.1. ....	51
<b>Figure 24</b> Michaelis-Menten kinetics for YihX ATP formation reaction. Data obtained by HPLC stop-assay.....	53
<b>Figure 25 a)</b> The YihX structure (PDB 2B0C). The covalent bond between Asp13 and the C2 hydroxyl group on $\alpha$ -D-glucose-1-phosphate is clearly shown. <b>b)</b> electron density map contoured at $0.22e - 1\text{\AA} - 3$ ; the electron density map does not support the fitting of an $\alpha$ G1P molecule, as shown by the lack of electron density around the bonds. ....	54
<b>Figure 26 a)</b> The PDB 2B0C structure refined. The $\text{Mn}^{2+}$ ion is now fitted in the structure and a single phosphate group is coordinated by as a ligand. <b>b)</b> electron density map contoured at $0.22e - 1\text{\AA} - 3$ ; the electron density is much better defined around the $\text{Mn}^{2+}$ ion with clear octahedral coordination and the phosphate ligand. .	55
<b>Figure 27</b> Octahedral trifluoromagnesate TSA complex of YihX. The electron density map is contoured at $0.18e - 1\text{\AA} - 3$ . ....	59
<b>Figure 28</b> $^{19}\text{F}$ NMR of YihX- $\text{MgF}_3(\text{H}_2\text{O})^-$ - $\text{H}_2\text{O}$ TSA complex. 0.3 mM YihX, 50 mM Tris-HCl pH 7, 200 mM NaCl, 10 mM $\text{MgCl}_2$ , 30 mM $\text{NH}_4\text{Cl}$ . The broad peak corresponds to the species $[\text{MgF}^+(\text{H}_2\text{O})_5](87)$ . ....	60
<b>Figure 29</b> Overlay of $^{19}\text{F}$ NMR of YihX- $\text{MgF}_3(\text{H}_2\text{O})^-$ - $\text{H}_2\text{O}$ TSA complex obtained in 10% $\text{D}_2\text{O}$ (red), and 90% $\text{D}_2\text{O}$ (blue). The SIIS is calculated by subtracting the $\delta$ value of the signal from the 90% $\text{D}_2\text{O}$ spectrum from the $\delta$ value of the 10% $\text{D}_2\text{O}$ spectrum. ....	61
<b>Figure 30</b> YihX-TSA crystal structure with fluorine atoms assigned through SIIS....	61
<b>Figure 31</b> DNA sequencing for SDM of YihX <sub>G8D</sub> and chromatography purification. .	63
<b>Figure 32</b> Michaelis-Menten kinetic characterisation of $\alpha$ G1P hydrolysis by YihX <sub>G8D</sub> and comparison with YihX <sub>WT</sub> . ....	64
<b>Figure 33</b> $^{19}\text{F}$ NMR spectrum of YihX <sub>WT</sub> TSA complex (red trace), and YihX <sub>G8D</sub> TSA complex (blue trace). The lack of peaks for the individual fluorine atoms in the YihX <sub>G8D</sub> trace clearly shows no formation of TSA complex.....	64
<b>Figure 34</b> Crystal structure of YihX <sub>WT</sub> TSA complex (green ribbon), AlfaFold predicted structure of YihX <sub>G8D</sub> aligned with $\text{MgF}_3(\text{H}_2\text{O})^-$ TSA (pink ribbon), crystal structure of Phosphoserine phosphatase aluminium fluoride TSA (PDB: 1L7N) (teal	

ribbon). Superimposition of YihX <sub>WT</sub> TSA complex and phosphoserine phosphatase aluminium fluoride TSA. ....	67
<b>Figure 35</b> Michaelis-Menten kinetic curves for YihX mutants. Note that parameter values are not correct due to not reaching plateau. ....	68
<b>Figure 36</b> The Sequence Similarity network for YihX. The localised cluster with no connections to neighbouring clusters denotes that this subclass of phosphatases does not share common structure and activity relationship with other enzyme classes. ....	70
<b>Figure 37</b> List of DxG enzymes that are present in both the SSN and a previous high-throughput screen conducted by K. Allen et al.(47) The table lists the UniProt ID, the gene identifier, the microorganism, the availability of a crystal structure in the PDB and the availability of an activity in the high-throughput screen. ....	71
<b>Figure 38</b> Sequence alignment of the genes identified by the SSN. The signature motifs are highlighted in the sequences. ....	72
<b>Figure 39</b> <sup>31</sup> P NMR spectrum of ADP phosphorylation reaction catalysed by PG_0725. The spectrum shows the timeframe t = 0 (red trace) and t = 24h (blue trace). Both samples contain 10% D <sub>2</sub> O. ....	72
<b>Figure 40</b> Comparison of the electrostatic surface of YihX (left) and PG_0725 (right). The top view highlights the loop region that restricts the pocket in PG_0725 while it is completely open and available in YihX. ....	73
<b>Figure 41</b> Michaelis-Menten kinetics for αG1P dephosphorylation catalysed by PG_0725. Data obtained by Malachite Green Phosphate Assay. ....	74
<b>Figure 42</b> The human mitochondrial DNA and the encoded genes in both the H-strand and the L-Strand. ....	75
<b>Figure 43</b> Schematic of nucleotide metabolism for mtDNA synthesis, repair, and most important genes in the pathways (in red) that have been associated with MDDS with corresponding gene mutation. ....	76
<b>Figure 44</b> Thymidine nucleoside analogues recognised by TK2 as substrates. ....	78
<b>Figure 45</b> A new series of compounds synthesised to more directly target TK2 independently. Chemical modifications on the furanose ring are key to this approach. ....	78
<b>Figure 46</b> The acyclic trityl derivatives. The commercial compound <b>7</b> is the starting point that will generate the entire series, including nitrogen substituted compounds. ....	80

<b>Figure 47</b> Thiourea and triazole derivatives (top) and tetrazole derivatives (bottom). .....	81
<b>Figure 48</b> Localisation of hTK1 (cytoplasm) and hTK2 (mitochondrial) and comparison with the <i>Drosophila Melanogaster</i> deoxyribonucleoside kinase (PDB 2VPP).....	82
<b>Figure 49</b> SDS-PAGE of <i>E. coli</i> BL21(DE3) transformed with pEXP-TK2 plasmid. Lanes 2 and 4 represent soluble and insoluble proteins, lanes 3 and 5 represent only soluble proteins. ....	84
<b>Figure 50 a)</b> Schematic of the new construct, <b>b)</b> Map of cloned plasmid, <b>c)</b> SDS- PAGE small-scale expression test using the new construct to probe different <i>E. coli</i> strains.....	84
<b>Figure 51 Left)</b> UV trace of Ni-NTA purification and <b>Right)</b> SDS-PAGE of TK2 peak that was collected.....	85
<b>Figure 52</b> SDS-PAGE of cleavage test reaction. The bands highlight the presence of all four expected species with respective molecular weights: uncleaved fusion protein, cleaved tag, deubiquitinase, TK2. ....	86
<b>Figure 53</b> SDS-PAGE of eluted fractions after Ni-NTA column. TK2 protein is eluting with other fragments from the cleavage reaction. L: protein ladder. Lane 1-11: collection of the flowthrough. ....	87
<b>Figure 54</b> SEC separation. UV trace (left) and SDS-PAGE analysis of the fractions (right).....	88
<b>Figure 55 Left)</b> SDS-PAGE of eluted fractions after IEX chromatography. L: protein ladder. Lanes 1-8: elution. <b>Right)</b> Isoelectric points of proteins and cleaved tag in solution. ....	88
<b>Figure 56</b> HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.05 mM .....	98
<b>Figure 57</b> HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.10 mM .....	99
<b>Figure 58</b> HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.25 mM .....	100

<b>Figure 59</b> HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.5 mM .....	101
<b>Figure 60</b> HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 1 mM .....	102
<b>Figure 61</b> HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 5 mM .....	103
<b>Figure 62</b> HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 10 mM. ....	104
<b>Figure 63</b> YihS purification by Ni-NTA chromatography. Lane 1 to 8: elution.....	128
<b>Figure 64</b> YihX purification by Ni-NTA chromatography. Lane 1: flowthrough. Lanes 2 to 12: elution.....	128
<b>Figure 65</b> YihV purification by Ni-NTA chromatography. Lanes 1 to 11: elution. ....	129
<b>Figure 66</b> YihX <sub>G8D</sub> after SEC purification. Lane 1: Protein ladder. Lane 2 to 12: elution.....	129
<b>Figure 67</b> YihX <sub>D6N</sub> purification by Ni-NTA chromatography. Lanes 1 to 12: elution.	130
<b>Figure 68</b> YihX <sub>N108L</sub> purification by Ni-NTA chromatography. Lanes 1 to 10: elution. ....	130
<b>Figure 70</b> TF-Ub-TK2 fusion protein expression and purification by Ni-NTA chromatography. Lane 1: before induction. Lane 2: after induction. Lanes 3 to 10: elution.....	131
<b>Figure 69</b> YihX <sub>K141L</sub> purification by Ni-NTA chromatography. Lanes 1 to 10: elution. ....	131

## V. List of Tables

<b>Table 1</b> List of primers.....	21
<b>Table 2</b> Standard curve for Malachite Green Phosphate Assay. ....	26
<b>Table 3</b> LC-MS elution program for protein Mass Spectrometry.....	28
<b>Table 4</b> Buffers and reagents used for SDS-PAGE .....	29

<b>Table 5</b> Data collection and refinement statistics for YihX-MgF <sub>3</sub> (H <sub>2</sub> O)-H <sub>2</sub> O. ....	57
<b>Table 6</b> Data collection and refinement statistics for apo YihX. ....	58
<b>Table 7</b> IC <sub>50</sub> values of small molecules using 2 μM [methyl-3H]dThd as substrate for kinase. All values are expressed in μM (70–73) .....	79
<b>Table 8</b> IC <sub>50</sub> values of small molecules using 1 μM [methyl-3H]dThd as substrate for kinase. All values are expressed in μM (75–77). ....	80
<b>Table 9</b> IC <sub>50</sub> values of small molecules using 1 μM [methyl-3H]dThd as substrate for kinase. All values are expressed in μM (80,81). ....	82
<b>Table 10</b> Best fit values for Michaelis-Menten kinetics for YihX phosphorylation of ADP. ....	105
<b>Table 11</b> Absorbance values and time points for αG1P hydrolysis catalysed by YihX during a Malachite Green Phosphate assay. Concentration of substrate varies from 0.005 mM to 5 mM. Inorganic phosphate concentration (μM) is plotted against time. ....	106
<b>Table 12</b> Best fit values for Michaelis-Menten kinetics for YihX hydrolysis of αG1P. ....	109
<b>Table 13</b> Integration values of SQ peak during whole-cell NMR time-course. ....	125

#### IV. List of Schemes

<b>Scheme 1 a)</b> Classical reaction mechanism for HAD hydrolase enzymes while performing phosphoryl transfer <b>b)</b> PTP canonical reaction mechanism .....	4
<b>Scheme 2</b> The series of enzymatic reactions required to produce SFP. First, isomerisation of SQ to SF by the action of YihS, then phosphorylation to SFP by YihV. Once the reaction is terminated, dephosphorylation will be performed by the addition of YihX to the reaction mixture. ....	15
<b>Scheme 3 top)</b> Reaction catalysed by dtd to salvage tRNA molecules charged with D-tyrosine <b>bottom)</b> Hypothetical reaction catalysed by YihX, dephosphorylating the tRNA charged with D-tyrosine. ....	16
<b>Scheme 4</b> SQ synthetic scheme used to prepare the substrate. ....	38
<b>Scheme 5</b> Expected reaction for YihX dephosphorylation of SFP. ....	49
<b>Scheme 6</b> HAD phosphatase mechanism of hydrolysis with a DxD motif I .....	55

**Scheme 7** Proposed catalytic mechanism for YihX which accounts for the lack of the GAGB residue in position  $D_{n+2}$  in motif I..... 62

**Scheme 8** Catalytic activity of TK2. dG and dU are substrates of TK2 as well. .... 77

## VI. Abbreviations

2-ME	2-mercaptoethanol
Å	Ångström (1 Å = 0.1 nm)
Amp	ampicillin
ANT1	adenine nucleotide translocase type 1
AraT	1-β-D-arabinofuranosylthimine
ATP	Adenosine triphosphate
ATPase	ATP hydrolase
AZT	3'-azido-3'-deoxythimidine
CADD	computer-aided drug design
Cam	chloramphenicol
CI	confidence interval
CV	column volume
dA	deoxyadenosine
dC	deoxycytidine
dG	deoxyguanosine
DGUOK	deoxyguanosine kinase
DHAP	dihydroxyacetone phosphate
DHPS	dihydroxypropane sulfonate
DNA	deoxyribonucleic acid
DNA2	DNA replication ATP-dependent helicase/nuclease
DNC	deoxynucleotide carrier, mitochondrial
DSS	sodium trimethylsilylpropanesulfonate
dT/dThd	deoxythymidine
DTT	dithiothreitol
dU	deoxyuridine
<i>E. coli</i>	Escherichia coli

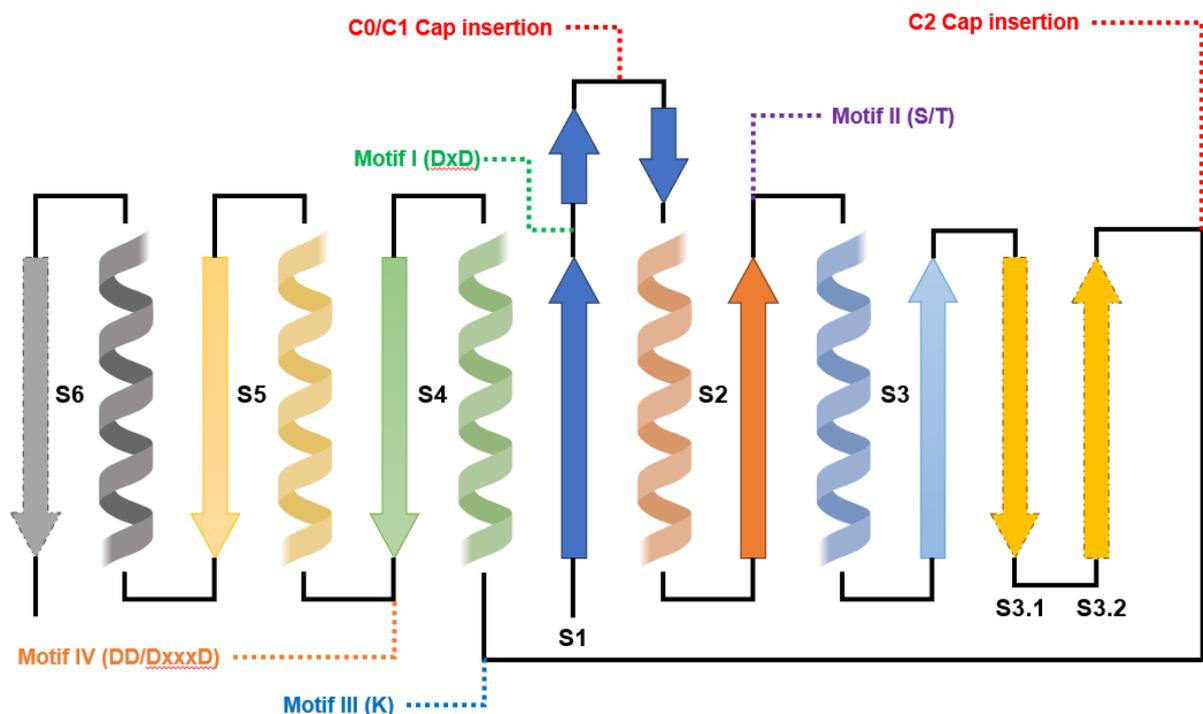
EDTA	ethylenediaminetetraacetic acid
F6P	fructose-6-phosphate
FBP	fructose-1,6-bisphosphate
FIAU	1- $\beta$ -D-(2'-fluoroarabinofuranosyl)-5-iodouracil
FPLC	fast protein liquid chromatography
G6P	glucose-6-phosphate
GAGB	general acid/general base
GAP	glyceraldehyde-3-phosphate
GTPase	GTP hydrolase
HAD	Haloacid dehalogenase
IC <sub>50</sub>	inhibitory concentration 50%
IMAC	immobilized metal affinity chromatography
Kan	kanamycin A
KDN-9-P	2-keto-3-deoxy-d-glycero-d-galacto-9-phosphonononic acid phosphatase
MDDS	mitochondrial DNA depletion syndrome
MF <sub>x</sub>	metal fluorides
MGME1	mitochondrial genome maintenance exonuclease 1
MPV17	protein MPV17
NAD <sup>+</sup>	nicotinamide adenine dinucleotide, oxidised
NADH	nicotinamide adenine dinucleotide, reduced
NTA	nitrilotriacetic acid
PCR	polymerase chain reaction
PNKP	polynucleotide kinase phosphatase
POLG	DNA polymerase subunit gamma
POLG2	DNA polymerase subunit gamma-2
PTP	protein tyrosine phosphatase
RNA	ribonucleic acid
RRM2B	ribonucleoside-diphosphate reductase subunit M2 B
rt-qPCR	real-time quantitative PCR
SAR	structure-activity relationship
SEC	size-exclusion chromatography
sEH	small epoxide hydrolase

SF	sulfofructose
SFP	sulfofructose phosphate
SLA	3-sulfolactaldehyde
SN	signal-noise ratio
S <sub>N</sub> 2	nucleophilic substitution, bimolecular
SQ	sulfoquinovose
SQDG	sulfoquinovose diacylglycerol
SUCLA2	succinyl-CoA ligase [ADP-forming] subunit beta, mitochondrial
SUCLG1	succinyl-CoA ligase [GDP-forming] subunit alpha, mitochondrial
TEMED	tetramethylethylenediamine
TK1	thymidine kinase 1
TK2	thymidine kinase 2
Tpi	triosephosphate isomerase
TS	transition state
TSA	transition state analogue
TWINK	TWINKLE protein, DNA helicase
TYMP	thymidine phosphorylase
YihS	sulfoquinovose isomerase
YihT	sulfofructose phosphate aldolase
YihU	3-sulfolactaldehyde reductase
YihV	sulfofructose kinase
YihW	sulfoglycolysis transcriptional repressor
αG1P	α-D-glucose-1-phosphate
γ	gyromagnetic ratio
ΔG	free energy (Gibbs)
ε	molar extinction coefficient

# 1. Introduction

## 1.1 The HAD superfamily: structure

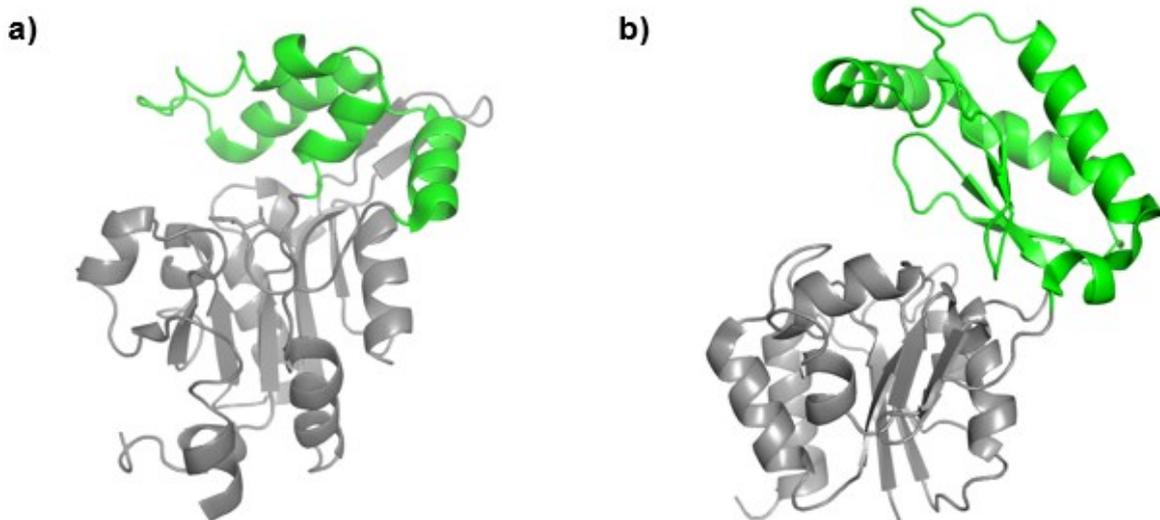
The haloacid dehalogenase (HAD) superfamily is a vast and complex superfamily that comprises multiple subclasses. Named after the archetypal dehalogenase enzymes present in bacteria that are able to cleave a carbon-halogen bond in substrates to exhibiting a detoxification activity towards many chlorinated compounds, a whole superfamily was identified through computer analysis<sup>1</sup>. This newly obtained superfamily comprised many different types of hydrolases: examples include phosphatase, phosphonatasases, phosphomutases, ATPases, etc. Most of the enzymes present in the HAD superfamily perform a phosphoryl group transfer reaction which is catalysed by acid residues present in different positions based on the type of phosphoryl transfer that is performed by each enzyme. The typical HAD enzyme has a very streamlined structure (Figure 1), composed of a Rossmann  $\alpha/\beta$  fold and a number of signature motifs that are present along the catalytic domain of the whole superfamily<sup>2</sup>. Motif I, present in S1, contains the DxD sequence which binds the metal



**Figure 1** A scheme representing the core structure of the typical HAD enzyme. Strands with dashed borders are not always present.

ion cofactor,  $Mg^{2+}$ , and harbours the nucleophilic Asp necessary for the first nucleophile attack when catalysing the phosphoryl transfer reaction, forming a phospho-aspartate intermediate<sup>3</sup>. The Asp<sub>n+2</sub> residue serves as a general acid/base in phosphatases and phosphomutases.

The second feature, Motif II, is located at the C-terminus of S2 and contains a serine or threonine residue, while Motif III is a highly conserved lysine present at the N-terminus of S4. These two motifs are responsible for the stabilisation of the substrate and the reaction intermediate, forming hydrogen bond interactions with oxygen atoms of the phosphoryl group. The last feature, Motif IV, is present at the end of S4 and is usually a DD or DxxxD motif but DxxxxD can be present too. These residues are required for  $Mg^{2+}$  binding and complete the arrangement of the binding pocket in HAD enzymes<sup>4-6</sup>.



**Figure 2** The two most common cap types in HAD superfamily. **a)** Crystal structure of Phosphoserine phosphatase (PDB: 1L8L) as a representative of C1 cap class and **b)** Crystal structure of Phosphomannomutase 1 (PDB: 2FUE) as a representative of C2 cap class. Cap domains are shown in green.

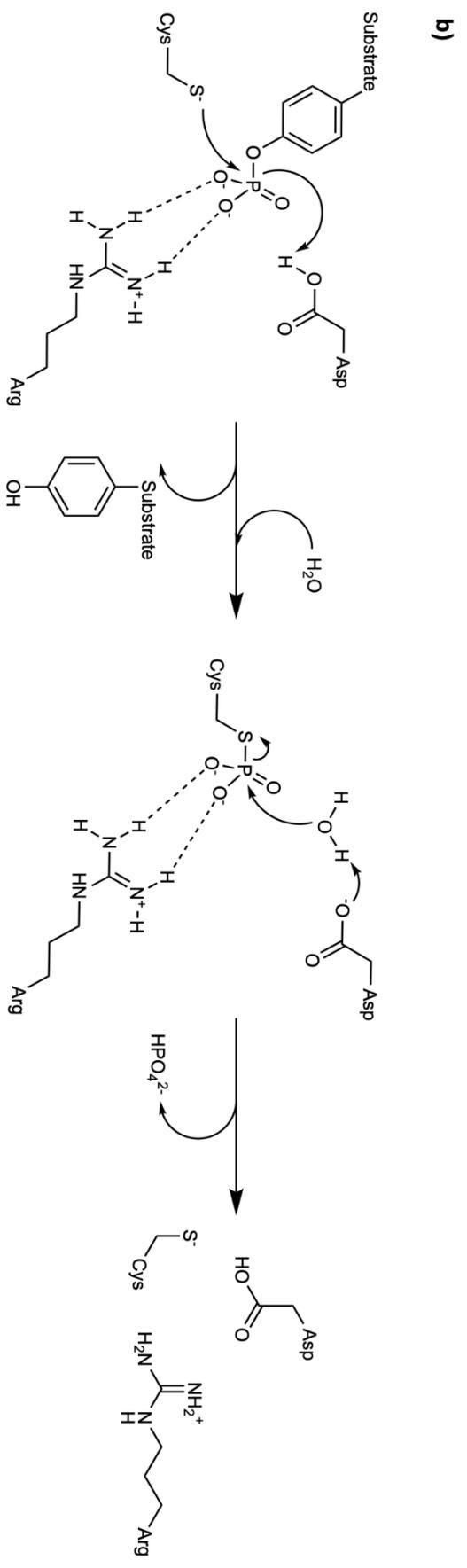
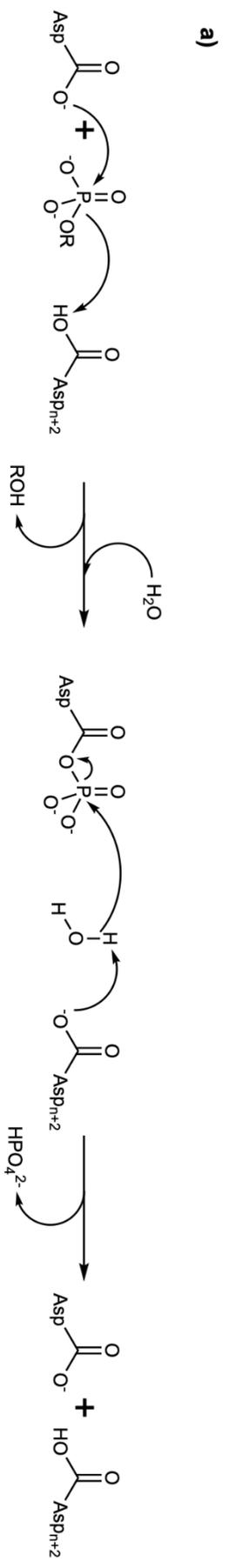
In addition to these four motifs, a subclassification of the enzymes is determined by the presence or absence of a cap domain defining the following: C0 class caps which present small inserts in the predefined locations and C1 and C2 class caps which exhibit large inserts (Figure 1 and 2).

## 1.2 Mechanism of reaction

The mechanism of the reaction has been extensively investigated for this major superfamily<sup>7</sup> and it consists of a nucleophilic attack of the Asp residue of motif I towards the phosphorylated substrate with an  $S_N2$  mechanism forming a phospho-aspartate intermediate and subsequent hydrolysis catalysed by an activated water molecule (Scheme **1a**). This mechanism is quite different compared to other well-known classes of phosphatases such as protein tyrosine phosphatases (PTPs) which makes HAD phosphatases weakly sensitive towards common phosphatase inhibitors in mammalian cells which target non-HAD phosphatases<sup>8-11</sup>. In PTP, the signature motif is represented by the conserved sequence HCxxxxxR where the cysteine acts as the nucleophile<sup>12</sup> (Scheme **1b**).

This mechanism involves the formation of two transition states: TS1 which leads to the formation of the phospho-aspartate intermediate and TS2 which liberates the enzyme and releases inorganic phosphate in solution. This mechanism is an example of general acid/general base (GAGB) catalysis. In the case of HAD phosphatases, the Asp<sub>n</sub> residue is the nucleophile while the Asp<sub>n+2</sub> represents the GAGB residue that executes the proton transfer<sup>13</sup>, donating the proton in the first step to create a better leaving group during the first step of nucleophilic attack and accepting the proton from the nucleophile (water) that hydrolyses the phospho-enzyme intermediate.

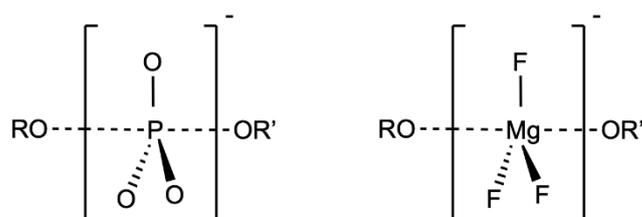
These transition states, TS1 and TS2, can be investigated with a very powerful and established technique which consists in the use of a transition state analogue (TSA) to trap the conformation adopted by the enzyme while the phosphoryl transfer is being performed during the transition state.



**Scheme 1 a)** Classical reaction mechanism for HAD hydrolase enzymes while performing phosphoryl transfer **b)** PTP canonical reaction mechanism

### 1.3 Metal fluorides as transition state analogue complex

When studying phosphoryl transfer as part of a reaction mechanism, the use of metal fluorides ( $\text{MF}_x$ ) as TSAs is remarkably powerful and useful tool of investigation. It involves the use of cheap reagents, such as  $\text{MgCl}_2$ ,  $\text{AlCl}_3$  and  $\text{NH}_4\text{F}$ ,  $\text{NaF}$  in the 10 - 100 mM range, and the protein complex in solution is readily formed and stable. This method has been used in the literature to investigate the catalysis of phosphoryl transfer of ATPases, phosphatases, GTPases, kinases, etc. To this date, there are multiple known metal fluoride complexes that serve different purposes.  $\text{MgF}_3^-$ , for example, is isoelectronic to phosphate and perfectly mimics the trigonal bipyramidal geometry of the transition state and  $\text{MgF}_3(\text{H}_2\text{O})^-$  and  $\text{AlF}_4^-$  in an octahedral geometry while still carrying -1 net charge<sup>14</sup>.



**Figure 3** Comparison between the natural transition state of phosphoryl transfer and the TSA  $\text{MgF}_3^-$ .

In using the protein TSA complex for studying phosphoryl transfer, aluminium fluoride and magnesium fluoride complexes are most commonly used. Members of the HAD superfamily have been crystallised as TSA complexes, such as phosphoserine phosphatase<sup>10,15</sup> and  $\beta$ -phosphoglucomutase<sup>11</sup>, with  $\text{AlF}_4^-$  and  $\text{MgF}_3^-$  respectively. In the case of the  $\text{MgF}_3(\text{H}_2\text{O})^-$ , this unusual TSA has been documented very recently just two times in structures published in the PDB in structures of Zika virus NS3 helicase (PDB 6S0J) and *Actinobacteria* Dop-PupE complex<sup>16,17</sup>.

## 1.4 <sup>19</sup>F NMR spectroscopy as a tool for characterization of MF<sub>x</sub> TSA complex

<sup>19</sup>F NMR spectroscopy is highly valuable for enzymology studies<sup>18</sup>. It is a very sensitive technique due to the 100% abundance of the isotope, the high gyromagnetic ratio ( $\gamma = 251.815 \frac{10^6 \text{rad}}{\text{s} \cdot \text{T}}$ ) and a wide chemical shift range over 800 ppm providing high dispersion of the signals. Additionally, very few compounds in biological systems are naturally fluorinated, giving a very high signal-to-noise (SN) ratio.

While X-Ray crystallography provides the structural data for the complex, <sup>19</sup>F NMR spectroscopy is essential for the confirmation and characterisation of the MF<sub>x</sub> complexes in solution, capable of assessing the electronic environment exposed to the MF<sub>x</sub> when binding to the protein with the conformation that is close to the one of the transition state<sup>19</sup>.

## 1.5 Importance of HAD superfamily enzymes

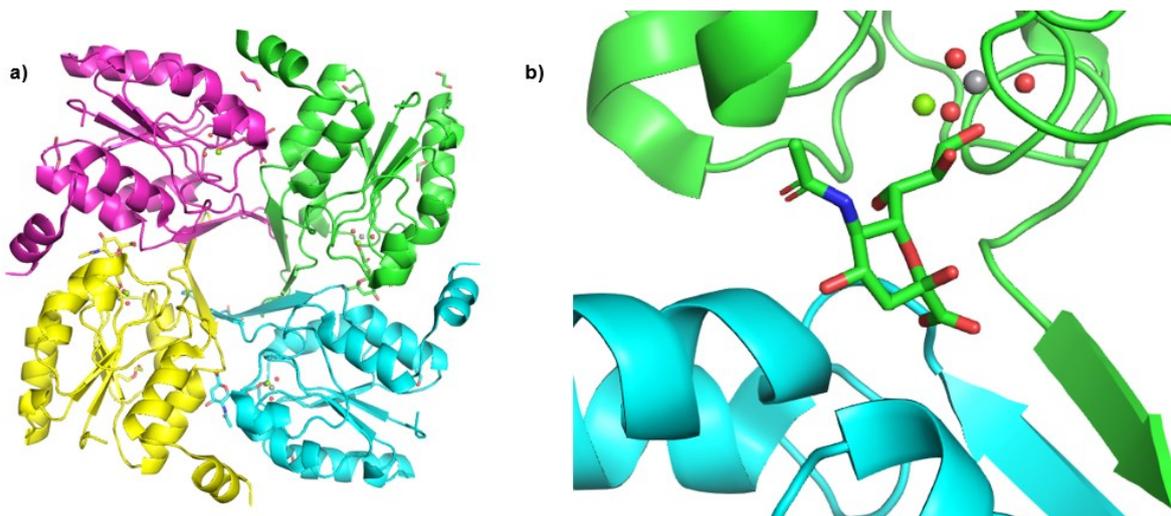
As extensively reported in the literature, HAD phosphatases represent, along with kinases, one of the most important reactions in biology — phosphoryl transfer. In plants, in rice, for example, HAD phosphatases are involved in the contrast of P<sub>i</sub> starvation in roots. A number of HAD phosphatases are actively expressed in stringent conditions to promote extracellular organic phosphate (P<sub>o</sub>) and inorganic phosphate in order to counteract the stress condition<sup>20–23</sup>. In humans, HAD phosphatases have undergone a big evolutionary change. While in procaryotic organisms HAD phosphatases are small proteins with a single domain, in humans the genes encoding these enzymes have been fused with other domains. This results in multi-domain enzymes capable of multiple activities. For example, the fusion of HAD phosphatase with polynucleotide kinase phosphatase (PNKP) and soluble epoxide hydrolase (I) has achieved very specialised enzymes for DNA repair and lipid metabolism<sup>24–26</sup>.

HAD P-type ATPases contribute, together with HAD phosphatases, to represent the majority of HAD enzymes involved in phosphoryl transfer. They are ubiquitous membrane proteins involved in active transport in all living organisms. They are responsible for pumping key small ions, H<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Ca<sup>2+</sup>, Mg<sup>2+</sup>, and transition metal ions which are toxic for the cell such as Cd<sup>2+</sup> and Pb<sup>2+</sup>. Together with ABC-type

transporters, they are responsible for the regulation of the homeostasis of many indispensable ions such as  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$  27–31.

## 1.6 Substrate specificity

The cap domains in HAD phosphatases play a big role in substrate recognition and it is also involved in oligomerisation. As discussed in section 1.1, the caps vary based on the region of insertion and size. While C0 caps are quite small, C1 and C2 caps are significant in size and often fold into domains that rule solvent access to the active site<sup>32,33</sup>. C1 cap is the most common and is composed of  $\alpha$ -helical domain which can vary in size between 2 to 8 helices. The other type, C2 cap, presents an  $\alpha/\beta$  fold and, in contrast to C1, can present more than one substrate specificity loop up to two<sup>34,35</sup>. Enzymes exhibiting a C0 cap domain usually are involved in reactions involving macromolecular substrates, which itself acts as a cap once bound to the active site to exclude the solvent. C0 class can also use small molecules as substrates while recruiting other enzyme molecules for oligomerisation, resulting in a “pseudocapping” that allows the binding of smaller substrates. This is the case of the enzyme 2-keto-3-deoxy-D-glycero-D-galacto-9-phosphonononic acid phosphatase (KDN-9-P) that forms a tetrameric complex to dephosphorylate the substrate<sup>36</sup> (Figure 4).



**Figure 4 a)** Crystal structure of KDN-9-P co-crystallised with N-acetyl- $\beta$ -neuraminic acid,  $\text{VO}_3^-$ ,  $\text{Mg}^{2+}$ , **b)** a zoomed-in view of the interface between two chains, highlighting the “pseudocapping” formed by the oligomerisation of the enzyme.

For C1 and C2 cap classes, loops in charge of the substrate specificity are present in these cap domains which are able to facilitate the entrance and binding of substrates due to the presence of specific amino acids. For example, phosphonoacetaldehyde hydrolase, which presents a C1 class cap domain, must alternate between an open and a closed conformation in order to bind the substrate and liberate the product, respectively. When switching to the close conformation, a specific residue evolved for substrate specificity Lys53 is inserted directly into the catalytic site creating a hydrogen bond network<sup>37-39</sup>.

The substrate of HAD phosphatases can vary from small molecules to proteins, specifically serine/threonine, tyrosine and histidine-phosphorylated proteins. Generally, C1 and C2 cap HAD phosphatases act on small molecules, while C0 (uncapped) HAD phosphatases can bind to phosphoproteins and DNA targets due to their exposed active site, although there are some capped HAD phosphatases that are able to perform dephosphorylation on phosphoproteins<sup>27-31</sup>.

## 1.7 The sulfoglycolytic pathway and comparison with glycolysis

More than 90% of the *E. coli* strains include a metabolic pathway encoded in their genome that is geared towards the importing, processing and excretion of sulfonated carbohydrates and sulfolipids<sup>40</sup>. The presence of this pathway allows growth of these microorganisms in niche environments where primary sources of carbon, namely glucose, glycerol and other high-energy molecules are not available. This pathway, since its discovery, has been named “sulfoglycolysis”, specifically the sulfo-EMP (Embden-Meyerhof-Parnas) pathway, due to the very high similarity to glycolysis for the substrates that are imported into the cell (glucose for glycolysis and sulfoquinovose for sulfoglycolysis) and the various enzymatic transformations that occur inside the cell to break down the 6-carbon molecules into two 3-carbon molecules (see Figure 5). This allows the metabolism of highly abundant substrates in the biosphere such as sulfoquinovose (SQ) and sulfoquinovose diacylglycerol (SQDG), which are ubiquitous structural components, as a membrane lipid, of photosynthetic organisms, including higher plants, algae, cyanobacteria<sup>41,42</sup>, providing the microorganism with a carbon source that can be used for energy production and proliferation.

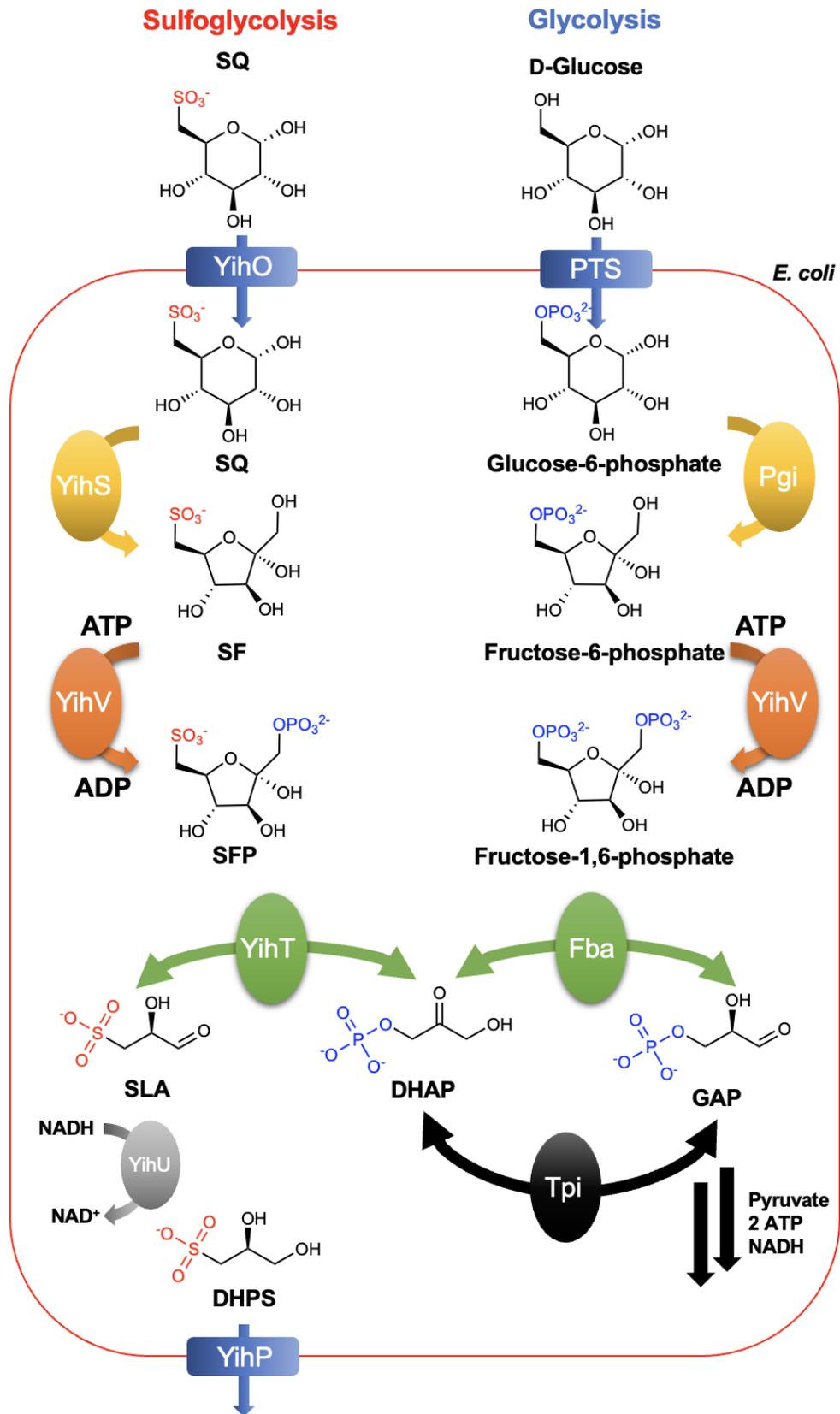
First, the substrate is imported inside the cytoplasm by the action of an active transporter. In the case of glycolysis, while the substrate is imported, the hexokinase phosphorylates glucose on the hydroxyl on position 6; this addition of net negative charge ensures that the molecule cannot diffuse back into the extracellular space. For the sulfoglycolytic pathway, the substrate already bears the sulfonic acid moiety. The next step follows an isomerisation of the 6-membered ring into a 5-membered ring by the action of the isomerase enzyme YihS, preparing the required substrate for the kinase reaction catalysed by the enzyme YihV, which represents the energy investment step of the pathway since a molecule of ATP is required to phosphorylate sulfofructose (SF) to sulfofructose phosphate (SFP). Once the high-energy intermediate is obtained, the aldolase YihT cleaves SFP into two 3-carbon molecules: sulfolactaldehyde (SLA) and dihydroxy acetone phosphate (DHAP). This latter metabolite is also produced in the glycolytic pathway so once formed it can be funnelled in the supply that feeds into the next steps of glycolysis, which after all the transformations will yield pyruvate, NADH and ATP (producing a net 1 molecule of ATP per each molecule of SQ consumed). The SLA will be further processed and reduced

by the enzyme YihU to produce dihydroxy propane sulfonate (DHPS) which represents the true byproduct of the pathway. Since *E. coli* is not able to further process this molecule, it is excreted into the extracellular space.

In early 2010s it was discovered that there are some strains of Bacteria (such as *Cupriavidus pinatubonensis* JMP134) that can actually process this byproduct molecule to extract the sulfono-group and liberate sulfate in the environment, so called “mineralisation”, which is ready for uptake by plants, closing the sulfur cycle<sup>43</sup>.

In glycolysis, the phosphorylation step of the phosphofructokinase, together with the hexokinase reaction and the pyruvate kinase (not shown for simplicity), represent the key control points of the pathway. The kinase-catalysed reactions are the only ones that have highly negative values of  $\Delta G$  compared to the other steps in the pathway that have values between 0 and 3 kJ/mol (close to equilibrium). Albeit extensive literature is available for glycolysis, the rate-limiting step of the sulfoglycolysis, being a much newer pathway, has not been established. Understanding the checkpoints for this pathway will shine light on pathway regulation, possible drug targets of infections caused by *E. coli* and other bacteria living in the human gastrointestinal tract that can use these substrates (such as *Eubacterium rectale*) introduced by a healthy diet which includes green vegetables, especially people that live under vegetarian and vegan diets, to proliferate and sometimes harm the host organism<sup>44</sup>.

In sulfoglycolysis, the only kinase reaction is the one catalysed by the enzyme YihV (performing the same role as the phosphofructokinase in glycolysis) but so far there has not been allosteric regulation being identified<sup>45</sup>. Our knowledge derived from intensive studies performed on the glycolytic pathway drives a very important question regarding sulfoglycolysis: how is the pathway regulated? As far as gene expression is concerned, it has already been shown in the literature that the sulfoglycolysis operon comprises 10 genes in the chromosome, encoding all the enzymes required for the various catalytic steps. A transcription factor has been identified in the gene *yihW*, which regulates the expression of the whole operon by repressing the latter, binding directly on the DNA in the binding site between the *yihV* and *yihU* genes. Once the bacterial cell is exposed to SQ present in the environment, the substrate binds to the repressor YihW which subsequently dissociates from the DNA strand allowing for the full transcription of the operon by the action of RNA polymerase.



**Figure 5** The sulfoglycolytic pathway (left) and the glycolytic pathway (right). Both pathways provide the key metabolite DHAP which can be interconverted into GAP by the activity of the Tpi enzyme. Further steps will yield the pyruvate which can be funnelled into the Krebs cycle, NADH and ATP.

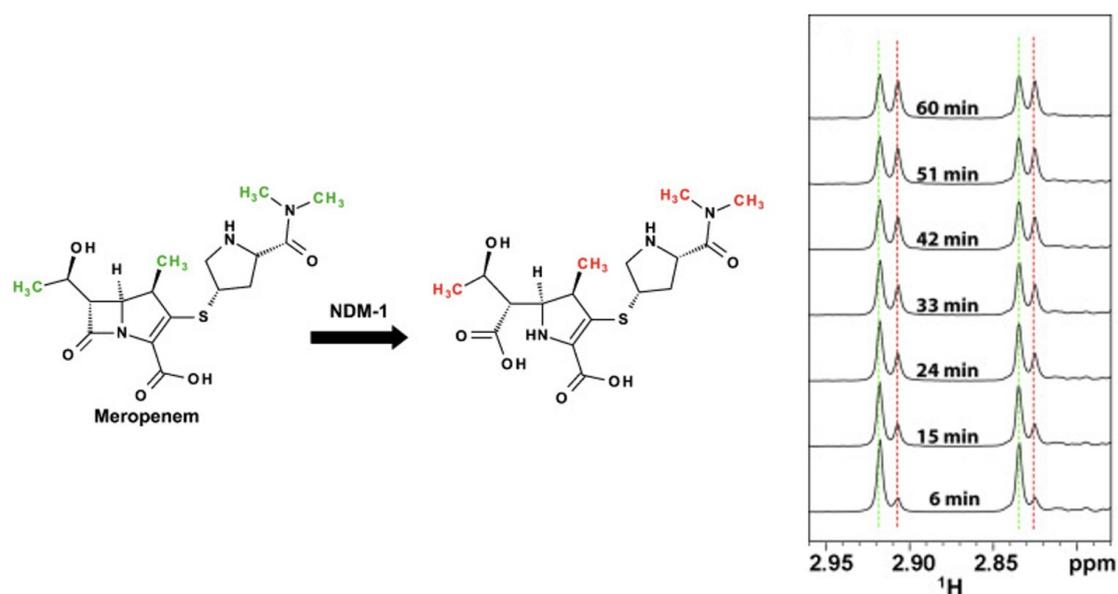
This evidence answers just half of the question. The remaining half consists of how the pathway regulates the direction of the SQ catabolism. As previously described, these reactions (apart from the kinase reaction catalysed by YihV), operate at near equilibrium. In the case of glycolysis, the pathway competes with gluconeogenesis, which basically reverses the direction of the reactions.

This is very useful, because it allows for fine control of each metabolite present in the pathway, regulating the concentration to meet the necessities of the cell in a very fast manner. Conveniently, during gluconeogenesis the organism uses the same enzymes used for glycolysis given that they operate at near equilibrium. The only enzymes that are replaced are the ones involved in the kinase-catalysed steps in glycolysis: hexokinase replaced by glucose-6-phosphatase, phosphofructokinase replaced by fructose-1,6-bisphosphatase and pyruvate kinase replaced by phosphoenolpyruvate carboxykinase. In sulfoglycolysis there is no current knowledge of such fine details. A more detailed understanding of the unique kinase-catalysed step is necessary, since it might be the only checkpoint of the pathway and, as of time of this work, no regulation of the direction of the pathway is known.

### **1.7.1 Possible Role of YihV as a Checkpoint in pathway regulation**

Expression of YihV *in vivo* represents a good starting point for the investigation, since it might represent the key checkpoint of the pathway, considering its similarity with glycolysis. In particular, it is possible to increase the expression of the *yihV* gene with standard recombinant DNA techniques. When the expression level of YihV is upregulated in this way, the pathway will accelerate the metabolism of SQ, depleting this substrate faster compared to the negative control. This will result in faster production of DHPS, the waste product of the pathway, which is excreted in the growth medium. To monitor all these concomitant processes in an *in vivo* manner, a non-invasive technique is necessary. <sup>1</sup>H NMR spectroscopy is perfect for this type of application, being able to detect SQ and DHPS in the live culture without perturbing the environment. This type of whole-cell NMR experiment has been already used in the literature to monitor hydrolysis of meropenem by the action of New Delhi metallo- $\beta$ -lactamase 1 (NDM-1) (Figure 6)<sup>46</sup>. Focusing on the characteristic peaks of SQ and

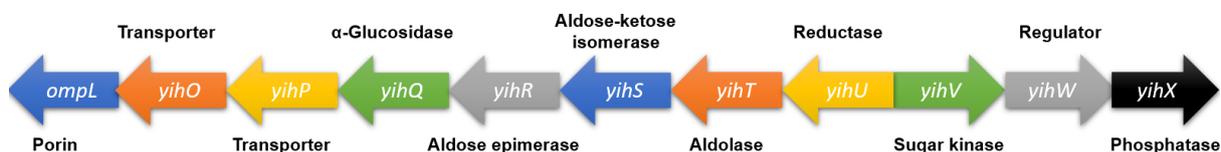
DHPS it is possible to quantitatively determine the overall speed of the pathway by scanning the sample at fixed timepoints and comparing the level of consumption of SQ when YihV is overexpressed and relating this value to the negative control which has native levels of YihV.



**Figure 6** Characteristic  $^1\text{H}$  NMR signals of Meropenem and the inactivated product formed after reaction with NDM-1. The analysis is performed with a live cell culture.

### 1.7.2 Identification of the *yihX* gene as a possible member of the sulfoglycolysis pathway

The *yihX* gene is located downstream of the sulfoglycolysis operon, which is presented as the 10-gene cluster that encodes the enzymes required for the pathway's functioning and regulation (Figure 7).



**Figure 7** The sulfoglycolysis operon and the position of the *yihX* gene relative to the pathway.

Previous studies performed on this enzyme by high throughput assays have highlighted how the enzyme, despite being annotated as an  $\alpha$ -D-glucose-1-phosphate

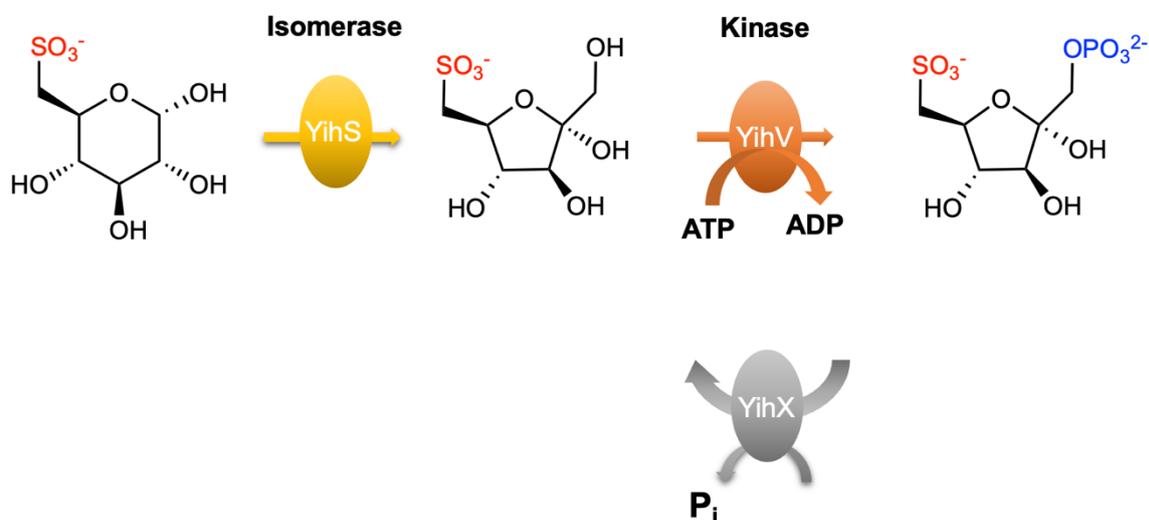
phosphatase, is particularly promiscuous towards a vast variety of different phosphorylated substrates<sup>47</sup> (Figure 8). Particularly, it is quite active towards compounds that present a 5-member ring moiety, such as 2'-deoxyribose-5-phosphate. Combined with the immediate proximity of the *yihX* gene to the sulfoglycolytic pathway, this led to the hypothesis that the YihX enzyme could catalyse the dephosphorylation reaction of SFP to SF, counterbalancing the action of the kinase YihV, providing the answer to the question about the regulation of the direction of the pathway, expanding the pathway to 11 genes and giving us a better understanding of the regulation of the pathway.

