Investigation of Oligomerisation in Transaminase Biocatalysis Using Structural Biology Approaches

Philip Smith

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

April 2023

School of Chemical and Environmental Engineering

University of Nottingham

Abstract

Transaminases (TAs) offer a sustainable, highly stereoselective alternative to existing chemocatalytic syntheses of chiral primary amines. Despite this the commercial utility of TAs is limited by poor stability under process conditions. Existing work towards engineering TAs for improved stability has adopted generally applicable protein engineering methods or has sought to stabilise the functionally critical intra-dimer interface present in all members of the especially promising class III group of TAs.

Recent work however, has shown that the higher oligomerisation state, specifically tetrameric, seen for some members of the class III family, may affect the enzyme function and stability, especially in regard to the retention of the cofactor PLP. This suggests that potentially the tetrameric TAs could offer improved biocatalysis, and engineering TAs to adopt the tetrameric state or stabilise the tetrameric state could potentially be another approach to engineering them for improved stability However, in total there is still relatively little information available on the effects arising specifically from tetramerization towards the functional properties.

In this work, a class III TA from *Pseudomonas fluorescens* (PfTA) was studied. Biophysical evidence was obtained and showed that it invariantly adopts the tetrameric state. A crystal structure was obtained, and by comparison against homologous TAs, mutations were designed to vitiate the tetramerization, in order to be able to compare dimeric and tetrameric forms with as minimal other structural differences as possible.

The computationally guided approach used to design mutations, made use of molecular dynamics simulations to attempt to infer the likely extent of interface disruption associated with each mutation. Whilst disappointingly, it was not possible to validate the computational approach, due to expression failure of most mutants, nevertheless one suitable mutant was obtained.

The successfully expressed PfTA R215E mutant showed an identical fold and active site structure but with a wholly dimeric oligomerisation state. In comparison to WT, it showed an inactivation rate more sensitive to high amine substrate loading conditions. Despite the lack of other mutants to reinforce the comparison and the validity of the results, this modestly evidences that tetramerization may help to improve cofactor retention and stability in TAs, as suggested previously in the literature.

Acknowledgements

Table of Contents

1 Introduction

1.1 Biocatalysis

In the cell, chemical transformations are catalysed predominantly by enzymes (except for a minority of RNA-catalysed reactions [1]). Biocatalysis is the utilisation of enzymes to catalyse chemical transformations to obtain products of interest, and as such, biocatalysts can be considered in contrast to chemical catalysts.

In specific regard to function, a major advantage of enzymes is that they routinely achieve good stereoselectivity, which is ultimately rooted in the inherent asymmetry of natural proteins, being constituted from L- amino acids only. Likewise, good regioselectivity is a common feature in enzymes, since enzymes *in vivo* are required not to interfere with metabolic pathways other that those to which they properly belong, and therefore must be able to distinguish between similar substrates.

In broader regard to production and disposal of the (bio)catalysts themselves, enzymes have a variety of advantages over chemical catalysts. Firstly, they fit within a "green" chemistry framework, especially in terms of avoiding hazardous transition metals frequently used in chemical catalysis, and also because their use implies aqueous (rather than organic) solvents. Secondly, they fit within a circular economy framework, since biocatalysts are expressed in living hosts in growth media derived from natural products such as yeast extract; and the inactivated biocatalyst itself poses little risk to the environment upon disposal [2].

Biocatalysis does also face challenges in competing with chemical catalysis. Frequently, enzyme stability is poor in the presence of a high concentration of substrate, and/or substrate solubility in aqueous solvent may be poor. The molecular mass of a typical enzyme is also much higher than that of a typical chemical catalyst. Therefore, biocatalysis struggles to compete in terms of the amount of catalyst required, and the overall amount of material (including solvent) required for a given amount of product. Biocatalysis therefore relies upon engineering of natural enzymes (most especially by directed evolution, and also by rational design) to obtain enzymes with optimised stability, activity, and specificity[3].