	1	2	3	4	5	6	7	8	9	10	11	12	
A	Blank	3 Carbon Acid Sugar			4 Carbon Acid Sugar			Di- and Tri-Phosphates					
B	2 Carbon	Alcohol	Aldolase	Ketose	Alcohol	Aldolase	Ketose						
C	5 Carbon Acid Sugar							Nucleotide Mono-Phosphates					
D	5 Carbon Alcohol Sugar												
E	5 Carbon Aldolase Sugar							Di-Saccharides	Phosphonates Bisphosphates				
F	5 Carbon	7 Carbon Sugar											
G	Ketose	6 Carbon Acid Sugar							Amine Sugars				
A	6 Carbon Alcohol Sugar							Amino Acids					
B	6 Carbon Aldolase Sugar												Easily Hydrolyzed
C	6 Carbon Ketose Sugar												
D	0.0265	0.1410	0.1760	0.4235	0.1285	0.2585	0.0830	0.0590	0.2430	0.3085	0.4420	0.2250	
E	0.1895	0.2330	0.0625	0.3400	0.5610	0.1675	0.3290	0.0545	0.2520	0.1495	0.1185	0.3180	
F	0.3425	0.3030	0.5555	0.3830	0.0640	0.0360	0.2250	0.2630	0.4375	0.4330	0.1770	0.2635	
G	0.1315	0.0170	0.4090	0.4745	0.4760	0.0845	0.4280	0.0600	0.1620	0.2030	0.1645	0.2210	
H	0.4170	0.1125	0.3960	0.5895	0.5460	0.2880	0.4425	0.0325	0.0375	0.1225	0.0775	0.1260	
I	0.0000	0.5370	0.4315	0.4620	0.4875	0.2675	0.0735	0.2185	0.1250	0.0735	0.0735	0.2040	
J	0.1835	0.4295	0.3860	0.1255	0.1950	0.0150	0.4905	0.1495	0.0645	0.1045	0.0900	0.3835	
K	0.2320	0.3405	0.2350	0.0930	0.1390	0.1245	0.2155	0.3920	0.0865	0.1880	0.1155	0.4370	
L	0.4870	0.1425	0.3955	0.4875	0.1170	0.1240	0.0000	0.1095	0.0295	0.0580	0.0450	0.0500	
M	0.3740	0.0415	0.4155	0.5115	0.1185	0.4820	0.2025	0.0405	0.3225	0.1365	0.0965	0.1520	
N	0.0860	0.1660	0.0000	0.0805	0.1505	0.0235	0.0370	0.0425	0.3630	0.2245	0.0655	0.1875	
O	0.0305	0.2435	0.5225	0.0575	0.4060	0.3125	0.1225	0.4230	0.1615	0.1585	0.0315	0.0000	
P	0.0000	0.3090	0.3825	0.5080	0.4350	0.3755	0.0255	0.0265	0.5160	0.0505	0.1480	0.0000	
Q	0.3630	0.5080	0.4365	0.4820	0.5990	0.3285	0.0295	0.1075	0.4190	0.1850	0.0510	0.1090	

**Figure 8** The high throughput assay performed by K. Allen et al. on the YihX enzyme. The top table shows the different substrates tested and the bottom table shows the activity towards the specific phosphorylated substrate<sup>47</sup>

To prove the newly formulated hypothesis, firstly, the SFP must be synthesized and then YihX activity towards the SFP dephosphorylation would be assessed. To this end, the same reaction cascade present in sulfoglycolysis for SFP production was constructed *in vitro*. To monitor each step, <sup>1</sup>H or <sup>31</sup>P NMR spectroscopy is used to observe the conversion of each pathway intermediate. The starting point will be the substrate SQ, which will be chemically synthesized starting from commercially available methyl  $\alpha$ -D-glucopyranose (see Scheme 4). Each enzyme involved in the transformation of SQ will be recombinantly expressed and purified with a combination

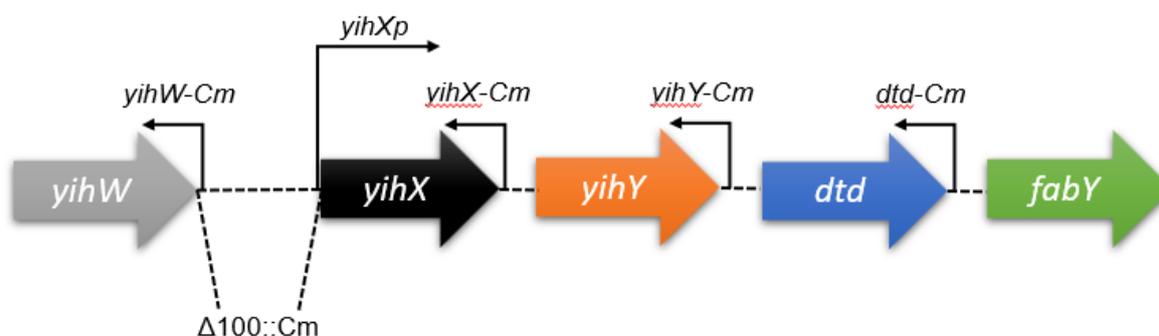
of IMAC and SEC to ensure high purity ( $\geq 95\%$ ). Once SQ and all the enzymes are obtained, the following cascade of reactions will be performed (see Scheme 2).



**Scheme 2** The series of enzymatic reactions required to produce SFP. First, isomerisation of SQ to SF by the action of YihS, then phosphorylation to SFP by YihV. Once the reaction is terminated, dephosphorylation will be performed by the addition of YihX to the reaction mixture.

### 1.7.3 A closer look at neighbouring genes: *yihX* operon

The *yihX* gene is directly downstream of a promoter sequence that suggests its transcription with the next operon: genes *yihX*, *yihY*, *dtd*, *fabY/yiiD* (see Figure 9), but without a clear definition of co-transcription<sup>48</sup>. The promoter sequence is just upstream of *yihX*. This other operon is related to D-tyrosine metabolism: it is responsible for salvaging the tRNA that is charged with D-tyrosine with the action of *dtd* that encodes for a D-aminoacyl-tRNA deacylase. Using this mechanism, bacteria are able to rescue

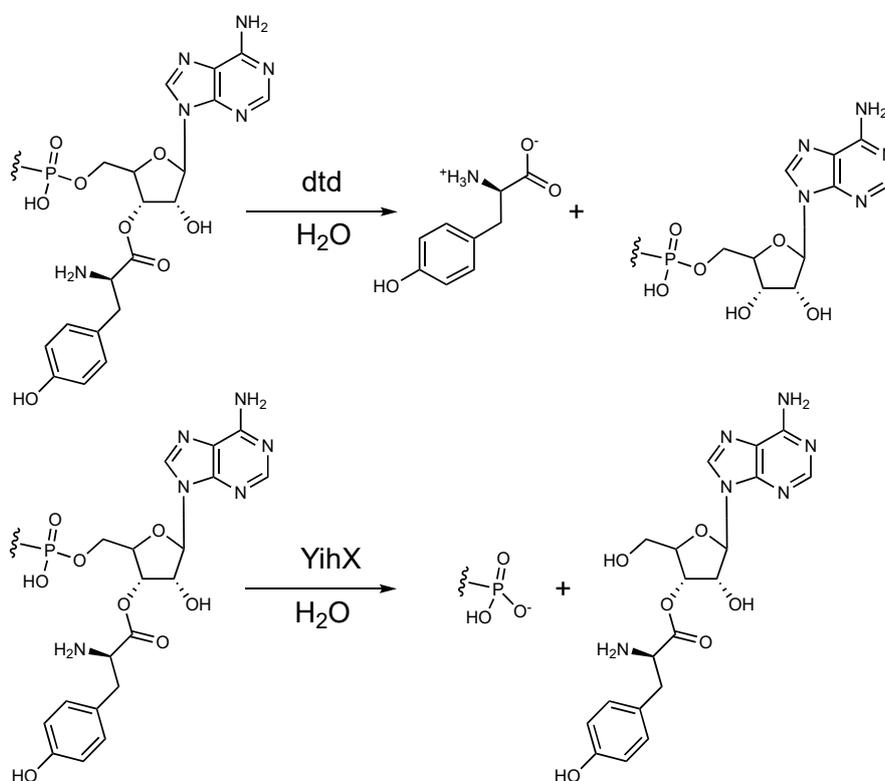


**Figure 9** The YihX operon with all the genes encoded and the position of the promoter sequence<sup>48</sup>.

the tRNA molecules charged with D-tyrosine while surviving in suboptimal conditions due to the lack of primary metabolites.

The other genes present in the operon are *yihX* which encodes a  $\alpha$ -D-glucose-1-phosphate hydrolase, *yihY* which encodes for a membrane protein and *fabY/yiiD* which encodes for an acyltransferase involved in fatty acid metabolism. The reactivity of *yihX* can hypothetically be linked to the rescue action of *dtd*, recovering essential tRNA molecules by dephosphorylation of mischarged tRNA molecules, but using a different mechanism compared to *dtd*.

YihX is a HAD phosphatase promiscuous towards a variety of phosphorylated substrates, including many 5-member ring carbohydrates (see Figure 8), bearing a wide open active site (despite being a C1 cap class HAD phosphatase, which generally provides specificity towards substrates with small size) with a positively charged groove. It is of interest to investigate if YihX can hydrolyse the phosphoester bond of the 5' position of the ribose ring of a potentially mischarged tRNA molecule with D-tyrosine. In this hypothetical case, the tRNA molecule would be hydrolysed and would not be available for protein synthesis purposes (see Scheme 3).



**Scheme 3 top)** Reaction catalysed by *dtd* to salvage tRNA molecules charged with D-tyrosine **bottom)** Hypothetical reaction catalysed by YihX, dephosphorylating the tRNA charged with D-tyrosine.

## 1.8 Aims and Objectives

1. The whole-cell NMR assay will elucidate whether YihV is the rate-limiting step in sulfoglycolysis. This will add to the current knowledge of the pathway, which could identify a possible drug target for those organisms that are able to metabolise SQ.
2. Determining if YihX is part of the sulphoglycolytic pathway or the D-tyrosine metabolism operon will finally allocate this enzyme in the correct operon if it is part of one of them. This consists in determining if YihX is capable of dephosphorylating SFP as a substrate thus being part of sulfoglycolysis and representing the enzyme able to counterbalance the action of the YihV kinase *de facto* controlling the direction of the pathway.
3. The crystallisation of YihX as a TSA complex will allow for new insights into the reaction mechanism. The metal fluoride to be used in this case is  $MgF_3^-$ .
4. Determination of kinetic parameters for YihX towards dephosphorylation reaction towards possible substrates other than  $\alpha$ G1P.
5. A new reaction mechanism for dephosphorylation will be proposed. This will reflect the structural implications derived from the new crystal structure.

## 2. Materials and Methods

### 2.1 Molecular Cloning

Gene oligonucleotides encoding for enzymes relevant for this work were purchased from Thermo Fischer Scientific as DNA strings.

All the primers used for PCR amplification were purchased from Merck Sigma Aldrich.

Restriction enzymes, T4 Ligase and compatible buffers for plasmid digestion and ligation protocols were purchased from Thermo Fischer Scientific.

PrimeStar MAX DNA polymerase and compatible buffers were purchased from Takara.

NEBuilder HiFi DNA assembly master mix was purchased from New England Biolabs.

### 2.2 List of *E. coli* Strains

XL1-Blue [*endA1 gyrA96(nal<sup>R</sup>) thi-1 recA1 relA1 lac gln'44 F'<sup>+</sup>:: Tn10 proAB<sup>+</sup> lacI<sup>q</sup> Δ(lacZ)M15] *hsdR17*(rK<sup>-</sup> mK<sup>+</sup>)*

BL21(DE3) [F<sup>-</sup> *ompT hsdS<sub>B</sub>* (rB<sup>-</sup> mB<sup>-</sup>) *gal dcm* (DE3)]

K-12 MG1655 [F<sup>-</sup> λ<sup>-</sup> *ilvG<sup>-</sup> rfb-50 rph-1*]

BW25113 [F<sup>-</sup> λ<sup>-</sup> *rrnB3 DElacZ4787 hsdR514 DE(araBAD)567 DE(rhaBAD)568 rph-1*]

JW5566 [F<sup>-</sup> λ<sup>-</sup> *rrnB3 DElacZ4787 hsdR514 DE(araBAD)567 DE(rhaBAD)568 rph-1 ΔyihX*]

JW3858 [F<sup>-</sup> λ<sup>-</sup> *rrnB3 DElacZ4787 hsdR514 DE(araBAD)567 DE(rhaBAD)568 rph-1 Δdtd*]

## 2.3 Buffers

All aqueous buffers were prepared using deionised H<sub>2</sub>O from a water purification system (Purelab Chorus 2) with resistivity levels of  $15 \cdot 10^6 \Omega \times \text{cm}$ . The desired pH of each buffer was achieved by titration with HCl or NaOH while constantly monitoring the solution with a pH meter (Jenway 3510) equipped with a SJ 113 pH electrode that was calibrated prior to measurements with a 3-point calibration curve (pH 4.00, pH 7.00, pH 10.00) using standard solutions (Fisher Scientific). All buffer components, other than where specified, were commercially available from Merck Sigma Aldrich and Thermo Fisher Scientific.

## 2.4 Plasmids

**pET29-YihX-His<sub>6</sub>** was cloned by digestion and ligation protocol. The vector was digested from a pET29-CK (human Creatine Kinase) plasmid available in the lab using NdeI and XhoI restriction enzymes. The YihX gene was digested from a pUC18-YihX plasmid available in the lab using NdeI and XhoI restriction enzymes. T4 Ligase was used to obtain the final construct.

**pET21b-YihS-His<sub>6</sub>** was kindly cloned and provided by Dr. Zuyan Wu.

**pET29-YihV-His<sub>6</sub>** Primers YihV-F and YihV-R were used to amplify the gene by PCR from the YihV synthetic gene fragment. Primers pET29-YihV-F and pET29-YihV-R were used to amplify the vector from pET29-YihX-His<sub>6</sub> plasmid. Gibson assembly of the PCR products yielded the final construct.

**pET29-YihX<sub>G8D</sub>-His<sub>6</sub>** was generated by SDM using primers YihX<sub>G8D</sub>-F and YihX<sub>G8D</sub>-R.

**pET29-YihX<sub>D6N</sub>-His<sub>6</sub>** was generated by SDM using primers YihX<sub>D6N</sub>-F and YihX<sub>D6N</sub>-R.

**pET29-YihX<sub>K141L</sub>-His<sub>6</sub>** was generated by SDM using primers YihX<sub>K141L</sub>-F and YihX<sub>K141L</sub>-R.

**pET29-YihX<sub>N108L</sub>-His<sub>6</sub>** was generated by SDM using primers YihX<sub>N108L</sub>-F and YihX<sub>N108L</sub>-R.

**pET29-PG\_0725-His<sub>6</sub>** was cloned by Gibson assembly using primers PG\_0725-F and PG\_0725-R to amplify the synthetic gene fragment. Primers pET29-PG\_0725-F and

pET29-PG\_0725-R were used to amplify the vector from pET29-YihX-His<sub>6</sub> plasmid. Gibson assembly of the PCR products yielded the final construct.

**pET15b-Usp2-cc-His<sub>6</sub>** was kindly provided by Dr. Patrick Baumann.

**pEXP-TK2** was present in the lab.

**pET28a-His<sub>6</sub>-TF-Ub-TK2** Primers TK2-F and Tk2-R were used to amplify the gene by PCR from the TK2 gene from pEXP-TK2 plasmid. Primers pET28a-TK2-F and pET28a-TK2-R were used to amplify the vector from pET28a-RhoA-His<sub>6</sub> plasmid which was kindly provided by Dr. Patrick Baumann. Gibson assembly of the PCR products yielded the final construct.

**pCA24N** vector was available in the lab.

**pCA24N-YihX** Primers T5YihX-F and T5YihX-R were used to amplify the gene by PCR from pET29-YihX-His<sub>6</sub> plasmid. Primers pCA24N-T5YihX-F and pCA24N-T5YihX-R were used to amplify the vector from pCA24N vector. Gibson assembly of the PCR products yielded the final construct.

**pCA24N-YihV** Primers T5YihV-F and T5YihV-R were used to amplify the gene by PCR from pET29-YihV-His<sub>6</sub> plasmid. Primers pCA24N-T5YihV-F and pCA24N-T5YihV-R were used to amplify the vector from pCA24N vector. Gibson assembly of the PCR products yielded the final construct.

pET vectors are used in combination with *E. coli* strains that express T7 RNA polymerase: for this work, *E. coli* BL21(DE3) is the strain used in conjunction with pET expressing systems. pCA24N vectors are used in combination with *E. coli* K-type strains that express T5 RNA polymerase: for this work, *E. coli* K-12 MG1655, parent strain *E. coli* BW25113, and knockout strains *E. coli* JW5566 and JW3858 are used in conjunction with pCA24N expressing systems.

## 2.5 Primers

**Table 1** List of primers.

<b>ID</b>	<b>5' - 3' nucleotide sequence</b>
YihV-F	GATCTCACCATCACCATCACCATATGATTCGTGTTGC
YihV-R	CTTAGCGGCCGCATAGGCCTTAAACAAACAGGCTC
YihX <sub>G8D</sub> -F	GAAGGAGATATACATATGCTCTATATCTTTGATTTAGATAA TGTGATTGTGCGATATCG
YihX <sub>G8D</sub> -R	TAAATCAAAGATATAGAGCATATGTATATCTCCTTC
YihX <sub>D6N</sub> -F	ATATACATATGCTCTATATCTTTAATTTAGGTAATGTGATTG TCGATATCGAC
YihX <sub>D6N</sub> -R	AAAGATATAGAGCATATGTATATCTCCTTCTTAAAG
YihX <sub>K141L</sub> -F	TGTCGCAAGATCTGGGGATGCGCCTGCCTGAAGCACG AATTTACCA
YihX <sub>K141L</sub> -R	ATCCCCAGATCTTGCGACAG
YihX <sub>N108L</sub> -F	ATCGCGTGGTGGTGCTTTCCCTTACCAACCGCCTGCAT AC
YihX <sub>N108L</sub> -R	AAAGCACCACCACGCGATG
PG_0725-F	ATGATTCGCAACATCGTGTTTG
PG_0725-R	TCGAGTTTCTGTTACGCGAG
pET29-PG_0725-F	TGCTGCGTGAACAGAACTCGAGCACCACCACCAC
pET29-PG_0725-R	TCAAACACGATGTTGCGAATCATATGTATATCTCCTTCTT AAAG
TK2-F	ATGCCATCAGTGATCTGTGTCGAG
TK2-R	TGGGCAATGCTTCCGATTCTCTGGAG
pET28a-TK2-F	AAGCATTGCCATAAGGATCCGAATTCGAG
pET28a-TK2-R	ACTGATGGCATGCCACCACG
T5YihX-F	CACCATCACCATATGCTCTATATCTTTGATTTAG
T5YihX-R	GCCGCATAGGCCGCATAACACCTTCGC
pCA24N-T5YihX-F	AAGGTGTTATGCGGCCTATGCGGCCGC
pCA24N-T5YihX-R	TAGAGCATATGGTGATGGTGATGGTGAGATCCTCTCATA G
T5YihV-F	GATCTCACCATCACCATCACCATATGATTCGTGTTGC
T5YihV-R	CTTAGCGGCCGCATAGGCCTTAAACAAACAGGCTC
pCA24N-T5YihV-F	CCTATGCGGCCGCTAAGGGTCGAC
pCA24N-T5YihV-R	GTGATGGTGAGATCCTCTCATAGTTAATTTTC
pET29-YihV-F	CCTATGCGGCCGCTAAGGGTCGAC
pET29-YihV-R	GTGATGGTGAGATCCTCTCATAGTTAATTTTC
LsrK-F	AGATATACATATGGCACGTCTGTTTACCCTGAGCGAAAG C
LsrK-R	AGCAGACCCGGTGCTTTCCACAGG

## 2.6 Growth Media

**LB medium:** 1 L of LB medium was prepared by adding 10 g tryptone, 5 g yeast extract and 5 g NaCl in water. The medium is sterilised by steam autoclaving at 121°C and a pressure of 2.0 bar for a minimum of 15 minutes.

**M9 minimum medium:** all solutions are prepared as following and sterilised by autoclaving at 121°C and a pressure of 2.0 bar for a minimum of 15 minutes. 10X M9 salts:

- $\text{Na}_2\text{HPO}_4$  33.7 mM
- $\text{KH}_2\text{PO}_4$  22.0 mM
- NaCl 8.55 mM
- $\text{NH}_4\text{Cl}$  9.35 mM

20% glucose

1 M  $\text{MgSO}_4$

1 M  $\text{CaCl}_2$

1 mg/mL Thiamin

1000X trace element solution:

EDTA 5 g/L 13.4 mM

$\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$  0.83 g/L 3.1 mM

$\text{ZnCl}_2$  84 mg/L 0.62 mM

$\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$  13 mg/L 76  $\mu\text{M}$

$\text{CoCl}_2 \cdot 2\text{H}_2\text{O}$  10 mg/L 42  $\mu\text{M}$

$\text{H}_3\text{BO}_3$  10 mg/L 162  $\mu\text{M}$

$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$  1.6 mg/L 8.1  $\mu\text{M}$

For 1 L of M9 medium add as following:

100 mL of 10X M9 salts, 20 mL of 20% glucose, 1 mL of 1 M  $\text{MgSO}_4$ , 0.3 mL of 1 M  $\text{CaCl}_2$ , 1 mL of 1 mg/mL Thiamin, 10 mL of 1000X trace element solution.

**SQ-M9 medium:** prepared adding 5 mM (final concentration) sulfoquinovose to M9 minimum medium.

**Antibiotic stocks:** 1000-fold stocks of each of the used antibiotics were prepared by dissolving the powder in pure water (Kanamycin and Ampicillin) or EtOH (Chloramphenicol), filter sterilised through a 0.22 µM cellulose syringe filter (Sartorius) and stored at -20 °C.

**LB agar plates:** 1 L of LB agar medium was prepared by adding 10 g tryptone, 5 g yeast extract, 5 g NaCl and 15 g of agar powder in water. The medium is sterilised by steam autoclaving at 121°C and a pressure of 2.0 bar for a minimum of 15 minutes. When the medium is cooled to approximately 40-50 °C, a volume of around 15 mL is poured into each of 90 mm Petri Dishes. If antibiotic selective plates were to be prepared, an appropriate quantity of 1000X stock of antibiotic would be added before pouring the medium.

## 2.7 Competent Cells

*E. coli* competent cells were made in-house from glycerol stocks for each strain. Cells were prepared according to literature procedure. Cells of interest were streaked on LB Agar plates and then a single colony was used to inoculate 10 mL of LB medium and grown overnight at 37 °C, 180 RPM. 100 mL of fresh LB medium was inoculated (1%) with overnight culture and grown at 37 °C, 180 RPM. When OD<sub>600</sub> reached 0.4 the culture was centrifuged for 15 min at 4 °C, 5500g. The LB was removed, and the cells were resuspended in 100 mL cold Transformation Buffer (10 mM MOPS, 25 mM CaCl<sub>2</sub>, 25 mM MnCl<sub>2</sub>, 25 mM MgCl<sub>2</sub>, pH 6.8). Cells were incubated on ice for 1 h. Cells were centrifuged for 15 min at 4 °C, 5500g. Transformation buffer was removed and 4 mL of cold Storage Buffer (10 mM MOPS, 25 mM CaCl<sub>2</sub>, 25 mM MnCl<sub>2</sub>, 25 mM MgCl<sub>2</sub>, pH 6.8, 7% v/v DMSO) was used to resuspend the cells. 100 µL aliquots of cells were prepared in sterile 1.5 mL microfuge tubes and flash-frozen in liquid nitrogen. Aliquots were stored at -80 °C. Cells were tested for contamination by plating on antibiotic-selective LB Agar plates and transformation efficiency was assayed by transformation with pUC18 plasmid.

## 2.8 Transformation Protocol

1  $\mu\text{L}$  of a 100 ng/ $\mu\text{L}$  plasmid is injected to a 100  $\mu\text{L}$  aliquot of *E. coli* cells of interest and incubated for 10 minutes on ice. Heat-shock is performed at 42 °C for 45 seconds and cells are placed on ice for 5 minutes. 700  $\mu\text{L}$  of LB medium are added to the aliquot and incubated at 37 °C for 45-60 minutes (Chloramphenicol and Kanamycin respectively) under mild shaking. 100  $\mu\text{L}$  of the *E. coli* culture is plated on selective agar plates (90 mm diameter) and incubated overnight at 37 °C (16 h).

## 2.9 Plasmid miniprep

Plasmids were prepared by transformation into *E. coli* XL-1 Blue cells. DNA was isolated from overnight cultures cell extracts using BioBasic EZ-10 Spin Column Plasmid DNA Miniprep Kit following manufacturer instructions.

## 2.10 PCR protocol

All PCR reactions are performed in a Techne TC-512 Thermocycler. The program consists of 30 cycles of the following steps: 10 seconds at 98 °C, 5 seconds at 55 °C, 5 seconds per kb at 72 °C. At the end, an additional 5 minutes at 72 °C are added for a final elongation step. The DNA polymerase employed was PRIMEStar MAX, following manufacturer instructions to perform the reaction. To amplify 100 ng of template DNA, primers were added at a 0.2  $\mu\text{M}$  concentration.

## 2.11 Site Directed Mutagenesis (SDM) protocol

All plasmids encoding for mutant proteins were generated with this protocol that employs partially overlapping primers for PCR. The reaction is performed following the PCR protocol mentioned in the section above to amplify the entire plasmid. After the PCR reaction is completed, FastDigest *DpnI* (Thermo Scientific) is used to digest methylated template DNA from the reaction and purified by Agarose gel electrophoresis. The resulting product is transformed into *E. coli* XL-1 Blue cells which will perform homologous recombination to produce the circular DNA of interest.

## **2.12 Gibson Assembly protocol**

DNA assemblies were obtained through Gibson Assembly using NEBuilder HiFi DNA (NEB) as per manufacturer protocols.

## **2.13 High Performance Liquid Chromatography**

HPLC analysis of samples was carried over an Agilent Infinity II HPLC system, using a SHIMADZU Shim-pack GIST C18 column (150 ×4.6 mm, 5 μm). The wavelength used to monitor nucleotide samples was 254 nm. For ATP, ADP, AMP and Adenosine, an unambiguous peak assignment was done by comparing retention time of samples with commercially available standards (Sigma Aldrich, Fluorochem).

### **2.13.1 Sample Preparation and Method for HPLC Analysis of ATP Formation and AMP Dephosphorylation**

Kinetic parameters characterization was performed through an HPLC stop-assay. Once the reaction is initiated, a sample is prepared at regular time intervals. Each sample contains: substrate (ADP or AMP in various concentrations), 100 μM YihX, 15 mM MgCl<sub>2</sub>, buffer (50 mM Tris-HCl, 200 mM NaCl, pH7.0), total volume 500 μL. 50 μL of reaction mixture is quenched with 0.5 μL of Trichloroacetic acid (TCA), spun down at 10000 g, 20 °C and the supernatant neutralised with the addition of 7 μL of 4 M NaOAc. 10 μL of the resulting mixture is injected in an Agilent Infinity II HPLC system. The program consists of a linear gradient of Buffer A (10 mM Tetrabutylammonium hydrogen sulfate, 30 mM Potassium Phosphate, pH 7.4) and Buffer B (Methanol), 92/8 to 50/50, 0.700 mL/min, 35 °C. Absorbance at 254 nM is monitored.

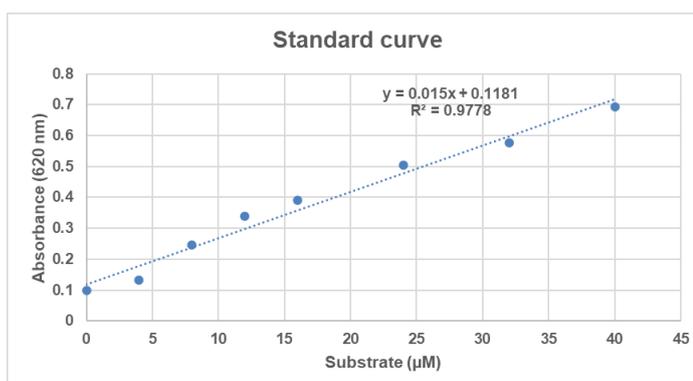
## 2.14 Malachite Green Phosphate Assay

An assay kit was purchased from Sigma-Aldrich. Working reagent was prepared as per manufacturer indication and a standard curve was scanned on a OmegaFluostar plate reader using an Endpoint absorbance protocol with 620 nm as the selected wavelength. Each well of a 96-well plate contained 20  $\mu\text{L}$  of coloring reagent and 80  $\mu\text{L}$  of sample. All the samples contained: 5 mM  $\text{MgCl}_2$ , 0.1  $\mu\text{M}$  Enzyme, Buffer (50 mM HEPES pH 7.5), substrate. A standard curve was plotted, using the phosphate standard provided with the kit. Linear regression analysis retrieves the function which will be used as a standard curve, with equation:

$$y = 0.015x + 0.1181$$

**Table 2** Standard curve for Malachite Green Phosphate Assay.

Substrate ( $\mu\text{M}$ )	Absorbance
0	0.098
4	0.132
8	0.246
12	0.338
16	0.391
24	0.504
32	0.577
40	0.694



## 2.15 Nuclear Magnetic Resonance (NMR) spectroscopy

All NMR spectra presented in this work has been acquired on Bruker spectrometers: Bruker DPX 400 MHz (9.3 T magnetic field  $^1\text{H}$  and  $^{13}\text{C}$ ), Bruker Avance 500 MHz (11.7 T magnetic field  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{31}\text{P}$ ) and a Bruker DRX 600 MHz with cryoprobe (14.0 T magnetic field  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{19}\text{F}$ ). Chemical shifts ( $\delta$ ) are expressed in parts per million (ppm) from the centre of the coupling pattern. Coupling constants are given in Hertz. Residual solvent peaks are used to calibrate  $^1\text{H}$  spectra ( $\text{D}_2\text{O}$   $^1\text{H}$ :  $\delta = 4.79$  ppm,  $\text{CDCl}_3$   $^1\text{H}$ :  $\delta = 7.26$ ,  $\text{CD}_3\text{OD}$   $^1\text{H}$ :  $\delta = 3.31$ ,  $\text{DMSO-}d_6$   $^1\text{H}$ :  $\delta = 2.50$ ).  $^{19}\text{F}$  NMR spectras were referenced to trichlorofluoromethane. Spectra are recorded at constant temperature (20  $^\circ\text{C}$ ) unless specified. For *E. coli* cultures spectra that required a water suppression experiment, a watergate experiment was used with custom power levels to reduce phase distortions. For  $^{19}\text{F}$  experiments, pre-saturation of  $\text{F}^-$  was used with a custom

sequence (see section 7.9). Starting concentration of ADP during phosphorylation experiment was 10 mM.

### 2.15.1 Culture growth and sample preparation for whole-cell NMR assay with M9-SQ

M9 minimal medium was used as a carbon-controlled medium which was supplemented with 5 mM SQ. After the bacteria were transformed with plasmid pCA24N-YihV, 5 mL of M9-SQ medium was inoculated and incubated at 30 °C for a time of 10 to 14 days (growth was very slow and without consistent lag phase) until the culture would reach saturation, which capped at the value of  $A_{600}= 0.30$ . Then, this culture was used to inoculate fresh medium, to free the new sample from any signal of DHPS produced during the growth. Once the  $A_{600}$  value reached 0.15, during the exponential growth phase, the culture was split into 2 different sealed NMR tubes; one sample was induced by the addition of IPTG and the other one was used as a negative control. A “watergate”  $^1\text{H}$  NMR experiment was used and the power level of the coils was adjusted in order to ensure minimum distortion around the 5 ppm region to provide a very accurate integration of the signal corresponding to the anomeric carbon proton of SQ (5.22 ppm). The samples were scanned every 1 h for 24 h (Figure 10).

The samples were run in duplicates. The optical density of the cultures was monitored at the start of the NMR-based analysis, which was always set to 0.15, and at the end of the 24 h period. When measured at the end, the negative control would always show a much higher OD value, with  $A_{600}$  values of 0.29 while the YihV-upregulated sample had an  $A_{600}$  of 0.24.



**Figure 10** Workflow for sample preparation for  $^1\text{H}$  NMR analysis of upregulated YihV cultures in M9-SQ medium.

## 2.16 Mass spectrometry

Protein Liquid Chromatography-Mass Spectrometry (LC-MS) was performed on a Synapt G2-Si quadrupole time-of-flight mass spectrometer (Waters) coupled to an Acquity H-Class ultraperformance liquid chromatography system (Waters); these samples are run by the School of Chemistry Analytical Services in Cardiff University. The column was an Acquity UPLC protein BEH C4 (300 Å, 1.7 µm × 2.1 mm × 100 mm) (Waters) installed in a controlled temperature oven at 60 °C. The elution program is listed in Table X. Spectra are obtained from positive ion mode and analysed using Waters MassLynx software version 4.1. Spectra deconvolution was performed using the maximum entropy 1 processing software.

**Table 3** LC-MS elution program for protein Mass Spectrometry.

Time (min)	A% (H <sub>2</sub> O + 0.1% HCOOH)	B% (MeCN + 0.1% HCOOH)
0	95	5
3	95	5
50	35	65
52	3	97
53	3	97
56	95	5
60	95	5

## 2.17 Fast protein liquid chromatography (FPLC)

Protein purification is carried out on an ÄKTA start FPLC system or a BioRad NGC Quest FPLC system. For IMAC, 5 mL HisTrap FF (GE Healthcare) columns were used; resin cleaning and regeneration was performed as per manufacturer instructions. For anion exchange, 5 mL Q Sepharose XL (GE Healthcare) or 25 mL Q Sepharose (GE Healthcare) columns were used; resin cleaning and regeneration was performed as per manufacturer instructions. Flow rate for IMAC and IEX was 5 mL/min. For SEC, Superdex 75 16/600, Superdex 75 26/600, Superdex 200 26/600 (GE Healthcare) columns were used; resin cleaning and regeneration was performed as per manufacturer instructions. Flow rate for SEC was 2 mL/min.

## 2.18 Gel Electrophoresis

Agarose Gel Electrophoresis was used to purify DNA samples. 1% agarose gels were hand casted using 0.40 g of Agarose dissolved in 40 mL TAE Buffer (40 mM Tris·HCl, pH 8.3, 20 mM AcOH, 1 mM EDTA) and heated until the solution became homogenous. The hot solution was stained by the addition of 2 µL of SYBR Safe DNA (Invitrogen) and poured into a tray with appropriate combs. After the gel solidified, DNA samples were loaded after addition of FastDigest Green Buffer (Thermo Scientific) and the electrophoresis was conducted in TAE Buffer with the following settings: 120 V, 110 mA, 35 min. Gels were imaged using a CHemiDoc XRS+ (Biorad) imaging system. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to screen protein-containing samples and assess their purity. Gels were hand cast using BioRad Mini-Protean Tetra Handcast Systems with buffers and quantities listed in Table X. The gels were allowed to harden at 4 °C for at least 24 h before use and stored for a maximum amount of 7 days. Samples were prepared mixing the correct amount of 4x SDS Loading Dye and heated at 98 °C for 5 minutes prior loading on gel. Gels were run on BioRad Mini-Protean electrophoretic cells using the following settings: 400 V, 35 mA, 40-50 min.

**Table 4** Buffers and reagents used for SDS-PAGE.

	<b>Component</b>	<b>Concentration</b>
<b>4x Resolving Buffer</b>	<ul style="list-style-type: none"> <li>• Tris·HCl pH 8.8</li> <li>• SDS</li> </ul>	1.5 M 0.4% (w/v)
<b>4x Stacking Buffer</b>	<ul style="list-style-type: none"> <li>• Tris·HCl pH 6.8</li> <li>• SDS</li> </ul>	0.5 M 0.4% (w/v)
<b>4x SDS Loading Dye</b>	<ul style="list-style-type: none"> <li>• Tris·HCl pH 6.8</li> <li>• DTT</li> <li>• SDS</li> <li>• Bromophenol Blue</li> <li>• Glycerol</li> </ul>	0.2 M 0.4 M 8 % (w/v) 0.4 % (w/v) 40%
<b>15% Resolving Buffer</b>	<ul style="list-style-type: none"> <li>• Acrylamide/Bisacrylamide (37.5:1)</li> <li>• 4x Resolving Buffer</li> <li>• APS</li> <li>• TEMED</li> </ul>	15% (w/v) 25% (v/v) 5 mM 7 mM
<b>12% Resolving Buffer</b>	<ul style="list-style-type: none"> <li>• Acrylamide/Bisacrylamide (37.5:1)</li> <li>• 4x Resolving Buffer</li> <li>• APS</li> <li>• TEMED</li> </ul>	12% (w/v) 25% (v/v) 5 mM 7 mM
<b>SDS Running Buffer</b>	<ul style="list-style-type: none"> <li>• Tris·HCl pH 6.8</li> <li>• Glycine</li> <li>• SDS</li> </ul>	0.025 M 0.02 M 0.1 % (w/v)

## 2.19 Spectrophotometric measurement of protein concentration

Protein concentration is determined with a NanoDrop One<sup>c</sup> (Thermo Scientific) using Lambert-Beer law equation, where  $c$  is the concentration,  $A$  is the absorbance,  $\epsilon$  is the extinction coefficient ( $\frac{1}{M \cdot cm}$ ) and  $l$  is the path length (1 cm).

$$c = \frac{A}{\epsilon \cdot l}$$

For accurate measurement, a baseline correction at 315 nm is used. Extinction coefficients are calculated by ExPASy-ProtParam using equation listed below, where  $\epsilon$  is the extinction coefficient ( $\frac{1}{M \cdot cm}$ ),  $nW$  is the number of tryptophane residues,  $nY$  is the number of tyrosine residues and  $nC$  is the number of cysteine residues.

$$\epsilon = (5500 \cdot nW) + (1490 \cdot nY) + (125 \cdot nC)$$

## 2.20 Thin Layer Chromatography (TLC)

TLC was performed on Merck 60 F<sub>254</sub> silica gel aluminium plates. The plates were dipped into a Cerium Ammonium Molybdate stain (0.5 g Ce(NH<sub>4</sub>)<sub>4</sub>(SO<sub>4</sub>)<sub>4</sub>·2H<sub>2</sub>O, 12 g (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>·4H<sub>2</sub>O, 235 mL H<sub>2</sub>O, 15 mL H<sub>2</sub>SO<sub>4</sub>) and subsequently heated.

## 2.21 Protein Expression and Purification

### 2.21.1 His<sub>6</sub>-YihS recombinant protein

pET21b-YihS-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 100 µg/mL Ampicillin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 10 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 2 shaking flasks (0.5%) each containing 1 L of LB medium supplied with 100 µg/mL Ampicillin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 18 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending

the pellets in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was filtered with 0.45 µm filters and centrifuged again for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter and loaded on a 5 mL HisTrap (GE Healthcare) column attached to an ÄKTA start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 300 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 25 mL 0% Buffer B, 75 mL 0-100% Buffer B, 30 mL 100% Buffer B; all buffers were filtered over 0.45 µm filters cellulose filters and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC and then purer fractions were concentrated using a Amicon Viva Spin 10kDa MWCO concentrator. Protein was buffer exchanged into 50 mM Ammonium Acetate pH 8.00.

### **2.21.2 YihX-His<sub>6</sub> recombinant protein**

pET29-YihX-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a shaker at 200 RPM. Overnight culture was used to inoculate 6 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50 µg/mL Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending the pellets in Buffer A (50 mM Tris-HCl, 300 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was filtered with 0.45 µm filters and centrifuged again for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter and loaded on a 5 mL HisTrap (GE Healthcare) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 300 mM NaCl, 400 mM Imidazole, pH 7.00) using a gradient elution: 25 mL 0% Buffer B, 75 mL 0-100% Buffer B, 30 mL 100% Buffer B; all buffers were filtered over 0.45 µm filters cellulose filters and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC and then purer fractions were concentrated using an Amicon Viva Spin 10kDa cut-off concentrator. Further purification was performed through SEC

using a Superdex 75 26/600 column using Buffer (50 mM Tris-HCl, 200 mM NaCl) at a flow rate of 1.0 mL/min, 0.30 MPa, 350 mL total elution volume. Fractions were screened on SDS-PAGE then combined and concentrated using a Amicon Viva Spin 10 kDa MWCO concentrator.

### **2.21.3 YihV-His<sub>6</sub> recombinant protein**

pET29-YihV-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 6 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50 µg/mL Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending the pellets in Buffer A (50 mM Tris-HCl, 200 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 200 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (15CV); all buffers were filtered over 0.45 µm cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa cut-off concentrator. Further purification was performed through SEC using a Superdex 75 26/600 (Cytiva) column using Buffer (50 mM Tris-HCl, 200 mM NaCl) at a flow rate of 1.0 mL/min, 0.30 MPa, 350 mL total elution volume on a BioRad Quest FPLC system. Fractions were screened on SDS-PAGE then combined and concentrated using an Amicon Viva Spin 10 kDa MWCO.

### **2.21.4 YihX<sub>G8D</sub>-His<sub>6</sub> recombinant protein**

pET29-YihX<sub>G8D</sub>-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL Kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 2 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50 µg/mL Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending the pellets in Buffer A (50 mM Tris-HCl, 200 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 200 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (15CV); all buffers were filtered over 0.45 µm cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa MWCO concentrator.

### **2.21.5 YihX<sub>D6N</sub>-His<sub>6</sub> recombinant protein**

pET29-YihX<sub>D6N</sub>-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL Kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 2 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50 µg/mL Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending the pellets in Buffer A (50 mM Tris-HCl, 200 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The

supernatant was passed through a 0.45  $\mu\text{m}$  filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 200 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (15CV); all buffers were filtered over 0.45  $\mu\text{m}$  cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa MWCO concentrator.

### **2.21.6 YihX<sub>K141L</sub>-His<sub>6</sub> recombinant protein**

pET29-YihX<sub>K141L</sub>-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50  $\mu\text{g}/\text{mL}$  Kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 2 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50  $\mu\text{g}/\text{mL}$  Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending the pellets in Buffer A (50 mM Tris-HCl, 200 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45  $\mu\text{m}$  filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 200 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (15CV); all buffers were filtered over 0.45  $\mu\text{m}$  cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa MWCO concentrator.

### **2.21.7 YihX<sub>N108L</sub>-His<sub>6</sub> recombinant protein**

pET29-YihX<sub>N108L</sub>-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL Kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 2 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50 µg/mL Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending the pellets in Buffer A (buffer A (NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>PO<sub>4</sub> 50 mM pH = 7.4, NaCl 300 mM, imidazole 12 mM, 2-ME 20 mM, glycerol 30% (v/v))) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 200 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (15CV); all buffers were filtered over 0.45 µm cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa MWCO concentrator.

### **2.21.8 PG\_0725-His<sub>6</sub> recombinant protein**

pET29-PG\_0725-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL Kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 2 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50 µg/mL Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending

the pellets in Buffer A (50 mM Tris-HCl, 200 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 200 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (15CV); all buffers were filtered over 0.45 µm cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa MWCO concentrator.

### **2.21.9 His<sub>6</sub>-Usp2-cc recombinant protein**

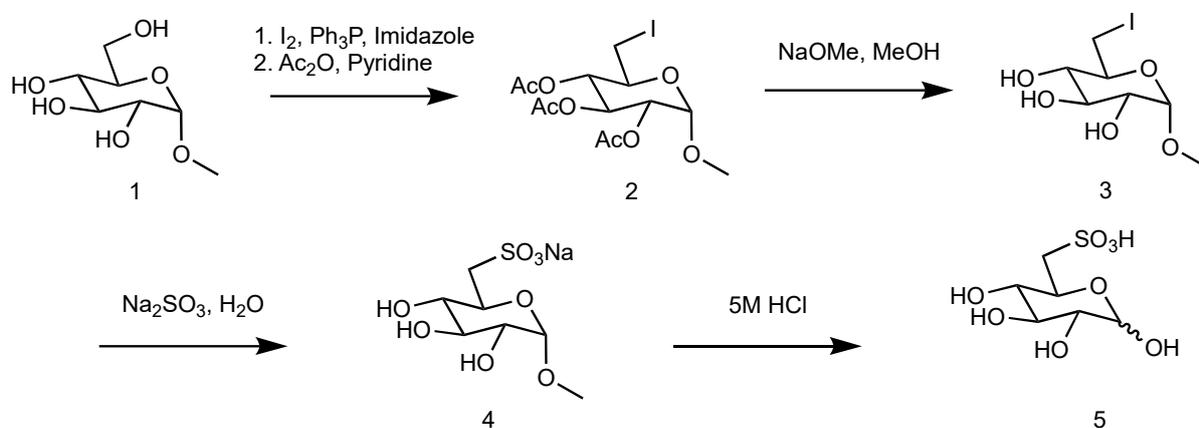
pET15b-Usp2-cc-His<sub>6</sub> expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL Kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 2 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 100 µg/mL Ampicillin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression was carried out for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000 g, draining the medium and resuspending the pellets in Buffer A (NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>PO<sub>4</sub> 50 mM pH = 7.4, NaCl 300 mM, imidazole 12 mM, 2-ME 20 mM, glycerol 30% (v/v)) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>PO<sub>4</sub> 50 mM pH = 7.4, NaCl 300 mM, imidazole 300 mM, 2-ME 20 mM, glycerol 30% (v/v)) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (25CV); all buffers were filtered over 0.45 µm cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa MWCO concentrator and buffer exchanged to Buffer C (buffer A without imidazole).

### 2.21.10 His<sub>6</sub>-TF-Ub-TK2 recombinant protein

pET28a-His<sub>6</sub>-TF-Ub-TK2 expression vector was used to transform BL21(DE3) chemically competent cells which were plated onto selective agar plates containing 50 µg/mL Kanamycin and incubated for 16 h at 37 °C. A single transformant was used to inoculate 20 mL of LB medium and incubated overnight at 37 °C shaking on a vertical shaker at 200 RPM. Overnight culture was used to inoculate 4 shaking flasks (0.3%) each containing 1 L of LB medium supplied with 50 µg/mL Kanamycin. At OD<sub>600</sub> 0.6 the cultures were induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the expression carried for 20 h at 20 °C. Cells were harvested by centrifugation for 30 minutes at 4 °C at 7000g, draining the medium and resuspending the pellets in Buffer A (50 mM Tris-HCl, 200 mM NaCl, 20 mM Imidazole, pH 7.00) and sonicated. The cell lysate was centrifuged for 20 minutes at 4 °C at 32000g. The supernatant was passed through a 0.45 µm filter (Sartorius) and loaded on a 5 mL HisTrap (Cytiva) column attached to an ÄKTA Start FPLC system. Elution was performed with Buffer B (50 mM Tris-HCl, 200 mM NaCl, 200 mM Imidazole, pH 7.00) using a gradient elution: 100% Buffer A (7CV), 100% Buffer B linear gradient (15CV); all buffers were filtered over 0.45 µm cellulose filters (Sartorius) and degassed prior to loading on the FPLC system. SDS-PAGE was used to check protein purity after IMAC, and then purer fractions were concentrated using an Amicon Viva Spin 10 kDa MWCO concentrator.

## 2.22 Chemical synthesis

### 2.22.1 Synthesis of SQ



**Scheme 4** SQ synthetic scheme used to prepare the substrate.

#### **Methyl 2,3,4-tri-O-acetyl-6-deoxy-6-iodo- $\alpha$ -D-glucopyranoside (2):**

2.00 g of methyl  $\alpha$ -D-glucopyranoside (1) are suspended in 200 mL of toluene at room temperature then 4.04 g of triphenylphosphine, 2.12 g of imidazole and 3.66 g of Iodine are added to the suspension and stirred at 70 °C for 2 h. The reaction is allowed to cool down to room temperature and 20 mL of water are added to the mixture and stirred vigorously for 10 minutes. The organic layer is extracted with 20 mL of water and the combined aqueous layers are evaporated to dryness under high vacuum. 21 mL of pyridine are added together with 0.128 g of DMAP. 6 mL of  $Ac_2O$  are added to the solution at room temperature and stirred for 6 h upon which another 3 mL of  $Ac_2O$  are added and the mixture let to stir overnight. The solvent is removed *in vacuo* and the residue is dissolved in 40 mL of toluene, washed with 40 mL of water and the combined organic layers evaporated to dryness under high vacuum to afford **2** as an off-white solid, 3.96 g (89.3%):  $^1H$  NMR (500 MHz,  $CDCl_3$ )  $\delta$  5.46 (t,  $J=9.8$  Hz, 1H), 4.95 (d,  $J = 3.75$  Hz, 1H), 4.89-4.84 (m, 2H), 3.80-3.76 (m, 1H), 3.47 (s, 3H), 3.29 (dd,  $J_1 = 11$  Hz,  $J_2 = 2.55$  Hz, 1H), 3.13 (dd,  $J_1 = 11$  Hz,  $J_2 = 8.3$  Hz, 1H), 2.07 (s, 3H), 2.05 (s, 3H), 2.00 (s, 3H).  $^{13}C$  NMR (125 MHz,  $CDCl_3$ )  $\delta$  170.2, 170.1, 169.7, 96.6, 72.4, 70.9, 69.6, 68.6, 55.8, 20.8, 20.8, 20.7, 3.7.

### **Methyl 6-deoxy-6-iodo- $\alpha$ -D-glucopyranoside (3):**

3.46 g of **2** are suspended in 27 mL of MeOH and 0.79 mL of NaOMe (1 M in MeOH) are added and stirred for 30 minutes at room temperature then neutralized by the addition of AcOH. The solvent is removed *in vacuo* and the residue was recrystallised from 95% ethanol to obtain 2.20 g of **3** as a white solid (90%).  $^1\text{H}$  NMR (400 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  4.65 (d,  $J = 3.9$  Hz, 1H), 3.60 (t,  $J = 8.8$  Hz, 2H), 3.44 (s, 3H, OMe), 3.40-3.34 (m, 2H), 3.28 (dd,  $J_1 = 10.3$ ,  $J_2 = 7.5$  Hz, 1H), 3.11 (t,  $J = 9.2$  Hz, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  100.0, 74.2, 73.2, 72.2, 71.0, 54.4, 5.7.

### **Methyl 6-deoxy-6-sulfonato- $\alpha$ -D-glucopyranoside (4):**

1.19 g of **3** are dissolved in 35 mL of water and 1.98 g of  $\text{Na}_2\text{SO}_3$  are added. The solution is stirred at 80 °C for 4 h then the solvent is removed *in vacuo*. Methanol (29 mL) is added, and the mixture is stirred at room temperature overnight. After filtration, the solution is evaporated to dryness under high vacuum and the residue is purified by flash chromatography (EtOAc:MeOH:H<sub>2</sub>O 9:2:1  $\rightarrow$  3:2:1) to obtain 0.850 g of **4** as a colourless solid (77%).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  4.05 (t,  $J = 10.0$  Hz, 1H), 3.69 (t,  $J = 9.1$  Hz, 1H), 3.61 (dd,  $J_1 = 9.6$  Hz,  $J_2 = 3.3$  Hz, 1H), 3.49 (s, 3H), 3.42 (d,  $J = 14.0$  Hz, 1H), 3.28 (t,  $J = 9.2$  Hz, 1H), 3.11 (dd,  $J_1 = 14.5$  Hz,  $J_2 = 9.9$  Hz, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  98.9, 73.0, 72.4, 71.1, 67.8, 55.2, 52.0.

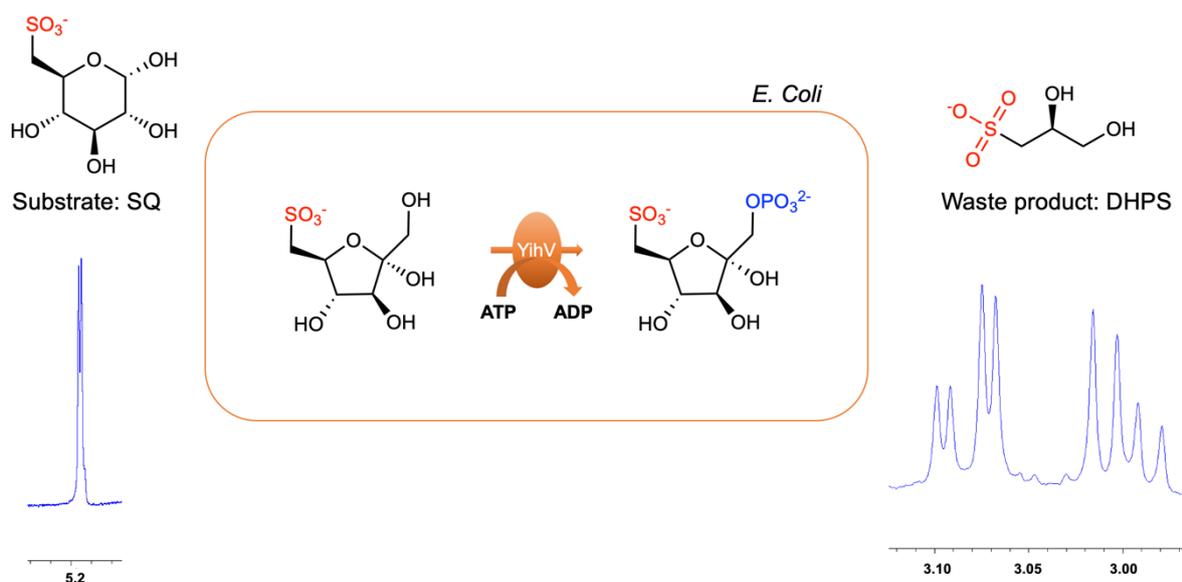
### **Sodium 6-deoxy-6-sulfonato-D-glucopyranose (5):**

0.850 g of **4** dissolved in H<sub>2</sub>O (5.12 mL) was treated with aq. HCl (5 M, 0.86 mL) and stirred at 100 °C for 80 h, then evaporated to dryness under reduced pressure. The residue was dissolved in H<sub>2</sub>O (33.37 mL) and stirred with IR-120 cation-exchange resin (17.86 g,  $\text{Na}^+$  form) for 1 h. The filtrate was evaporated to dryness under high vacuum and the residue was purified by flash chromatography (EtOAc:MeOH:H<sub>2</sub>O 9:2:1  $\rightarrow$  3:2:1) to afford **5** as a brown solid (with slight impurities) (0.660 g, 82%).  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ )  $\delta$  5.17 (d,  $J_{a,b} = 3.8$  Hz, 1H, Ha- $\alpha$ ), 4.63 (d,  $J_{a,b} = 7.9$  Hz, 1H, Ha- $\beta$ ), 4.19 (t,  $J_{d,e} = J_{e,f} = 10.2$  Hz, 1H, He- $\alpha$ ), 3.75 (t,  $J_{d,e} = J_{e,f} = 9.2$  Hz, 1H, He- $\beta$ ), 3.69 (t,  $J_{b,c} = J_{c,d} = 9.2$  Hz, 1H, Hc- $\alpha$ ), 3.51 (dd,  $J_{b,c} = 9.8$ ,  $J_{a,b} = 3.4$  Hz, 1H, Hb- $\alpha$ ), 3.46 (t,  $J_{b,c} = J_{c,d} = 9.6$  Hz, 1H, Hc- $\beta$ ), 3.36 (d,  $J_{f,f'} = 14.7$  Hz, 1H, Hf- $\beta$ ), 3.26–3.20 (m, 3H, Hd- $\alpha$ , b- $\beta$ , d- $\beta$ ), 3.07–3.00 (m, 3H, Hf'- $\alpha$ , f- $\alpha$ , f'- $\beta$ ).  $^{13}\text{C}$  NMR (100 MHz,  $\text{D}_2\text{O}$ )  $\delta$  95.9, 92.0, 75.5, 74.0, 72.7, 72.6, 72.3, 72.1, 71.4, 67.7, 52.3, 52.2.

### 3. Results and discussion

#### 3.1 Investigating YihV as a checkpoint of sulfoglycolysis

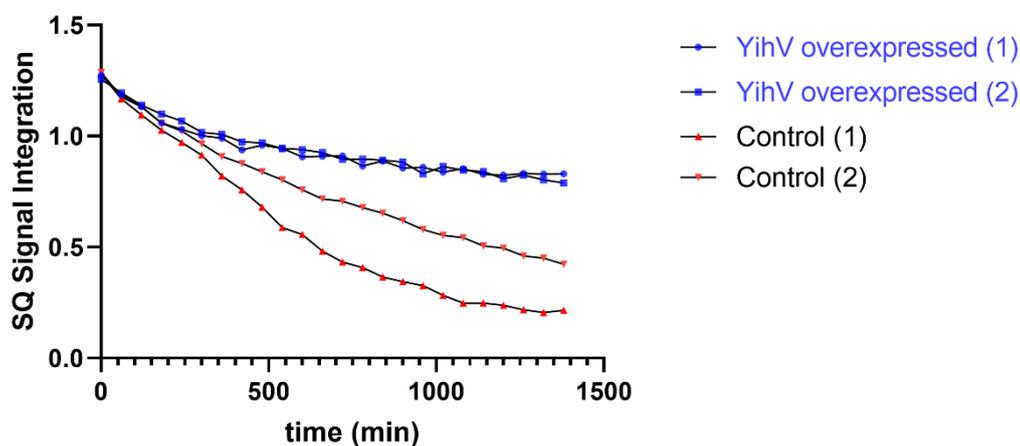
YihV kinase could represent an important check point in the sulphoglycolytic pathway, thus identifying whether it represents the rate-limiting step in the pathway would grant additional knowledge in the regulation of the pathway. The strategy chosen to investigate whether the hypothesis is correct is to employ a whole-cell NMR assay where *E. coli* transformants bearing a plasmid that can be used to upregulate the expression level of YihV are grown in a controlled carbon source environment.  $^1\text{H}$  NMR spectroscopy can be used to identify characteristic peaks for SQ, the substrate used by *E. coli* as a carbon source, and characteristic peaks for DHPS, the by-product of the pathway which is excreted in the media (Figure 11).



**Figure 11** Characteristic  $^1\text{H}$  NMR signals of SQ and DHPS. YihV is upregulated by induction through IPTG addition.

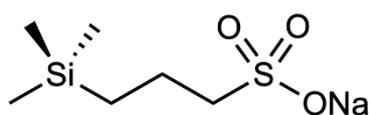
To activate the pathway, SQ must be present in the growth medium in order to bind to the repressor that blocks the expression of the *yih* operon. More importantly, SQ must be the only carbon source present in the growth medium, due to the much greater preference of bacteria to other standard sources such as glucose or glycerol that would activate the main glycolytic pathway while keeping sulfoglycolysis inactive. The strength of this *in vivo* experiment is that *E. coli* is monitored in a non-intrusive manner

while it is cultured in a controlled carbon source medium. The strain used was *E. coli* K-12 MG1655, which presents the sulfoglycolytic pathway in its genome.



**Figure 12** Depletion of SQ displayed using the integration of SQ signal compared to the internal standard.

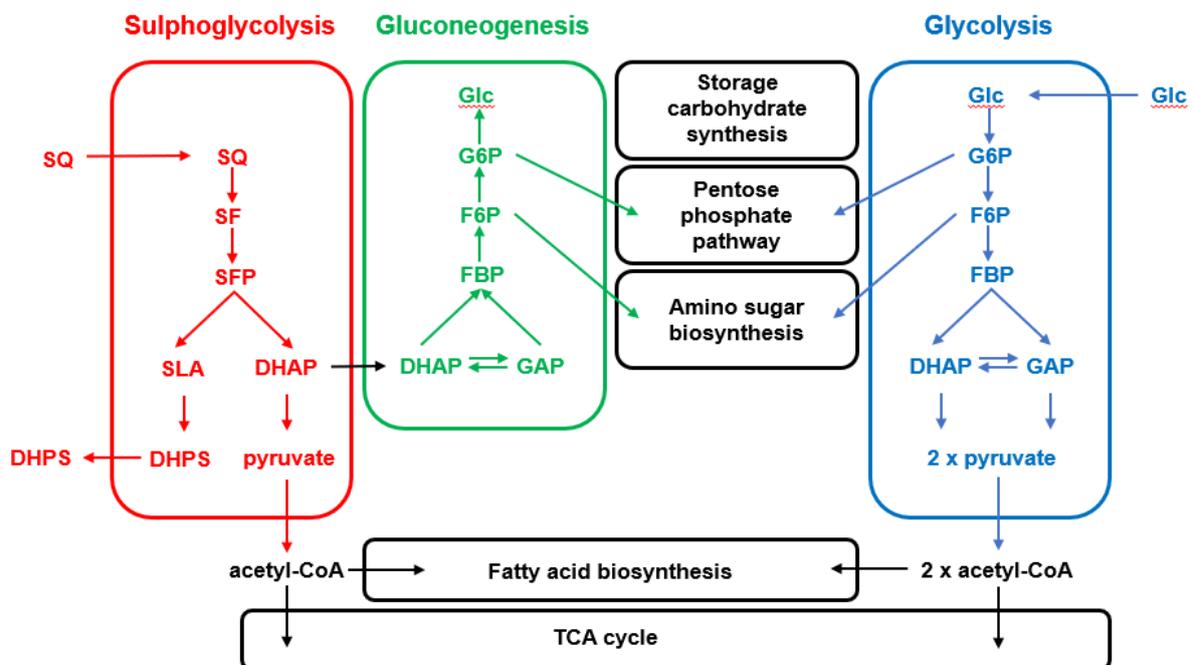
As shown in Figure 12, the data shows an unexpected trend. The data reports the integration of the proton of the anomeric carbon of SQ compared to an internal standard that is non-toxic nor a metabolite for the bacteria, so that its concentration does not change over the course of the 24h NMR experiment; the compound used was sodium trimethylsilylpropanesulfonate (DSS).



**Figure 13** Sodium trimethylsilylpropanesulfonate (DSS).

When YihV is upregulated, the rate of SQ depletion is much slower compared to the control. This behaviour is quite odd and is non-conclusive towards supporting the hypothesis that a higher expression level of the enzyme that represents the only plausible checkpoint in the sulfoglycolytic pathway could enhance the overall rate of the pathway.

Recent work published by the Williams group highlighted that while sulfoglycolysis indeed produces ATP, NADH and pyruvate (although in reduced quantities compared to glycolysis) the pathway does not produce key intermediates that are important for other metabolic pathways, namely G6P and F6P. In this case, while the *E. coli* is able to grow in SQ-only medium through this metabolic pathway, it is lacking key metabolites simply because they are not intermediates of the sulfo-EMP pathway. To compensate, some of the 3-carbon products, DHAP and GAP must be relocated to the gluconeogenic pathway to produce acceptable levels of G6P and F6P. This constitutes a massive drawback for the pathway and would explain why the growth rate is much slower than glucose-grown bacteria. Additionally, it was proven that during the logarithmic phase, the *E. coli* would produce glycogen and trehalose storage, to switch its metabolism back to glucose when the primary source, SQ, would reach depletion in the environment. This behaviour is quite unique and shows that sulfoglycolysis is strictly an emergency pathway<sup>49</sup> (Figure 14). In addition, stimulating protein expression by the induction of T7 expressing system while the bacteria are cultured in minimum medium (M9 supplied with SQ in this case) might trigger metabolic burden in the cell, slowing down the growth rate and justifying a lower optical density value at the end of the experiment<sup>50</sup>.



**Figure 14** Sulfoglycolysis, gluconeogenesis, glycolysis, and other major metabolic pathways. Pathway crosstalk points are highlighted by arrows connecting metabolites to multiple pathways. Adapted from Janice W.-Y. Mui et al<sup>49</sup>.

### 3.2 YihX as a backup enzyme for Dtd

The role of YihX could be very important if it is capable of acting as an alternative salvage path for D-tyrosine. To test this hypothesis, a tailored experiment was designed. First, the bacteria must be put in conditions of exposure to high concentrations of D-tyrosine in the growth medium so that importing it inside the cell is much more favourable than activating its own biosynthetic pathway. Second, a set of knockout and parent strains was obtained from the Keio collection (code BW25113 for the parent strain, code JW5566 for the  $\Delta yihX$  strain and code JW3858 for the  $\Delta dtd$  strain). This allows for independent evaluation of the effects of both enzymes.

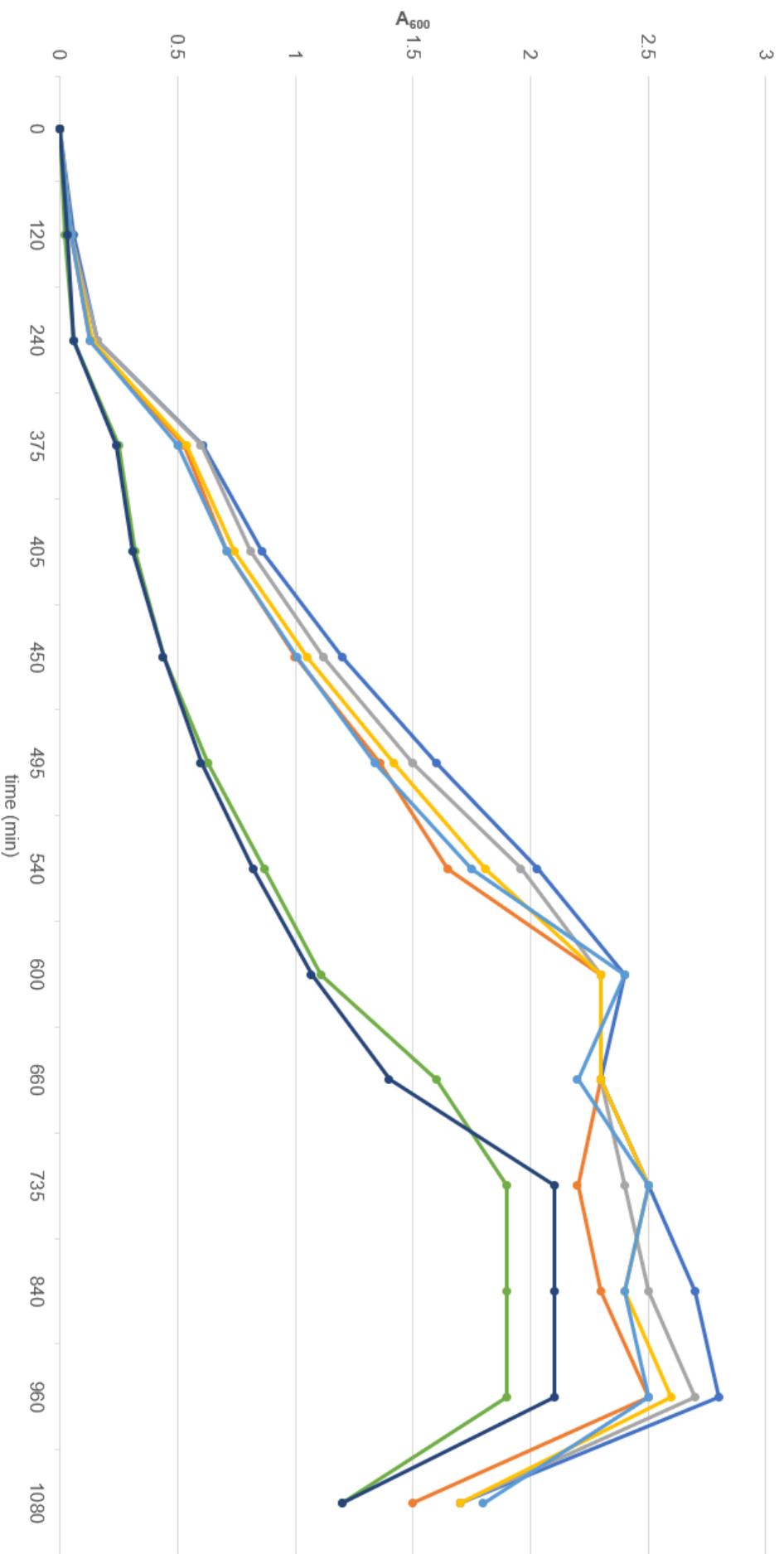
The YihX expression levels can be manipulated by the addition of an inducible plasmid, pCA24N-YihX in this case. An empty pCA24N vector is used as a negative control.

The hypothesis is that when exposed to high concentrations of D-tyrosine, the *dtd* knockout strain will be unable to salvage enough tRNA molecules charged with D-tyrosine, exhibiting little to no growth in D-tyrosine-containing medium but when the YihX enzyme is artificially upregulated in the complementary assay, this toxicity effect of D-tyrosine would be mitigated by the side action of the dephosphorylation performed by YihX.

The medium used for this experiment is M9-glucose minimum medium supplied with 0.4 mg/mL D-tyrosine (2.2 mM), which is very close to the solubility limit of D-tyrosine in water (0.45 mg/mL). The cultures are constantly monitored by measuring the  $A_{600}$ .

Interestingly, all the cultures were able to grow, including the JW3858 although the growth rate in this case is much slower than the rest of the cultures which exhibited the same overall growth rate.

The results undoubtedly disprove the hypothesis, given the fact that there is no benefit nor malus in upregulating or completely knocking out YihX in the cultures (see Figure 15). These results also reinforce the hypothesis that YihX is a non-essential gene for bacterial metabolism. The next section will explore the possible connections between YihX and the sulfoglycolysis operon, the other neighbouring operon in the genome.



**Figure 15** Growth curve of parent strain and single-knockout mutants in M9 minimal medium supplied with D-tyrosine.

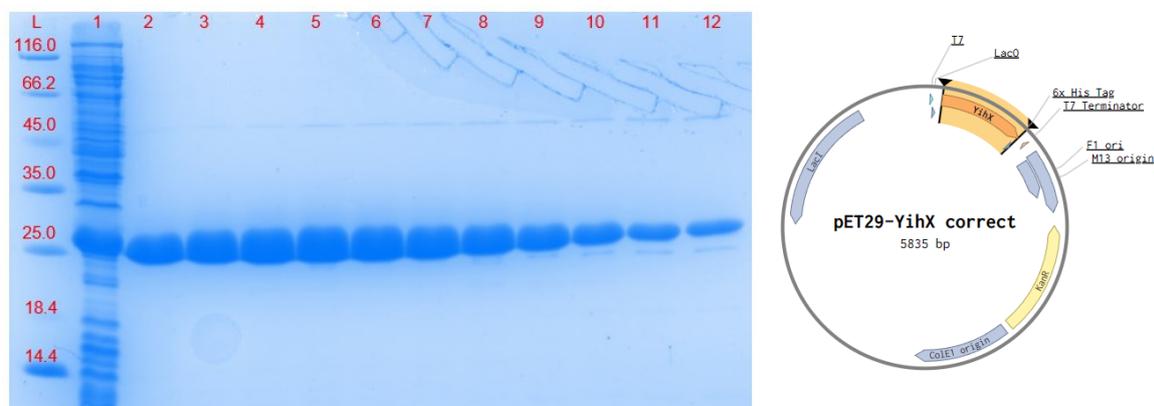
### 3.3 YihX as putative SFP phosphatase

#### 3.3.1 SQ synthesis

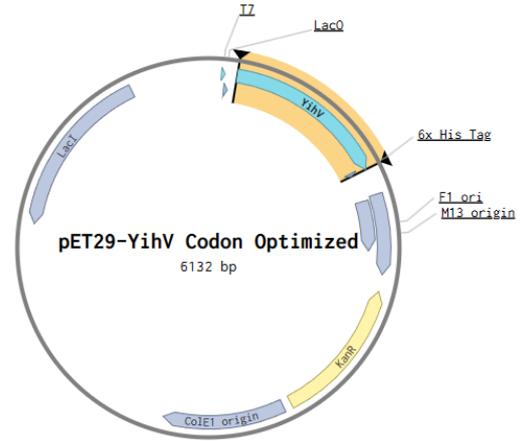
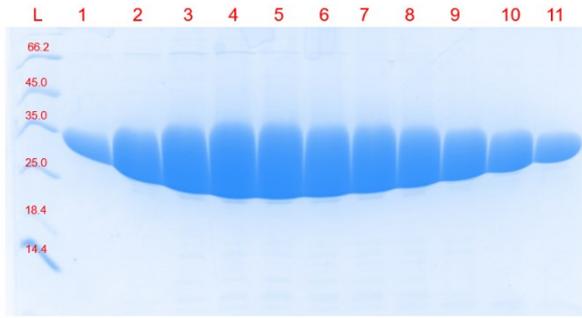
The procedure used was based on the work published by the Williams group<sup>51</sup> (Scheme 4). It consists of a one-pot iodination/peracetylation based on the Appel reaction followed by deprotection by basic hydrolysis. This apparent redundant step of protection/deprotection is very useful since **3** is quite crystalline and can be isolated pure without the need of a chromatographic step. Sodium sulfite is then used as the nucleophile to instal the sulfonic acid moiety on the glucose ring yielding compound **4**. Hydrolysis under strong acid and high temperature then deprotects the anomeric carbon yielding SQ. Recrystallisation of the product from the crude, as described in the literature procedure, was attempted but unsuccessful. For this reason, flash chromatography was performed to obtain pure SQ from the crude mixture (see **2.22.1** for experimental).

#### 3.3.2 YihS, YihV and YihX recombinant proteins

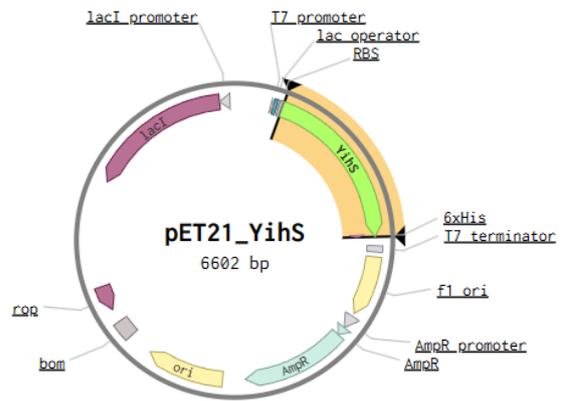
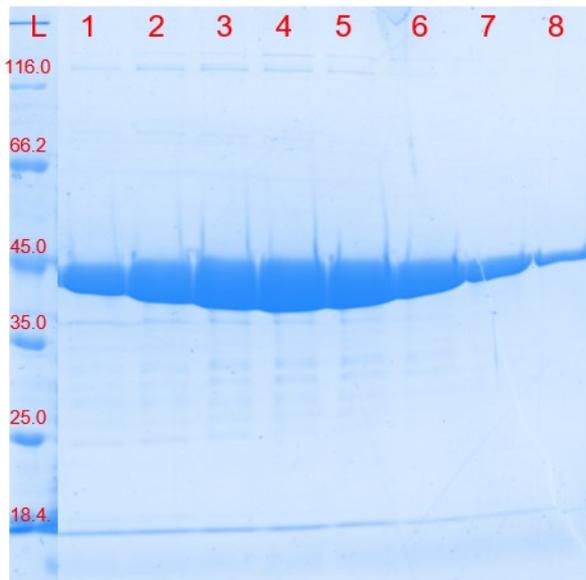
The genes encoding for the enzymes required for the production of SFP and subsequent dephosphorylation were cloned into pET21b or pET29 vectors downstream a T7 promoter sequence as a His<sub>6</sub> fusion protein. Once obtained, the plasmids were used to transform *E. coli* BL21(DE3). After protein production was complete, the protein was purified by Ni-NTA resin IMAC and subsequently by SEC (see section **2.21** for experimental).



**Figure 16 Left)** YihX purification by Ni-NTA chromatography. L: protein ladder. Lane 1: flowthrough. Lanes 2 to 12: elution with imidazole gradient. **Right)** Plasmid map



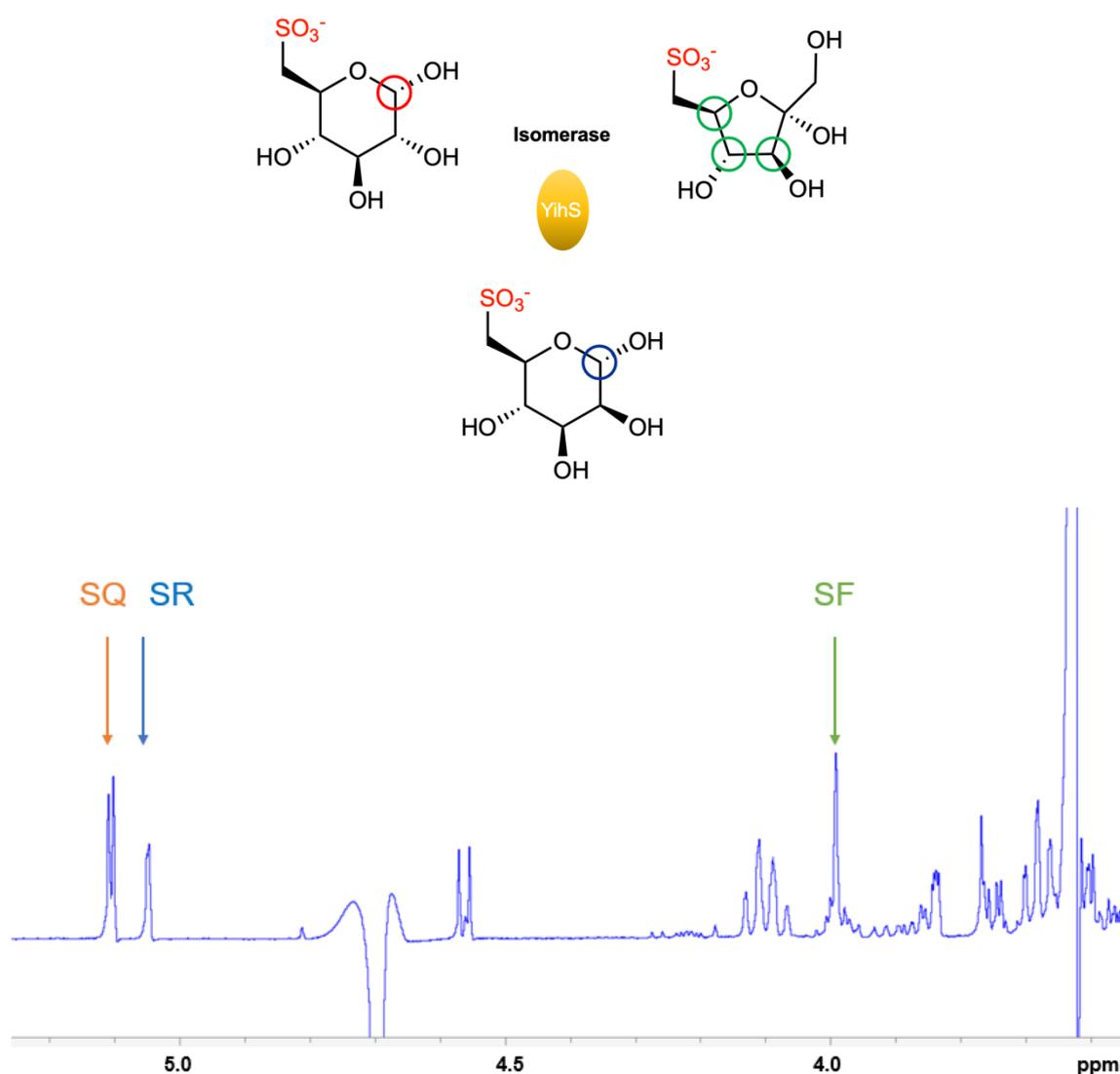
**Figure 17 Left)** YihV purification by Ni-NTA chromatography. L: protein ladder. Lanes 1 to 11: elution with imidazole gradient. **Right)** Plasmid map.



**Figure 18 Left)** YihS purification by Ni-NTA chromatography. L: protein ladder. Lane 1 to 8: elution with imidazole gradient. **Right)** Plasmid map.

### 3.3.3 YihS reaction: isomerisation of SQ to SF

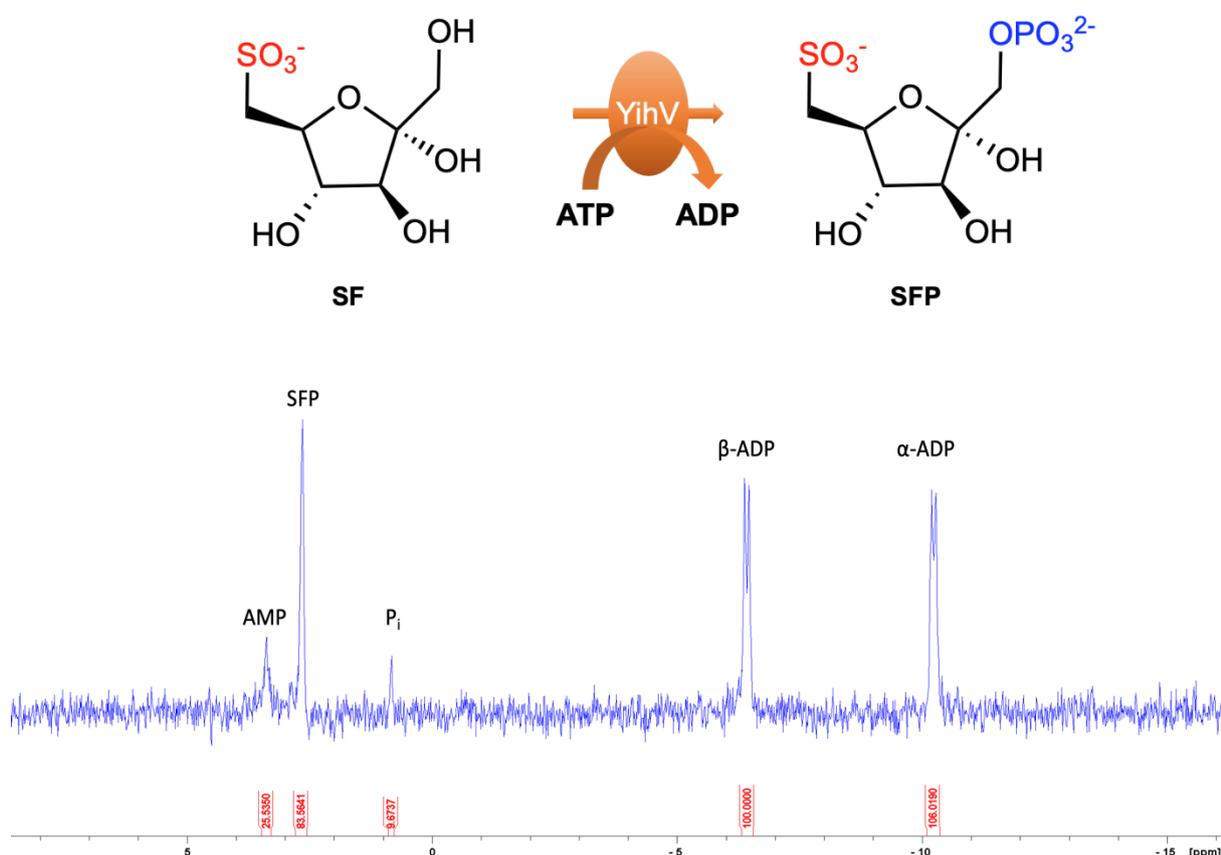
The isomerisation reaction was performed by the addition of isomerase YihS (50  $\mu$ M) to a solution of SQ (50 mM) dissolved in protein buffer (50 Mm Tris-HCl, 200 Mm NaCl, pH 7.0, rt). The reaction reached equilibrium very quickly at room temperature after 10 minutes. When monitoring the progress of the reaction by  $^1\text{H}$  NMR, characteristic peaks for SQ and SF (colour coded) could be easily assigned but another species was present in the reaction mixture. This third compound would then be assigned as sulforhamnose (SR) as shown by the Williams group<sup>45</sup>. The three isomers are present in an equilibrium where the distribution of the species is as follows determined by NMR integrations: SQ  $\approx$  50%, SR  $\approx$  20%, SF  $\approx$  30%.



**Figure 19** YihS isomerisation reaction and  $^1\text{H}$  NMR spectrum of the reaction after reaching equilibrium with the presence of characteristic signals of SQ, SR and SF.

### 3.3.4 YihV reaction: SFP synthesis

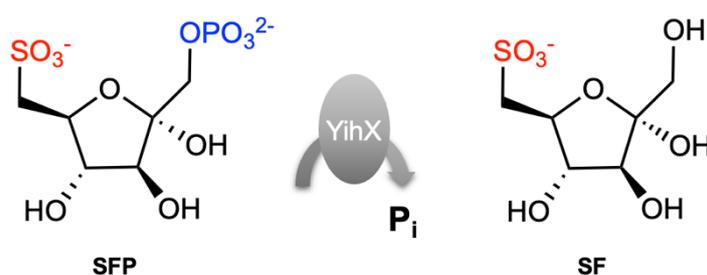
The reaction mixture obtained by the action of YihS is used without purification for the phosphorylation step. After adding ATP and MgCl<sub>2</sub> (25 mM and 50 mM respectively) to the solution, the reaction is initiated by the addition of YihV kinase followed by incubation at 30 °C for 24 h. A <sup>31</sup>P NMR spectrum was acquired to monitor the reaction progress which shows the complete depletion of ATP in solution and the formation of SFP. This system of cascading the reactions allows for the higher conversion rate of SQ and SR into SF because the kinase reaction catalysed by YihV predominantly phosphorylates SF thus funnels more SQ from the equilibrium to the formation of SFP. The formation of SFP is confirmed by LC-MS analysis (see section 7.5.3).



**Figure 20** YihV reaction scheme (top) and <sup>31</sup>P NMR spectra of the mixture after reaction completion.

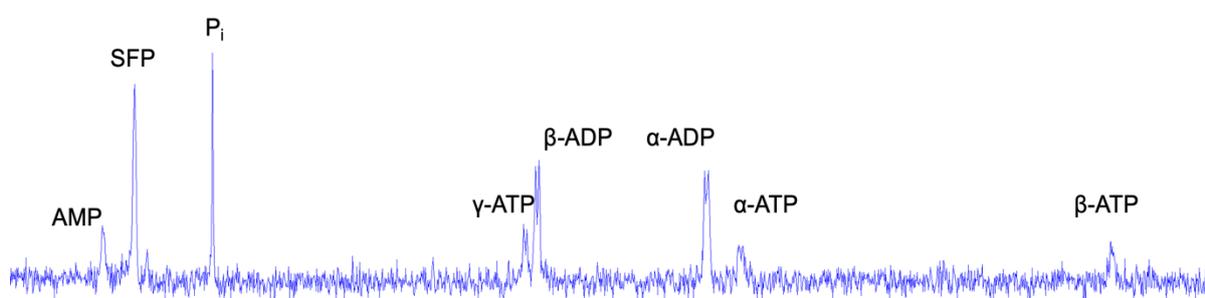
### 3.3.5 YihX reaction: SFP dephosphorylation to SF

The YihV reaction mixture containing SFP is thermally inactivated by heating at 95 °C for 5 minutes, precipitating both YihS and YihV proteins. This will leave only SFP and minor quantities of SQ, SR, and SF in solution. At this point, 50 μM YihX phosphatase is added (the catalytic metal Mg<sup>2+</sup> is already present in the solution) and the reaction is incubated for 24 h at 37 °C (Scheme 5).



**Scheme 5** Expected reaction for YihX dephosphorylation of SFP.

Unexpectedly, the reaction complicates the mixture by forming other species (e.g., ATP, discussed in the next chapter), complicating the accurate assignment of all the resonances the <sup>31</sup>P NMR spectrum. While monitoring the reaction for an extended amount of time (24 h) the SFP peak does not change its integration significantly (less than 5% difference in the integration value after 24 h of incubation time), so this evidence will conclude that SFP is not a substrate of YihX (Figure 21). This result prompted the investigation of this new, never reported before, phosphorylating activity.

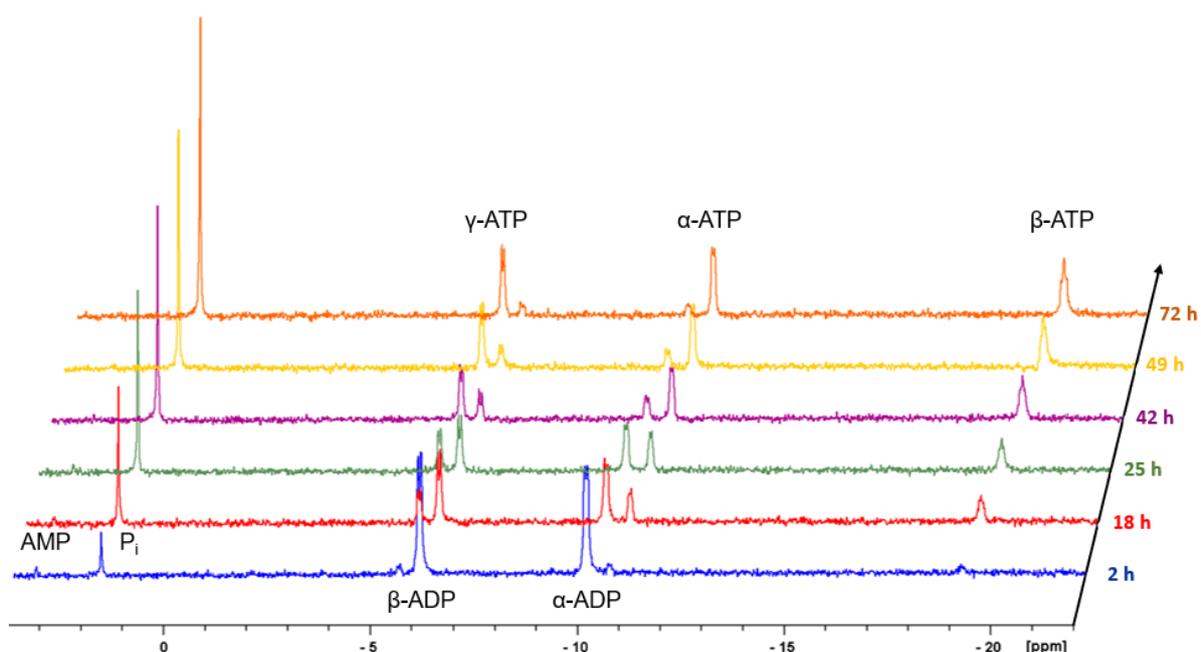


**Figure 21** <sup>31</sup>P NMR spectra of reaction mixture after addition of YihX phosphatase. ATP peaks are clearly produced from the reaction, since there was no unreacted ATP left from the previous reaction.

### 3.4 A new activity for YihX: ADP phosphorylation to ATP

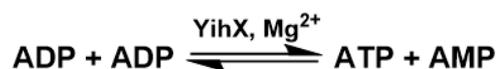
While performing the enzymatic reactions for the production of SFP discussed in the previous section, the appearance of ATP-related peaks in the  $^{31}\text{P}$  NMR was very unexpected and so it was important to dig deeper into this experimental evidence. First of all, it is of primary importance to define which chemical species is acting as the phosphoryl group donor.

This is because the enzymatic reaction mixture has many species present that could cover this role. Investigations to identify the phosphoryl group donor were conducted, in which  $\text{P}_i$ ,  $\alpha\text{G1P}$  and were added in solution with ADP to be tested as donors but no increase in the ATP synthesis reaction rate was observed when compared to just ADP (see Appendix for  $^{31}\text{P}$  NMR spectra). It was finally concluded that ADP was indeed the phosphoryl group donor and acceptor at the same time. Performing the same reaction with a non-hydrolysable ADP analogue such as AMP- $\text{CH}_2\text{-P}$  did not yield ATP production, thus supporting the previous evidence of ADP acting as both donor and acceptor. Performing the reaction with ADP as donor and acceptor of phosphoryl group for an extended period and continuous monitoring of the reaction with  $^{31}\text{P}$  NMR spectroscopy highlights how the ADP is constantly being phosphorylated to ATP and



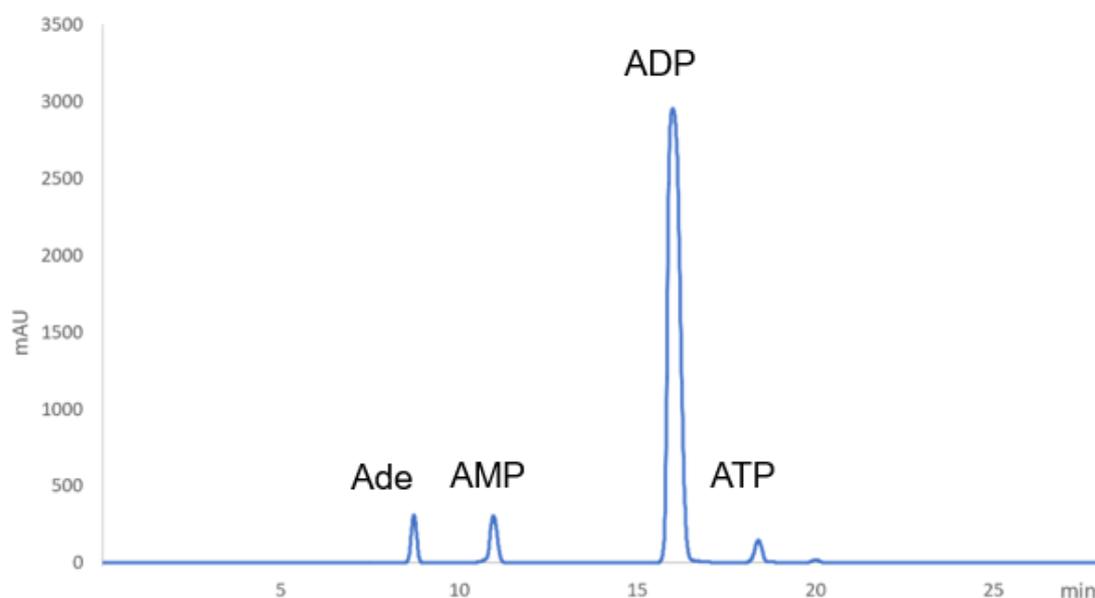
**Figure 22**  $^{31}\text{P}$  NMR time-course experiment highlighting ATP formation over an extended period. The reaction was carried in 50 mM Tris-HCl buffer pH 7.00, 200 mM NaCl, 10 mM ADP, 15 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  YihX, 37  $^\circ\text{C}$ .

driving the reaction almost to completion (Figure 22). This experimental evidence clearly shifted the hypothesis to a reaction described by the following equation:



### 3.4.1 Result validation: HPLC characterisation

To confirm that the NMR data was indeed correct, a cross validation was performed with an HPLC characterisation of the reaction mixture. The trace shows 3 peaks expected for the 3 species formed by the hypothesized reaction (AMP, ADP and ATP) and another peak which denotes the presence of a UV-active specie which is not phosphorylated due to the fact that no other signals are present in the NMR spectrum. The first hypothesis is that the peak consists of Adenosine and indeed, after calibration with a commercially available standard, the hypothesis can be confirmed (Figure 23).



**Figure 23** HPLC trace of ADP phosphorylation reaction mixture. The retention time of all the species is validated with commercially available standards. Run conditions as explained in section 2.13.1.

The presence of adenosine in the sample remarks the fact that this is not a common enzyme; it is a dual activity enzyme, able to perform both phosphorylation (phosphoryl transfer) and dephosphorylation (hydrolysis). On top of this, the enzyme only performs phosphorylation with ADP and only dephosphorylation with AMP (see section 7.1 and

**7.3.5).** This experimental evidence leads to a reformulation of the hypothesised reaction which can now be fully written in the following equation:

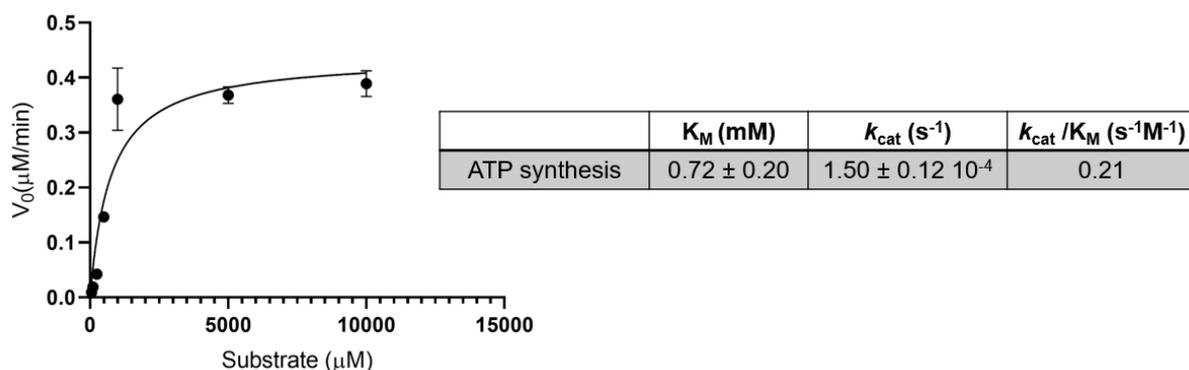


This phosphotransferase activity is quite unique, especially for a hydrolase class enzyme like YihX. Other enzymes are capable of doing the first half of the reaction (ATP and AMP production from ADP) such as Adenylate Kinase. The problem with these enzymes is that they are equilibrium bound; the equilibrium constant is very close to 1. The uniqueness of YihX being able to perform at the same time ADP phosphorylation and AMP hydrolysis means that AMP is continuously formed and consumed, thus constantly subtracting it from the equilibrium and driving the reaction almost to completion. Determining the kinetic parameters of the overall reaction can reveal how relevant this activity can be in a biological context.

### **3.4.2 Kinetic parameters determination for ATP synthesis activity**

To fully characterize this newly discovered activity, an HPLC stop-assay was used to determine the kinetic parameters. Initially, a thermal inactivation step was used to precipitate the enzyme in solution to stop the reaction and make the sample ready for an HPLC injection, but this method showed its limitations due to the fact that heat would promote hydrolysis and is not fast enough to immediately inactivate the enzyme present in solution. For this purpose, a new method of quenching the reaction was employed which consists of immediate precipitation of enzyme with the addition of trichloroacetic acid (TCA). This method was completely reliable and consistent compared to the older method. After removal of precipitated enzyme, the supernatant was neutralised with NaOAc to prime the sample for HPLC injection. This methodology ensured an accurate and precise measurement of  $k_{cat}$ ,  $K_M$ ,  $k_{cat}/K_M$  (Figure 24). To address the question raised in the section above about the relevancy of this ATP synthesis reaction in a biological context, the values suggests that it is not relevant due to the poor catalytic efficiency shown by the low value of  $k_{cat}/K_M$  and millimolar

range of  $K_M$ . This does not exclude the possibility that, given that this enzyme can still provide a little boost to ATP production levels in cases where the energy source is scarce.



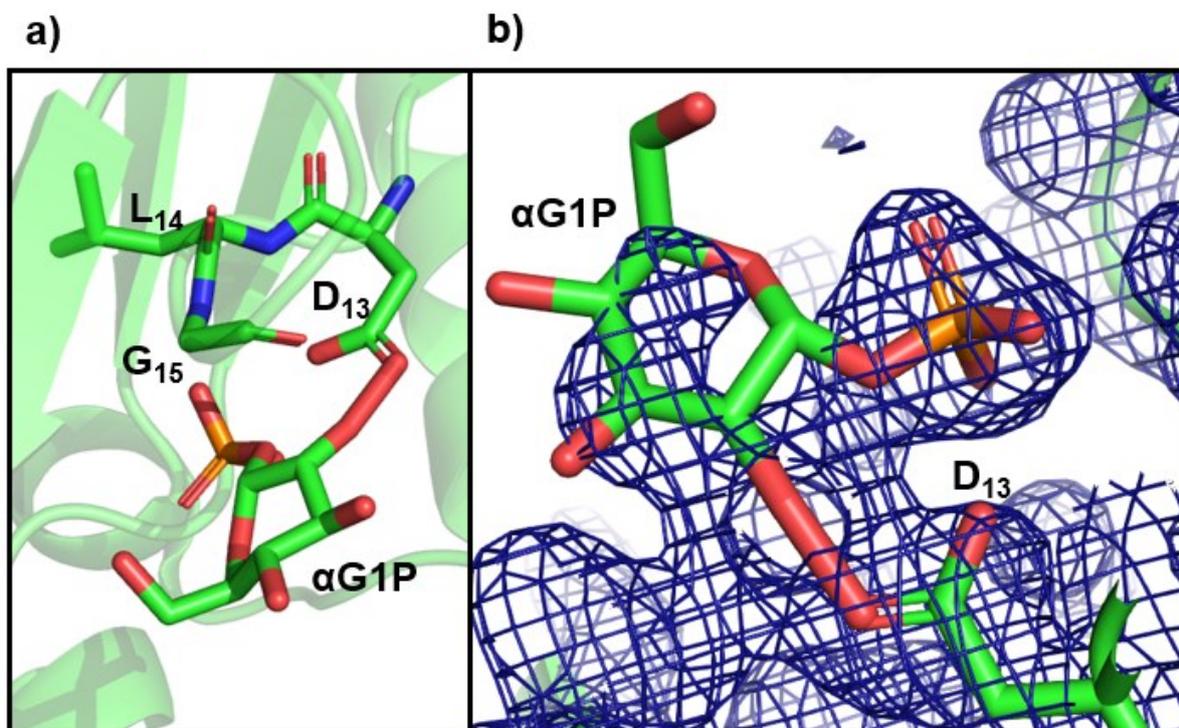
**Figure 24** Michaelis-Menten kinetics for YihX ATP formation reaction. Data obtained by HPLC stop-assay.

### 3.5 Structural and kinetic characterisation of YihX

#### 3.5.1 Mistakes in the PDB 2B0C structure and its value

To explore the specificity of YihX and probe into how the enzyme can catalyse phosphoryl transfer reaction, the crystal structure of YihX was inspected. Currently, only one crystal structure is available on the PDB server, code 2B0C (Figure 25). The structure is resolved at 2.0 Å and it shows the presence of  $\alpha$ -D-glucose-1-phosphate bound in the active site and a  $Mg^{2+}$  ion in the surface of the protein, very distant from the active site either. More importantly, a covalent bond is clearly observed between the carboxylate of Asp13 and the oxygen of the hydroxyl at the C2 position of the carbohydrate. This rings an alarm since the bonding is at least questionable by the chemical properties of the atoms participating in this interaction. The active site also

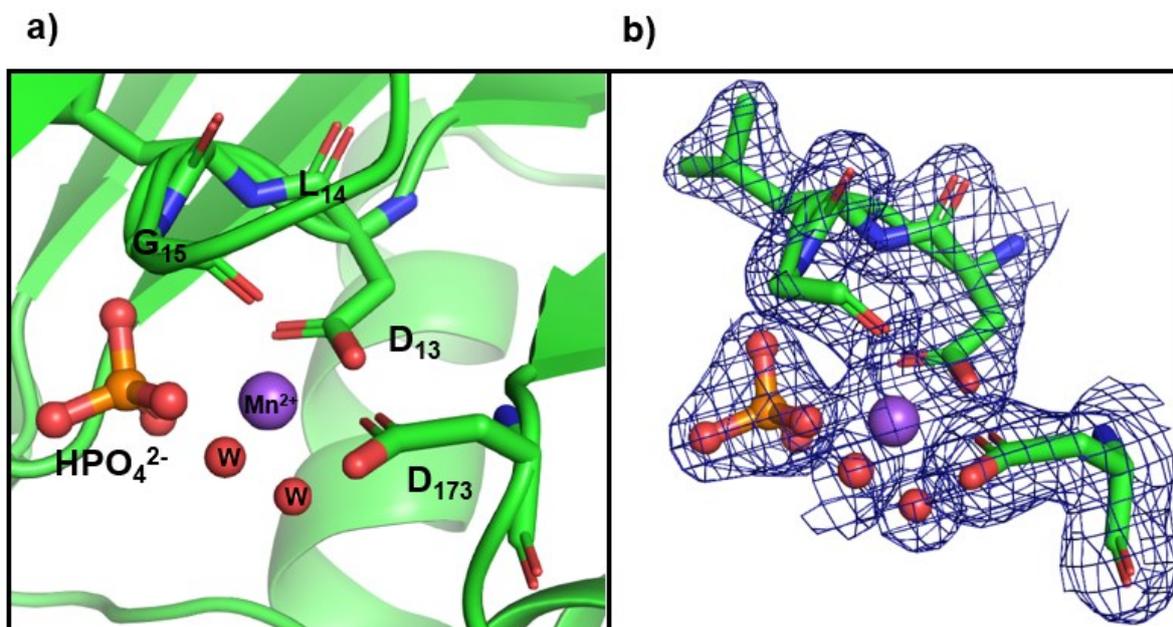
lacks the catalytic metal,  $Mg^{2+}$ , although the standard HAD-like family catalysis mechanism requires the divalent cation to coordinate the phosphate group.



**Figure 25** **a)** The YihX structure (PDB 2B0C). The covalent bond between Asp13 and the C2 hydroxyl group on  $\alpha$ -D-glucose-1-phosphate is clearly shown. **b)** electron density map contoured at  $0.22e^{-1}\text{\AA}^{-3}$ ; the electron density map does not support the fitting of an  $\alpha$ G1P molecule, as shown by the lack of electron density around the bonds.

### 3.5.2 Refinement of YihX structure from PDB 2B0C

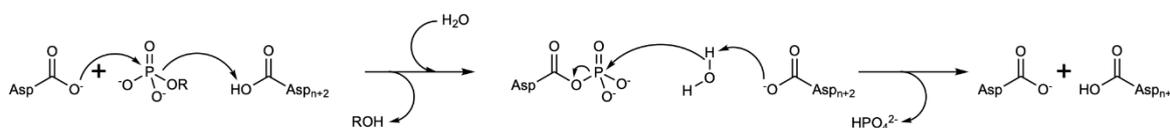
With all the inaccuracies present in the 2B0C structures, a new refinement of the deposited structure factor was performed using cycles of REFMAC after removing the  $\alpha$ G1P as the ligand to generate the omit map. A  $Mn^{2+}$  ion is fitted into the catalytic metal site. This resulted in a much better-defined electron density around the  $Mn^{2+}$ , showing a clear octahedral geometry. This also improved the tetrahedral density coordinated to this  $Mn^{2+}$ , which could be either a phosphate from the soaked  $\alpha$ G1P or sulfate from the crystallisation conditions. After fitting and refining both, phosphate is a better ligand for the density according to the B-factors values because they are more consistent with the B-factors from the rest of the protein backbone (Figure 26).



**Figure 26 a)** The PDB 2B0C structure refined. The  $Mn^{2+}$  ion is now fitted in the structure and a single phosphate group is coordinated by as a ligand. **b)** electron density map contoured at  $0.22e^{-1}\text{\AA}^{-3}$ ; the electron density is much better defined around the  $Mn^{2+}$  ion with clear octahedral coordination and the phosphate ligand.

### 3.5.3 YihX catalytic mechanism: a deviation from canonical HAD phosphatase

As previously mentioned, the mechanism of the HAD superfamily for the phosphatase subclass is strongly dependent on the signature DxD motif I in S1 (Scheme 6). First, the  $D_n$  carboxylate nucleophilic attacks the phosphate of the substrate to generate a phosphoenzyme intermediate. Subsequently, the  $D_{n+2}$  carboxylate activates a water molecule that will nucleophilic attack the phosphoenzyme intermediate to release the enzyme as the leaving group and the hydrolysed substrate. The presence of a DxG motif replacing the DxD motif suggests a different mechanism.



**Scheme 6** HAD phosphatase mechanism of hydrolysis with a DxD motif I

To investigate the latter, a new crystal structure of the enzyme, free of ambiguity in the electron density, is required as the starting point for the structural characterisation of the mechanism.

#### **3.5.4 Crystallisation of YihX Apo and YihX-MgF<sub>3</sub>(H<sub>2</sub>O)<sup>-</sup>-H<sub>2</sub>O transition state analogue complex**

A C-terminal His<sub>6</sub>-tagged fusion protein was expressed in *E. coli* BL21(DE3) and purified with Ni-NTA affinity chromatography and size exclusion chromatography to set up crystallisation trays. The crystal trials were conducted by Dr. Patrick Baumann at the University of Manchester. A series of commercially available crystallisation screens were tested using the sitting-drop method. The protein stock and the precipitant were mixed in a 1:1 ratio for all conditions. For both the apo protein and the TSA complex, a hit was observed with the following precipitant: 55.5 mM MES pH 6.5, 44.5 mM Imidazole, 30 mM MgCl<sub>2</sub>, 30 mM CaCl<sub>2</sub>, 20% glycerol, 10% PEG4000, 20% PEG500. The protein buffer contained 20 mg/mL YihX, 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM MgCl<sub>2</sub>, 30 mM NaF. Protein crystals were sent to Diamond Light Source for X-Ray diffraction and the structure was solved by molecular replacement using PDB 2B0C as a template. However, no clear density of TSA was observed even when the resolution is better than 1.5 Å, thus further soaking of 100 mM NaF had to be carried out to increase the occupancy of the metal fluoride moiety. The final structure was solved at 2.20 Å resolution.

**Table 5** Data collection and refinement statistics for YihX-MgF<sub>3</sub>(H<sub>2</sub>O)·H<sub>2</sub>O.

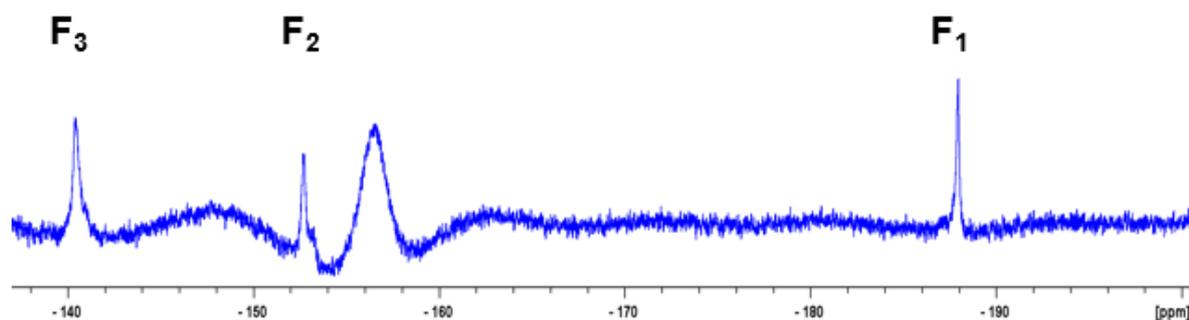
	YihX-MgF <sub>3</sub> (H <sub>2</sub> O)·H <sub>2</sub> O
<b>Data collection</b>	
Wavelength	0.9762
Spacegroup	P 1 21 1
<i>a</i> , <i>b</i> , <i>c</i> (Å)	67.259, 79.325, 80.297
<i>a</i> , <i>b</i> , <i>c</i> (°)	90.000, 108.516, 90.000
Resolution (Å)	59.04-2.20, 2.27-2.20
<i>R</i> <sub>merge</sub>	0.225
<i>I</i> / <i>σI</i>	5.8 / 1.0
CC(1/2)	0.992 / 0.664
Completeness (%)	100.0 / 99.7
<b>Refinement</b>	
Resolution (Å)	59.04-2.20
No. reflections	40788 / 2023
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.214 / 0.257
No. atoms	
Protein	3210
Ligan/ion	64 / 4
Water	218
<i>B</i> -factors	
Protein	41.82
Ligan/ion	51.2 / 33.92
Water	40.82
RMS deviations	
Bond lengths (Å)	0.0075
Bond angles (°)	1.462

**Table 6** Data collection and refinement statistics for apo YihX.

	apo YihX
<b>Data collection</b>	
Wavelength	0.9762
Spacegroup	1 2, 2, 2
<i>a, b, c</i> (Å)	62.216, 79.567, 148.188
<i>a, b, c</i> (°)	90.000, 90.000, 90.000
Resolution (Å)	70.10 / 1.90
<i>R</i> <sub>merge</sub>	0.225
<i>I</i> / $\sigma$ <i>I</i>	12.9 / 1.3
CC(1/2)	0.998 / 0.827
Completeness (%)	100.0 / 99.8
<b>Refinement</b>	
Resolution (Å)	70.20-1.90
No. reflections	31733
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.204 / 0.233
No. atoms	
Protein	1617
Ligan/ion	32 / 6
Water	174
<i>B</i> -factors	
Protein	48.02
Ligan/ion	61.65 / 54.36
Water	50.58
RMS deviations	
Bond lengths (Å)	0.0652
Bond angles (°)	1.275



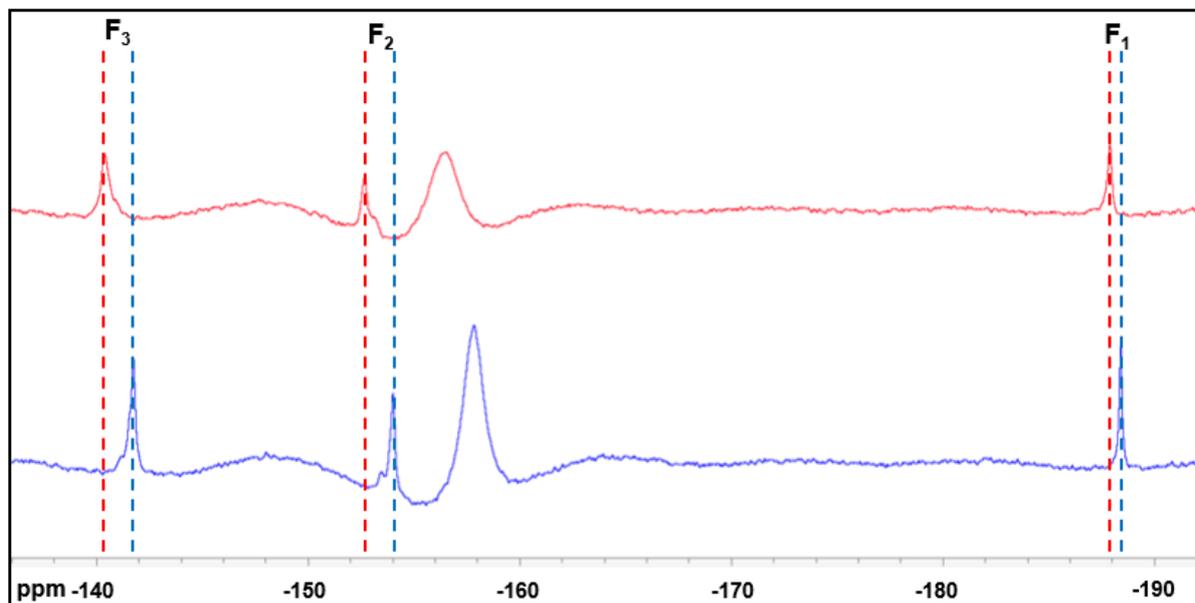
A confirmation of the formation of the TSA complex in solution can be obtained through  $^{19}\text{F}$  NMR spectroscopy. A sample of the protein in the same crystallisation protein buffer is added with 10 mM  $\text{MgCl}_2$  and 30 mM  $\text{NH}_4\text{F}$  and 10%  $\text{D}_2\text{O}$  at room temperature. Performing a  $^{19}\text{F}$  experiment (2048 scans) with presaturation on fluoride's resonance (-119 ppm) retrieves a spectrum that highlights the presence of the complex due to the presence of the three fluorine's composing the TSA (Figure 28).



**Figure 28**  $^{19}\text{F}$  NMR of YihX- $\text{MgF}_3(\text{H}_2\text{O})\text{-H}_2\text{O}$  TSA complex. 0.3 mM YihX, 50 mM Tris-HCl pH 7, 200 mM NaCl, 10 mM  $\text{MgCl}_2$ , 30 mM  $\text{NH}_4\text{Cl}$ . The broad peak corresponds to the species  $[\text{MgF}^+(\text{H}_2\text{O})_5]^{87}$ .

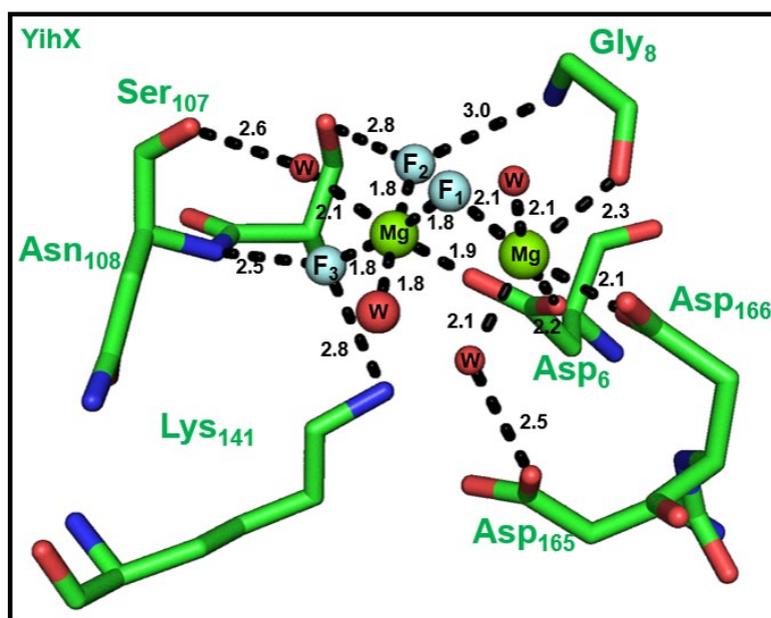
Assignment of the fluorine's signals to the structure can be achieved by performing the same experiment with 90%  $\text{D}_2\text{O}$ . This will allow the measurement of the Solvent Isotope-Induced Shift (SIIS). In principle, the hydrogen bond network that is involved in coordinating the  $\text{MF}_x$  group comprising  $\text{NH}\text{---}\text{F}$  and  $\text{OH}\text{---}\text{F}$  interactions are susceptible to the specific isotope, in this case  $\text{F}\text{---}\text{D}$  and  $\text{F}\text{---}\text{H}$ . This is due to transmission in the electric field when comparing  $^1\text{H}$  to  $^2\text{D}^{14,52}$ . The resulting effect is that fluorine atoms that are coordinated by more and shorter H-bonds will show a bigger value of SIIS (Figure 29). Once the SIIS value is obtained, each fluorine peak can be assigned on the structure considering that the more hydrogen bonds are formed with the fluorine atom, the higher the SIIS effect will be.  $\text{F}_1$  is always assigned to the peak with the highest negative  $\delta$  and corresponds to the fluorine atom that is bridged between the two magnesium ions.  $\text{F}_2$  and  $\text{F}_3$  in this case are harder to assign due to the fact that the SIIS values are very close and both of them are involved in two hydrogen bond interactions. A univoque assignment is not possible in this case, but judging from the difference in chemical shift, angles and bond lengths, it can be

plausible to associate  $F_2$  with the fluorine interacting with Gly<sub>8</sub> and Ser<sub>107</sub> and associating  $F_3$  with the fluorine interacting with Asn<sub>108</sub> and Lys<sub>141</sub> (Figure 30).



	ppm in 10% D <sub>2</sub> O	ppm in 90% D <sub>2</sub> O	$\Delta$ SIIS
F <sub>3</sub>	-140.41	-141.78	1.37
F <sub>2</sub>	-152.70	-154.06	1.36
F <sub>1</sub>	-187.94	-188.45	0.51

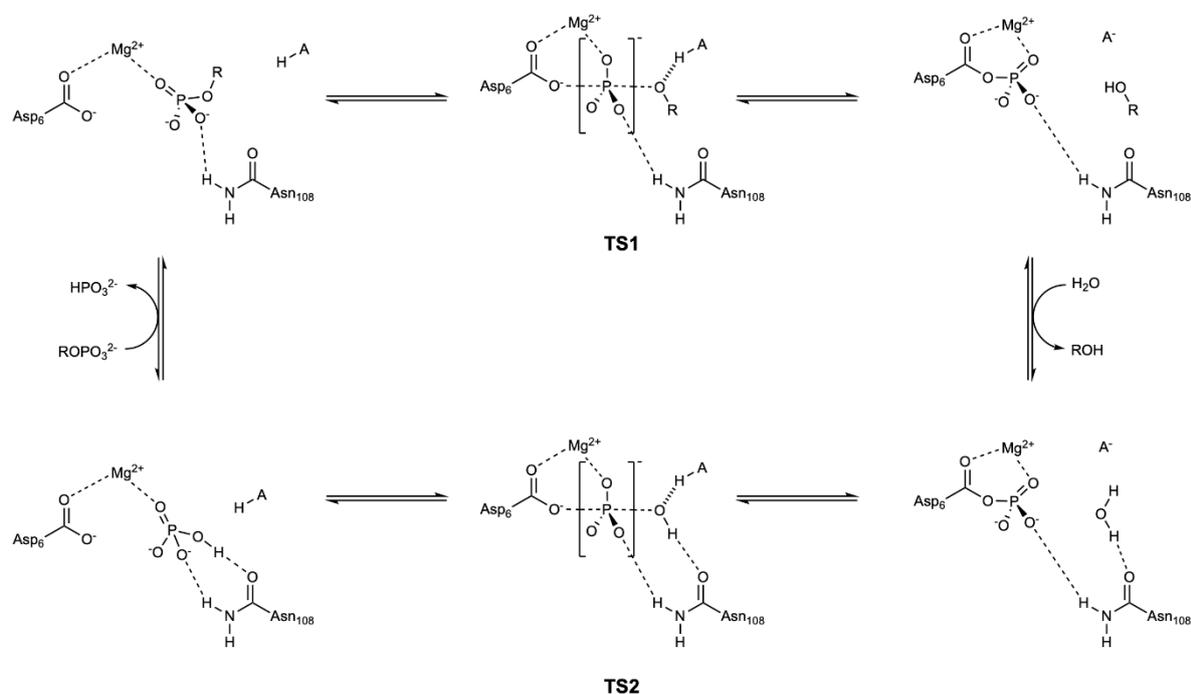
**Figure 29** Overlay of  $^{19}\text{F}$  NMR of YihX-MgF<sub>3</sub>(H<sub>2</sub>O)-H<sub>2</sub>O TSA complex obtained in 10% D<sub>2</sub>O (red), and 90% D<sub>2</sub>O (blue). The SIIS is calculated by subtracting the  $\delta$  value of the signal from the 90% D<sub>2</sub>O spectrum from the  $\delta$  value of the 10% D<sub>2</sub>O spectrum.



**Figure 30** YihX-TSA crystal structure with fluorine atoms assigned through SIIS.

### 3.5.6 A new mechanism for YihX-catalysed hydrolysis: solvent assisted catalysis

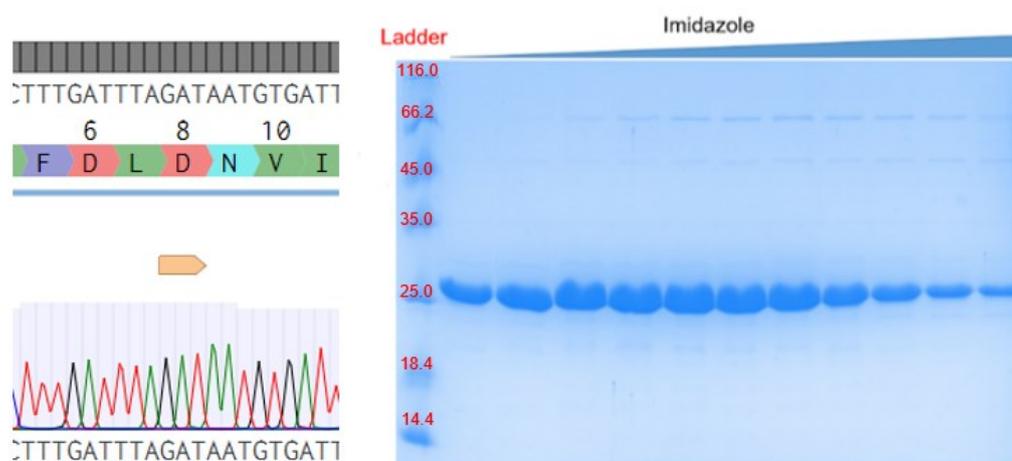
With the new insights obtained by the crystallisation of a TSA complex, new hypothesis can be advanced towards the mechanism used by YihX and in general DxG motif I class of HAD phosphatases. Substrate or solvent-assisted catalysis does not require the presence of the general acid/base (GAGB) residue to deprotonate a water molecule that will ultimately hydrolyse the phosphoenzyme intermediate to restore the enzyme in the free state and liberate inorganic phosphate in solution. In this case, the proton spontaneously transfers from the water molecule to the phosphoryl group, increasing the nucleophilicity of the water molecule that will ultimately hydrolyse the phospho-enzyme intermediate (Scheme 7).



**Scheme 7** Proposed catalytic mechanism for YihX which accounts for the lack of the GAGB residue in position  $\text{D}_{n+2}$  in motif I.

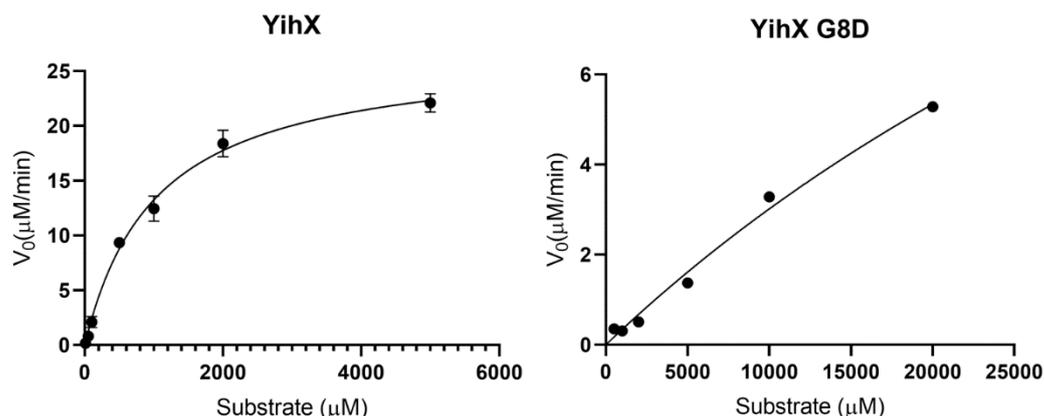
### 3.5.7 DxD knock-in mutant

In order to test whether the standard HAD mechanism is still a viable option, a mutant exposing a DxD canonical HAD motif I was introduced into position 8 using site-directed mutagenesis, expressed in *E. coli* BL21(DE3) cells and purified with standard chromatographic techniques (Figure 31). This variant will elucidate whether the DxD motif enhances the reaction rate or is detrimental to the activity since it follows a completely different mechanism compared to the canonical HAD mechanism.



**Figure 31** DNA sequencing for SDM of YihX<sub>G8D</sub> and chromatography purification.

With the newly obtained YihX<sub>G8D</sub> variant, activity towards  $\alpha$ G1P was assayed with a Malachite Green Phosphate Assay to obtain Michaelis-Menten kinetic parameters while the curve never reached plateau even at very high concentrations of substrate. Therefore, in this case, the  $K_M$  value is underestimated, and the real value will lie above 67.33 mM. (Figure 32). When comparing the kinetics values to YihX<sub>WT</sub>, it is very clear that the introduction of the DxD motif, compared to the native DxG, is detrimental to the activity of the enzyme. A big increase in  $K_M$  values denotes how the enzyme ability to bind the substrate is substantially hindered by the presence of the D<sub>n+2</sub> residue thus supporting the newly proposed mechanism and definitively rejecting the possibility of a standard HAD superfamily mechanism of reaction for YihX.

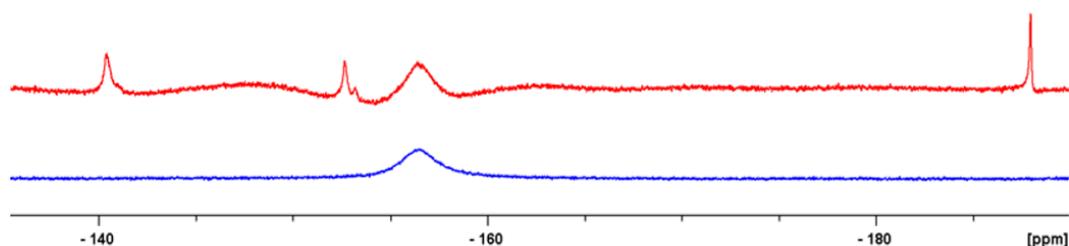


	$K_M$ ( $\mu\text{M}$ )	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$k_{\text{cat}}/K_M$ ( $\text{s}^{-1}\text{M}^{-1}$ )
YihX	$1034 \pm 96$	$4.49 \pm 0.15$	$4.34 \cdot 10^3$
YihX <sub>G8D</sub>	67334	3.88	$5.77 \cdot 10$

**Figure 32** Michaelis-Menten kinetic characterisation of  $\alpha\text{G1P}$  hydrolysis by YihX<sub>G8D</sub> and comparison with YihX<sub>WT</sub>.

### 3.5.8 YihX<sub>G8D</sub> TSA complex

To support the previous statement, a TSA complex formation was attempted with YihX<sub>G8D</sub> with the addition of  $\text{MgCl}_2$  and  $\text{NH}_4\text{F}$  in the protein solution to form the magnesium fluoride TSA complex to mimic the phosphoryl transfer transition state of the  $\alpha\text{G1P}$  hydrolysis reaction.  $^{19}\text{F}$  NMR spectroscopy was employed to monitor the formation of the complex but even after extended incubation time, the complex did not form (Figure 33).

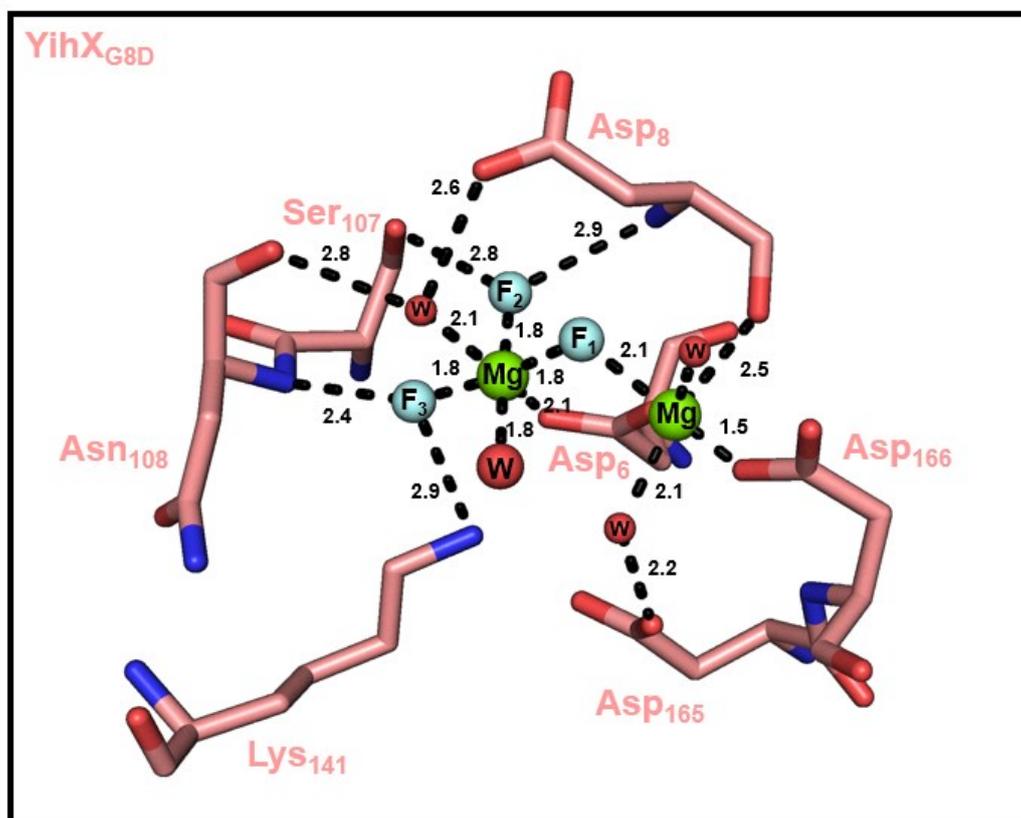
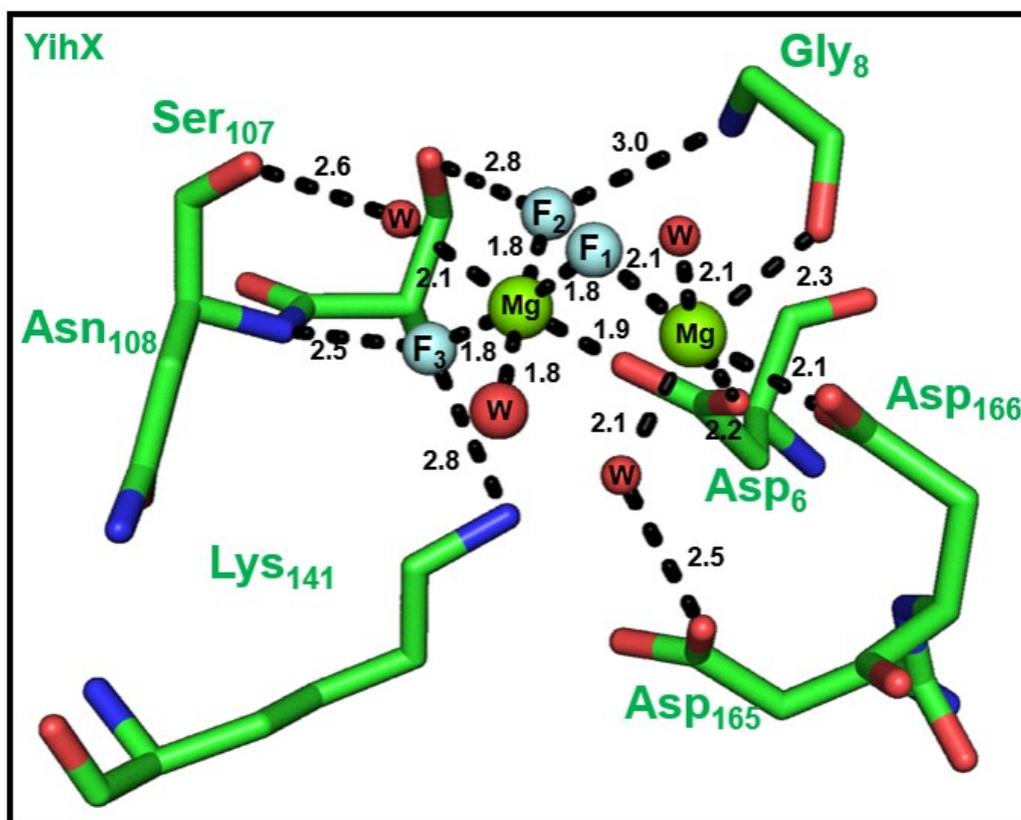


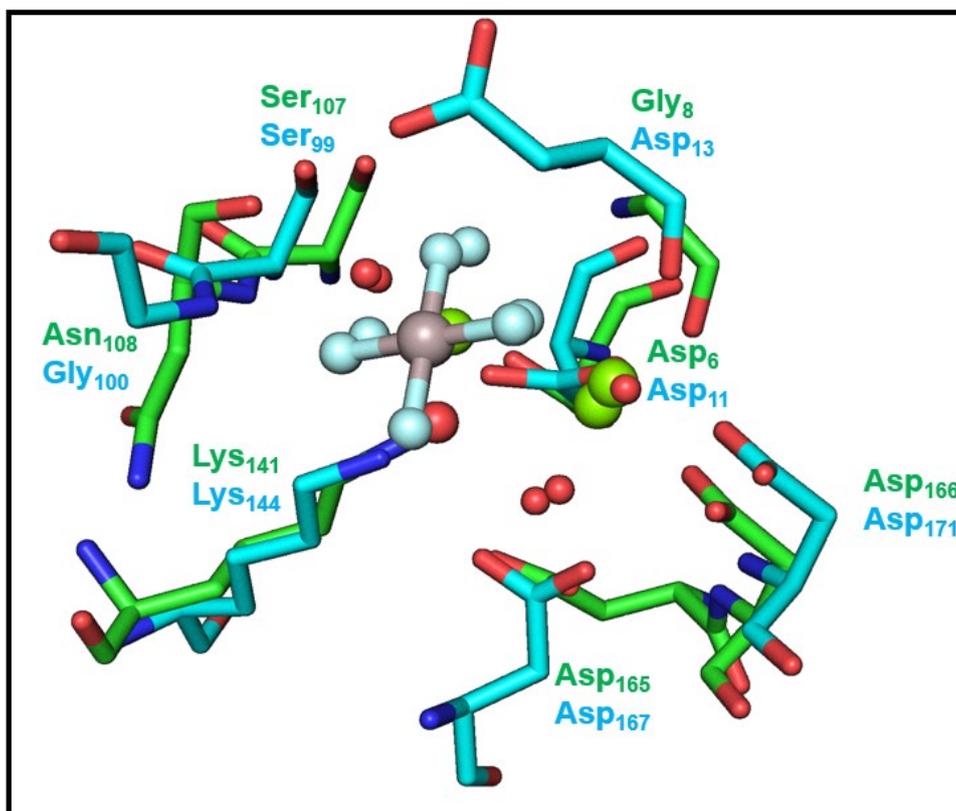
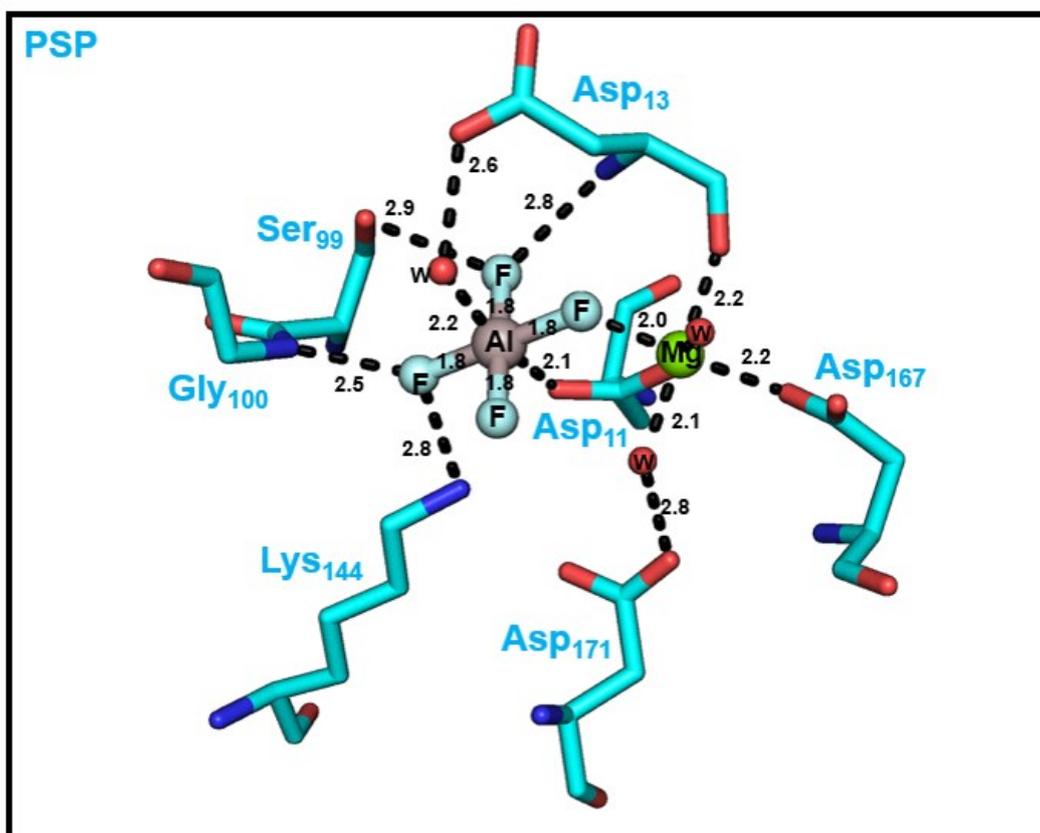
**Figure 33**  $^{19}\text{F}$  NMR spectrum of YihX<sub>WT</sub> TSA complex (red trace), and YihX<sub>G8D</sub> TSA complex (blue trace). The lack of peaks for the individual fluorine atoms in the YihX<sub>G8D</sub> trace clearly shows no formation of TSA complex.

The reason why the TSA complex did not form when using YihX<sub>G8D</sub> could be speculated when aligning the TSA complex structure of YihX<sub>WT</sub> with the crystal structure of YihX<sub>G8D</sub>. However, the crystallisation of this latter was attempted but could not be achieved so the next step was to align the previously obtained with the AlphaFold predicted structure of the mutant YihX<sub>G8D</sub> (Figure **34**).

While investigating the YihX<sub>G8D</sub> predicted structure, no easy conclusion can be drawn out since the alignment with the MgF<sub>3</sub>(H<sub>2</sub>O)<sup>-</sup>-H<sub>2</sub>O does not highlight any obvious structural clashing. There is one difference that is directly related to the presence of the G8D mutation: a new hydrogen bond can be formed between the Asp<sub>8</sub> residue and the axial water that caps the octahedron in the MgF<sub>3</sub>(H<sub>2</sub>O)<sup>-</sup>-H<sub>2</sub>O. This interaction is not present in the WT enzyme, which instead exposes the same side of the active site to the solvent. When comparing YihX with a DxD-containing HAD phosphatase, such as phosphoserine phosphatase (PSP) that is crystallised as a TSA complex with AlF<sub>4</sub><sup>-</sup> (PDB 1L7N) in the same octahedral geometry, this difference can lead to a possible hypothesis. When the axial water molecule is coordinated by a single residue, backbone carbonyl of Asn<sub>108</sub> in YihX or carboxylate of Asp<sub>13</sub> in PSP, it can be held in place at the correct distance so that the complex can be formed. In the case of the YihX<sub>G8D</sub> mutant, the possibility of the nucleophilic water molecule interacting with two hydrogen bond acceptors, backbone carbonyl of Asn<sub>108</sub> or carboxylate of Asp<sub>8</sub>, might

misalign the orbitals for the nucleophilic attack in the transition state and for the correct TSA complex formation.

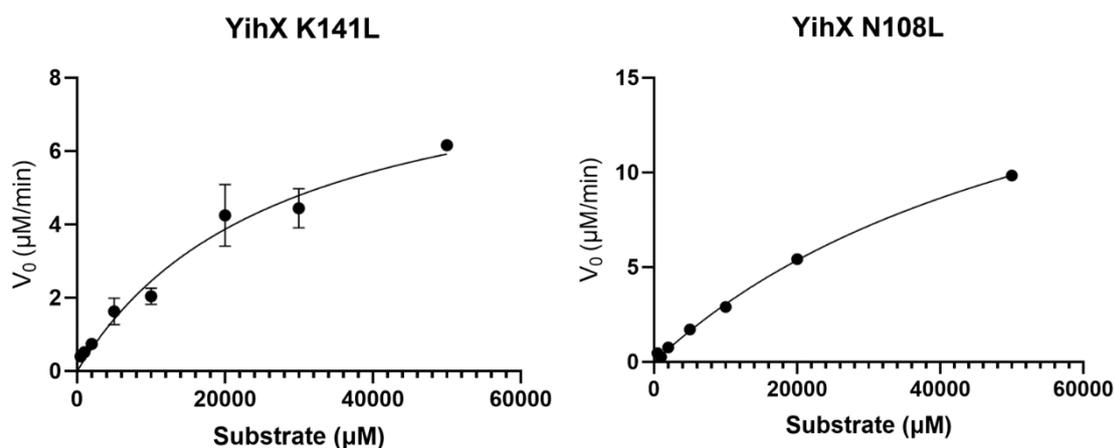




**Figure 34** Crystal structure of YihX<sub>WT</sub> TSA complex (green ribbon), AlfaFold predicted structure of YihX<sub>G8D</sub> aligned with MgF<sub>3</sub>(H<sub>2</sub>O)<sup>-</sup> TSA (pink ribbon), crystal structure of Phosphoserine phosphatase aluminium fluoride TSA (PDB: 1L7N) (teal ribbon). Superimposition of YihX<sub>WT</sub> TSA complex and phosphoserine phosphatase aluminium fluoride TSA.

### 3.5.9 Site-directed mutagenesis of catalytic site residues

Several residues in the YihX active site were mutated in order to confirm their involvement in the binding of the substrate and catalysis. For this reason, YihX<sub>N108L</sub>, YihX<sub>K141L</sub> and YihX<sub>D6N</sub> were cloned, expressed, and purified (see section 2.21). In the case of the D6N mutation no activity is expected since the residue that performs the nucleophilic attack on the phosphorylated substrate is completely missing. Indeed, after incubating  $\alpha$ G1P with YihX<sub>D6N</sub> almost no reaction was observed after an extended period of incubation. For YihX<sub>N108L</sub> and YihX<sub>K141L</sub> it was possible to perform a Malachite Green Phosphate Assay to determine kinetic parameters, although the Michaelis-Menten curve did not reach plateau even at high substrate concentrations in both cases. The  $K_M$  values are much higher for the mutants compared to the WT, confirming the hypothesis that hydrogen bond donor residues are highly important for substrate binding, stabilising the negative charge of the phosphate group in the active site.



	$K_M$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_M$ (s <sup>-1</sup> M <sup>-1</sup> )
YihX	1.03 ± 0.1	4.49 ± 0.15	4.34 · 10 <sup>3</sup>
YihX <sub>K141L</sub>	27.34	1.53	5.58 · 10
YihX <sub>N108L</sub>	62.52	3.69	5.91 · 10
YihX <sub>D6N</sub>	ND	ND	ND

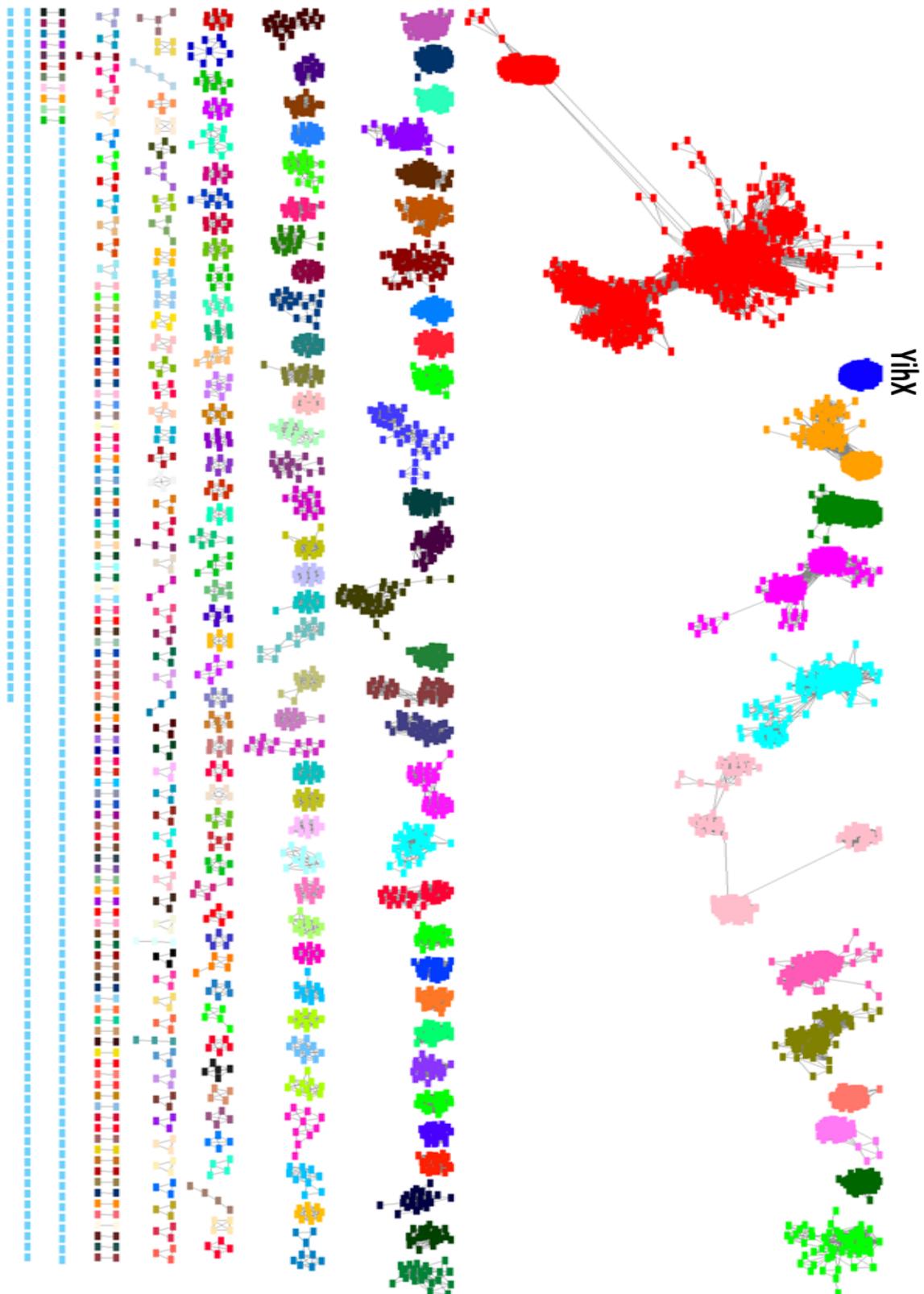
**Figure 35** Michaelis-Menten kinetic curves for YihX mutants. Note that parameter values are not correct due to not reaching plateau.

### 3.5.10 The PG\_0725 gene: another DxG motif HAD phosphatase

A very common bioinformatic tool to perform genome mining and identify possible relationships between protein sequences is the the **Sequence Similarity Network (SSN)** (University of Illinois). This tool, in combination with another software, Cytoscape, utilize an input sequence (in this case the YihX sequence) to identify all the genes that share sequence similarities (which can be modulated by applying cutoff values to exclude sequences with lower scores). The use of this tool is established in bioinformatics to find and assign genes that are not fully annotated in the genomes of microorganisms to a specific activity or pathway. For example, it has been used to annotate enzymes present in the glycol radical enzyme superfamily (IPR004184) of human gut microbiome. This greatly improves the ability to screen enzymes that are not fully characterized in a very fast and efficient way. The key feature that was identified for YihX was the DxG motif I sequence, deviating from canonical HAD hydrolase enzymes. With this tool, it was possible to create an SSN in collaboration with Dr. Remi Zallot (Manchester Metropolitan University) that identified 10000 sequences which shared similarities with YihX at the given cutoff.

Among a list of over 10000 sequences from the PF13419 protein family, the SSN approach found similar enzymes that share the key DxG motif in the same cluster that is different from the canonical HAD-like enzymes (Figure **36**). Among these having the same DxG motif, 17 HAD enzymes have been tested and identified with high throughput screening by Allen et al<sup>47</sup> (present in both the SSN and the activity tables) (Figure **37**). From these 17 sequences, only 3 enzymes have a crystal structure deposited in the PDB (UniProt IDs: Q8A4Q5, Q8A947 and Q7MWA6) and the signature motifs (I, II, III and IV) were identified in all sequences (Figure **38**).

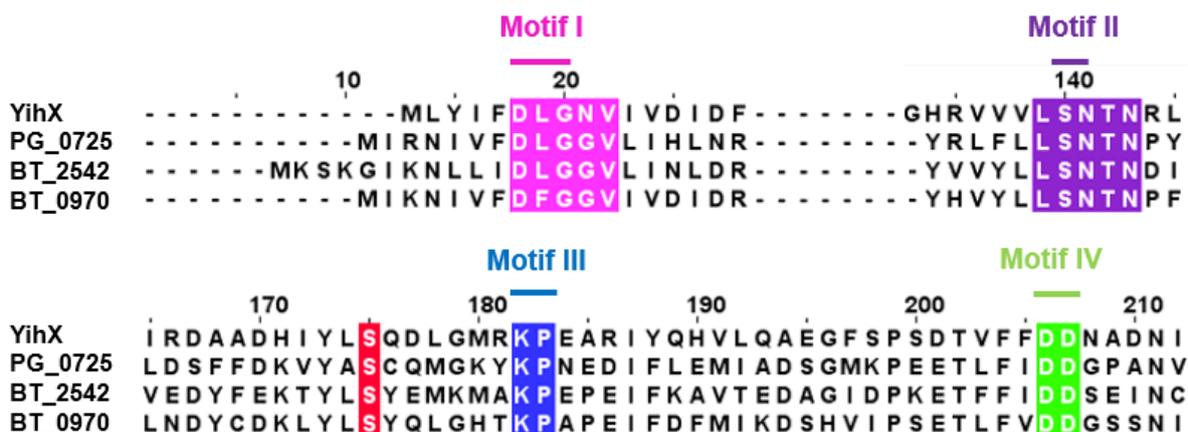
The protein encoded by the Uniprot ID Q7MWA6 (gene PG\_0725) has not been characterised biochemically before, thus was chosen as a test enzyme to investigate the possibility of other DxG motif I HAD phosphatases that can catalyse both ADP phosphorylation to ATP and dephosphorylation through the newly proposed mechanism. The sequence similarity is 33% with YihX but all the key residues (D6, L7, G8, N108, K141, D172, D173 in YihX sequence numbering) are completely conserved in the PG\_0725 enzyme (PDB 2I6X).



**Figure 36** The Sequence Similarity network for YihX. The localised cluster with no connections to neighbouring clusters denotes that this subclass of phosphatases does not share common structure and activity relationship with other enzyme classes.

UniProt ID	UniProt Submitted Name	Gene	Organism	Crystal Structure	K. Allen Activity
A5E9S3	Putative hydrolase (Epoxide hydrolase domain -like phosphatase)	<b>BbtA_0649</b>	<i>Pseudothermotoga lettingae</i> (strain ATCC BAA -301 / DSM 14385 / NBRC 107922 / TMO) ( <i>Thermotoga lettingae</i> )	NO	NO
ABF8V2	HAD-superfamily hydrolase, subfamily IA, variant 3	<b>Tlet_2032</b>	<i>Pseudothermotoga lettingae</i> (strain ATCC BAA -301 / DSM 14385 / NBRC 107922 / TMO) ( <i>Thermotoga lettingae</i> )	NO	NO
CGXTU6	HAD-superfamily hydrolase, subfamily IA, variant 3	<b>Phep_1518</b>	<i>Pedobacter heparinus</i> (strain ATCC 13125 / DSM 2366 / CIP 104194 / JCM 7457 / NBRC 12017 / NCIMB 9290 / NRRL B -14731 / HIM 762 -3)	NO	NO
POA8Y3	Alpha-D-glucose 1-phosphate phosphatase YihX	<b>yihX</b>	<i>Escherichia coli</i> (strain K12)	2B0C	501054
Q133V2	Epoxide hydrolase-like phosphatase	<b>RPD_3413</b>	<i>Rhodopseudomonas palustris</i> (strain BisB5)	NO	NO
Q21J32	HAD-superfamily hydrolase subfamily IA, variant 3	<b>Sde_2037</b>	<i>Saccharophagus degradans</i> (strain 2 -40 / ATCC 43961 / DSM 17024)	NO	NO
Q47M05	HAD-superfamily hydrolase subfamily IA, variant 3	<b>Tfu_2484</b>	<i>Thermobifida fusca</i> (strain YX)	NO	NO
Q5FK60	Putative HAD superfamily hydrolase	<b>LBA_1068</b>	<i>Lactobacillus acidophilus</i> (strain ATCC 700396 / NCK56 / N2 / NCFM)	NO	NO
Q6N3Z3	Possible hydrolase	<b>RPA_3549</b>	<i>Rhodopseudomonas palustris</i> (strain ATCC BAA -98 / CGA009)	NO	900082
Q6N5S2	Haloacid dehalogenase-like hydrolase	<b>RPA_2898</b>	<i>Porphyromonas gingivalis</i> (strain ATCC BAA -308 / W83)	NO	NO
Q7MWA6	Hydrolase, haloacid dehalogenase-like family	<b>PG_0725</b>	<i>Porphyromonas gingivalis</i> (strain ATCC BAA -308 / W83)	2I6X	NO
Q82NCO	Putative hydrolase	<b>SAVERM_1383</b>	<i>Streptomyces avermitilis</i> (strain ATCC 31267 / DSM 46492 / JCM 5070 / NBRC 14893 / NCIMB 12804 / NRRL 8165 / MA -4680)	NO	NO
Q8A4Q5	Putative haloacid dehalogenase-like hydrolase	<b>BT_2542</b>	<i>Bacteroides thetaiotaomicron</i> (strain ATCC 29148 / DSM 2079 / JCM 5827 / CCUG 10774 / NCTC 10582 / VPI -5482 / E50)	4DFD	501088
Q8A947	D-ribitol-5-phosphate phosphatase	<b>BT_0970</b>	<i>Bacteroides thetaiotaomicron</i> (strain ATCC 29148 / DSM 2079 / JCM 5827 / CCUG 10774 / NCTC 10582 / VPI -5482 / E50)	4I83	501083
Q8XG41	Glucose-1-phosphatase YihX	<b>yihX</b>	<i>Salmonella typhi</i>	NO	501052
Q97D99	Glucose-1-phosphatase	<b>CA_C3581</b>	<i>Clostridium acetobutylicum</i> (strain ATCC 824 / DSM 792 / JCM 1419 / LMG 5710 / VKM B -1787)	NO	900152
Q9KR15	Uncharacterized protein	<b>VC_1645</b>	<i>Vibrio cholerae serotype O1</i> (strain ATCC 39315 / El Tor Inaba N16961)	NO	NO

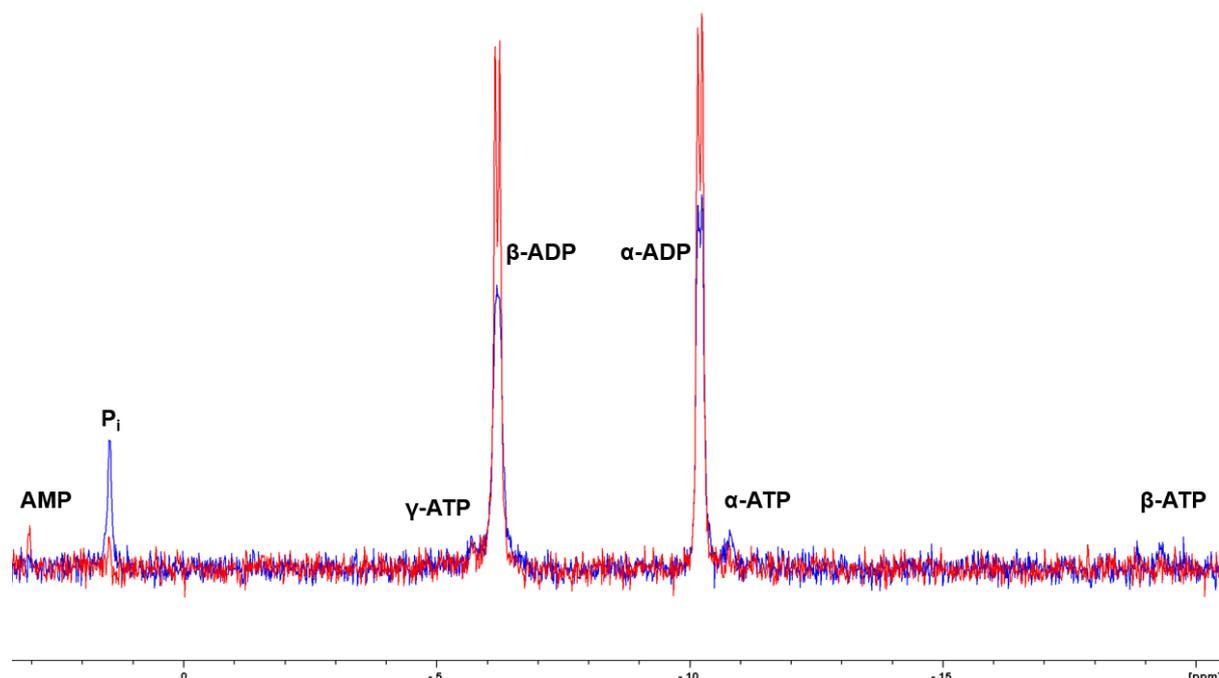
**Figure 37** List of DxG enzymes that are present in both the SSN and a previous high-throughput screen conducted by K. Allen et al.<sup>47</sup> The table lists the UniProt ID, the gene identifier, the microorganism, the availability of a crystal structure in the PDB and the availability of an activity in the high-throughput screen.



**Figure 38** Sequence alignment of the genes identified by the SSN. The signature motifs are highlighted in the sequences.

### 3.5.11 Testing PG\_0725 for ADP phosphorylation

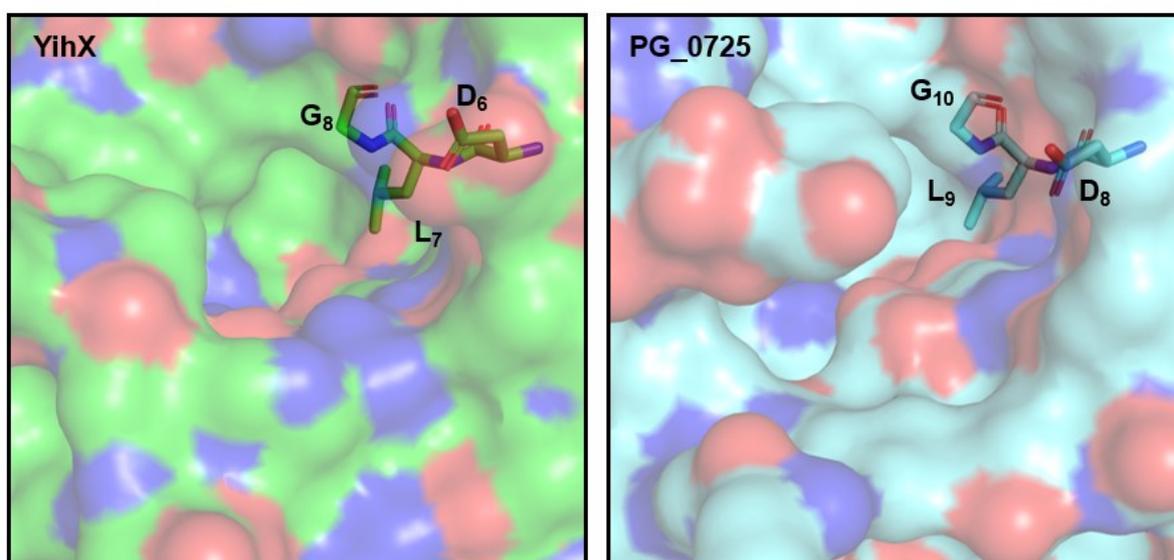
The PG\_0725 gene was cloned into a pET29 vector, expressed and purified with standard chromatographic techniques as described in section 2.21.8. The purified protein was used to perform the ADP phosphorylation reaction. The same operating conditions were used as for YihX catalysed phosphorylation. After the reaction mixture was prepared, the enzyme was added to initiate the reaction and incubated for 24h at



**Figure 39**  $^{31}\text{P}$  NMR spectrum of ADP phosphorylation reaction catalysed by PG\_0725. The spectrum shows the timeframe  $t = 0$  (red trace) and  $t = 24\text{h}$  (blue trace). Both samples contain 10%  $\text{D}_2\text{O}$ .

room temperature.  $^{31}\text{P}$  NMR spectroscopy was used to monitor the reaction progress. When PG\_0725 was mixed with ADP, the phosphorylation reaction was completely negligible (Figure 39).

This experiment highlights with a high degree of confidence that YihX is more of a unique case as per dual activity enzyme in the class of DxG HAD phosphatases. The reason might lie in the accessibility of the active site pocket of PG\_0725 compared to YihX. As shown in Figure 40, the PG\_0725 (PDB 2I6X) has a loop region (residues 39-43) that restricts the access to the pocket where the active site is located.



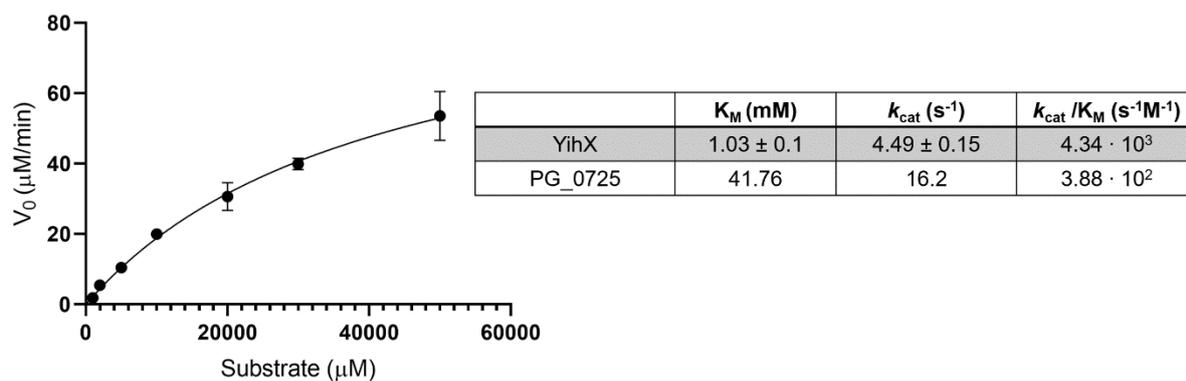
**Figure 40** Comparison of the electrostatic surface of YihX (left) and PG\_0725 (right). The top view highlights the loop region that restricts the pocket in PG\_0725 while it is completely open and available in YihX.

### 3.5.12 $\alpha\text{G1P}$ dephosphorylation catalysed by PG\_0725

After testing ADP phosphorylation, it is of great value to understand whether this DxG motif HAD phosphatase, selected with the EFI bioinformatic tool, is capable of performing the dephosphorylation reaction of  $\alpha\text{G1P}$ , as YihX, and if the newly proposed mechanism (see scheme 7) can be applied to this enzyme which presents the same DxG motif I pattern.

$\alpha\text{G1P}$  dephosphorylation was assayed with the Malachite Green Phosphate Assay (see section 2.14). Reaction conditions and timings for the end-point assay are kept consistent with the ones used for the same reaction catalysed by YihX. As shown in

Figure 41, the Michaelis-Menten graph and parameters for PG\_0725 show higher  $K_M$  and  $k_{cat}$  values and lower  $k_{cat}/K_M$  value compared to YihX, thus suggesting that the latter has higher affinity for the substrate and exhibits an overall higher catalytic efficiency for  $\alpha$ G1P dephosphorylation. The lower value of  $k_{cat}/K_M$ , compared to YihX, is a clear hint that  $\alpha$ G1P might not be a suitable substrate for PG\_0725.

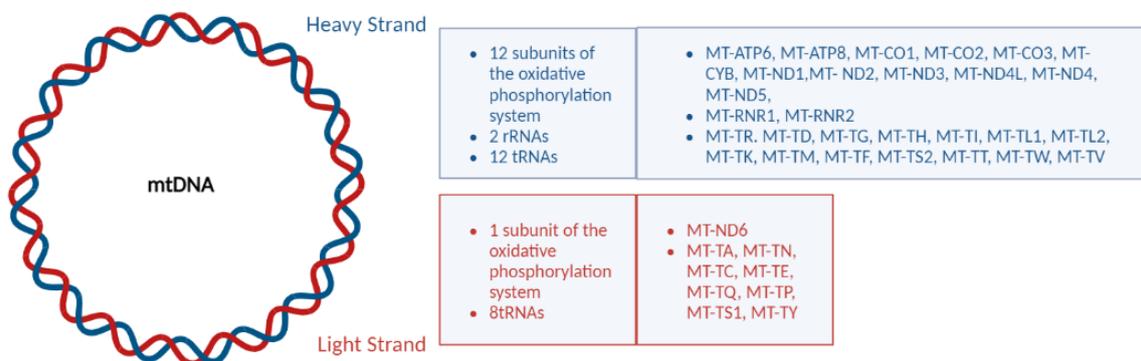


**Figure 41** Michaelis-Menten kinetics for  $\alpha$ G1P dephosphorylation catalysed by PG\_0725. Data obtained by Malachite Green Phosphate Assay.

## 4 Introduction to Thymidine Kinase 2

### 4.1 Mitochondrial DNA

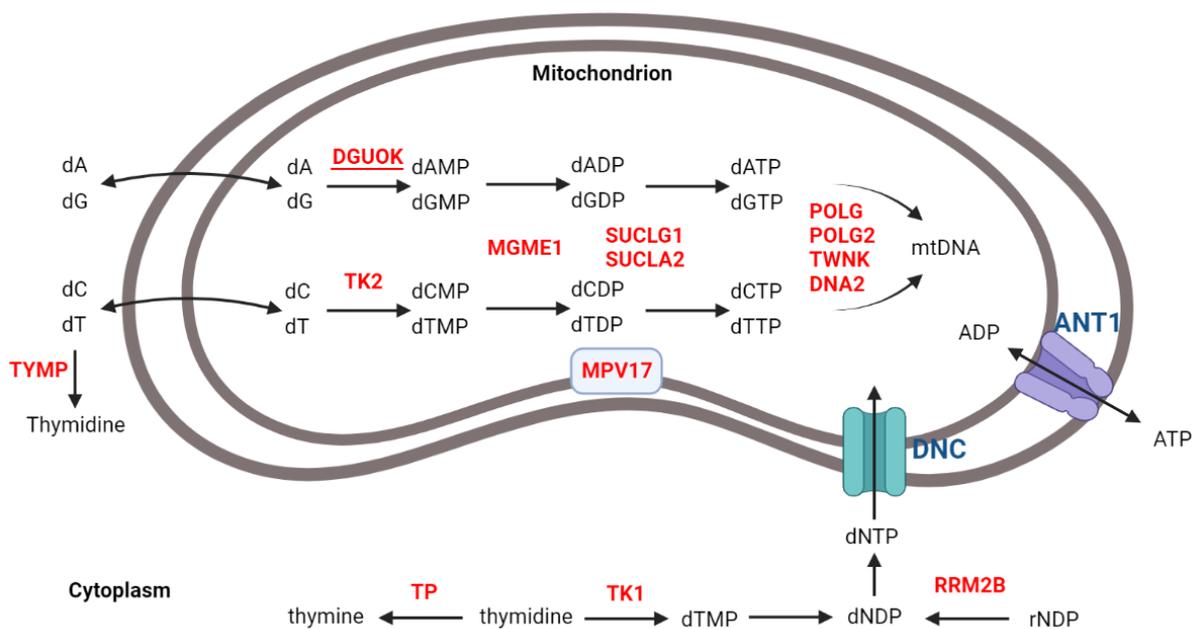
In eukaryotic cells, the organelle responsible for cellular respiration is the mitochondrion<sup>53</sup>. They possess their own DNA, namely mitochondrial DNA (mtDNA), which has some peculiar features: unlike its nuclear counterpart it is circular, it replicates independently from cell cycle, and it is present in multiple copies<sup>54</sup>. In humans, it is responsible for encoding a total of 37 genes (Figure 42): 13 subunits of the oxidative phosphorylation complex, 2 rRNAs and 22 tRNAs<sup>55</sup>. Replication of mtDNA is critical for the mitochondrion and the process is completely localised inside the double membrane. The process is performed by DNA polymerase in cooperation with TWINKLE, a DNA helicase that unwinds the double stranded structure of mtDNA, and mitochondrial single-stranded DNA-binding proteins (mtSSB) which prevents RNA synthesis and nuclease activity on the single-strand DNA<sup>56</sup>.



**Figure 42** The human mitochondrial DNA and the encoded genes in both the H-strand and the L-Strand.

Genome repair and synthesis are both very important processes and the building blocks, deoxynucleotides, must be available whenever one of the processes is being started. This implies that the nucleotides pool inside the mitochondrion must be always balanced or deletions, mutations and mtDNA depletion will occur. The first phosphorylation step is catalysed by deoxyguanosine kinase (DGUOK) for purine nucleotides and Thymidine Kinase 2 (TK2) for pyrimidine nucleotides. Formation of deoxynucleotide diphosphate (dNDP) and subsequent phosphorylation to deoxynucleotide triphosphate (dNTP) will form the required building blocks for

mtDNA<sup>57</sup>. This pathway of salvaging valuable dNTPs is extremely important because the contribution of active transport of dNTPs from the cytoplasm to the mitochondrion is not always available. For example, when cells are not actively dividing, TK1 activity is low<sup>58</sup> and dNTP synthesis is downregulated<sup>59</sup> (Figure 43). Mutations arising in any of the genes controlling the metabolism of the nucleotide pool (highlighted in red) have been associated with different forms of Mitochondrial DNA Depletion Syndrome (MDDS) based on which gene has been modified<sup>57</sup>.



**Figure 43** Schematic of nucleotide metabolism for mtDNA synthesis, repair, and most important genes in the pathways (in red) that have been associated with MDDS with corresponding gene mutation.

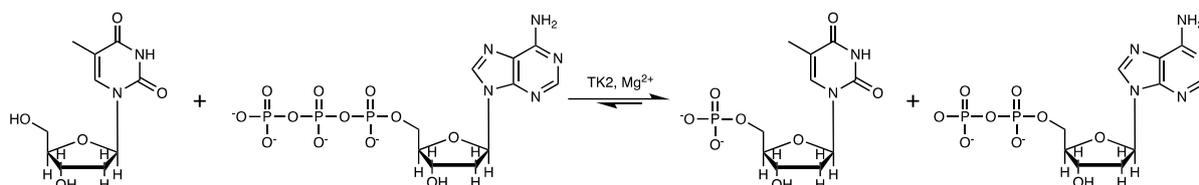
## 4.2 Mitochondrial DNA Depletion Syndrome (MDDS)

This family of syndromes is characterised by a major reduction of mtDNA, and it manifests itself in four major different forms, depending on the gene mutation that has occurred: myopathic form, due to *TK2* gene mutation, encephalomyopathic form, related to *FBXL-4* and *SUCLG1* gene mutation, hepatocerebral form, correspondent to mutation of *POLG1*, *DGUOK*, *TWINK* and *MPV17* and lastly the neurogastrointestinal form caused by *TYMPY* gene mutation<sup>60</sup>. All these are autosomal recessive diseases, and generally impact the early life of new-borns with individuals usually dying during the first 2 years of life<sup>61,62</sup>. These syndromes are rare, impacting between 1:5000 to

1:10000<sup>63</sup>. Diagnosis is generally quite accurate depending on the technique, ranging from blood samples, urine samples, spinal fluid, mtDNA rt-qPCR. The metabolites monitored are usually lactate, pyruvate and urine organic acids. The symptoms are severe and range from ataxia and respiratory failure (for example in the TK2 mutation), to liver failure (in case of the DGUOK mutation)<sup>64</sup>.

### 4.3 Evolution of therapeutics for human TK2

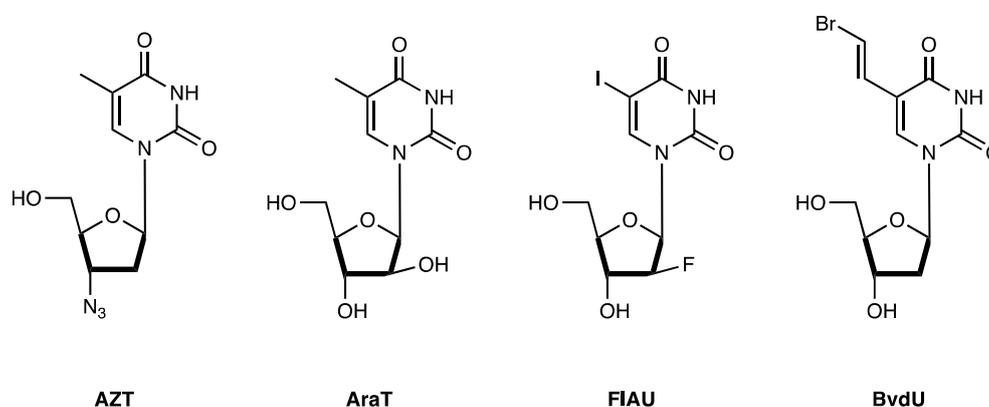
TK2 is the enzyme responsible for catalysing the phosphorylation of thymidine monophosphate, using a  $Mg^{2+}$  ion as a cofactor and a molecule of ATP as a phosphoryl group donor. The reaction is depicted in Scheme 8:



**Scheme 8** Catalytic activity of TK2. dG and dU are substrates of TK2 as well.

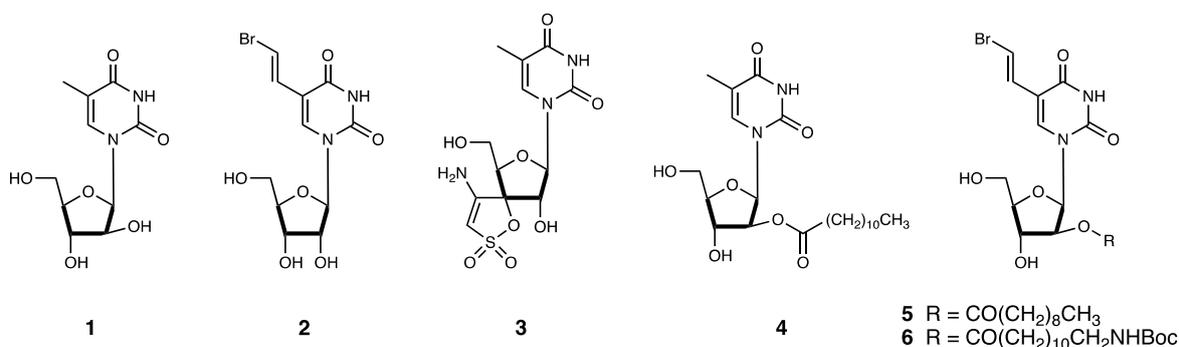
The reaction performed is the same as TK1, but the localisation is key: TK1 is a cytosolic enzyme and TK2 is a mitochondrial enzyme<sup>65</sup>. As previously mentioned, TK1 is cell cycle dependent, meaning that during cell division the levels of TK1 are very high. This determines very high production of dTMP which can be imported inside the mitochondrion. When the cells are resting (not dividing), TK1 is not expressed and the synthesis of dNTPs in the cytoplasm is not sufficient to supply enough nucleotides to the mitochondrion that is continuously replicating and repairing its own genome. In this scenario, the only possibility for the mitochondrion to replenish the nucleotide pool is to use its own salvage pathways<sup>66</sup>: phosphorylation of purine bases by DGUOK, and phosphorylation of pyrimidine bases by TK2. Mutations occurring in the TK2 gene (present on chromosome 16q22 of the human genome), can greatly affect the activity levels of TK2 which directly translate in low rates of deoxynucleotide salvage and mtDNA depletion which will determine the MDDS. To target this problem, small molecules have played a huge role in the therapeutics of this syndrome since it was discovered that TK2 can recognise substrate analogues with antiviral and anticancer

properties: these includes molecules such as 1-β-D-arabinofuranosulthimine (AraT), 1-β-D-(2'-fluoroarabinofuranosyl)-5-iodouracil (FIAU), (E)-5-(2-bromovinyl)-2'-dUrd (BvdU) and the HIV drug 3'-azido-3'-deoxythymidine (AZT)<sup>67-69</sup> (Figure 44).



**Figure 44** Thymidine nucleoside analogues recognised by TK2 as substrates.

These analogues can be recognised by both TK1 and TK2 so the options for therapeutic use are very limited due to very poor specificity so the first problem to address was clearly identified. In the early 2000s, a new series of inhibitors were synthesised by research groups to target TK2 more specifically<sup>70-73</sup>(Figure 45).



**Figure 45** A new series of compounds synthesised to more directly target TK2 independently. Chemical modifications on the furanose ring are key to this approach.

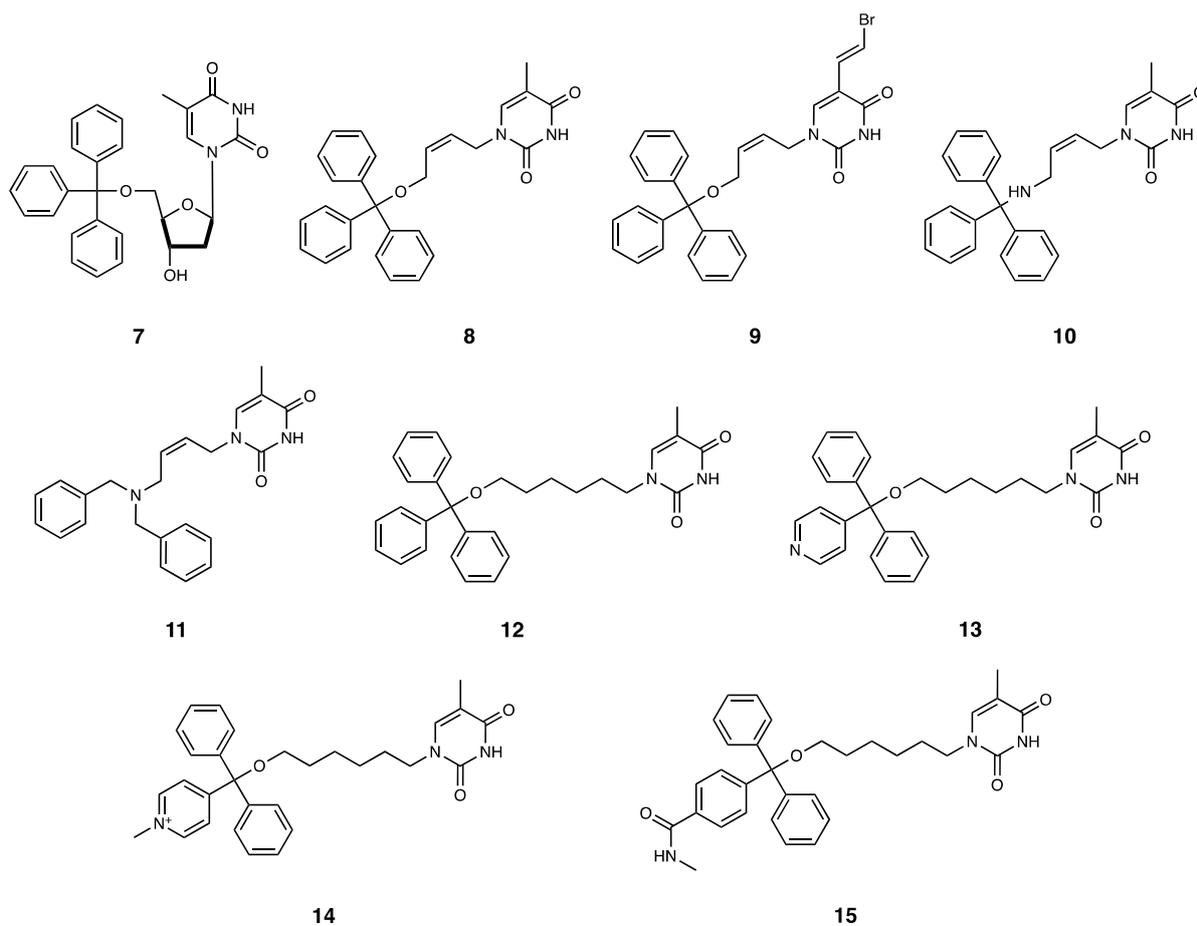
These new compounds were tested for specificity with 4 different targets: TK1, TK2, *Drosophila Melanogaster* deoxynucleotide kinase (*dm*-nDK) and Herpes Simplex Virus Type 1 thymidine kinase (HSV-1 TK). The inclusion of the latter 2 proteins is based on the high homology with TK2, around 48% for the *Drosophila* equivalent, and high 3D structure similarity for the HSV. As shown in Table 1, determining the IC<sub>50</sub>

values of these new compounds (compared to the previous inhibitor AraT) highlights the much higher affinity of the small molecules to TK2 compared to TK1<sup>74</sup>.

**Table 7** IC<sub>50</sub> values of small molecules using 2 μM [methyl-3H]dThd as substrate for kinase. All values are expressed in μM<sup>70-73</sup>

ID	TK1	TK2	<i>Dm</i> -dNK	HSV-1 TK
<b>1 (AraT)</b>	>1000	285 ± 1	24.0 ± 3.1	64 ± 28
<b>2</b>	>1000	19.0 ± 1.1	-	-
<b>3</b>	>1000	4.6 ± 0.35	-	-
<b>4</b>	>1000	28 ± 2	>1000	>1000
<b>5</b>	>1000	6.8 ± 0.7	163	>1000
<b>6</b>	ND	3.8 ± 0.2	>500	>500

Another approach used for small molecule design is to use an acyclic inhibitor. The work of Balzarini and Pérez-Pérez clearly shows how the omission of the pentose ring in the small molecule design can achieve very high selectivity towards TK2 and sub-micromolar range IC<sub>50</sub> values. The trityl series, originated from compound **7**, was further explored with modifications of the phenyl rings, carbon chain linker and nitrogen derivatives<sup>75-77</sup>.

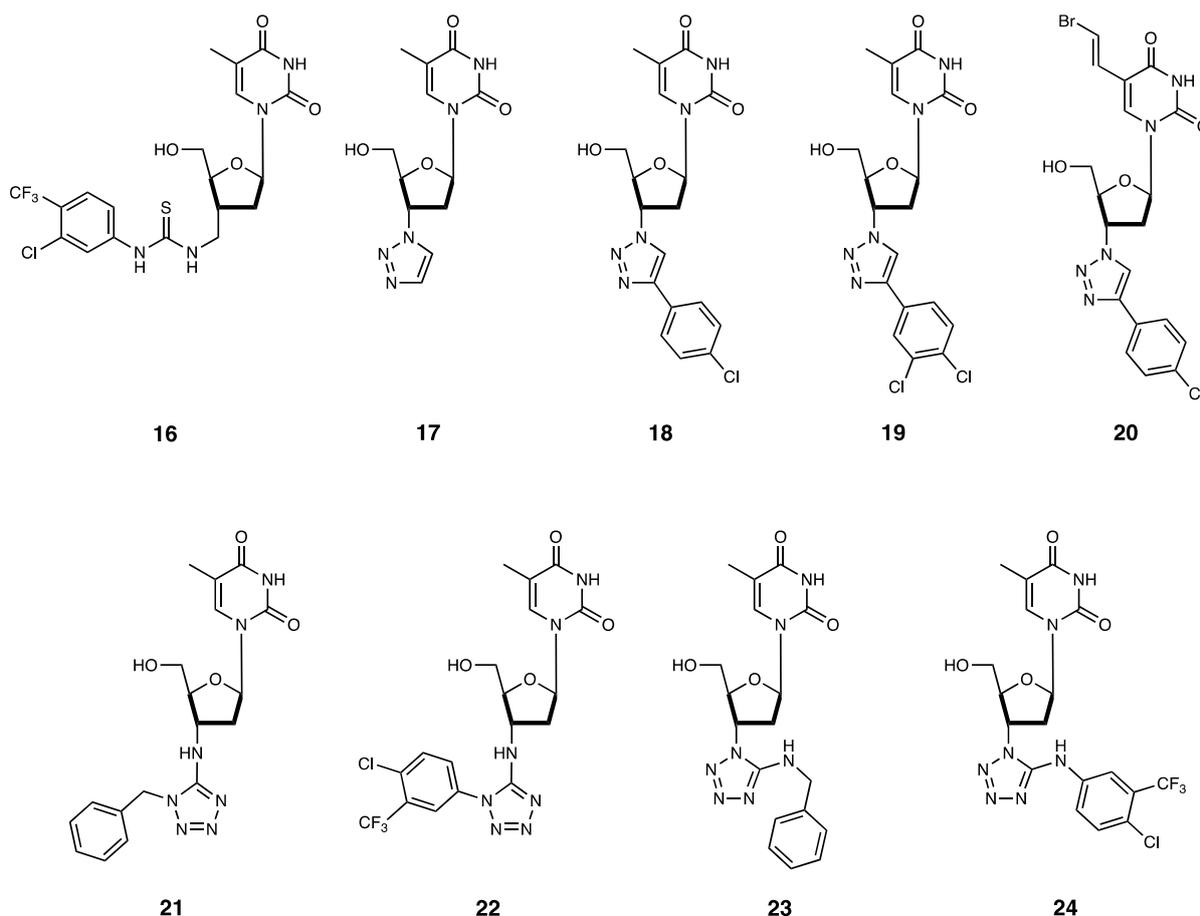


**Figure 46** The acyclic trityl derivatives. The commercial compound **7** is the starting point that will generate the entire series, including nitrogen substituted compounds.

**Table 8** IC<sub>50</sub> values of small molecules using 1 μM [methyl-3H]dTd as substrate for kinase. All values are expressed in μM<sup>75-77</sup>.

.ID	TK1	TK2	<i>Dm</i> -dNK	HSV-1 TK
<b>7</b>	>500	33 ± 20	12 ± 1	7.8 ± 0.3
<b>8</b>	>500	1.5 ± 0.2	3.3 ± 0.9	45 ± 1
<b>9</b>	>500	1.3 ± 1.1	>500	>500
<b>10</b>	>500	2.3 ± 0.4	4.4 ± 0.4	26 ± 4
<b>11</b>	>500	3.5 ± 0.5	16 ± 11	>500
<b>12</b>	>500	0.50 ± 0.01	17 ± 10	3.7 ± 0.5
<b>13</b>	>500	0.47 ± 0.03	2.7 ± 0.2	2.0 ± 0.4
<b>14</b>	>500	23 ± 1	29 ± 1	41 ± 6
<b>15</b>	>500	0.39 ± 0.03	3.5 ± 0.1	0.7 ± 0.4

Modifications of the 3' position of the furanose ring is among the newest approach of designing small molecule inhibitors for TK2. One of the promising derivations was the inclusion of a thiourea moiety on position 3'. Position 5' was also substituted with this approach, generating a library that exhibited very high specificity for TK2 and sub  $\mu\text{M}$   $\text{IC}_{50}$  values<sup>78,79</sup>. Iterating on this idea, substituting the thiourea moiety with a triazole and tetrazole groups created a new class of compounds (Figure 47).



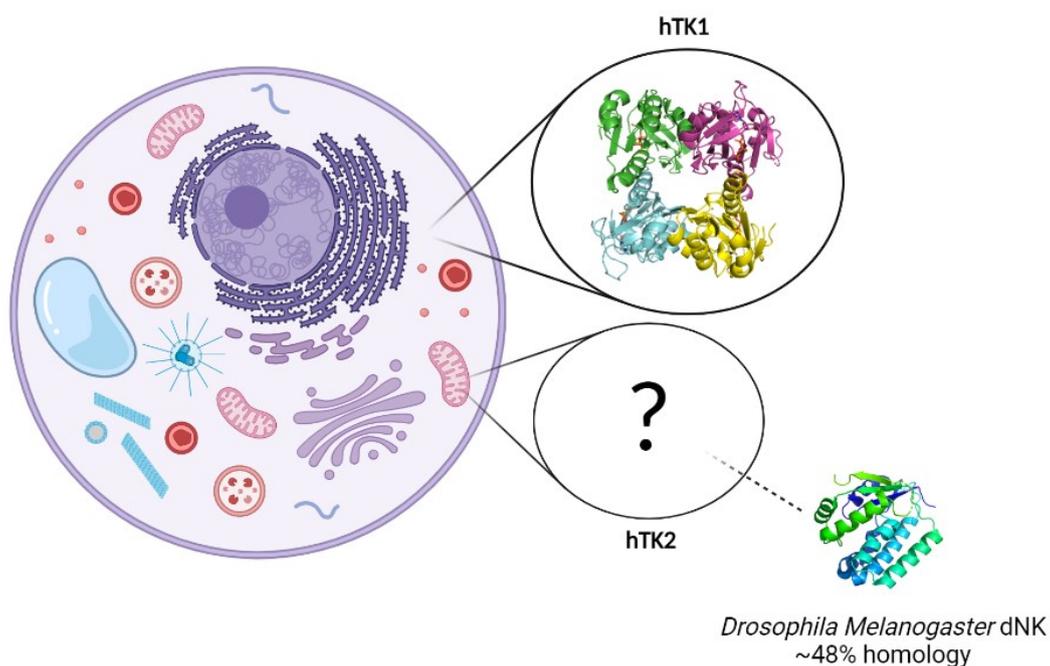
**Figure 47** Thiourea and triazole derivatives (top) and tetrazole derivatives (bottom).

The improvements observed in the reduction of  $\text{IC}_{50}$  values this time are counterbalanced by a minor loss in specificity<sup>80,81</sup>: TK1  $\text{IC}_{50}$  values are decreasing when compared to acyclic derivatives, but not as significantly as when using the thiourea moiety<sup>78,79</sup>(Table 9).

**Table 9** IC<sub>50</sub> values of small molecules using 1 μM [methyl-3H]dThd as substrate for kinase. All values are expressed in μM<sup>80,81</sup>.

ID	TK1	TK2	<i>Dm</i> -dNK	HSV-1 TK
16	>500	38 ± 1.0	-	472 ± 40
17	>500	4.7 ± 2.1	40 ± 1	212 ± 17
18	71 ± 0	0.046 ± 0.002	1.8 ± 0.9	27 ± 5
19	75 ± 49	0.042 ± 0.008	0.99 ± 0.53	30 ± 2
20	>500	0.036 ± 0.003	0.46 ± 0.06	4.5 ± 0.6
21	>500	0.90 ± 0.01	10 ± 6	>500
22	382 ± 52	0.035 ± 0.007	1.2 ± 0.0	316 ± 39
23	222 ± 112	2.6 ± 1.2	31 ± 4	39 ± 1
24	29 ± 1	0.59 ± 0.19	3.6 ± 0.7	40 ± 0

### 4.3 The lack of a TK2 crystal structure



**Figure 48** Localisation of hTK1 (cytoplasm) and hTK2 (mitochondrial) and comparison with the *Drosophila Melanogaster* deoxyribonucleoside kinase (PDB 2VPP).

Current and past design of inhibitors for TK2 has always been based on homology models and docking studies since a true X-Ray crystal structure of the protein is not

available on the PDB (Figure **48**). Despite the very high quality of models built for *in silico* design<sup>79,80,82</sup>, access to a high resolution human TK2 structure would give a much higher degree of precision while performing CADD and SAR studies to design small molecule inhibitors.

As of today, recombinant expression and purification of TK2 is not a very efficient procedure. The expression level using standard plasmids is not very high and most of the protein will be present in the inclusion body, complicating purification procedures.

#### **4.4 Project aims**

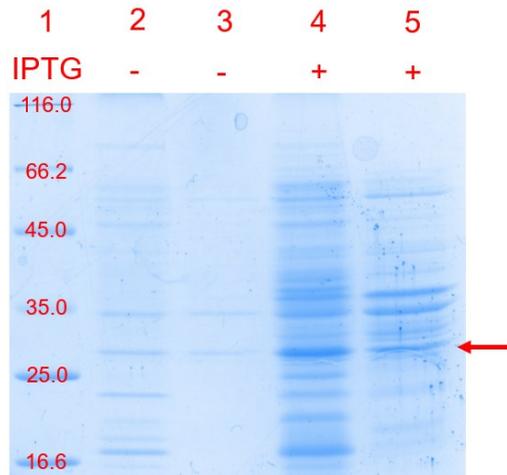
This project will focus on the major objective to create a new protocol for high level of soluble recombinant hTK2 expression and purification. This will create the conditions for cheaper and easier setup of crystal trays to screen conditions for protein crystallisation. The crystallisation will be attempted with the use of metal fluorides (as previously described in this work) to trap a TSA complex of TK2.

#### **4.5 Results**

##### **4.5.1 A new expression system for TK2**

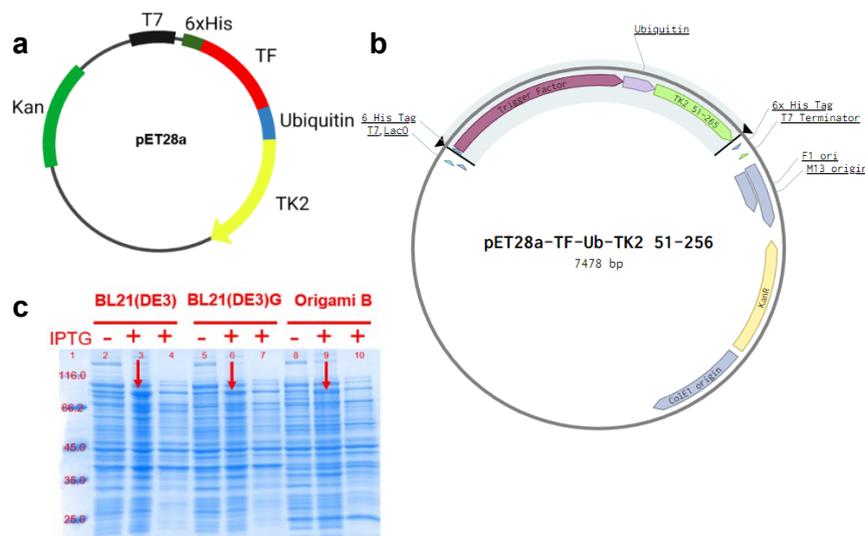
TK2 sequence is fused to a signalling peptide that is required in the cell for mitochondrial localisation of the protein. Cutting this portion from the sequence yields a fully functional form of TK2. The sequence can be further shortened and still retain all activity<sup>83,84</sup> (see TK2 sequence in section **7.7**).

The use of a plasmid composed of a pEXP vector, and this shortened sequence was investigated. *E. coli* BL21(DE3) cells were used as the candidate strain to conduct expression and levels of expression of TK2 protein were analysed by SDS-PAGE (previous testing for different strains and different conditions were screened in the research group). The SDS-PAGE (Figure **49**) clearly shows how the levels of soluble TK2 are not very high and for this reason a new strategy was developed.



**Figure 49** SDS-PAGE of *E. coli* BL21(DE3) transformed with pEXP-TK2 plasmid. Lanes 2 and 4 represent soluble and insoluble proteins, lanes 3 and 5 represent only soluble proteins.

A new construct was built, using a pET28a as a vector. Downstream of the T7 promoter region, the TK2 gene was fused to a folding chaperon, Trigger Factor (TF)<sup>85</sup>, and a Ubiquitin tag<sup>86</sup> (Figure 50 a and b). The design of this new construct will improve protein expression, protein solubility and ease of cleavage of the TF-Ub assembly with the use of Deubiquitinase Usp2-cc<sup>86</sup> to release pure TK2. The construct was obtained by Gibson Assembly and the map is illustrated in Figure x. Once the new construct was available, a protein expression test was performed using *E. coli* strains BL21(DE3), BL21(DE3)Gold, Origami B(DE3) (Figure 50 c).

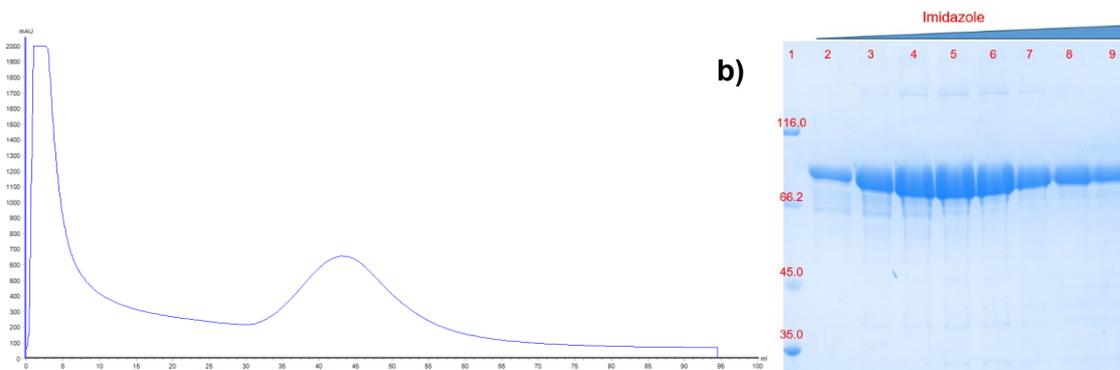


**Figure 50 a)** Schematic of the new construct, **b)** Map of cloned plasmid, **c)** SDS-PAGE small-scale expression test using the new construct to probe different *E. coli* strains.

The best results for soluble TK2 protein were obtained with BL21(DE3), so the large-scale expression was based on this specific strain.

#### 4.5.2 Recombinant TK2 purification

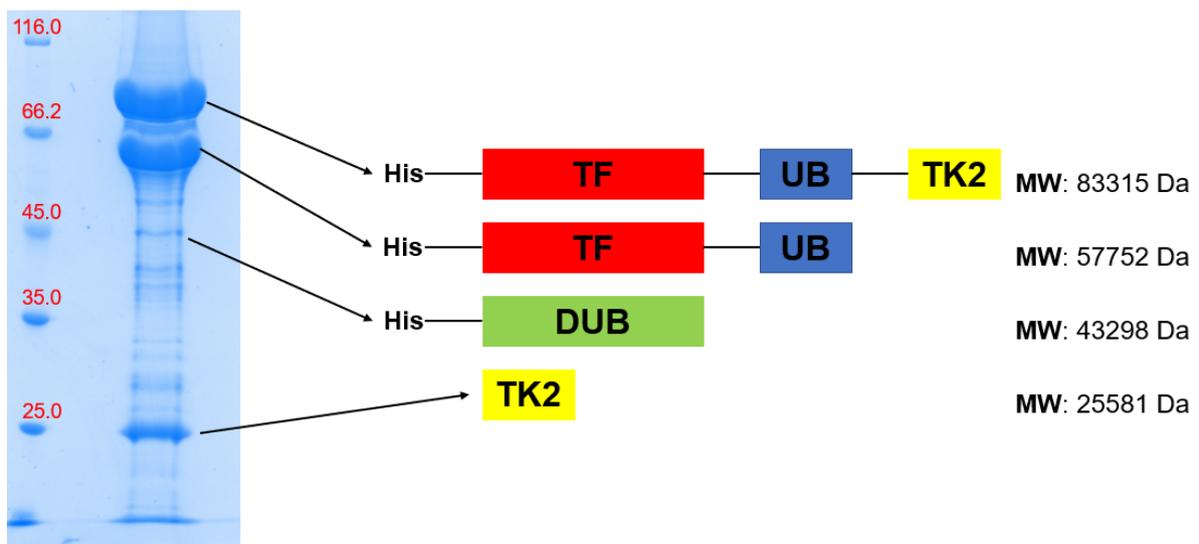
A large-scale expression was set up using BL21(DE3) and a first purification step was performed. Ni-NTA affinity chromatography was the technique of choice, yielding very high amounts of TK2 fusion protein of reasonable purity, considering that this purification step is not the final one. Once the protein was concentrated to a small volume and buffer exchanged to an imidazole free buffer, it was ready for the next step: TF-Ub assembly cleavage by the action of Usp2-cc protein.



**Figure 51 Left)** UV trace of Ni-NTA purification and **Right)** SDS-PAGE of TK2 peak that was collected.

### 4.5.3 Cleavage of TK2 fusion protein

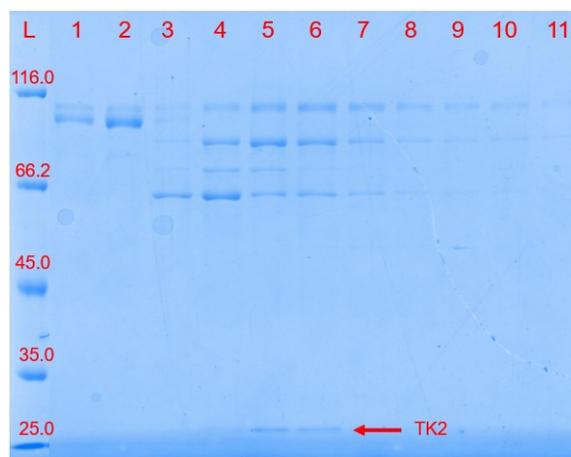
Usp2-cc is the deubiquitinase chosen for the cleavage of the TF-Ub assembly. The protein was expressed and purified as described in section 2.21.10. Once the protein was obtained pure, a trial cleavage reaction was set up. The TF-Ub-TK2 protein was incubated in buffer (50 mM Tris-HCl, 200 mM NaCl, pH 8.00) with deubiquitinase Usp2-cc in a 1:100 (dUb:protein) ratio at room temperature for 24 h. As shown in Figure 52, the cleavage was not complete but a band corresponding to the TK2 molecular weight was observed and before optimising reaction conditions for a much more efficient cleavage of the tag, confirmation of TK2 identity with Mass Spectrometry was necessary. For this reason, the next step was to purify the mixture to obtain pure TK2 protein.



**Figure 52** SDS-PAGE of cleavage test reaction. The bands highlight the presence of all four expected species with respective molecular weights: uncleaved fusion protein, cleaved tag, deubiquitinase, TK2.

#### 4.5.4 Purification of cleavage mixture

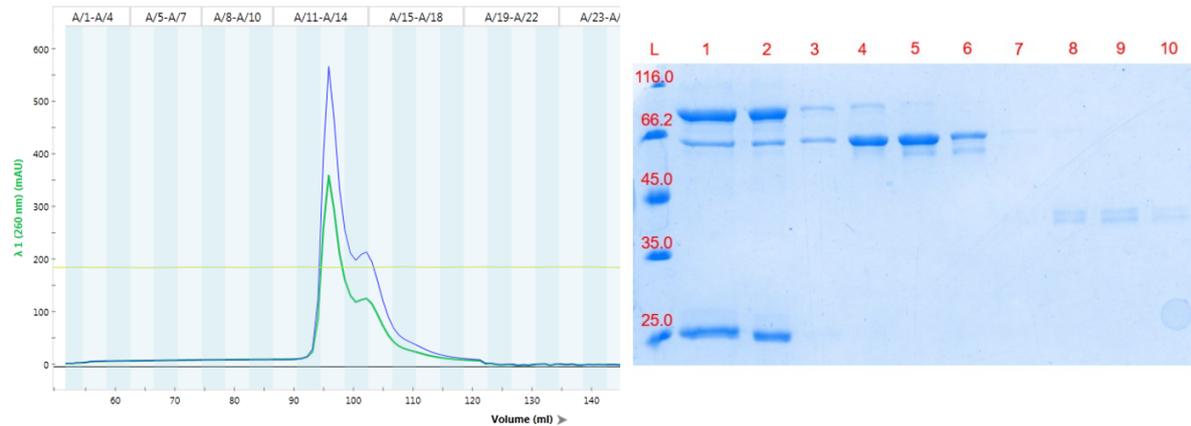
The fusion protein construct was built with ease-of-purification in mind. All the fragments generated by the cleavage of the full-length protein have an N-terminal His<sub>6</sub> tag, except for the TK2 fragment. The strategy consists of loading the cleavage mixture on a Ni-NTA affinity chromatography column and run buffer A (50 mM Tris-HCl, 200 mM NaCl, pH 8.00) while collecting the flowthrough. In principle, the full-length protein, the cleaved-off tag and the deubiquitinase should bind to the column and the TK2 protein will elute in the flow-through fraction, allowing for easy and quick separation and purification of the TK2 protein. For this purpose, the reaction mixture containing the products of the full-length protein cleavage was loaded on a Ni-NTA column and the flow-through was collected. The fractions were then analysed by SDS-PAGE to confirm the presence of the target protein, TK2, with relatively little impurities carried over from the previous purification step. As the gel clearly shows in Figure 53, coelution of all fragments is observed. The same purification was repeated increasing the concentration of NaCl to 1 M to disrupt any possible molecular interaction between macromolecules in solution, resulting in the same problem of coelution. Addition of Triton X-100 (1%) in the buffer solution did not improve the situation, still observing



**Figure 53** SDS-PAGE of eluted fractions after Ni-NTA column. TK2 protein is eluting with other fragments from the cleavage reaction. L: protein ladder. Lane 1-11: collection of the flowthrough.

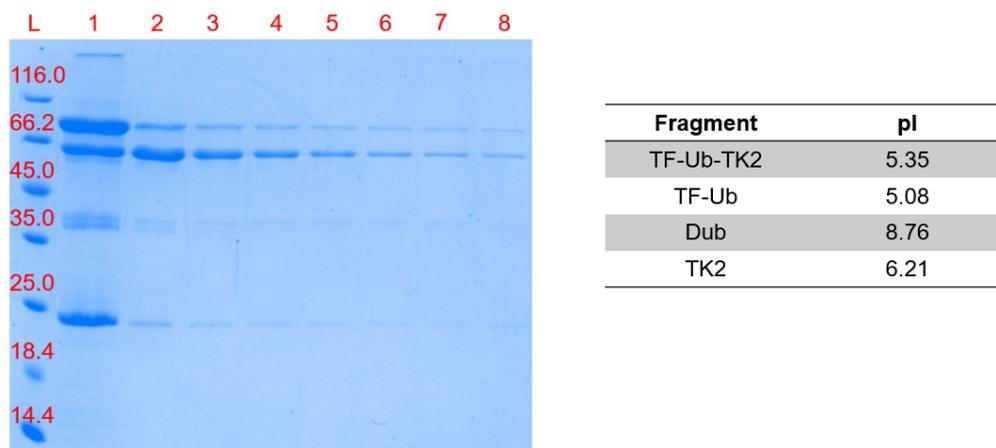
coelution while eluting the column. SEC was employed after the efforts of purifying the reaction mixture to obtain TK2 using IMAC did not yield the desired results. Unfortunately, SEC purification was not able to achieve separation, because coelution is observed in the UV trace when analysed by SDS-PAGE analysis of the fractions.

The column employed for this separation, a Superdex75 16/600, should have more than enough resolving power to separate the fragments present in the mixture. This evidence concludes that TK2 has very strong interactions with the tag, making it very hard to separate (Figure 54).



**Figure 54** SEC separation. UV trace (left) and SDS-PAGE analysis of the fractions (right). L: protein ladder. Lanes 1 to 10 represent fractions 11 to 21 which contain all the elution.

Switching to IEX chromatography, choosing Q-Sepharose as the resin that composes the stationary phase, the same problem arises (Figure 55). The isoelectric points of the 4 major macromolecules present in solution are listed below. The elution program used a NaCl gradient from 0 to 1 M in the same Tris-HCl buffer used for the previous steps. Other attempts to try to purify the mixture featured different additives to try to disrupt as much as possible protein protein interactions, such as adding 10% glycerol, 1% Triton X-100, 1 mM DTT; they all failed to achieve the objective. For this reason, the project was abandoned in order to dedicate time and effort to other projects.



**Figure 55** Left) SDS-PAGE of eluted fractions after IEX chromatography. L: protein ladder. Lanes 1-8: elution. Right) Isoelectric points of proteins and cleaved tag in solution.

## 5 Summary

YihV was recombinantly expressed and purified. The whole-cell NMR assay proved to be a very powerful technique and it was able to monitor SQ depletion from the samples during an extended period. Unfortunately, the NMR data showing in-cell SQ depletion and DHPS formation did not confirm the hypothesis that YihV upregulation would lead to increased levels of DHPS in solution due to the acceleration of the sulfoglycolytic pathway.

YihX was recombinantly expressed and purified. It was tested as an SFP phosphatase and tested as part of the D-tyrosine salvage pathway with no meaningful results arising. Crystallisation efforts led to two new structures for YihX: an apo structure and a TSA complex structure. This is a major improvement compared to the only structure published in the PDB (which has no associated publication), featuring, in the case of the TSA complex structure, an uncommon octahedral magnesium trifluoromagnesate of which there are only two other documented cases in the PDB. This evidence, combined with the  $^{19}\text{F}$  NMR characterisation, was used to propose a new enzymatic mechanism of catalysis for phosphoryl transfer. In this case, both substrate-assisted and solvent-assisted catalysis fit the hypothesis but more evidence is needed to fully determine which mechanism the reaction undergoes. Site-directed mutagenesis and kinetic parameters determination highlighted the importance of residues N<sub>108</sub> and Lys<sub>141</sub> in the binding and stabilisation of the substrate. The attempt of producing a knock-in mutant, YihX<sub>G8D</sub>, to restore the characteristic DxD motif resulted in slower kinetics supporting the newly proposed mechanism compared to the standard HAD phosphatase mechanism:  $^{19}\text{F}$  NMR showed no formation of TSA complex of YihX<sub>G8D</sub> when incubated with  $\text{Mg}^{2+}$  and  $\text{F}^-$  adding experimental evidence supporting the statement.

Serendipitous discovery of ATP synthesis activity by YihX was characterised by a combination of  $^{31}\text{P}$  NMR experiments and HPLC stop-assay to determine the overall reaction equation and the kinetic parameters governing the phosphoryl transfer.

Bioinformatic investigation identified a list of DxG motif I uncharacterised enzymes. PG\_0725 was chosen as a test candidate to prove whether this newly discovered ATP synthesis activity is a peculiarity of just YihX or other DxG HAD phosphatases could

perform the same reaction. Results highlighted how YihX most likely represents a one-off scenario for ATP synthesis, due to a combination of being a C1 class HAD enzyme with an unusually wide-open active site that can accept more bulky substrates.

A new construct for human TK2 expression was cloned. The resulting fusion protein proved to greatly increase the yield of soluble TK2 protein and that the target protein could be obtained by a simple cleavage reaction by the action of deubiquitinase Usp2-cc. However, a major problem was encountered while purifying the reaction mixture after the cleavage. Despite the fact that construct design accounted for easy and quick removal of His-tagged fragments to obtain pure TK2, the target protein could not be separated by IMAC, nor by other chromatographic techniques such as SEC and IEX, despite extensive efforts and screening of multiple buffering conditions to obtain pure TK2. For these reasons, the project was stopped to continue working on different projects.

## 6 References

1. Koonin E V., Tatusove RL. Computer Analysis of Bacterial Haloacid Dehalogenases Defines a Large Superfamily of Hydrolases with Diverse Specificity. *Journal of Molecular Biology*. **1994**;244(1):125–32.
2. Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes. *Journal of Molecular Biology*. **2006**;361(5):1003–34.
3. Morais MC, Zhang W, Baker AS, Zhang G, Dunaway-Mariano D, Allen KN. The Crystal Structure of Bacillus cereus Phosphonoacetaldehyde Hydrolase: Insight into Catalysis of Phosphorus Bond Cleavage and Catalytic Diversification within the HAD Enzyme Superfamily†,‡. *Biochemistry*. **2000**;39(34):10385–96.
4. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science*. **2003**;299(5615):2067–71.
5. Wang W, Kim R, Jancarik J, Yokota H, Kim SH. Crystal structure of phosphoserine phosphatase from Methanococcus jannaschii, a hyperthermophile, at 1.8 Å resolution. *Structure*. **2001**;9(1):65–71.
6. Peisach E, Selengut JD, Dunaway-Mariano D, Allen KN. X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily. *Biochemistry*. **2004**;43(40):12770–9.
7. Collet JF, Stroobant V, Pirard M, Delpierre G, Van Schaftingen E. A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif. *Journal of Biological Chemistry*. **1998**;273(23):14107–12.
8. Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. *Nature reviews Molecular cell biology*. **2006**;7(11):833–46.
9. Alonso A, Sasin J, Bottini N, Friedberg I, Friedberg I, Osterman A, et al. Protein tyrosine phosphatases in the human genome. *Cell*. **2004**;117(6):699–711.
10. Wang W, Cho HS, Kim R, Jancarik J, Yokota H, Nguyen HH, et al. Structural Characterization of the Reaction Pathway in Phosphoserine Phosphatase: Crystallographic “snapshots” of Intermediate States. *Journal of Molecular Biology*. **2002**;319(2):421–31.
11. Griffin JL, Bowler MW, Baxter NJ, Leigh KN, Dannatt HRW, Hounslow AM, et al. Near attack conformers dominate  $\beta$ -phosphoglucomutase complexes where geometry and charge distribution reflect those of substrate. *Proceedings of the National Academy of Sciences of the United States of America*. **2012**;109(18):6910–5.
12. Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. *Nature Reviews Molecular Cell Biology* 2006 7:11. **2006**;7(11):833–46.
13. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN. Caught in the Act: The Structure of Phosphorylated  $\beta$ -Phosphoglucomutase from Lactococcus lactis†,‡. *Biochemistry*. **2002**;41(26):8351–9.
14. Jin Y, Richards NG, Waltho JP, Blackburn GM. Metal Fluorides as Analogues for Studies on Phosphoryl Transfer Enzymes. *Angewandte Chemie (International ed in English)*. **2017**;56(15):4110–28.

15. Cho H, Wang W, Kim R, Yokota H, Damo S, Kim SH, et al. BeF<sub>3</sub><sup>-</sup> acts as a phosphate analog in proteins phosphorylated on aspartate: Structure of a BeF<sub>3</sub><sup>-</sup> complex with phosphoserine phosphatase. *Proceedings of the National Academy of Sciences of the United States of America.* **2001**;98(15):8525–30.
16. Ge M, Molt RW, Jenkins HT, Blackburn GM, Jin Y, Antson AA. Octahedral Trifluoromagnesate, an Anomalous Metal Fluoride Species, Stabilizes the Transition State in a Biological Motor. *ACS Catalysis.* **2021**;11(5):2769–73.
17. Cui H, Müller AU, Leibundgut M, Tian J, Ban N, Weber-Ban E. Structures of prokaryotic ubiquitin-like protein Pup in complex with depupylase Dop reveal the mechanism of catalytic phosphate formation. *Nature Communications* 2021 12:1. **2021**;12(1):1–12.
18. Chen H, Viel S, Ziarelli F, Peng L. 19 F NMR: a valuable tool for studying biological events. *Chemical Society Reviews.* **2013**;42(20):7971–82.
19. Brünger AT. X-ray crystallography and NMR reveal complementary views of structure and dynamics. *Nature Structural Biology.* **1997**;4 Suppl(SUPPL.):862–5.
20. Angkawijaya AE, Ngo AH, Nguyen VC, Gunawan F, Nakamura Y. Expression profiles of 2 phosphate starvation-inducible phosphocholine/phosphoethanolamine phosphatases, PECP1 and PS2, in arabidopsis. *Frontiers in Plant Science.* **2019**;10:461509.
21. Hanchi M, Thibaud MC, Légeret B, Kuwata K, Pochon N, Beisson F, et al. The Phosphate Fast-Responsive Genes PECP1 and PPsPase1 Affect Phosphocholine and Phosphoethanolamine Content. *Plant Physiology.* **2018**;176(4):2943–62.
22. Angkawijaya AE, Nakamura Y. Arabidopsis PECP1 and PS2 are phosphate starvation-inducible phosphocholine phosphatases. *Biochemical and Biophysical Research Communications.* **2017**;494(1–2):397–401.
23. Deng S, Lu L, Li J, Du Z, Liu T, Li W, et al. Purple acid phosphatase 10c encodes a major acid phosphatase that regulates plant growth under phosphate-deficient conditions in rice. *Journal of Experimental Botany.* **2020**;71(14):4321–32.
24. Bernstein NK, Williams RS, Rakovszky ML, Cui D, Green R, Karimi-Busheri F, et al. The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Molecular cell.* **2005**;17(5):657–70.
25. Newman JW, Morisseau C, Harris TR, Hammock BD. The soluble epoxide hydrolase encoded by EPXH2 is a bifunctional enzyme with novel lipid phosphate phosphatase activity. *Proceedings of the National Academy of Sciences of the United States of America.* **2003**;100(4):1558–63.
26. Seifried A, Schultz J, Gohla A. Human HAD phosphatases: structure, mechanism, and roles in health and disease. *The FEBS Journal.* **2013**;280(2):549–71.
27. Lutsenko S, Petris MJ. Function and regulation of the mammalian copper-transporting ATPases: Insights from biochemical and cell biological approaches. *Journal of Membrane Biology.* **2003**;191(1):1–12.
28. Nelson N. Metal ion transporters and homeostasis. *The EMBO Journal.* **1999**;18(16):4361–71.

29. Rosen BP. Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology.* **2002**;133(3):689–93.
30. Okkeri J, Haltia T. Expression and Mutagenesis of ZntA, a Zinc-Transporting P-Type ATPase from Escherichia coli. *Biochemistry.* **1999**;38(42):14109–16.
31. Rensing C, Fan B, Sharma R, Mitra B, Rosen BP. CopA: An Escherichia coli Cu(I)-translocating P-type ATPase. *Proceedings of the National Academy of Sciences of the United States of America.* **2000**;97(2):652–6.
32. Kamenski T, Heilmeyer S, Meinhart A, Cramer P. Structure and mechanism of RNA polymerase II CTD phosphatases. *Molecular Cell.* **2004**;15(3):399–407.
33. Zhang Y, Kim Y, Genoud N, Gao J, Kelly JW, Pfaff SL, et al. Determinants for Dephosphorylation of the RNA Polymerase II C-Terminal Domain by Scp1. *Molecular cell.* **2006**;24(5):759.
34. Tremblay LW, Dunaway-Mariano D, Allen KN. Structure and activity analyses of Escherichia coli K-12 NagD provide insight into the evolution of biochemical function in the haloalkanoic acid dehalogenase superfamily. *Biochemistry.* **2006**;45(4):1183–93.
35. Lu Z, Dunaway-Mariano D, Allen KN. HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry.* **2005**;44(24):8684–96.
36. Lu Z, Wang L, Dunaway-Mariano D, Allen KN. Structure-Function Analysis of 2-Keto-3-deoxy-D-glycero-D-galactononate-9-phosphate Phosphatase Defines Specificity Elements in Type C0 Haloalkanoate Dehalogenase Family Members. *The Journal of Biological Chemistry.* **2009**;284(2):1224.
37. Allen KN, Dunaway-Mariano D. Markers of fitness in a successful enzyme superfamily. *Current opinion in structural biology.* **2009**;19(6):658–65.
38. Rinaldo-Matthis A, Rampazzo C, Reichard P, Bianchi V, Nordlund P. Crystal structure of a human mitochondrial deoxyribonucleotidase. *Nature structural biology.* **2002**;9(10):779–87.
39. Lahiri SD, Zhang G, Dunaway-Mariano D, Allen KN. Diversification of Function in the Haloacid Dehalogenase Enzyme Superfamily: The Role of the Cap Domain in Hydrolytic Phosphorus-Carbon Bond Cleavage. *Bioorganic chemistry.* **2006**;34(6):394.
40. Denger K, Weiss M, Felux A-K, Schneider A, Mayer C, Spitteller D, et al. Sulphoglycolysis in Escherichia coli K-12 closes a gap in the biogeochemical sulphur cycle. *Nature.* **2014**;507(7490):114–7.
41. Benning C. BIOSYNTHESIS AND FUNCTION OF THE SULFOLIPID SULFOQUINOVOSYL DIACYLGLYCEROL. *Annual review of plant physiology and plant molecular biology.* **1998**;49:53–75.
42. Endo K, Kobayashi K, Wada H. Sulfoquinovosyldiacylglycerol has an Essential Role in Thermosynechococcus elongatus BP-1 Under Phosphate-Deficient Conditions. *Plant and Cell Physiology.* **2016**;57(12):2461–71.
43. Denger K, Huhn T, Hollemeyer K, Schleheck D, Cook AM. Sulfoquinovose degraded by pure cultures of bacteria with release of C3-organosulfonates: complete degradation in two-member communities. *FEMS microbiology letters.* **2012**;328(1):39–45.

44. Hanson BT, Dimitri Kits K, Löffler J, Burrichter AG, Fiedler A, Denger K, et al. Sulfoquinovose is a select nutrient of prominent bacteria and a source of hydrogen sulfide in the human gut. *The ISME Journal* 2021 15:9. **2021**;15(9):2779–91.
45. Sharma M, Abayakoon P, Epa R, Jin Y, Lingford JP, Shimada T, et al. Molecular Basis of Sulfosugar Selectivity in Sulfoglycolysis. *ACS Central Science*. **2021**;7(3):476–87.
46. Ma J, McLeod S, Maccormack K, Sriram S, Gao N, Breeze AL, et al. Real-Time Monitoring of New Delhi Metallo- $\beta$ -Lactamase Activity in Living Bacterial Cells by  $^1\text{H}$  NMR Spectroscopy. *Angewandte Chemie International Edition*. **2014**;53(8):2130–3.
47. Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L, et al. Panoramic view of a superfamily of phosphatases through substrate profiling. *Proceedings of the National Academy of Sciences of the United States of America*. **2015**;112(16):E1974–83.
48. Hussain T, Kruparani SP, Pal B, Dock-Bregeon AC, Dwivedi S, Shekar MR, et al. Post-transfer editing mechanism of a D-aminoacyl-tRNA deacylase-like domain in threonyl-tRNA synthetase from archaea. *The EMBO journal*. **2006**;25(17):4152–62.
49. Mui JWY, De Souza DP, Saunders EC, McConville MJ, Williams SJ. Remodeling of Carbon Metabolism during Sulfoglycolysis in Escherichia coli. *Applied and Environmental Microbiology*. **2023**;89(2).
50. WE B, N M, DC A, RH D, DS K. Plasmid-encoded protein: the principal factor in the 'metabolic burden' associated with recombinant bacteria. *Biotechnology Bioengineering*. **1990**. *Biotechnology and bioengineering*. **2009**;102(5):1283.
51. Zhang Y, Mui JWY, Arumaperuma T, Lingford JP, Goddard-Borger ED, White JM, et al. Concise synthesis of sulfoquinovose and sulfoquinovosyl diacylglycerides, and development of a fluorogenic substrate for sulfoquinovosidases. *Organic & Biomolecular Chemistry*. **2020**;18(4):675–86.
52. Hansen PE. Isotope Effects on Chemical Shifts in the Study of Hydrogen Bonds in Small Molecules. *Molecules*. **2022**;27(8).
53. Osellame LD, Blacker TS, Duchon MR. Cellular and molecular mechanisms of mitochondrial function. *Best Practice & Research Clinical Endocrinology & Metabolism*. **2012**;26(6):711.
54. Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. **1999**;1410(2):103–23.
55. Barchiesi A, Vascotto C. Transcription, Processing, and Decay of Mitochondrial RNA in Health and Disease. *International Journal of Molecular Sciences* 2019, Vol 20, Page 2221. **2019**;20(9):2221.
56. Falkenberg M. Mitochondrial DNA replication in mammalian cells: overview of the pathway. *Essays in Biochemistry*. **2018**;62(3):287.
57. Dolce V, Fiermonte G, Runswick MJ, Palmieri F, Walker JE. The human mitochondrial deoxynucleotide carrier and its role in the toxicity of nucleoside antivirals. *Proceedings of the National Academy of Sciences of the United States of America*. **2001**;98(5):2284–8.
58. Munch-Petersen B, Cloos L, Jensen HK, Tyrsted G. Human thymidine kinase 1. Regulation in normal and malignant cells. *Advances in Enzyme Regulation*. **1995**;35(C).

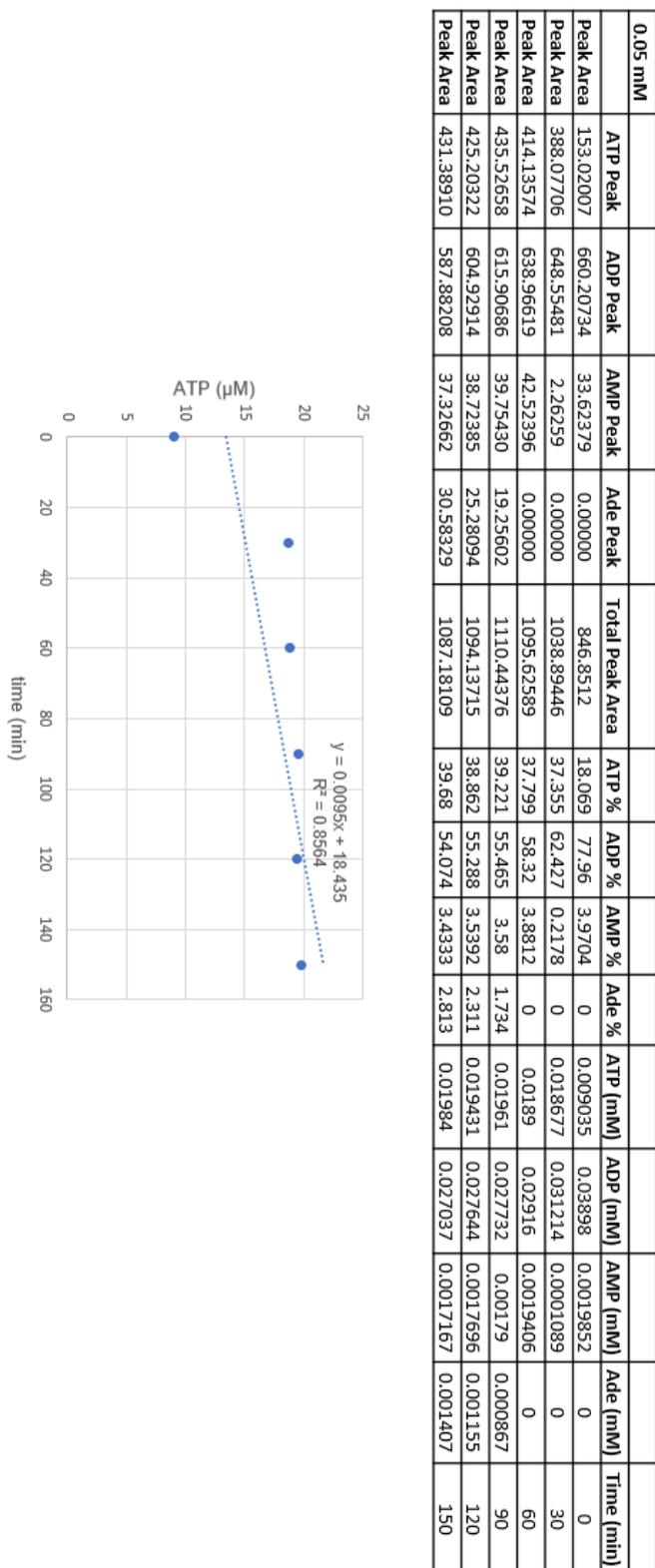
59. Stillman B. Deoxynucleoside triphosphate (dNTP) synthesis and destruction regulate the replication of both cell and virus genomes. *Proceedings of the National Academy of Sciences of the United States of America.* **2013**;110(35):14120.
60. Wang H, Han Y, Li S, Chen Y, Chen Y, Wang J, et al. Mitochondrial DNA Depletion Syndrome and Its Associated Cardiac Disease. *Frontiers in Cardiovascular Medicine.* **2022**;8:2200.
61. Chanprasert S, Wang J, Weng SW, Enns GM, Boué DR, Wong BL, et al. Molecular and clinical characterization of the myopathic form of mitochondrial DNA depletion syndrome caused by mutations in the thymidine kinase (TK2) gene. *Molecular Genetics and Metabolism.* **2013**;110(1–2):153–61.
62. Landsverk ML, Zhang VW, Wong LJC, Andersson HC. A SUCLG1 mutation in a patient with mitochondrial DNA depletion and congenital anomalies. *Molecular Genetics and Metabolism Reports.* **2014**;1:451.
63. Nogueira C, Almeida LS, Nesti C, Pezzini I, Videira A, Vilarinho L, et al. Syndromes associated with mitochondrial DNA depletion. *Italian Journal of Pediatrics.* **2014**;40(1):34.
64. Parikh S, Goldstein A, Koenig MK, Scaglia F, Enns GM, Saneto R, et al. Diagnosis and management of mitochondrial disease: a consensus statement from the Mitochondrial Medicine Society. *Genetics in medicine : official journal of the American College of Medical Genetics.* **2015**;17(9):689.
65. Arnér ESJ, Eriksson S. Mammalian deoxyribonucleoside kinases. *Pharmacology & Therapeutics.* **1995**;67(2):155–86.
66. Rampazzo C, Ferraro P, Pontarin G, Fabris S, Reichard P, Bianchi V. Mitochondrial deoxyribonucleotides, pool sizes, synthesis, and regulation. *The Journal of biological chemistry.* **2004**;279(17):17019–26.
67. Susan-Resiga D, Bentley AT, Lynx MD, LaClair DD, McKee EE. Zidovudine Inhibits Thymidine Phosphorylation in the Isolated Perfused Rat Heart. *Antimicrobial Agents and Chemotherapy.* **2007**;51(4):1142.
68. Wang L, Munch-Petersen B, Herrström Sjöberg A, Hellman U, Bergman T, Jörnvall H, et al. Human thymidine kinase 2: molecular cloning and characterisation of the enzyme activity with antiviral and cytostatic nucleoside substrates. *FEBS letters.* **1999**;443(2):170–4.
69. Lynx MD, McKee EE. 3'-Azido-3'-deoxythymidine (AZT) is a competitive inhibitor of thymidine phosphorylation in isolated rat heart and liver mitochondria. *Biochemical pharmacology.* **2006**;72(2):239–43.
70. Ciliberti N, Manfredini S, Angusti A, Durini E, Solaroli N, Vertuani S, et al. Novel selective human mitochondrial kinase inhibitors: design, synthesis and enzymatic activity. *Bioorganic & medicinal chemistry.* **2007**;15(8):3065–81.
71. Manfredini S, Baraldi PG, Durini E, Porcu L, Angusti A, Vertuani S, et al. Design, synthesis and enzymatic activity of highly selective human mitochondrial thymidine kinase inhibitors. *Bioorganic and Medicinal Chemistry Letters.* **2001**;11(10):1329–32.
72. Balzarinia J, Degrève B, Zhu C, Durini E, Porcu L, De Clercq E, et al. 2'-O-Acyl/alkyl-substituted arabinosyl nucleosides as inhibitors of human mitochondrial thymidine kinase. *Biochemical Pharmacology .* **2001**;61(6):727–32.

73. Balzarini J, Zhu C, De Clercq E, Perez-Perez MJ, Chamorro C, Camarasa MJ, et al. Novel ribofuranosyl nucleoside lead compounds for potent and selective inhibitors of mitochondrial thymidine kinase-2. *Biochemical Journal*. **2000**;351(Pt 1):167.
74. Priego E-M, Karlsson A, Gago F, Camarasa M-J, Balzarini J, Perez-Perez M-J. Recent advances in thymidine kinase 2 (TK2) inhibitors and new perspectives for potential applications. *Current pharmaceutical design*. **2012**;18(20):2981–94.
75. Priego EM, Balzarini J, Karlsson A, Camarasa MJ, Pérez-Pérez MJ. Synthesis and evaluation of thymine-derived carboxamides against mitochondrial thymidine kinase (TK-2) and related enzymes. *Bioorganic & medicinal chemistry*. **2004**;12(19):5079–90.
76. Hernández AI, Familiar O, Negri A, Rodríguez-Barrios F, Gago F, Karlsson A, et al. N1-substituted thymine derivatives as mitochondrial thymidine kinase (TK-2) inhibitors. *Journal of Medicinal Chemistry*. **2006**;49(26):7766–73.
77. Balzarini J, Hernández AI, Roche P, Esnouf R, Karlsson A, Camarasa MJ, et al. Non-nucleoside inhibitors of mitochondrial thymidine kinase (TK-2) differentially inhibit the closely related herpes simplex virus type 1 TK and Drosophila melanogaster multifunctional deoxynucleoside kinase. *Molecular pharmacology*. **2003**;63(2):263–70.
78. Van Daele I, Munier-Lehmann H, Froeyen M, Balzarini J, Van Calenbergh S. Rational design of 5'-thiourea-substituted alpha-thymidine analogues as thymidine monophosphate kinase inhibitors capable of inhibiting mycobacterial growth. *Journal of medicinal chemistry*. **2007**;50(22):5281–92.
79. Balzarini J, Van Daele I, Negri A, Solaroli N, Karlsson A, Liekens S, et al. Human mitochondrial thymidine kinase is selectively inhibited by 3'-thiourea derivatives of beta-thymidine: identification of residues crucial for both inhibition and catalytic activity. *Molecular pharmacology*. **2009**;75(5):1127–36.
80. Van Poecke S, Negri A, Janssens J, Solaroli N, Karlsson A, Gago F, et al. Synthesis, modeling and evaluation of 3'-(1-aryl-1H-tetrazol-5-ylamino)-substituted 3'-deoxythymidine derivatives as potent and selective human mitochondrial thymidine kinase inhibitors. *Organic & biomolecular chemistry*. **2011**;9(3):892–901.
81. Van Poecke S, Negri A, Gago F, Van Daele I, Solaroli N, Karlsson A, et al. 3'-[4-Aryl-(1,2,3-triazol-1-yl)]-3'-deoxythymidine analogues as potent and selective inhibitors of human mitochondrial thymidine kinase. *Journal of Medicinal Chemistry*. **2010**;53(7):2902–12.
82. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*. **2009**;4(3):363–73.
83. Wang L, Munch-Petersen B, Herrström Sjöberg A, Hellman U, Bergman T, Jörnvall H, et al. Human thymidine kinase 2: molecular cloning and characterisation of the enzyme activity with antiviral and cytostatic nucleoside substrates. *FEBS Letters*. **1999**;443(2):170–4.
84. Wang L, Saada A, Eriksson S. Kinetic Properties of Mutant Human Thymidine Kinase 2 Suggest a Mechanism for Mitochondrial DNA Depletion Myopathy. *Journal of Biological Chemistry*. **2003**;278(9):6963–8.
85. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. *Nature* 2011 475:7356. **2011**;475(7356):324–32.

86. Catanzariti A-M, Soboleva TA, Jans DA, Board PG, Baker RT. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Science*. **2004**;13(5):1331–9.

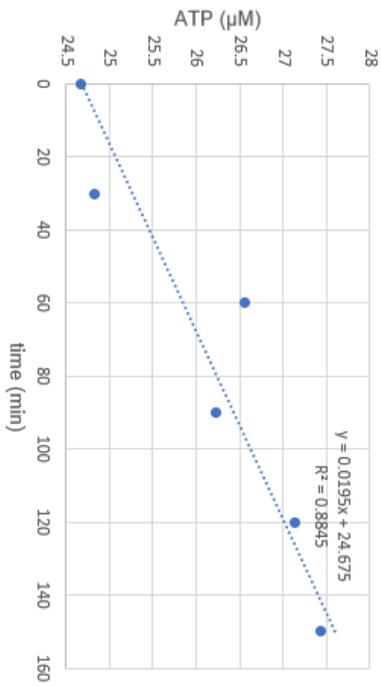
## 7 Appendix

### 7.1 HPLC stop-assay for ADP phosphorylation by YihX



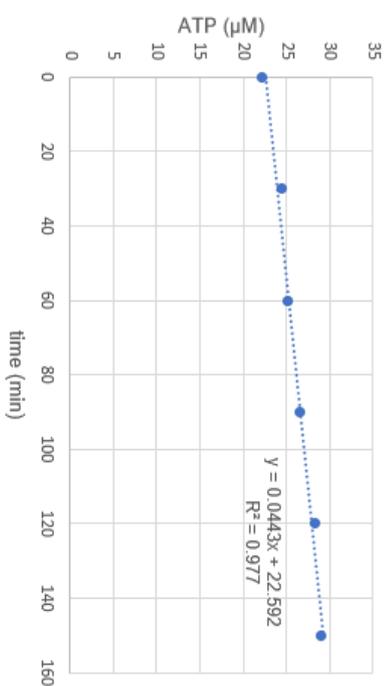
**Figure 56** HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.05 mM

0.1 mM														
Peak Area	ATP Peak	ADP Peak	AMP Peak	Ade Peak	Total Peak Area	ATP %	ADP %	AMP %	Ade %	ATP (mM)	ADP (mM)	AMP (mM)	Ade (mM)	Time (min)
439.13785	439.13785	1265.41028	76.04285	0.00000	1780.59098	24.662	71.067	4.2707	0	0.024662	0.071067	0.0042707	0	0
437.65482	437.65482	1251.02869	74.60003	0.00000	1763.28354	24.82	70.949	4.2307	0	0.02482	0.070949	0.0042307	0	30
475.90067	475.90067	1219.03857	71.99879	25.29464	1792.23267	26.554	68.018	4.0173	1.411	0.026554	0.068018	0.0040173	0.001411	60
459.74832	459.74832	1186.42627	69.87448	37.46243	1753.5115	26.219	67.66	3.9848	2.136	0.026219	0.06766	0.0039848	0.002136	90
474.14685	474.14685	1156.21204	67.89818	48.84357	1747.10064	27.139	66.179	3.8863	2.796	0.027139	0.066179	0.0038863	0.002796	120
477.80200	477.80200	1137.62219	66.06150	60.12883	1741.61452	27.434	65.32	3.7931	3.452	0.027434	0.06532	0.0037931	0.003452	150



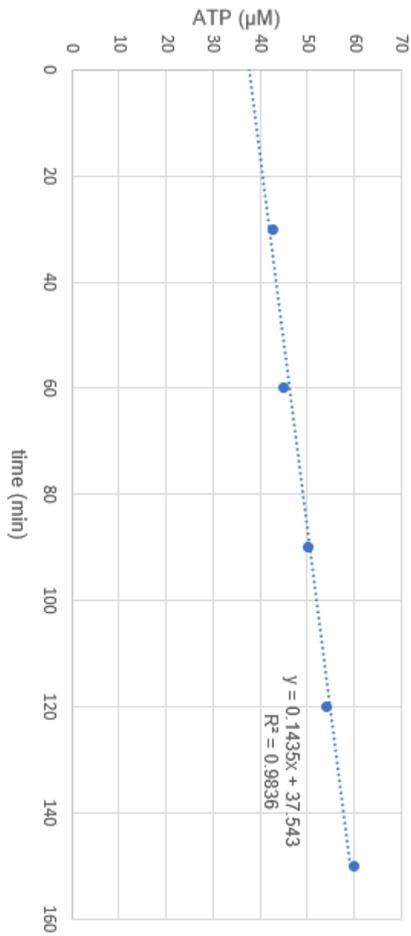
**Figure 57** HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.10 mM

0.25 mM														
	ATP Peak	ADP Peak	AMP Peak	Ade Peak	Total Peak Area	ATP %	ADP %	AMP %	Ade %	ATP (mM)	ADP (mM)	AMP (mM)	Ade (mM)	Time (min)
Peak Area	322,39423	3140,35767	169,71837	0,00000	3632,47027	8,8753	86,452	4,6723	0	0,022188	0,216131	0,0116806	0	0
Peak Area	352,43130	3046,32935	165,71198	30,29690	3594,76953	9,804	84,743	4,6098	0,843	0,02451	0,211858	0,0115245	0,002107	30
Peak Area	361,75821	3015,91309	167,95633	61,22499	3606,85262	10,03	83,616	4,6566	1,697	0,025074	0,20904	0,0116415	0,004244	60
Peak Area	380,00302	2945,39209	165,60115	91,19063	3582,18689	10,608	82,223	4,6229	2,546	0,02652	0,205558	0,0115573	0,006364	90
Peak Area	401,18381	2866,84473	162,79097	119,84320	3550,66271	11,299	80,741	4,5848	3,375	0,028247	0,201853	0,011462	0,008438	120
Peak Area	410,26721	2819,02246	162,20721	148,98543	3540,48231	11,588	79,623	4,5815	4,208	0,02897	0,199056	0,0114538	0,01052	150



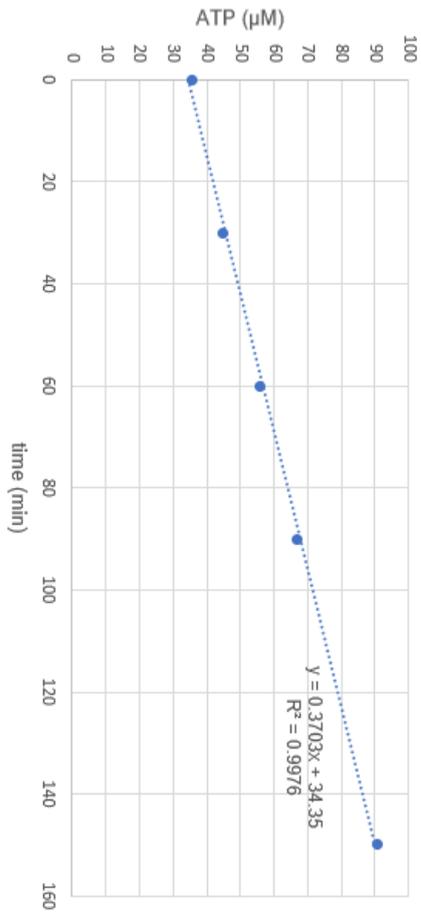
**Figure 58** HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.25 mM

0.5 mM														
	ATP Peak	ADP Peak	AMP Peak	Ade Peak	Total Peak Area	ATP %	ADP %	AMP %	Ade %	ATP (mM)	ADP (mM)	AMP (mM)	Ade (mM)	Time (min)
Peak Area	631.50201	2877.81860	143.56786	0.00000	3652.88847	17.288	78.782	3.9303	0	0.086439	0.39391	0.0196513	0	0
Peak Area	280.70154	2766.33740	179.89854	46.44432	3273.3818	8.5753	84.51	5.4958	1.419	0.042876	0.42255	0.027479	0.007094	30
Peak Area	287.40030	2632.65088	168.44139	94.05117	3182.54374	9.0305	82.722	5.2927	2.955	0.045153	0.413608	0.0264633	0.014776	60
Peak Area	322.75452	2574.95557	173.29083	145.34770	3216.34862	10.035	80.058	5.3878	4.519	0.050174	0.400292	0.0269391	0.022595	90
Peak Area	352.52692	2525.73828	174.86038	198.68150	3251.80708	10.841	77.672	5.3773	6.11	0.054205	0.388359	0.0268866	0.030549	120
Peak Area	387.74176	2427.28662	174.33319	248.76031	3238.12188	11.974	74.96	5.3838	7.682	0.059871	0.374799	0.0269189	0.038411	150



**Figure 59** HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 0.5 mM

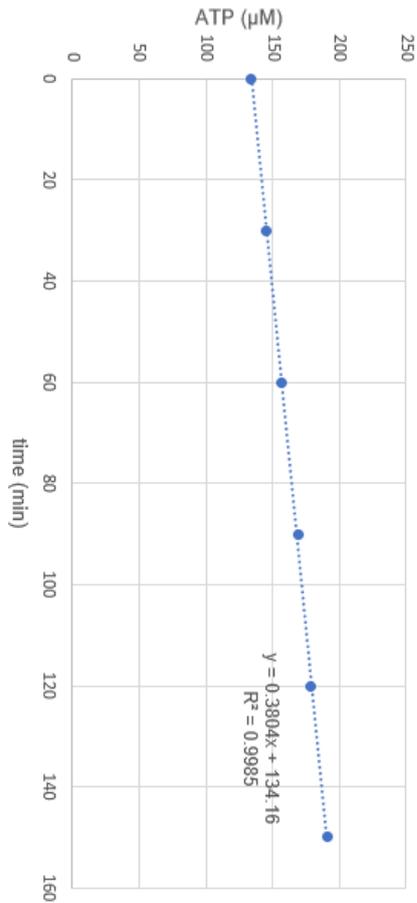
1 mM													
ATP Peak	ADP Peak	AMP Peak	Ade Peak	Total Peak Area	ATP %	ADP %	AMP %	Ade %	ATP (mM)	ADP (mM)	AMP (mM)	Ade (mM)	
224.06396	5757.79492	297.56708	0.00000	6279.42596	3.5682	91.693	4.7388	0	0.035682	0.91693	0.0473876	0	
Peak Area	281.81274	5628.73145	297.15482	88.34419	6296.0432	4.476	89.401	4.7197	1.403	0.04476	0.894011	0.0471971	0.014032
Peak Area	346.07324	5381.81934	290.46631	173.41228	6191.77117	5.5892	86.919	4.6912	2.801	0.055892	0.869189	0.0469117	0.028007
Peak Area	414.01202	5238.82031	287.58572	260.42975	6200.8478	6.6767	84.486	4.6378	4.2	0.066767	0.844855	0.0463785	0.041999
Peak Area	954.24634	5234.62500	290.10126	353.43524	6832.40784	13.966	76.615	4.246	5.173	0.139665	0.766146	0.0424596	0.051729
Peak Area	587.91211	5142.22510	296.95224	443.54993	6470.63938	9.0858	79.47	4.5892	6.855	0.090858	0.794701	0.0458923	0.068548
													Time (min)
													0
													30
													60
													90
													120
													150



**Figure 60** HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 1 mM



10 mM														
	ATP Peak	ADP Peak	AMP Peak	Ade Peak	Total Peak Area	ATP %	ADP %	AMP %	Ade %	ATP (mM)	ADP (mM)	AMP (mM)	Ade (mM)	Time (min)
Peak Area	808.19220	56314.90000	3003.35034	0.00000	60126.44254	1.3442	93.661	4.9951	0	0.134415	9.366079	0.4995057	0	0
Peak Area	865.34131	55498.80000	2569.16626	783.26819	59716.57576	1.4491	92.937	4.3023	1.312	0.144908	9.293701	0.4302267	0.131164	30
Peak Area	947.51514	55693.60000	2304.35620	1502.63296	60448.1043	1.5675	92.135	3.8121	2.486	0.156749	9.213457	0.3812123	0.248582	60
Peak Area	1018.67322	54792.70000	2066.02856	2089.21362	59966.6154	1.6987	91.372	3.4453	3.484	0.169873	9.137201	0.3445298	0.348396	90
Peak Area	1101.63000	55806.30000	1964.41077	2694.45239	61566.79316	1.7893	90.644	3.1907	4.376	0.178932	9.064351	0.3190699	0.437647	120
Peak Area	1175.90637	55276.30000	1819.85022	3207.45142	61479.50801	1.9127	89.91	2.9601	5.217	0.191268	8.991012	0.2960092	0.521711	150



**Figure 62** HPLC stop-assay for ADP phosphorylation. Integration values are used to calculate the concentration of all the species involved in the reaction. The graph shows ATP production. ADP concentration = 10 mM.

**Table 10** Best fit values for Michaelis-Menten kinetics for YihX phosphorylation of ADP.

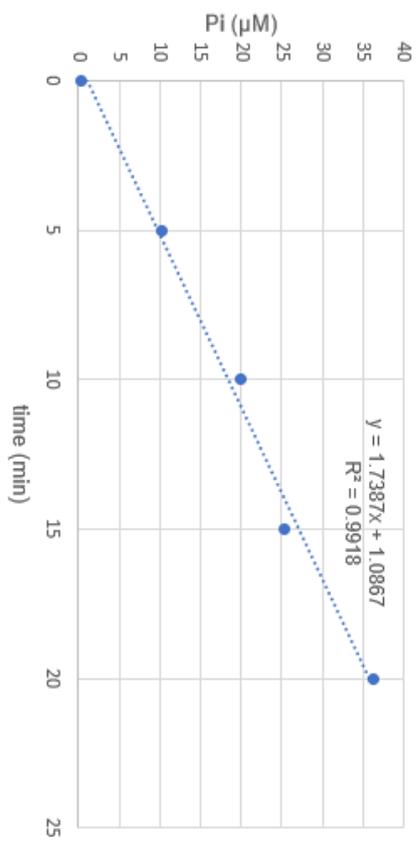
<b>Michaelis-Menten Best-fit values</b>	
Vmax	0.4491
Km	723.7
Std. Error	
Vmax	0.03567
Km	199.0
<b>95% CI (profile likelihood)</b>	
Vmax	0.3829 to 0.5223
Km	440.5 to 1193
<b>Goodness of Fit</b>	
Degrees of Freedom	19
R squared	0.8719
Sum of Squares	0.07789
Sy.x	0.06403
<b>Constraints</b>	
Km	Km > 0
<b>Number of points</b>	
# of X values	21
# Y values analyzed	21

## 7.2 Malachite Green Phosphate assay for $\alpha$ G1P hydrolysis

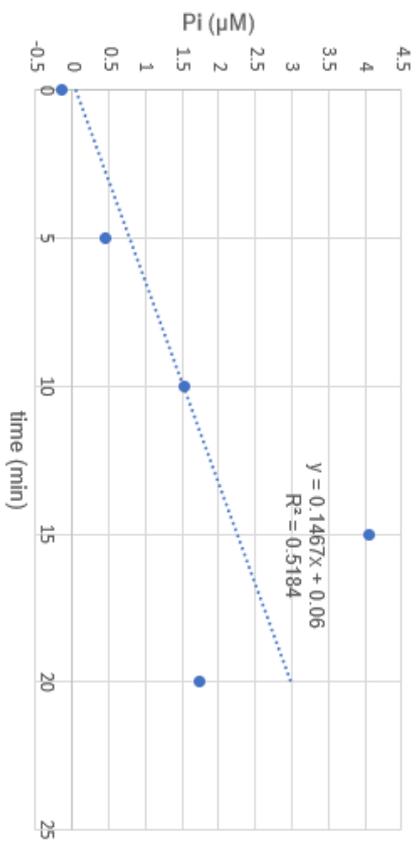
**Table 11** Absorbance values and time points for  $\alpha$ G1P hydrolysis catalysed by YihX during a Malachite Green Phosphate assay. Concentration of substrate varies from 0.005 mM to 5 mM. Inorganic phosphate concentration ( $\mu$ M) is plotted against time.

<b>5 mM</b>			<b>2 mM</b>		
<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>	<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>
0	0.113	-0.34	0	0.098	-1.34
5	0.264	9.72667	5	0.205	5.79333
10	0.433	20.9933	10	0.342	14.9267
15	0.632	34.26	15	0.485	24.46
20	0.793	44.9933	20	0.621	33.5267
<b>0.1 mM</b>			<b>0.05 mM</b>		
<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>	<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>
0	0.124	0.39333	0	0.113	-0.34
5	0.273	10.3267	5	0.186	4.52667
10	0.417	19.9267	10	0.255	9.12667
15	0.499	25.3933	15	0.282	10.9267
20	0.663	36.3267	20	0.356	15.86
<b>1 mM</b>			<b>0.5 mM</b>		
<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>	<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>
0	0.081	-2.4733	0	0.081	-2.4733
5	0.167	3.26	5	0.141	1.52667
10	0.283	10.9933	10	0.232	7.59333
15	0.396	18.5267	15	0.312	12.9267
20	0.494	25.06	20	0.381	17.5267
<b>0.01 mM</b>			<b>0.005 mM</b>		
<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>	<b>time (min)</b>	<b>Absorbance</b>	<b>Pi (<math>\mu</math>M)</b>
0	0.116	-0.14	0	0.11	-0.54
5	0.125	0.46	5	0.121	0.19333
10	0.141	1.52667	10	0.124	0.39333
15	0.179	4.06	15	0.133	0.99333
20	0.144	1.72667	20	0.172	3.59333

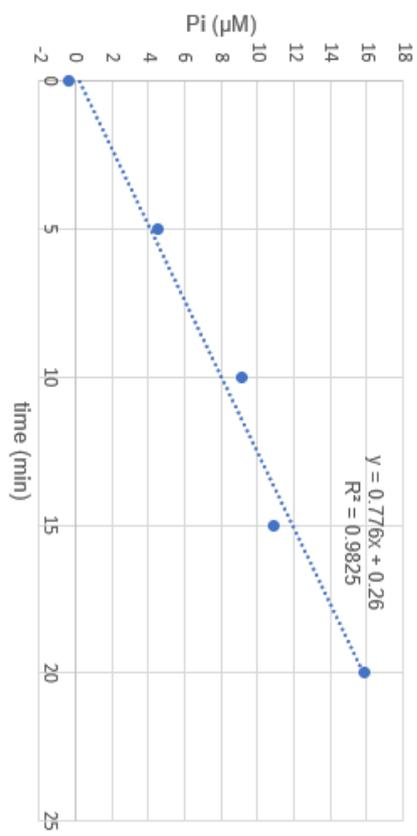
0.1 mM G1P



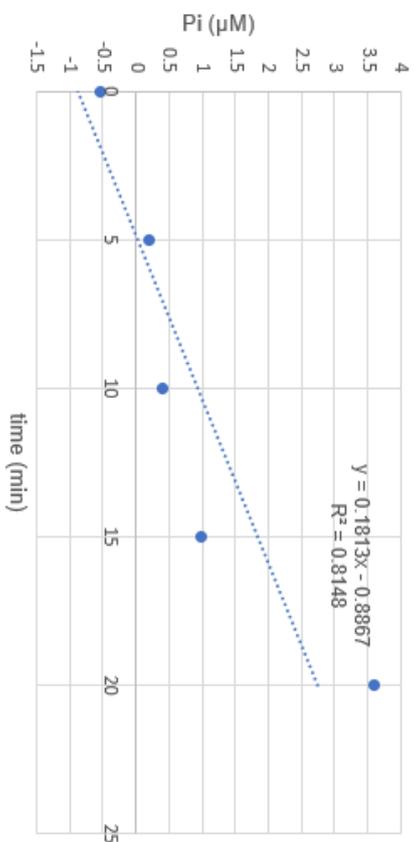
0.01 mM G1P

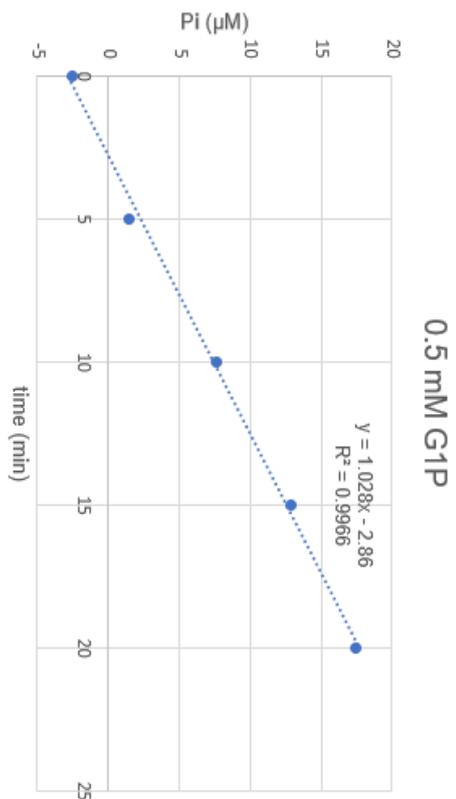
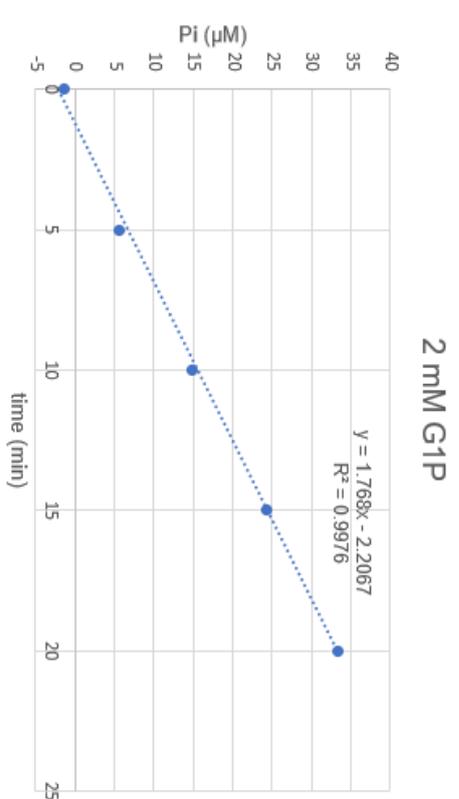
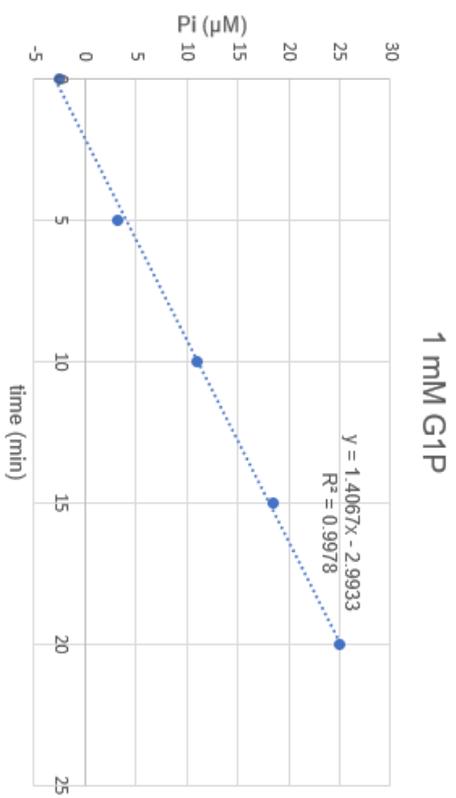
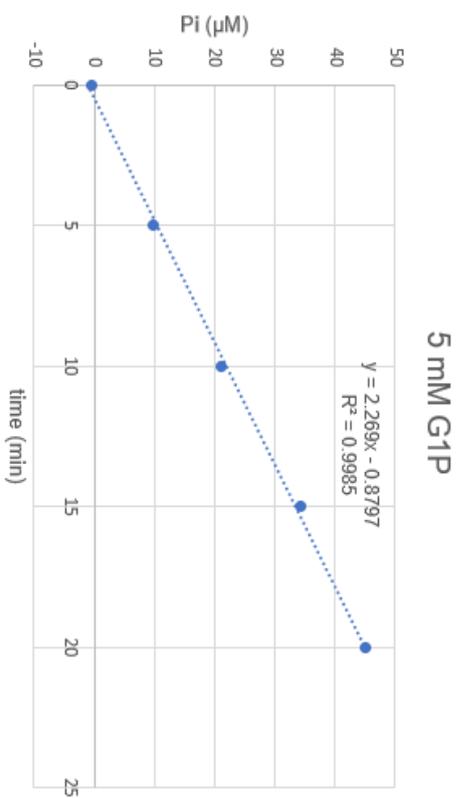


0.05 mM G1P



0.005 mM G1P





**Table 12** Best fit values for Michaelis-Menten kinetics for YihX hydrolysis of  $\alpha$ G1P.

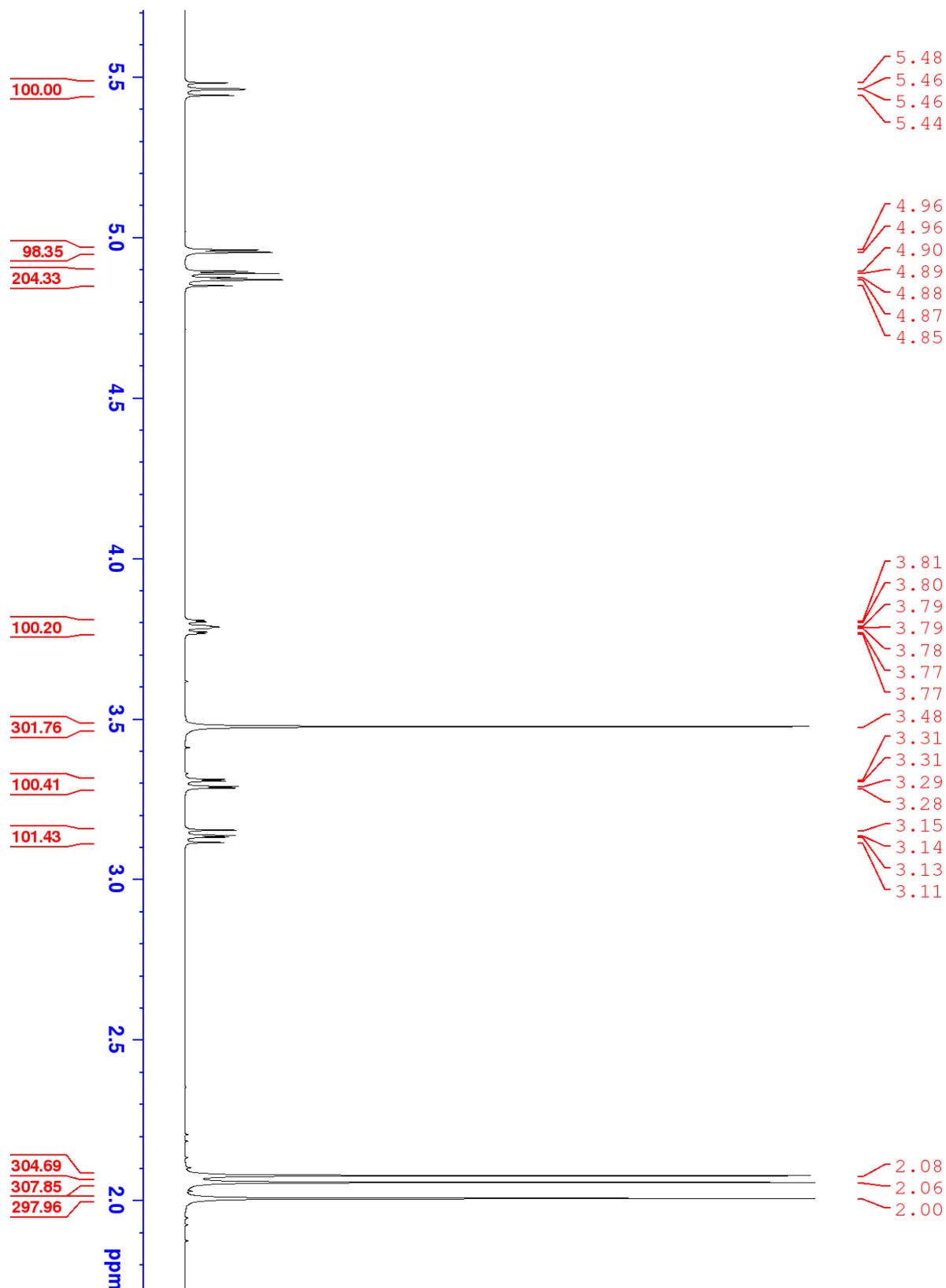
<b>Michaelis-Menten</b>	
<b>Best-fit values</b>	
Vmax	26.92
Km	1034
Std. Error	
Vmax	0.8994
Km	96.41
<b>95% CI (profile likelihood)</b>	
Vmax	25.10 to 28.95
Km	847.4 to 1260
<b>Goodness of Fit</b>	
Degrees of Freedom	14
R squared	0.9933
Sum of Squares	7.210
Sy.x	0.7177
<b>Constraints</b>	
Km	Km > 0
<b>Number of points</b>	
# of X values	24
# Y values analyzed	16

## 7.3 NMR spectra

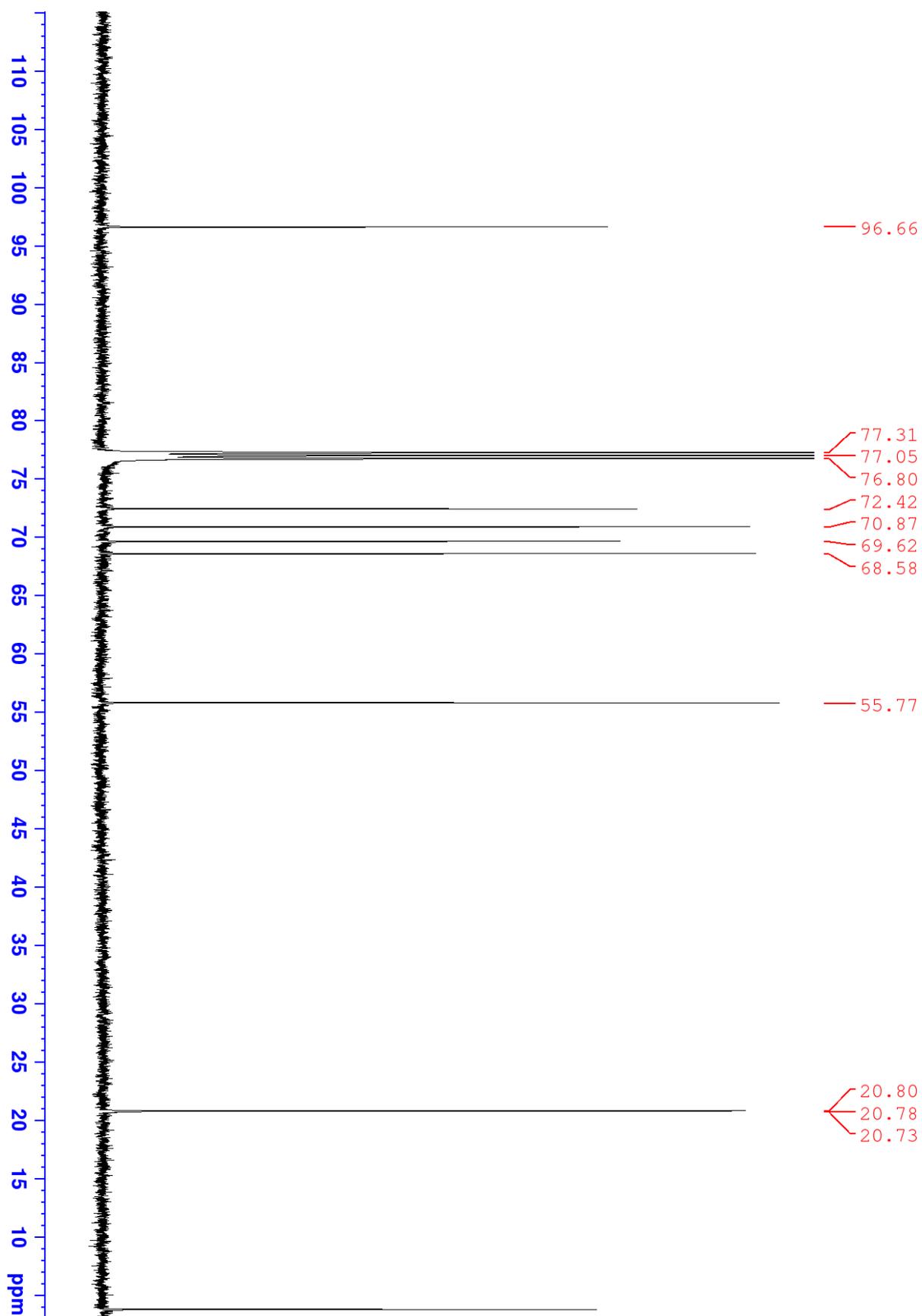
### 7.3.1 SQ synthesis spectra

#### 7.3.1.1 Methyl 2,3,4-tri-O-acetyl-6-deoxy-6-iodo- $\alpha$ -D-glucopyranoside (2):

$^1\text{H}$  NMR

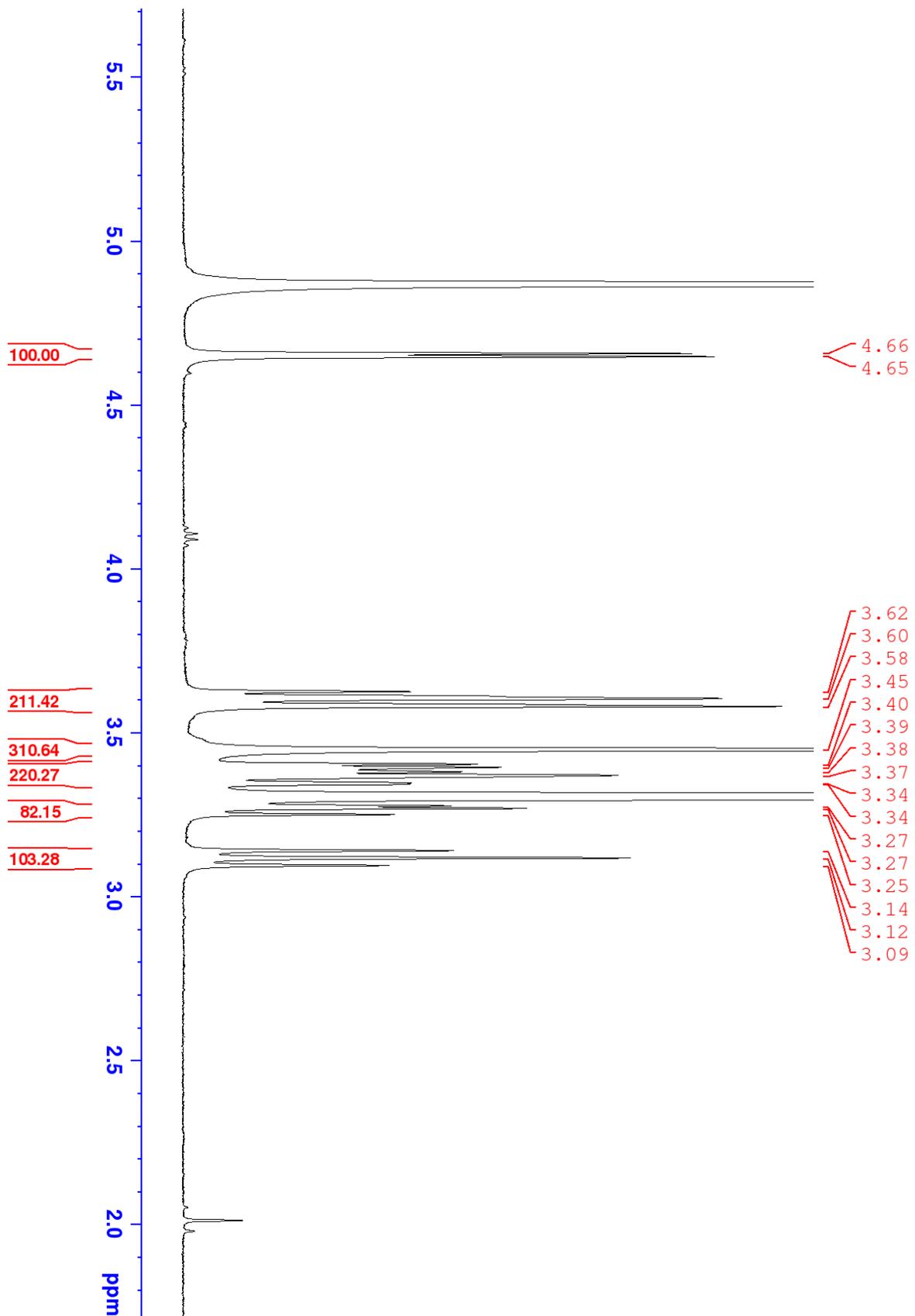


<sup>13</sup>C NMR

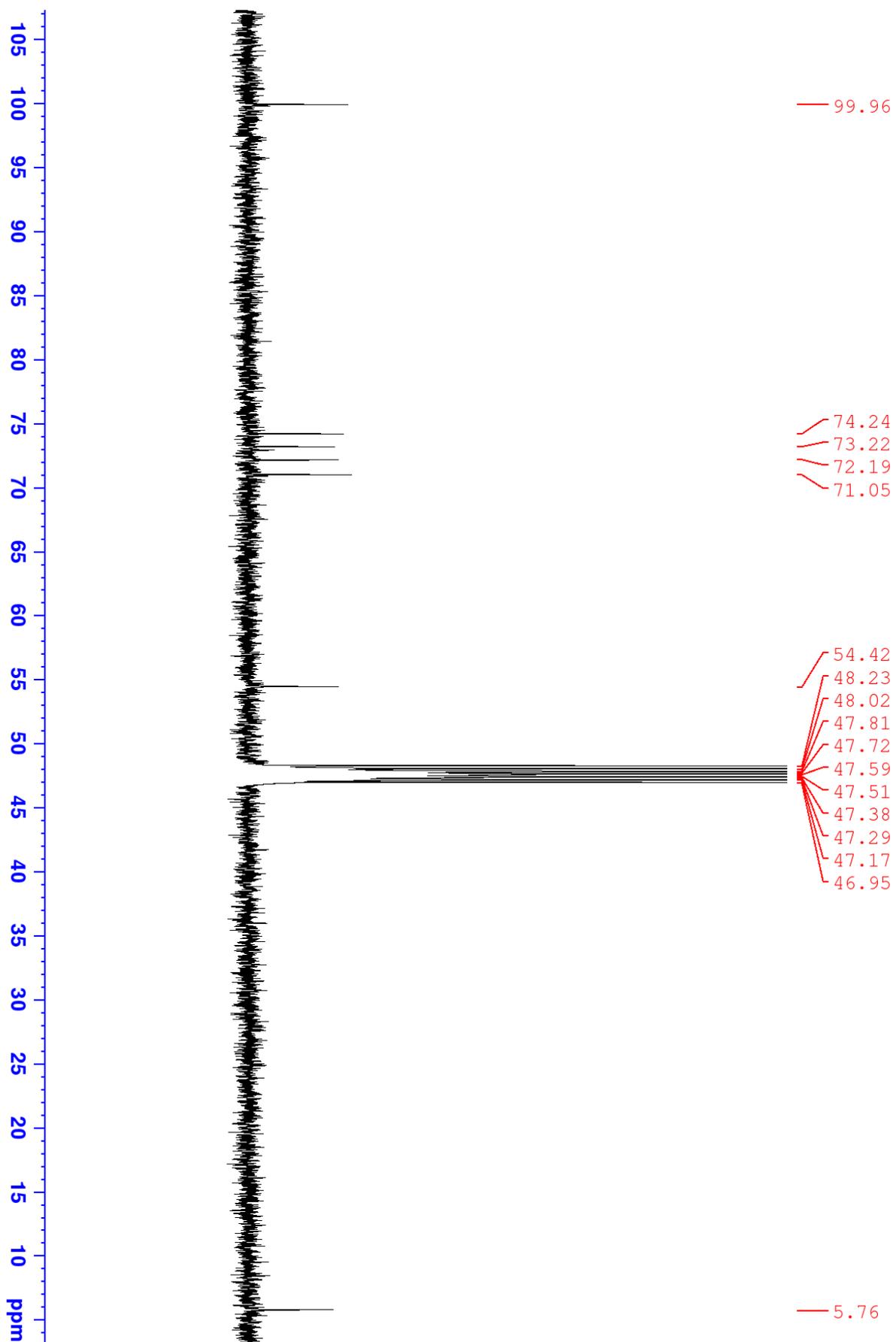


### 7.3.1.2 Methyl 6-deoxy-6-iodo- $\alpha$ -D-glucopyranoside (3):

$^1\text{H}$  NMR

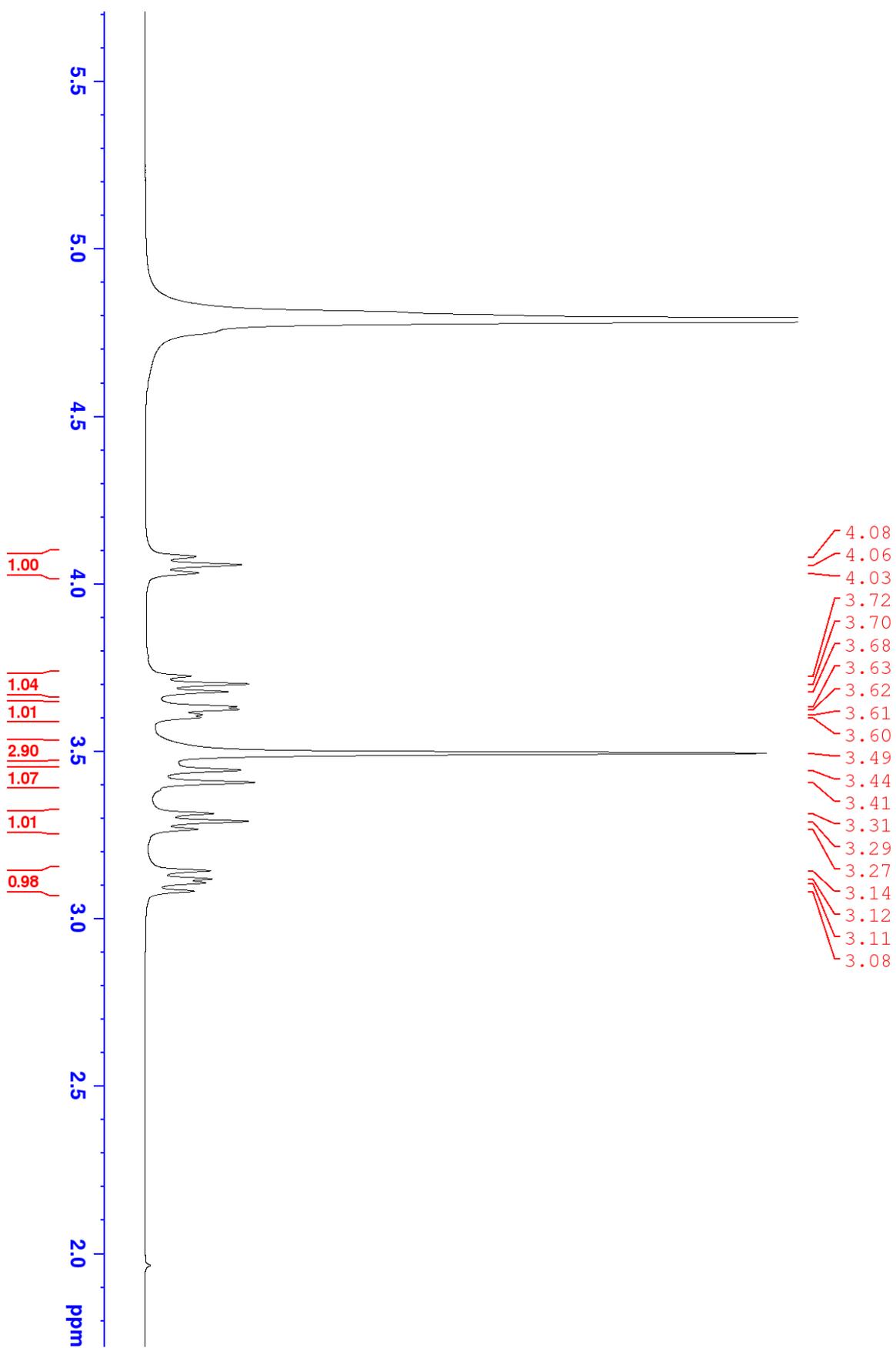


$^{13}\text{C}$  NMR

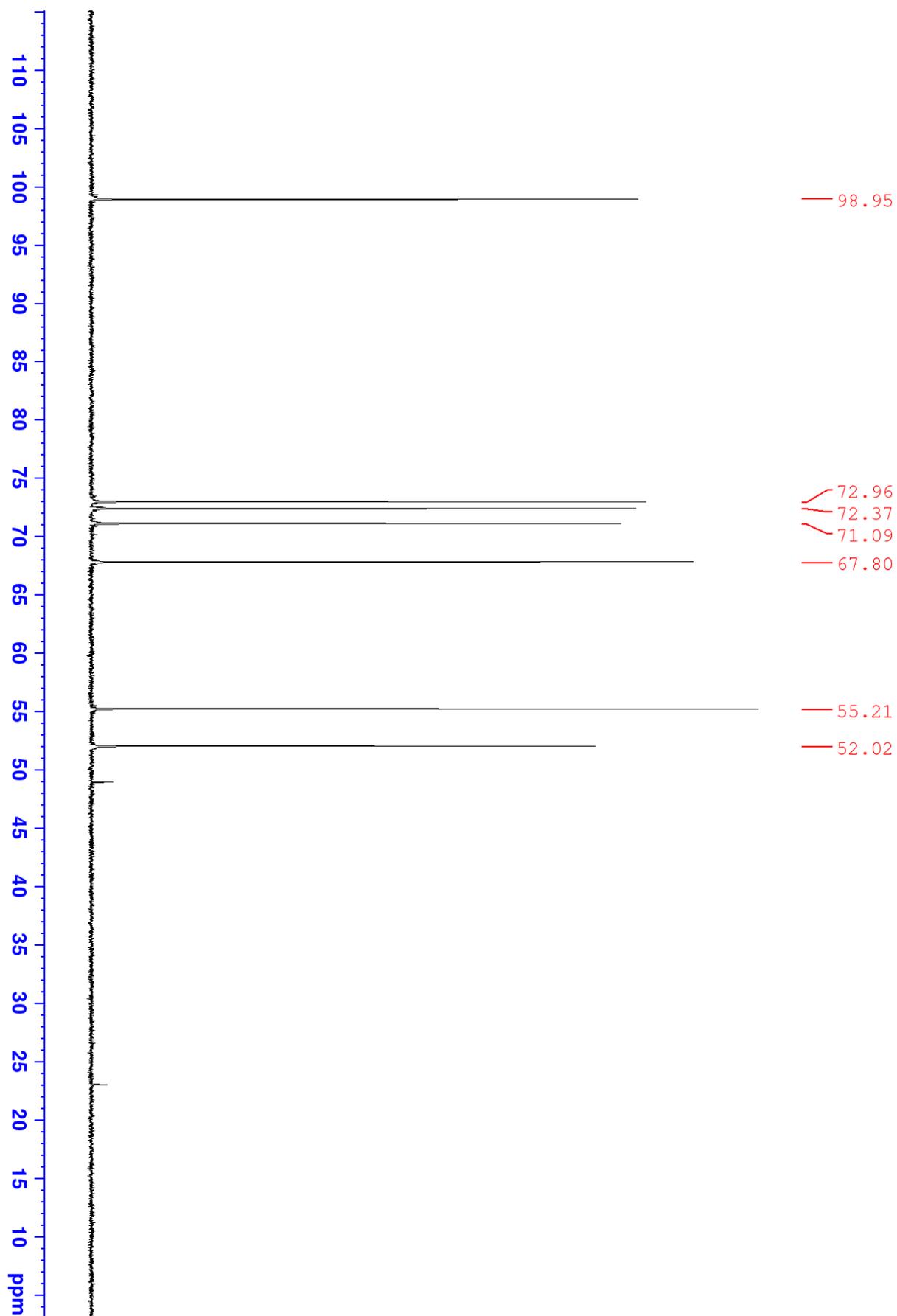


### 7.3.1.3 Methyl 6-deoxy-6-sulfonato- $\alpha$ -D-glucopyranoside (4):

$^1\text{H}$  NMR

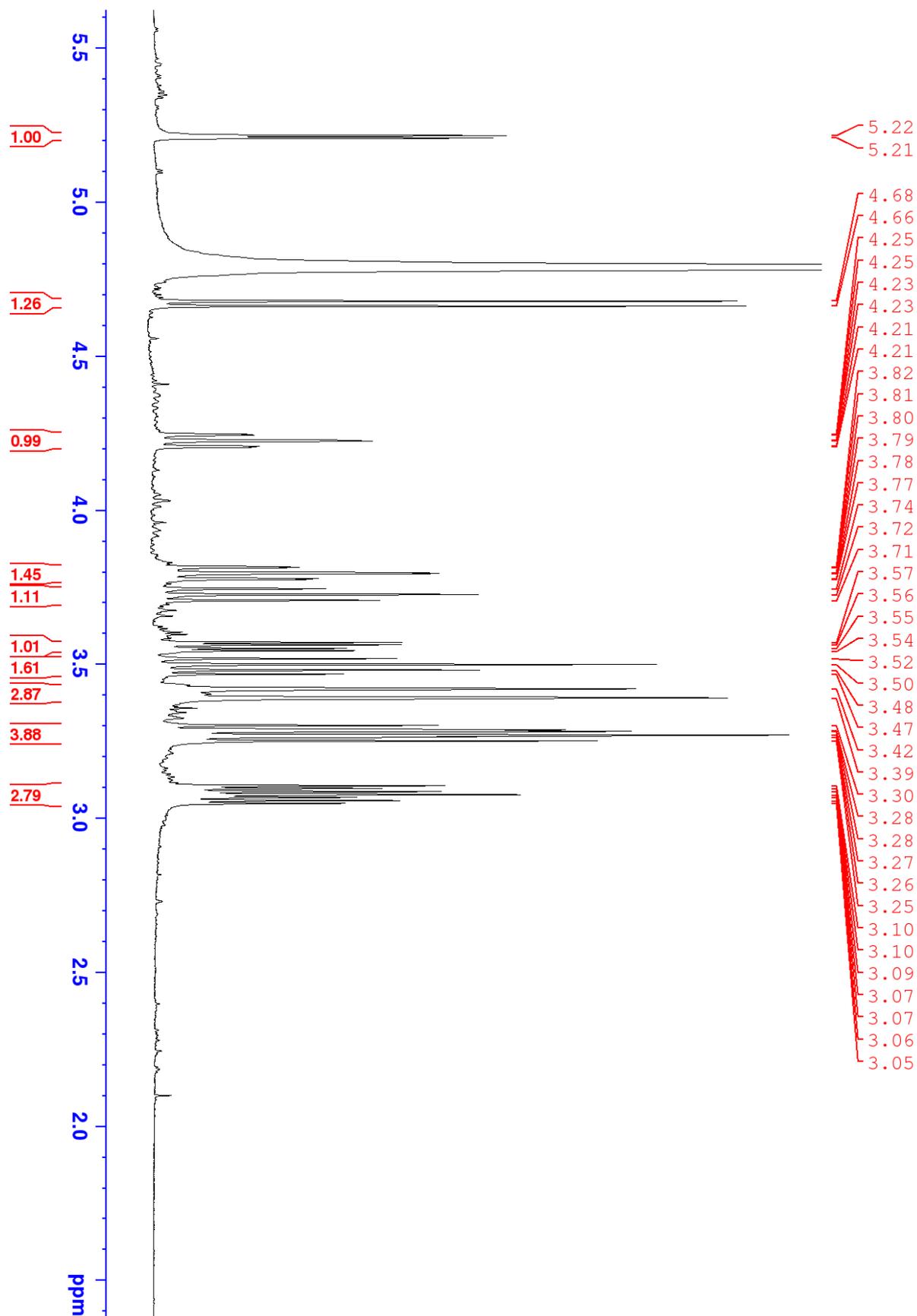


<sup>13</sup>C NMR

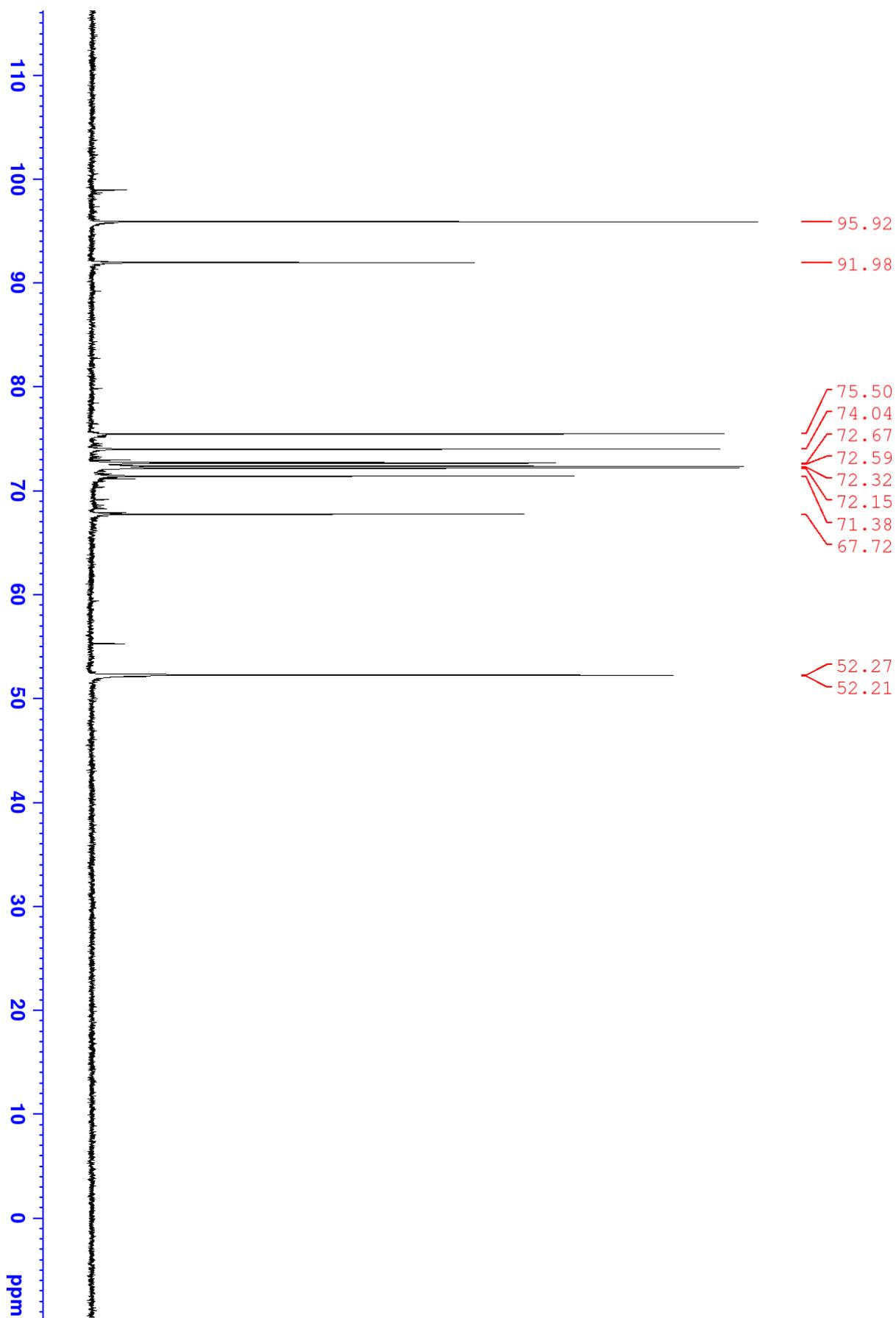


### 7.3.1.4 Sodium 6-deoxy-6-sulfonato- $\alpha$ -D-glucopyranoside (1):

$^1\text{H}$  NMR

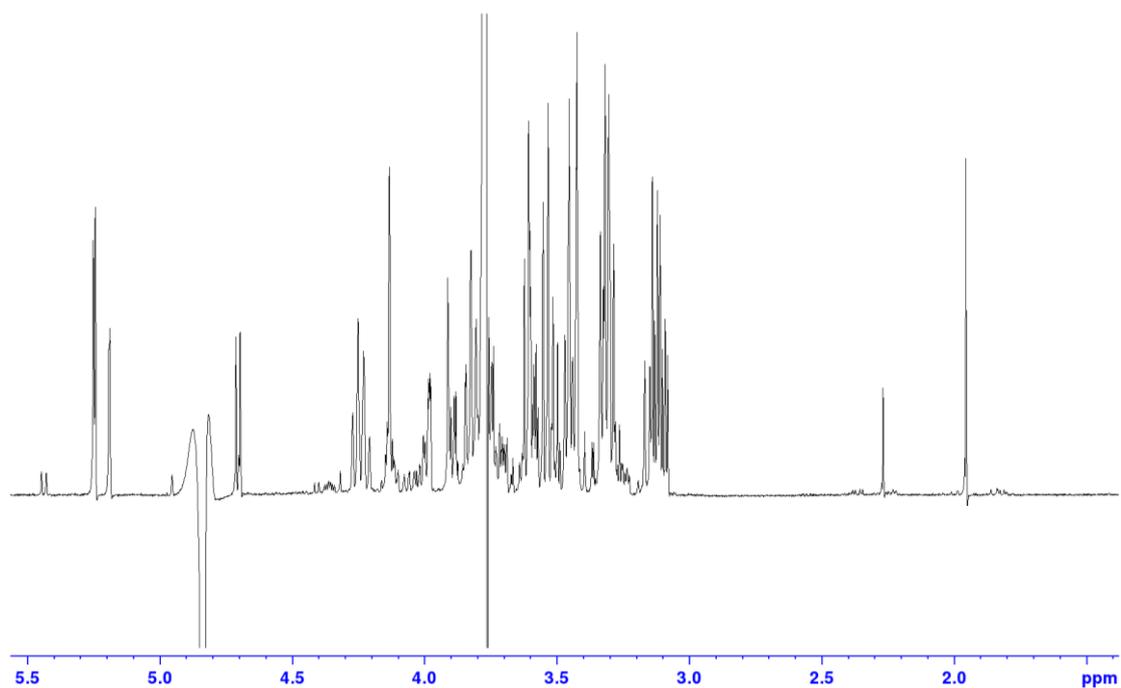


$^{13}\text{C}$  NMR



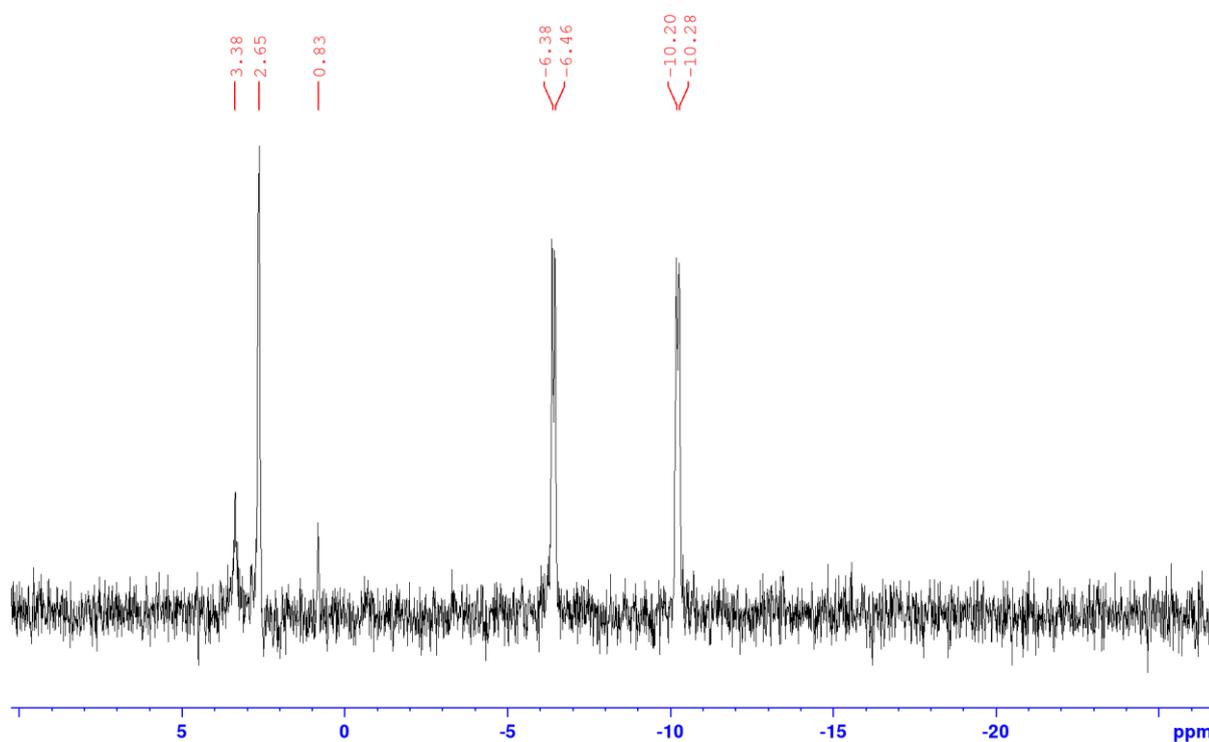
## 7.3.2 SFP synthesis spectra

### 7.3.2.1 $^1\text{H}$ NMR of YihS isomerisation of SQ



SQ: 20 mM, YihS: 50  $\mu\text{M}$ , Buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.0), 10%  $\text{D}_2\text{O}$

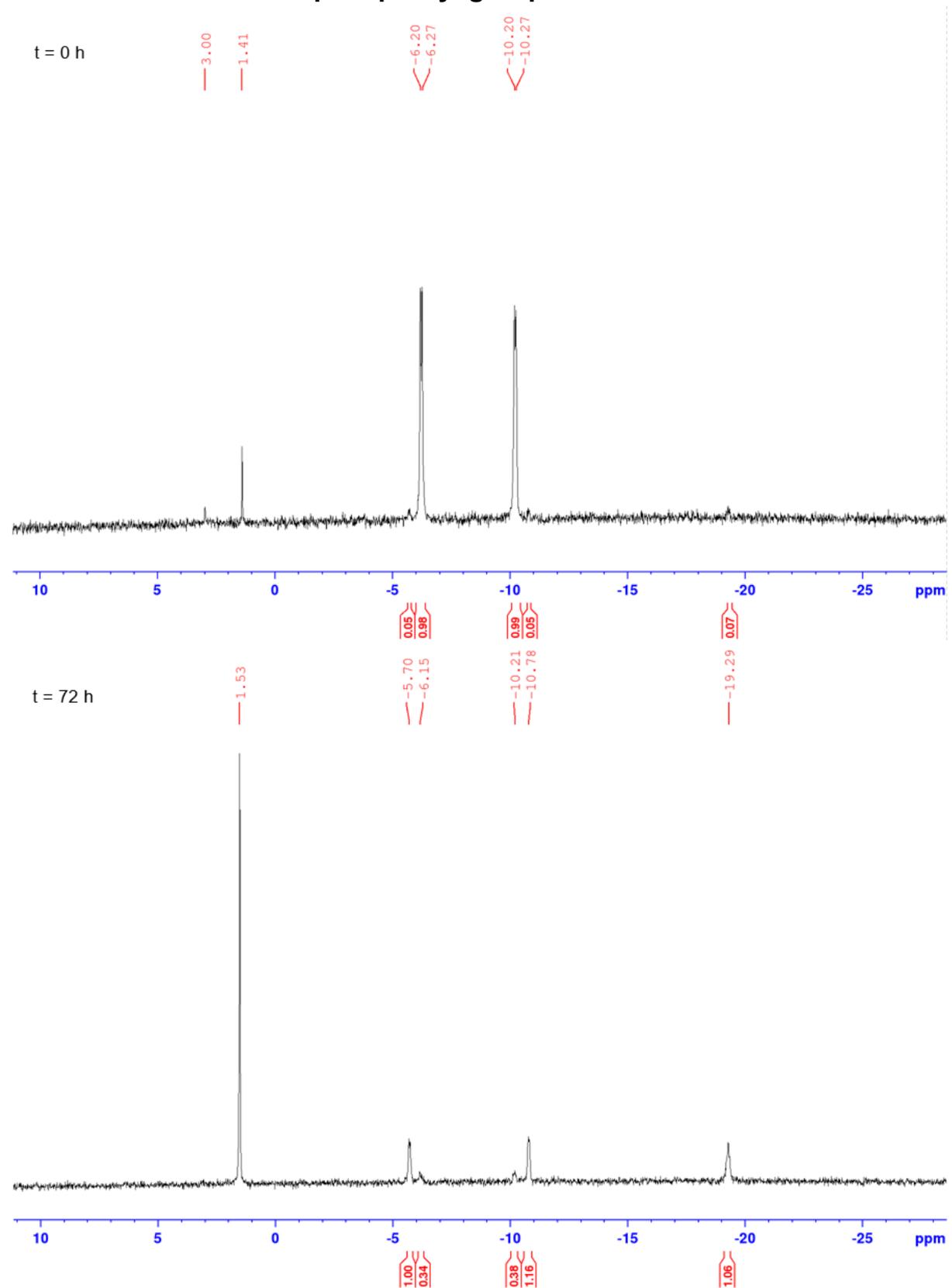
### 7.3.2.2 $^{31}\text{P}$ NMR of YihV phosphorylation



YihS reaction mixture +  $\text{MgCl}_2$ : 50 mM, YihV: 0.15 mg/mL, ATP: 25 mM, Buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.0), 10%  $\text{D}_2\text{O}$

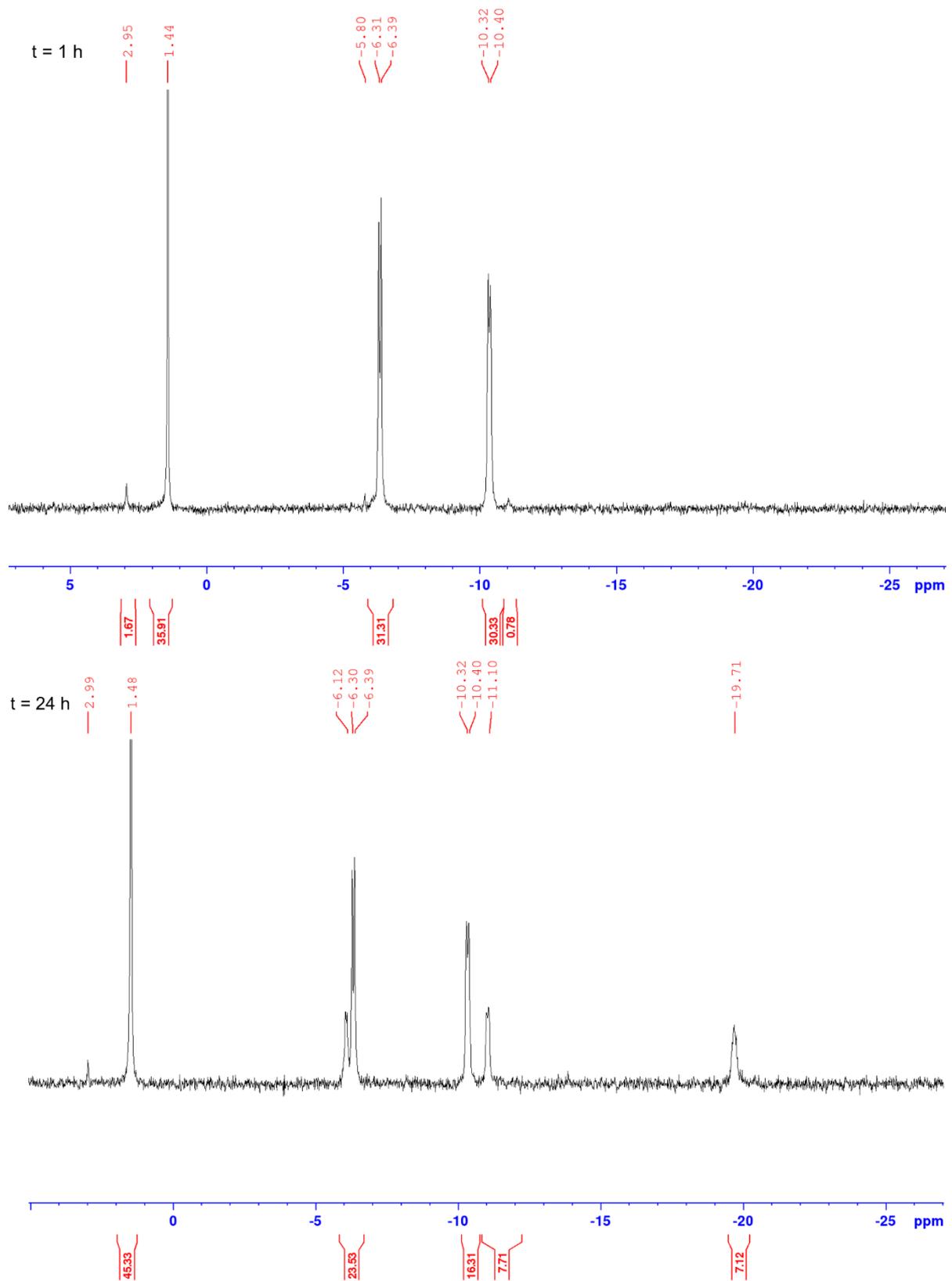
### 7.3.3 YihX ADP synthesis spectra

#### 7.3.3.1 $^{31}\text{P}$ NMR ADP as phosphoryl group donor



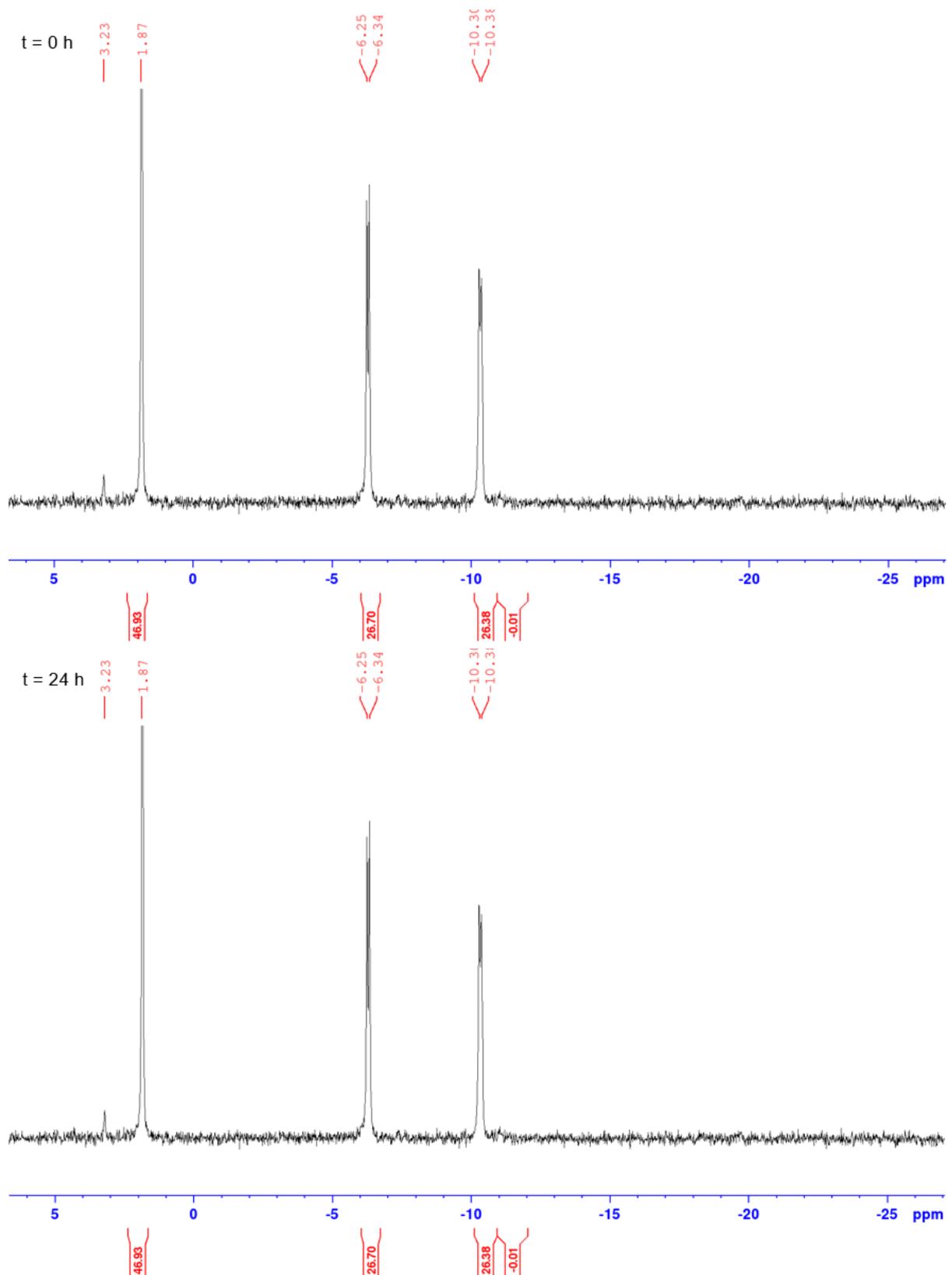
ADP: 10 mM, YihX: 100  $\mu\text{M}$ ,  $\text{MgCl}_2$ : 50 mM, Buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.0), 10%  $\text{D}_2\text{O}$ . Time = 0 h (top) and 72 h (bottom)

### 7.3.3.2 $^{31}\text{P}$ NMR $\text{P}_i$ as phosphoryl group donor



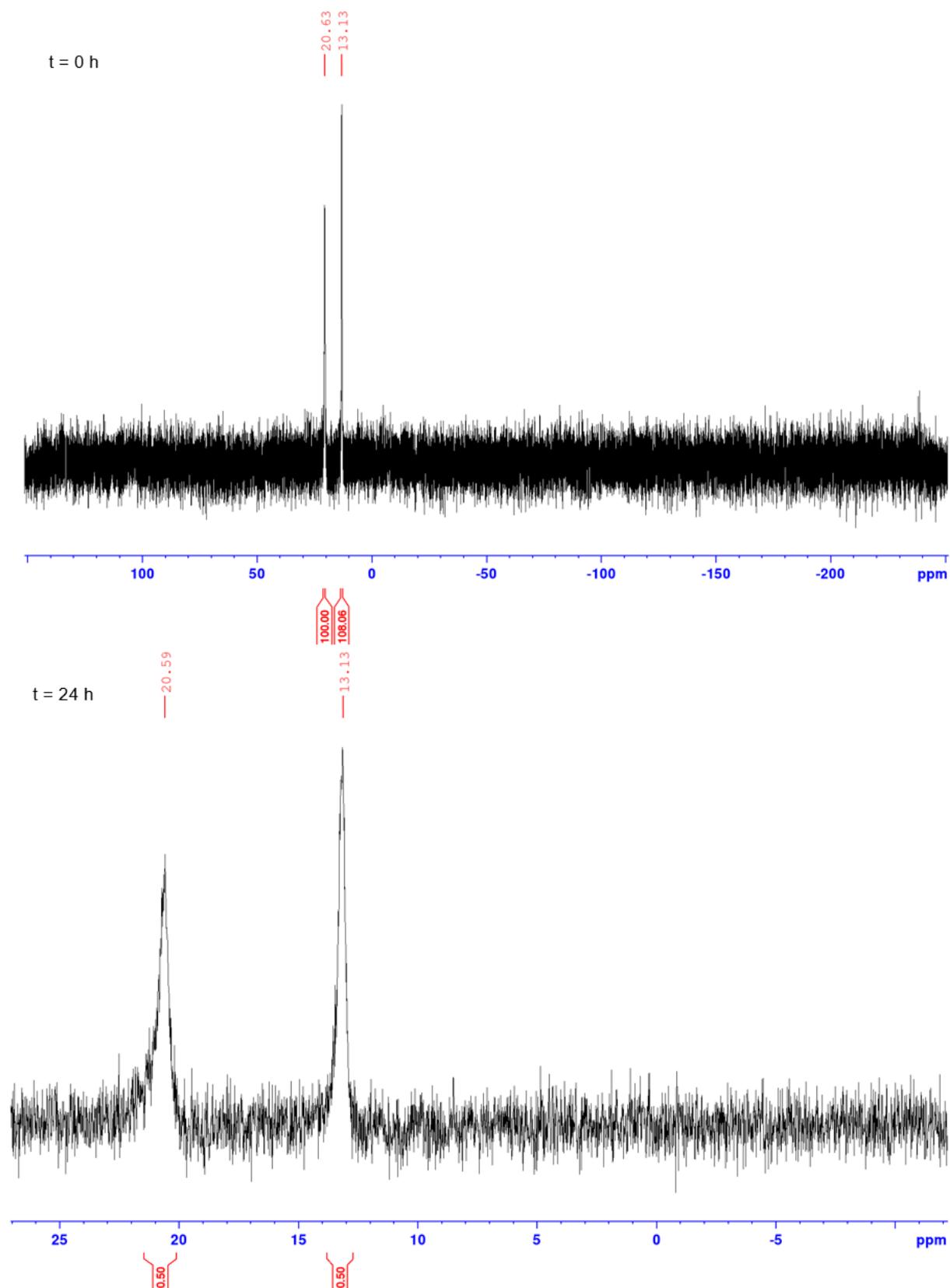
ADP: 10 mM,  $\text{P}_i$ : 10 mM, YihX: 100  $\mu\text{M}$ ,  $\text{MgCl}_2$ : 50 mM, Buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.0), 10%  $\text{D}_2\text{O}$ . Time = 0 h (top) and 24 h (bottom)

### 7.3.3.3<sup>31</sup>P NMR αG1P as phosphoryl group donor



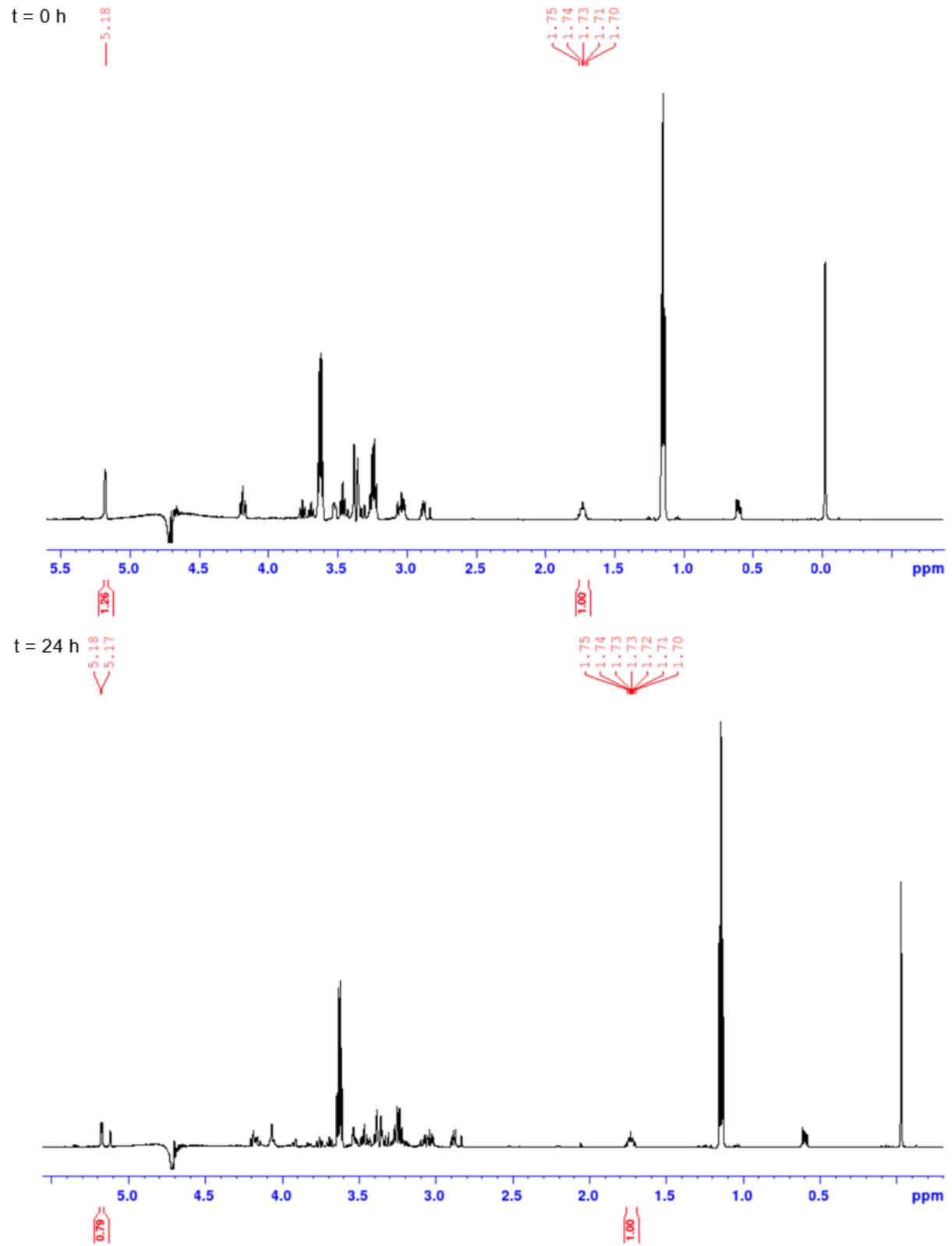
ADP: 10 mM, αG1P: 10 mM, YihX: 100 μM, MgCl<sub>2</sub>: 50 mM, Buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.0), 10% D<sub>2</sub>O. Time = 0 h (top) and 24 h (bottom)

### 7.3.3.4 $^{31}\text{P}$ NMR AMP-CH<sub>2</sub>-P non hydrolysable analogue



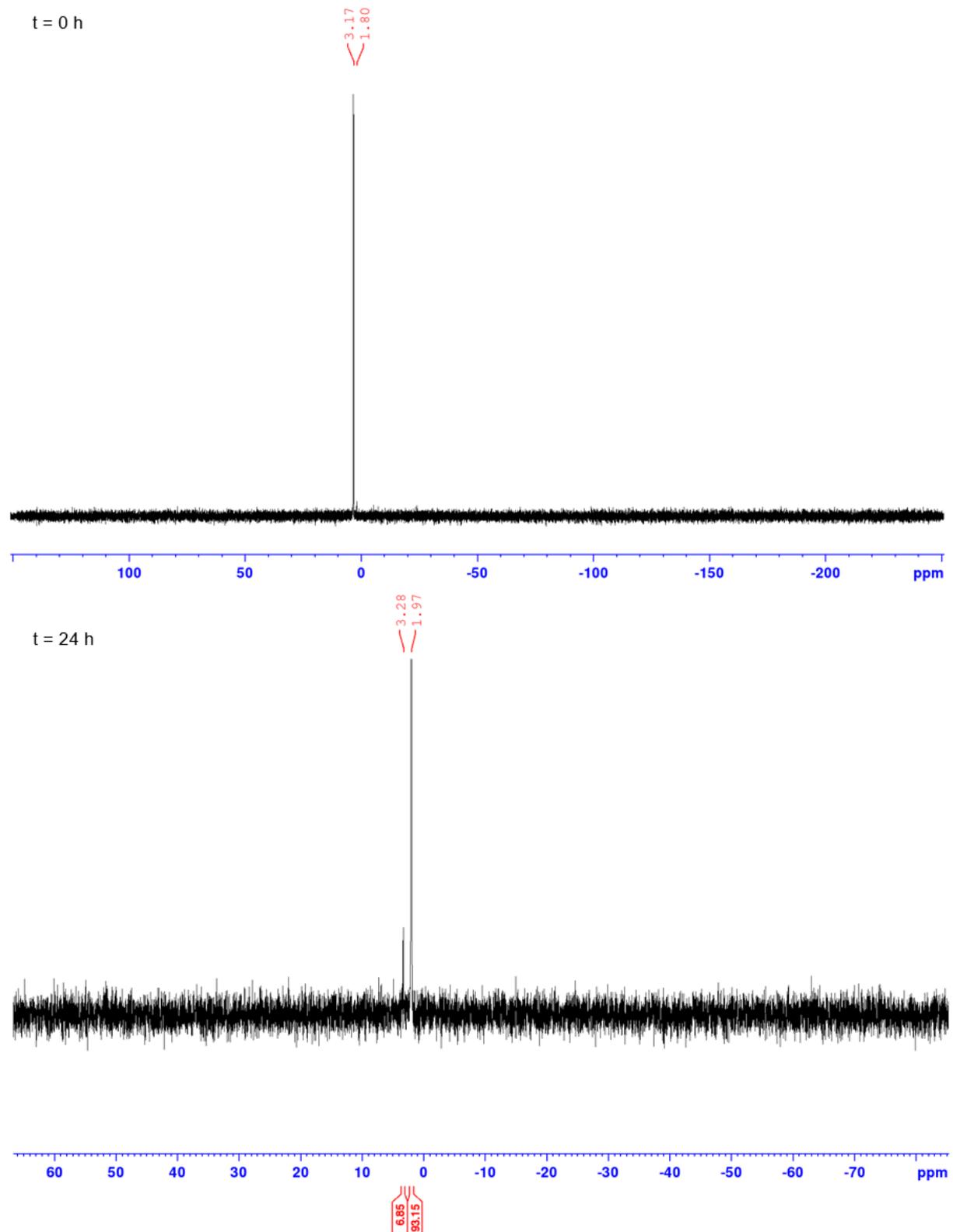
AMP-CH<sub>2</sub>-P: 10 mM, YihX: 100  $\mu\text{M}$ , MgCl<sub>2</sub>: 50 mM, Buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.0), 10% D<sub>2</sub>O. Time = 0 h (top) and 24 h (bottom)

### 7.3.4 $^1\text{H}$ NMR whole-cell NMR assay



SQ-M9: 5 mM (SQ), 10%  $\text{D}_2\text{O}$ . Time = 0 h (top) and 24 h (bottom)

### 7.3.5 $^{31}\text{P}$ NMR AMP hydrolysis



AMP: 10 mM, YihX: 50  $\mu\text{M}$ ,  $\text{MgCl}_2$ : 50 mM, Buffer (50 mM Tris-HCl, 200 mM NaCl, pH 7.0), 10%  $\text{D}_2\text{O}$ . Time = 0 h (top) and 24 h (bottom)

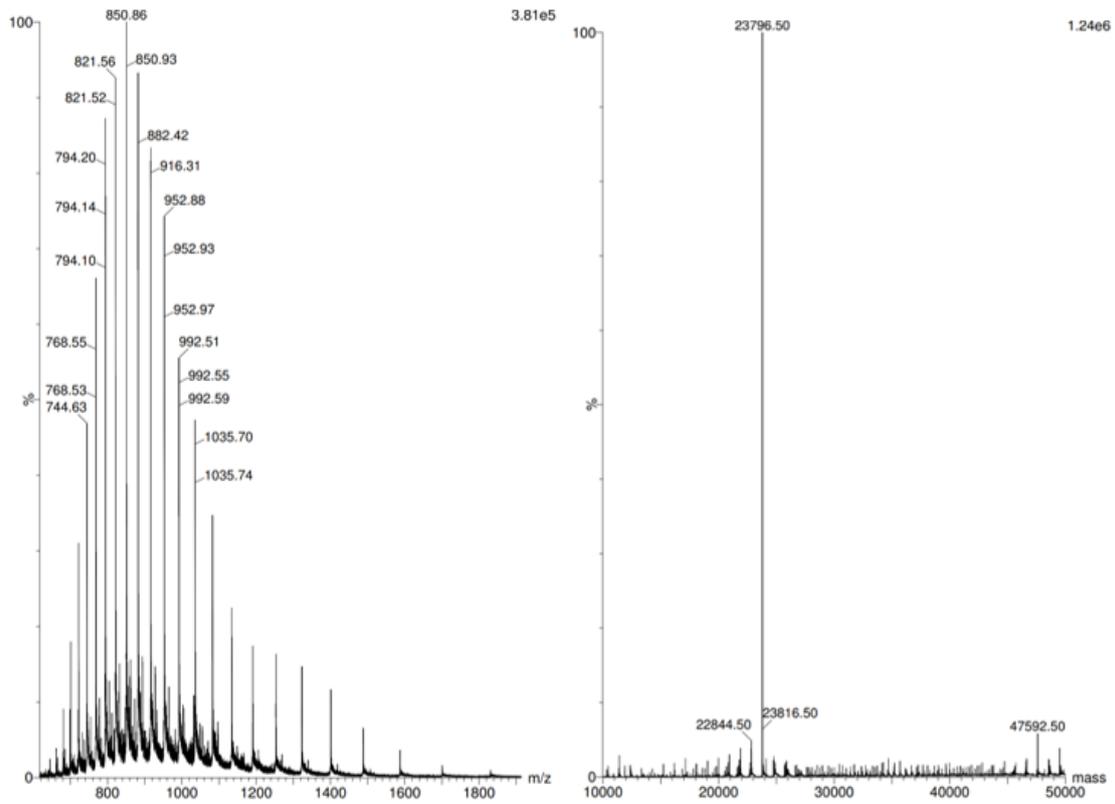
## 7.4 Extended data for YihV whole-cell NMR assay

**Table 13** Integration values of SQ peak during whole-cell NMR time-course.

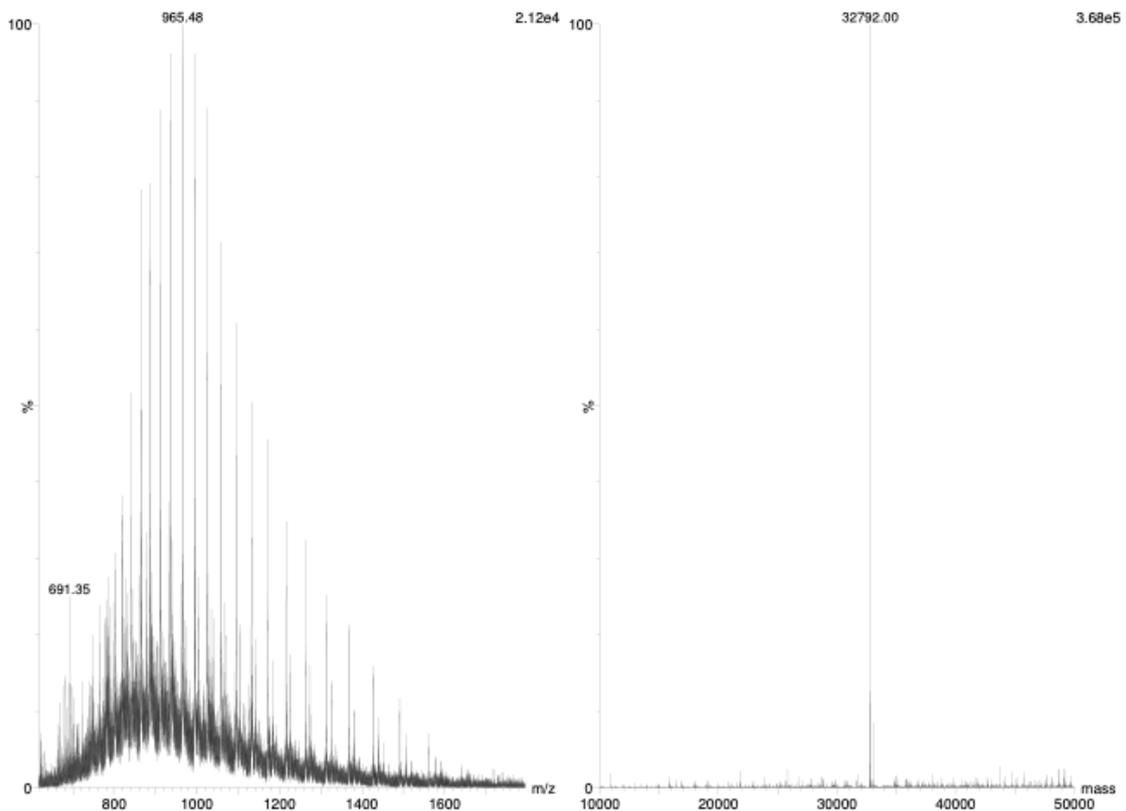
	YihV (1)	YihV (2)	Control (1)	Control (2)
time (min)	SQ peak	SQ peak	SQ peak	SQ peak
0	1.2743	1.2551	1.2798	1.2869
60	1.1865	1.1955	1.1684	1.1767
120	1.1335	1.1395	1.0957	1.1319
180	1.0597	1.0988	1.0264	1.0582
240	1.031	1.0685	0.9721	1.0237
300	1.002	1.0166	0.9145	0.9662
360	0.9908	1.0082	0.8204	0.9086
420	0.938	0.9748	0.7583	0.8763
480	0.9599	0.9699	0.6801	0.8396
540	0.9447	0.9444	0.5883	0.8037
600	0.9073	0.939	0.5567	0.7586
660	0.9097	0.9262	0.482	0.7164
720	0.9104	0.8961	0.433	0.7065
780	0.8657	0.8963	0.4086	0.6783
840	0.888	0.8921	0.3658	0.6526
900	0.8569	0.8832	0.3441	0.6191
960	0.8599	0.8309	0.3265	0.5796
1020	0.8387	0.8628	0.2832	0.5537
1080	0.8537	0.8471	0.2486	0.5427
1140	0.8281	0.8407	0.2482	0.5053
1200	0.8247	0.8084	0.2387	0.4958
1260	0.8324	0.8246	0.2187	0.4611
1320	0.8298	0.8025	0.2051	0.4501
1380	0.8302	0.7893	0.216	0.4229

## 7.5 Mass Spectrometry

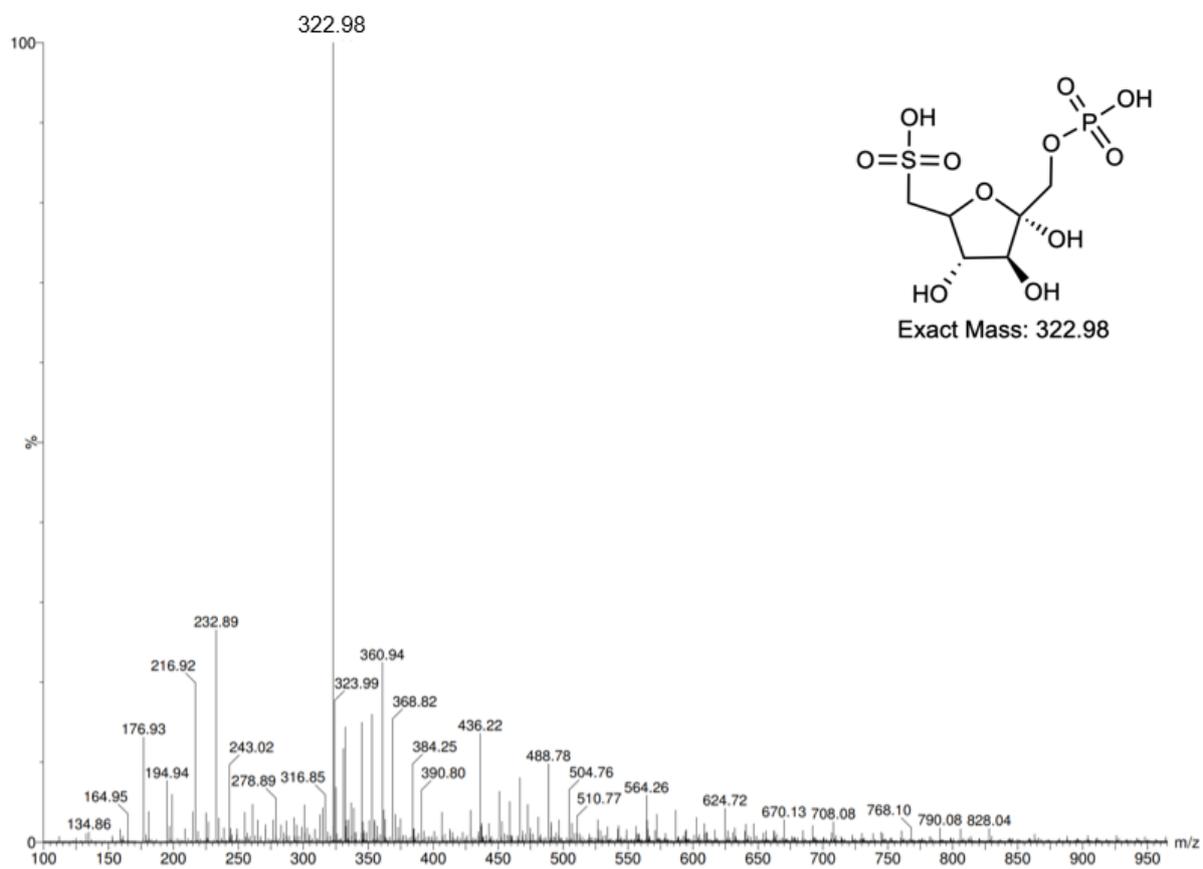
### 7.5.1 YihX protein



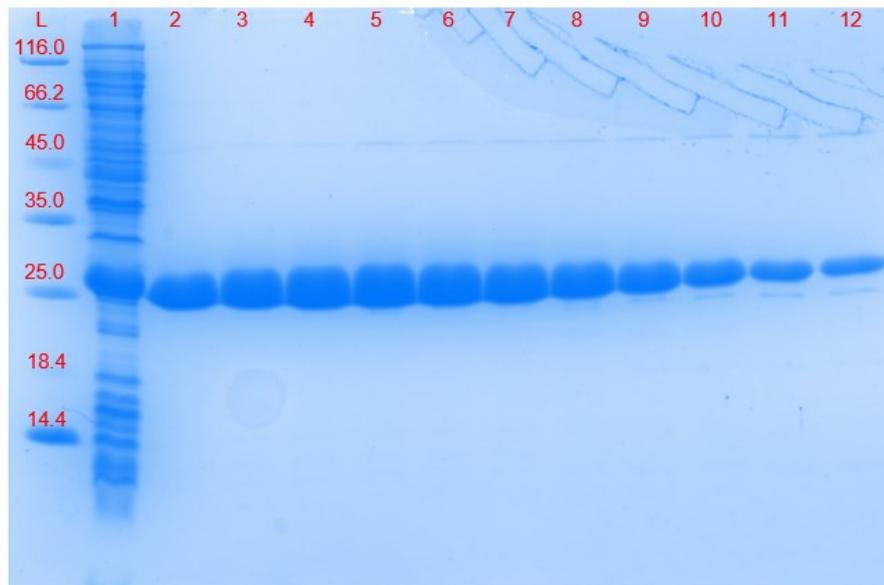
### 7.5.2 YihV protein



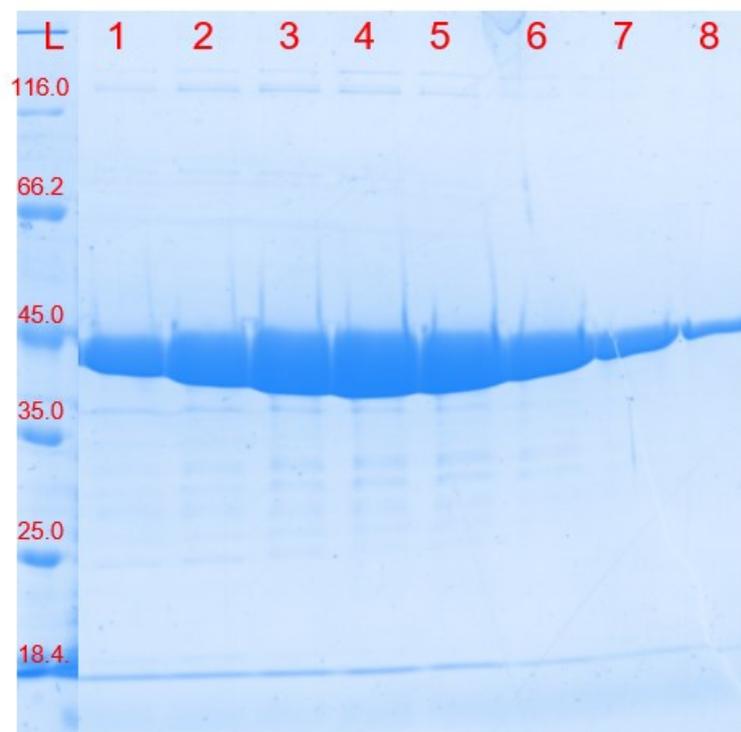
### 7.5.3 SFP



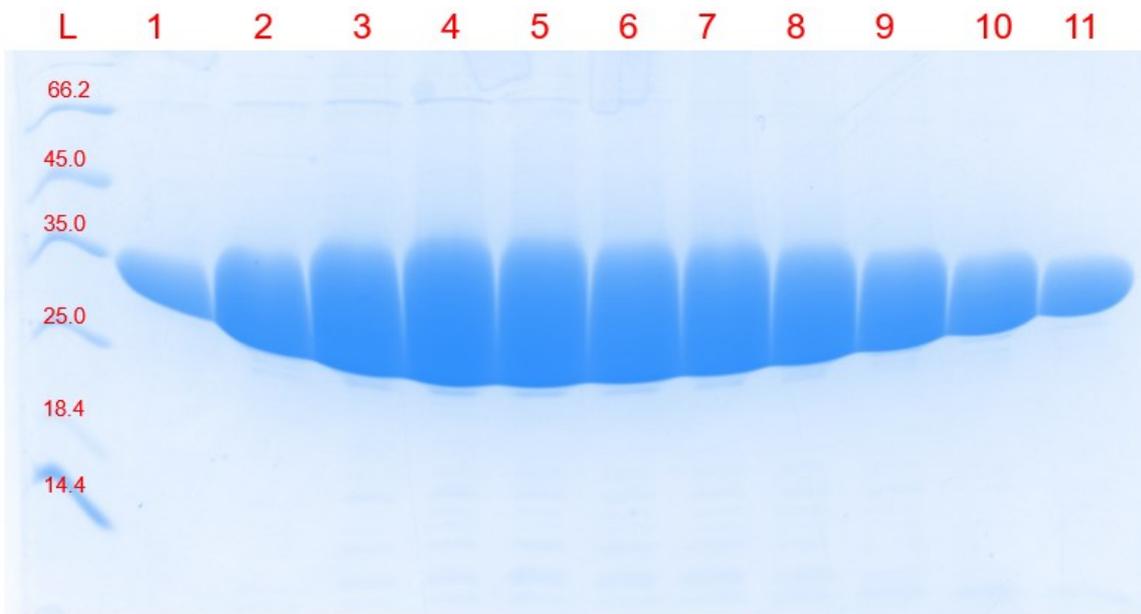
## 7.6 SDS-PAGE



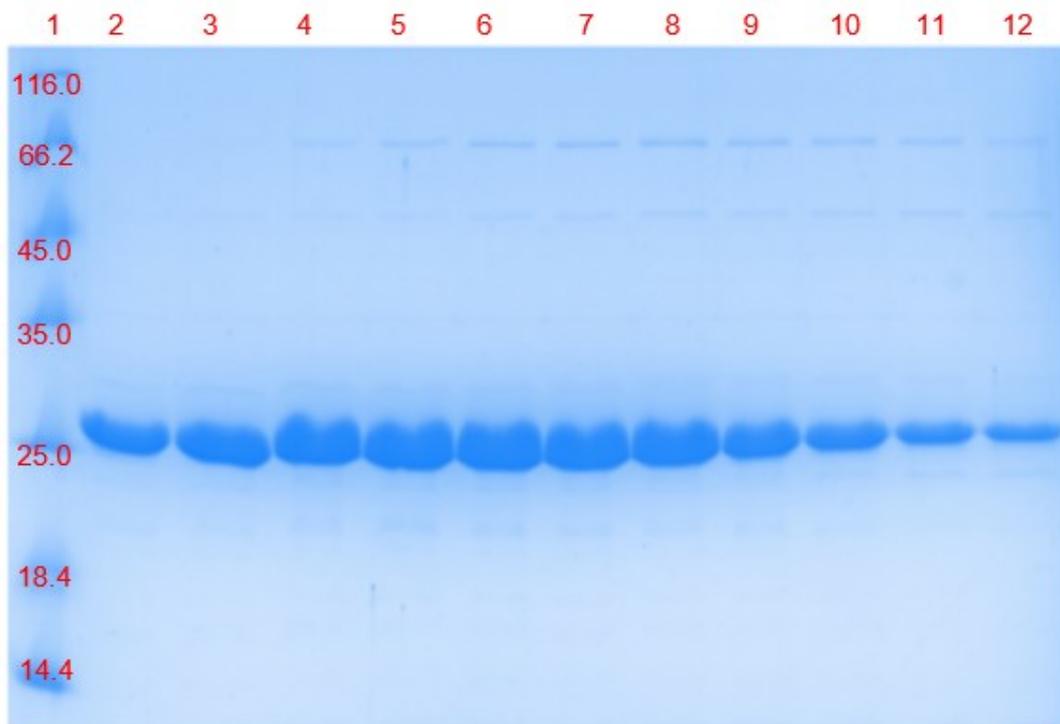
**Figure 64** YihX purification by Ni-NTA chromatography. Lane 1: flowthrough. Lanes 2 to 12: elution.



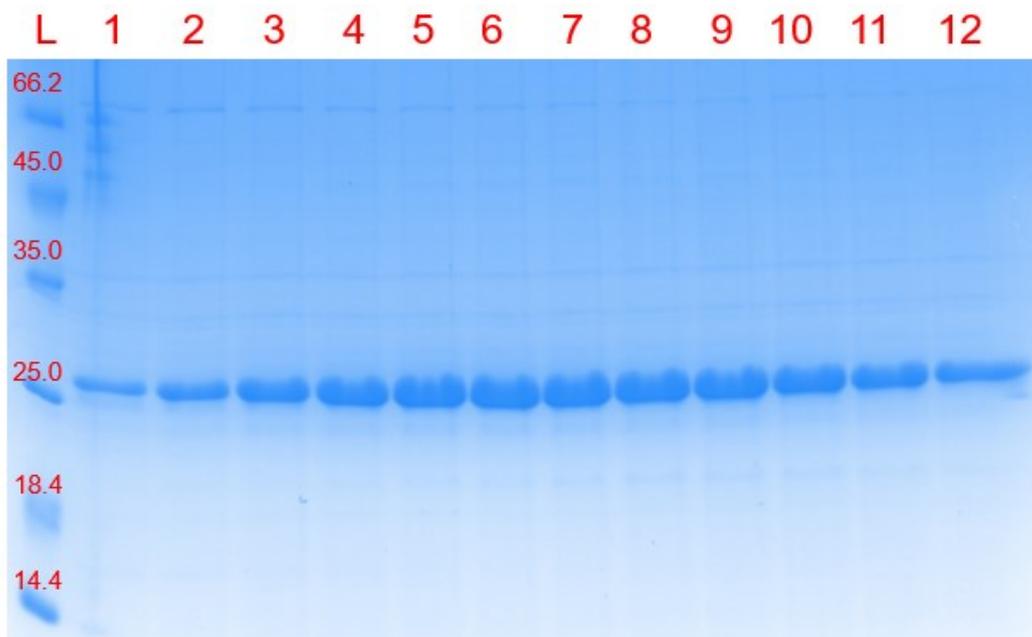
**Figure 63** YihS purification by Ni-NTA chromatography. Lane 1 to 8: elution.



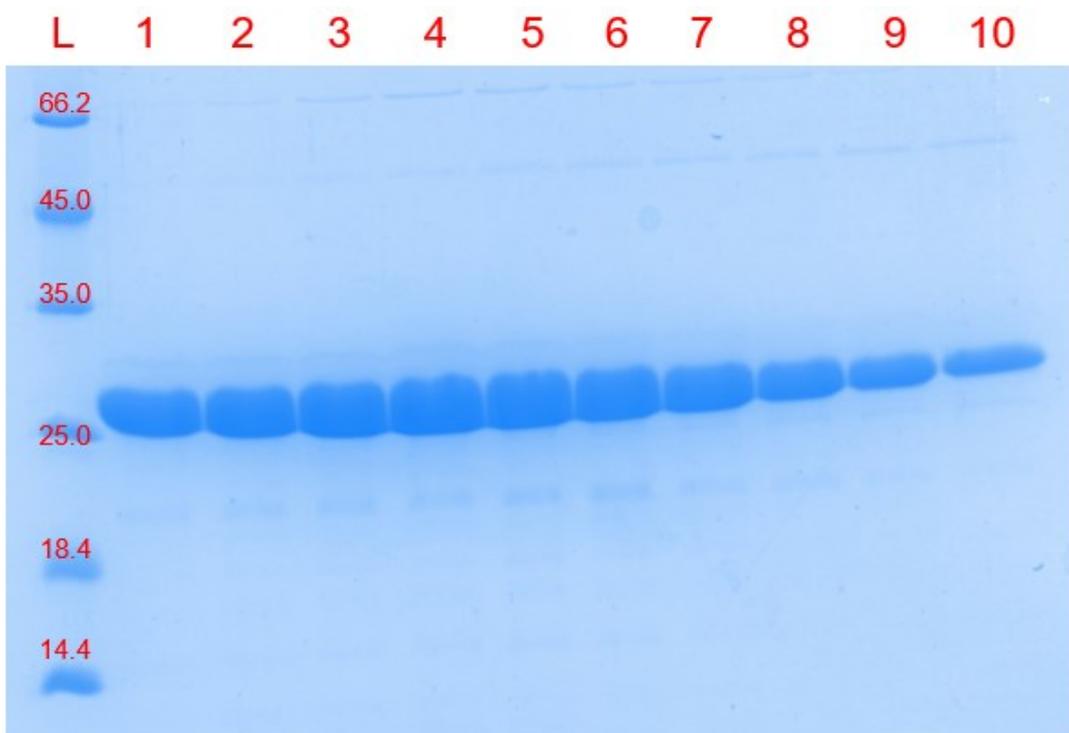
**Figure 65** YihV purification by Ni-NTA chromatography. Lanes 1 to 11: elution.



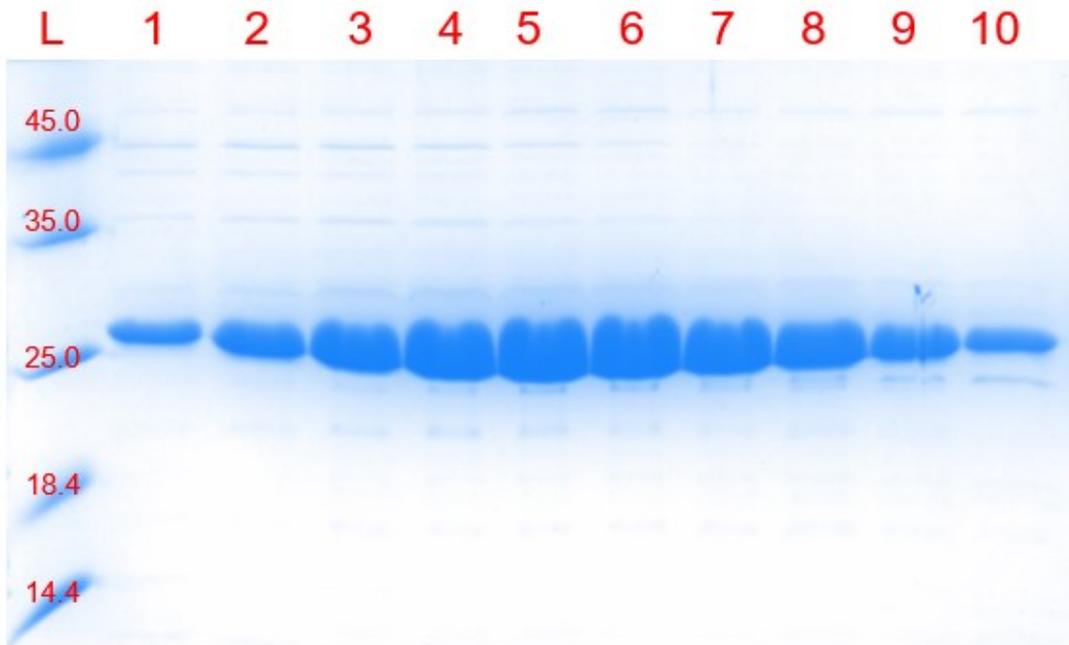
**Figure 66** YihX<sub>G8D</sub> after SEC purification. Lane 1: Protein ladder. Lane 2 to 12: elution.



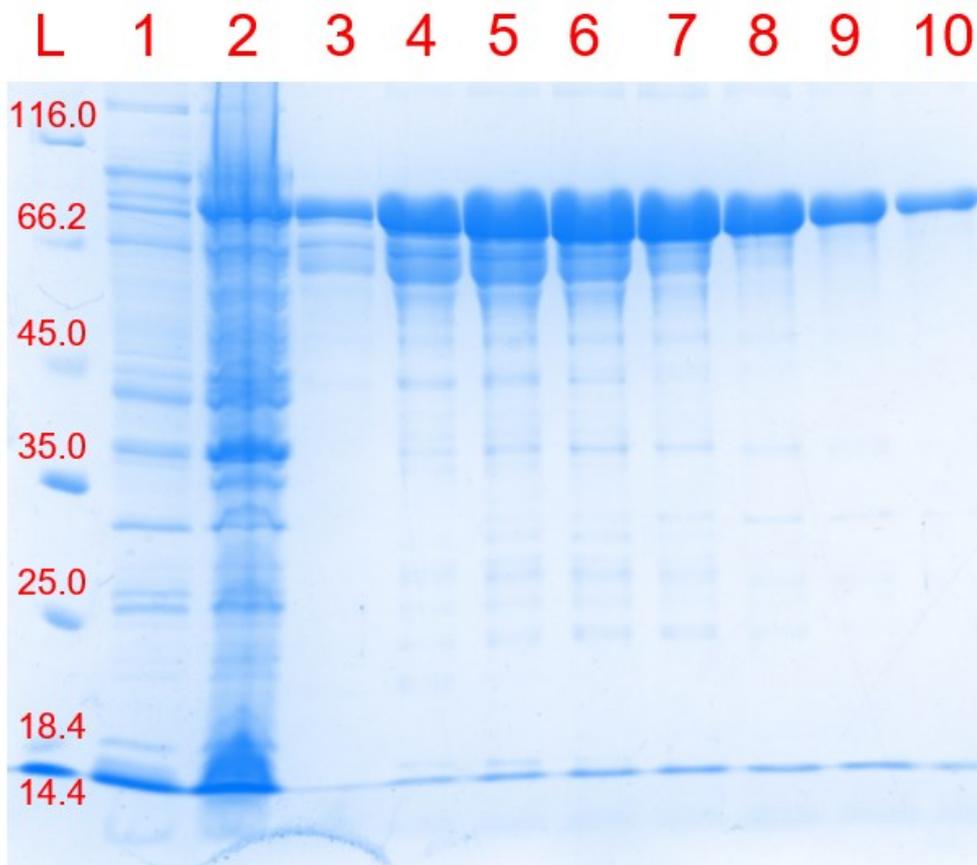
**Figure 67** YihX<sub>D6N</sub> purification by Ni-NTA chromatography. Lanes 1 to 12: elution.



**Figure 68** YihX<sub>N108L</sub> purification by Ni-NTA chromatography. Lanes 1 to 10: elution.



**Figure 70** YihX<sub>K141L</sub> purification by Ni-NTA chromatography. Lanes 1 to 10: elution.



**Figure 69** TF-Ub-TK2 fusion protein expression and purification by Ni-NTA chromatography. Lane 1: before induction. Lane 2: after induction. Lanes 3 to 10: elution.

## 7.7 Recombinant Protein Sequences

### YihX

MLYIFDLGNVIVDIDFNRVLGAWSDLTRIPLASLKKSFHMGEAFHQHERGEISDEAFA  
EALCHEMALPLSYEQFSHGWQAVFVALRPEVIAIMHKLREQGHRVVLSNTNRLHT  
TFWPEEYPEIRDAADHIYLSQDLGMRKPEARIQHVLAEGFSPSDTVFFDDNADNI  
EGANQLGITSILVKDKTTIPDYFAKVLICLEHHHHHH

### YihV

MRGSHHHHHHMIRVACVGITVMDRIYYVEGLPTESGKYVARNYTEVGGGPAATAAV  
AAARLGAQVDFIGRVGDDDTGNSLLAELESWGVNTRYTKRYNQAKSSQSAIMVDT  
KGERIIINYPSPDLLPDAEWLEEIDFSQWDVVLADVRWHDGAKKAFTLARQAGVMT  
VLDGDITPQDISELVALSDHAAFSEPGLARLTGVKEMASALKQAQTLTNGHVYVTQG  
SAGCDWLENGGRQHQPFAKVDVDTTGAGDVFHGALAVLATSGDLAESVRFASG  
VAALKCTRPGGRAGIPDCDQTRSFLSLFV

### YihS

MKWFNTLSHNRWLEQETDRIFDFGKNSVVPTGFGWLGNKGQIKEEMGTHLWITAR  
MLHVYSVAAAMGRPGAYSLVDHGKAMNGALRDKKYGGWYACVNDEGVVDASKQ  
GYQHFFALLGAASAVTTGHPEARKLLDYTIEIIEKYFWSEEEQMCLESWDEAFSKTE  
EYRGGNANMHAVEAFLIVYDVTHDKKWLDRAIRVASVIIHDVARNNHYRVNEHFDT  
QWNPLPDYNKDNPAHRFRAFGGTPGHWIEWGRLMLHIHAALEARCEQPPAWLLE  
DAKGLFNATVRDAWAPDGADGIVYTVDWEGKPVVRRERVRWPIVEAMGTAYALYTV  
TGDRQYETWYQTWWEYCIKYLMDYENGSWWQELDADNKVTTKVWDGKQDIYHL  
LHCLVIPRIPLAPGMAPAVAAGLLDINAKLEHHHHHH

### TF-Ub-hTK2

MGSSHHHHHHMQVSVETTQGLGRRVTITVAADSIETAVKSELVNVAKKVRIDGFRK  
GKVPMNIVAQRYGASVRQDVLGDLMSRNFIDAIIEKINPAGAPTYVPGEYKLGEDF  
TYSVEFEVYPEVELQGLEAIEVEKPIVEVTDADVDGMLDTLRKQQATWKEKDGAVE  
AEDRVTIDFTGSVDGEEFEGGKASDFVLAMGQGRMIPGFEDGKIGHKAGEEFTIDV  
TFPEEYHAESLKGKAAKFAINLKKVEERELPELTAEFIKRFGVEDGSVEGLRAEVRK  
NMERELKS AIRNRVKSQAIEGLVKANDIDVPAALIDSEIDVLRQAQRFGGNEKQA  
LELPRELFEEQAKRRVVVGLLLGEVIRTNELKADEERVKGLIEEMASAYEDPKEVIEF  
YSKNKELMDNMRNVALEEQAQAVEAVLAKAKVTEKETTFNELMNQQAQIFVKTLTGKTI  
TLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYNIQKESTLHLVL  
RLRGGMPVICVEGNIASGKTTCLEFFSNATDVEVLTEPVSKWRNVRGHNPLGLMY  
HDASRWGLTLQTYVQLTMLDRHTRPQVSSVRLMERSIHSARYIFVENLYRSGKMPE  
VDYVVLSEWFDWILRNMDVSVDLIVYLRTNPETCYQRLKKRCREEEKVIPLEYLEAI  
HHLHEEWLIKSLFPMAAPVLVIEADHHMERMLELFEQNRDRILTPENRKHCP

### PG\_0725

MIRNIVFDLGGVLIHLNREESIRRFKAIGVADIEEMLDPYLQKGLFLDLESGRKSEEEF  
RTELSRYIGKELTYQQVYDALLGFLEEISAEKFDYIDSLRPDYRLFLLSNTNPYVLDLA  
MSPRFLPSGRTLDSFFDKVYASCQMKGKPNEDIFLEMIADSGMKPEETLFIDGDP  
ANVATAERLGFHTYCPDNGENWIPAITRLLREQKLEHHHHHH

### **Usp2-cc**

MGSSHHHHHHSSGLVPRGSHMLEDPLLTAKNSKSAQGLAGLRNLGNTCFMNSIL  
QCLSNTRERLDYCLQRLYMRDLGHTSSAHTALMEEFAKLIQTIWTSSPNDVVPSE  
FKTQIQRYAPRFMGYNQQDAQEFLRFLLDGLHNEVNRVAARPKASPETLDHLPDEE  
KGRQMWRKYLEREDSRIGDLFVGQLKSSLTCTDCGYCSTVFDPFWDLSLPIAKRG  
YPEVTLMDCMRLFTKEDILDGDEKPTCCRCRARKRCIKKFSVQRFPKILVLHLKRFS  
ESRIRTSKLTTFVNFPLRDLDLREFASENTNHAVYNLYAVSNHSGTTMGGHYTAYCR  
SPVTGEWHTFNDSSVTPMSSSQVRTSDAYLLFYELASPPSRM

## **7.8 Gene sequences**

### **YihX**

ATGCTCTATATCTTTGATTTAGGTAATGTGATTGTCGATATCGACTTTAACCGTGTG  
CTGGGAGCCTGGAGCGATTTAACGCGTATTCCGCTGGCATCGCTTAAGAAGAGT  
TTTCACATGGGGGAGGCGTTTCATCAGCATGAGCGTGGGGAAATTAGCGACGAA  
GCGTTCGCAGAGGCGCTGTGTCATGAGATGGCTCTACCGCTAAGCTACGAGCA  
GTTCTCTCACGGCTGGCAGGCGGTGTTTGTGCGCTGCGCCCGGAAGTGATCG  
CCATCATGCATAAACTGCGTGAGCAGGGGCATCGCGTGGTGGTGTCTTCCAATA  
CCAACCGCCTGCATACCACCTTCTGGCCGGAAGAATACCCGGAATTCGTGATG  
CTGCTGACCATATCTATCTGTGCAAGATCTGGGGATGCGCAAACCTGAAGCAC  
GAATTTACCAGCATGTTTTGCAGGCGGAAGGTTTTTACCCAGCGATACGGTCTT  
TTTCGACGATAACGCCGATAATATAGAAGGAGCCAATCAGCTGGGCATTACCACT  
ATTCTGGTGAAAGATAAAACCACCATCCCGGACTATTTGCGGAAGGTGTTATGCC  
TCGAGCACCACCACCACCACCAC

### **YihV**

ATGAGAGGATCTCACCATCACCATCACCATATGATTCGTGTTGCCTGCGTTGGTA  
TTACCGTTATGGATCGTATTTACTATGTGGAAGGTCTGCCGACCGAAAGCGGTAA  
ATATGTTGCACGTAATTATAACCGAAGTTGGTGGTGGTCCGGCAGCAACCGCAGC  
AGTTGCAGCAGCACGTCTGGGTGCACAGGTTGATTTTATTGGTCGTGTTGGTGA  
TGATGATACCGGTAATAGCCTGCTGGCAGAAGCTGGAAAGCTGGGGTGTTAATAC  
CCGTTATACCAAACGTTATAATCAGGCAAAAAGCAGCCAGAGCGCAATTATGGTT  
GATACCAAAGGTGAACGCATTATCATCAATTATCCGAGTCCGGATCTGCTGCCTG  
ATGCAGAATGGCTGGAAGAAATTGATTTTAGCCAGTGGGATGTTGTTCTGGCAG  
ATGTTTCGTTGGCATGATGGTGCAAAAAAAGCATTTACCCTGGCACGTCAGGCAG  
GCGTTATGACCGTTCTGGATGGTGTATTACACCGCAGGATATTAGCGAACTGGT

TGCACTGAGCGATCATGCAGCATTAGCGAACCAGGTCTGGCACGCCTGACCG  
GTGTTAAAGAAATGGCAAGCGCACTGAAACAGGCACAGACCCTGACCAATGGT  
CATGTTTATGTTACCCAGGGTAGCGCAGGTTGTGATTGGCTGGAAAATGGTGGT  
CGTCAGCATCAGCCTGCATTTAAAGTTGATGTTGTGGATAACCACGGTGCCGGT  
GATGTTTTTCATGGTGCAGTGGCAGTTGCCCTGGCAACCAGCGGTGATCTGGCA  
GAAAGCGTTCGTTTTGCAAGCGGTGTTGCAGCACTGAAATGTACCCGTCAGGT  
GGTCGTGCAGGTATTCCGGATTGTGATCAGACCCGTAGCTTTCTGAGCCTGTTT  
GTT

## TF-Ub-TK2

ATGGGCAGCAGCCATCATCATCATCATATGCAAGTTTCAGTTGAAACCACTC  
AAGGCCTTGGCCGCCGTGTAACGATTACTGTCGCTGCTGACAGCATCGAGACC  
GCTGTTAAAAGCGAGCTGGTCAACGTTGCGAAAAAAGTACGTATTGACGGCTTC  
CGCAAAGGCAAAGTGCCAATGAATATCGTTGCTCAGCGTTATGGCGCGTCTGTA  
CGCCAGGACGTTCTGGGTGACCTGATGAGCCGTAACCTCATTGACGCCATCATT  
AAAGAAAAAATCAATCCGGCTGGCGCACCCGACTTATGTTCCGGGCGAATACAAG  
CTGGGTGAAGACTTCACTTACTCTGTAGAGTTTGAAGTTTATCCGGAAGTTGAAC  
TGCAGGGTCTGGAAGCGATCGAAGTTGAAAAACCGATCGTTGAAGTGACCGAC  
GCTGACGTTGACGGCATGCTGGATACTCTGCGTAAACAGCAGGCGACCTGGAA  
AGAAAAAGACGGCGCTGTTGAAGCAGAAGACCGCGTAACCATCGACTTCACCG  
GTTCTGTAGACGGCGAAGAGTTTGAAGGCGGTAAAGCGTCTGATTTTCGTAAGT  
CGATGGGCCAGGGTCGTATGATCCCGGGCTTTGAAGACGGTATCAAAGGCCAC  
AAAGCTGGCGAAGAGTTCACCATCGACGTGACCTTCCCGGAAGAATACCACGC  
AGAAAGCCTGAAAGGTAAAGCAGCGAAATTCGCTATCAACCTGAAGAAAGTTGA  
AGAGCGTGAAGTCCGGAAGTACTGACTGCAGAATTCATCAAACGTTTTCGGCGTTGA  
AGATGGTTCGTTAGAAAGTCTGCGCGCTGAAGTGCGTAAAAACATGGAGCGCG  
AGCTGAAGAGCGCCATCCGTAACCGCGTTAAGTCTCAGGCGATCGAAGGTCTG  
GTAAAAGCTAACGACATCGACGTACCGGCTGCGCTGATCGACAGCGAAATCGAC  
GTTCTGCGTCCGAGGCTGCACAGCGTTTCGGTGGCAACGAAAAACAAGCTCT  
GGAAGTCCCGCGCGAAGTTCGAAGAACAGGCTAAACGCCGCGTAGTTGTTG  
GCCTGCTGCTGGGCGAAGTTATCCGCACCAACGAGCTGAAAGCTGACGAGGAG  
CGCGTGAAAGGCCTGATCGAAGAGATGGCTTCTGCGTACGAAGATCCGAAAGA  
AGTTATCGAGTTCTACAGCAAAAACAAGAAGTACTGATGGACAACATGCGCAATGTT  
GCTCTGGAAGAACAGGCTGTTGAAGCTGTACTGGCGAAAGCGAAAGTGACTGA  
AAAAGAAACCACTTTCAACGAGCTGATGAACCAGCAGGCGCAGATCTTCGTTAA  
AACCTGACCGGCAAGACCATTACCCTGGAAGTGGAAACCGAGCGACACCATCG  
AGAACGTGAAAGCGAAGATCCAAGACAAAGAAGGTATTCCGCCGGATCAGCAAC  
GTCTGATTTTTGCGGGCAAGCAGCTGGAGGACGGTCGTACCCTGAGCGATTAC  
AACATCCAAAAAGAAAGCACCTGCATCTGGTGCTGCGTCTGCGTGGTGGCAT  
GCCATCAGTGATCTGTGTGAGGGCAATATTGCAAGTGGGAAGACGACATGCCT  
GGAATTCTTCTCCAACGCGACAGACGTCGAGGTGTTAACGGAGCCTGTGTCCA  
AGTGGAGAAATGTCCGTGGCCACAATCCTCTGGGCCTGATGTACCACGATGCCT  
CTCGCTGGGGTCTTACGCTACAGACTTATGTGCAGCTCACCATGCTGGACAGGC  
ATACTCGTCCTCAGGTGTCATCTGTACGGTTGATGGAGAGGTGATTACAGCG  
CAAGATACATTTTTGTAGAAAACCTGTATAGAAGTGGGAAGATGCCAGAAGTGGA

CTATGTAGTTCTGTCTCGGAATGGTTTGGACTGGATCTTGAGGAACATGGACGTGTCT  
GTTGATTTGATAGTTTACCTTCGGACCAATCCTGAGACTTGTTACCAGAGGTTAA  
AGAAGAGATGCAGGGAAGAGGAGAAGGTCATTCCGCTGGAATACCTGGAAGCA  
ATTCACCATCTCCATGAGGAGTGGCTCATCAAAGGCAGCCTTTTCCCATGGCA  
GCCCTGTTCTGGTGATTGAGGCTGACCACCACATGGAGAGGATGTTAGAACTC  
TTTGAACAAAATCGGGATCGAATATTA ACTCCAGAGAATCGGAAGCATTGCCCA

## PG\_0725

ATGATTCGCAACATCGTGTTTGGATTTAGGTGGTGTCTGATTCATCTGAATCGCG  
AAGAAAGCATTTCGTCGTTTTAAAGCAATTGGTGTGGCCGATATTGAAGAAATGCT  
GGATCCGTATCTGCAGAAAGGTCTGTTTCTGGATCTGGAAAGCGGTCGTAAAAG  
CGAAGAAGAATTCGTACCGAACTGAGCCGTTATATTGGTAAAGAACTGACCTAT  
CAGCAGGTTTATGATGCACTGCTGGGTTTTCTGGAAGAAATTAGCGCAGAGAAA  
TTCGATTATATCGATAGCCTGCGTCCGGATTATCGCCTGTTTCTGCTGAGCAATAC  
CAATCCGTATGTTCTGGACCTGGCAATGAGTCCGCGTTTTCTGCCGAGTGGTCG  
TACCCTGGATAGCTTTTTTCGATAAAGTTTATGCAAGCTGCCAGATGGGTAAATACA  
AACCGAACGAAGATATTTTCTGGAAATGATTGCAGACAGCGGTATGAAACCGG  
AAGAAACCCTGTTTATTGATGATGGTCCGGCAAATGTTGCAACCGCAGAACGTC  
TGGGCTTTCATACCTATTGTCCGGATAATGGTGAAAATTGGATTCGGGCAATTACC  
CGTCTGCTGCGTGAACAGAACTCGAGCACCACCACCACCAC

## 7.9 Pulse sequence for F<sup>-</sup> suppression in <sup>19</sup>F NMR

;zgpr

;avance-version (12/01/11)

;1D sequence with f1 presaturation

;\$CLASS=HighRes

;\$DIM=1D

;\$TYPE=

;\$SUBTYPE=

;\$COMMENT= #include <Avance.incl>

"d12=20u"

"acqt0=-p1\*2/3.1416"

1 ze

2 30m

d12 pl9:f1

d1 cw:f1 ph29

4u do:f1

d12 pl1:f1

p1 ph1

go=2 ph31

30m mc

#0 to 2 F0(zd)

exit

ph1=0 2 2 0 1 3 3 1

ph29=0

ph31=0 2 2 0 1 3 3 1

;pl1 : f1 channel - power level for pulse (default)

;pl9 : f1 channel - power level for presaturation

;p1 : f1 channel - 90 degree high power pulse

;d1 : relaxation delay; 1-5 \* T1

;d12: delay for power switching 20 usec]

;ns: 1 \* n, total number of scans: NS \* TD0

;\$ld: zgpr,v 1.11 2012/01/31 17:49:32 ber Exp \$