1.2 Background to transaminases and potential for biocatalytic applications

Transaminases (TAs) catalyse a bi-bi ping-pong reaction which effectively results in the exchange of a carbonyl group and an amine group between a pair of substrates (Figure 1.1). Mechanistically, only the amine nitrogen is transferred between substrates, whereas the

carbonyl oxygen is gained from (or lost to) the solvent. With potential for enantioselective amination of prochiral carbonyls using cheap amine donors, TAs have gathered interest for application as biocatalysts. An implementation of a TA from *Arthrobacter sp.* in the production of diabetes drug Sitagliptin became a prominent example of commercial success, using iPrNH$_2$ as an amine donor, and achieving a better stereoselectivity than the chemical catalyst it replaced[4].
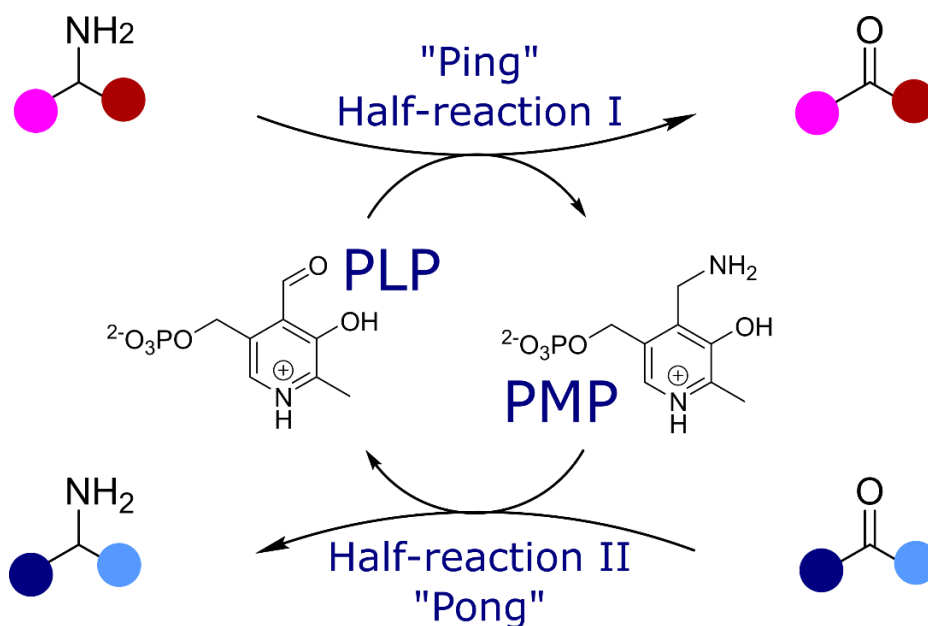
Figure 1.1 Schematic overview of the TA catalytic cycle exhibiting a bi-bi ping-pong mechanism. Deamination of a first substrate (red/pink) with concomitant conversion of pyridoxal 5'-phosphate (PLP) to pyridoxamine 5'-phosphate (PMP), is followed by amination of the second substrate (blue/navy) with re-formation of PLP.

Though occurring in a variety of contexts, TAs are an essential part of cell growth[5] and feature especially in amino acid metabolism, with annotated examples of this such as alanine TA, glutamate-pyruvate TA, etc., available in the UniProt database. TAs are ubiquitous across the kingdoms of life, but novel TAs for biocatalysis have traditionally come from bacteria obtained by enrichment culture[6], in which a selection pressure is applied by limiting the source of nitrogen to one compound which the bacteria must be able to metabolise using transamination.

1.2.1 Transaminase catalytic mechanism

The catalytic mechanism of TAs (Figure 1.2) begins with a lysine residue of the active site attacking the carbonyl moiety of the cofactor pyridoxal 5' phosphate (PLP), leading to loss of the aldehydic oxygen atom, and the formation of an imine bond between the cofactor and the enzyme, a state referred to as the internal aldimine state. The carbon atom of the imine moiety

is subsequently attacked by the amine group of the amino acid substrate, displacing the lysine residue as a new imine bond is formed, between the cofactor and the substrate, a state referred to as the external aldimine state.

In this state, the catalytic lysine residue is able to deprotonate the Cα position of the substrate, since the nascent Cα carbocation is stabilised by resonance with a quinonoid structure which enables the negative charge to be distributed through the conjugated π-bonding system of the pyridinyl moiety.

It is this step of the catalytic cycle which determines the stereoselectivity of the enzyme. In the reverse direction of the reaction, the Cα carbocation structure is prochiral, and the face which is protonated by the catalytic lysine, determines the resulting chirality at the Cα position. Protonation on the Re- side leads to an (*S*)- chirality, corresponding to an L-amino acid (such as the natural canonical amino acids), whereas protonation on the Si- side leads to (*R*)- chirality, corresponding to a D-amino acid.

In the forward direction of the reaction, the resonance-stabilised carbocation/quinonoid form is protonated at the C4 position. Water in the active site is then able to make nucleophilic attack on the Cα position of the resulting structure. This hydrolysis leads to a keto-acid product being released, whilst the cofactor is left behind in its pyridoxamine 5' phosphate (PMP) form. Binding of a carbonyl substrate is then required to start the reverse half reaction of the catalytic cycle and reform PLP.
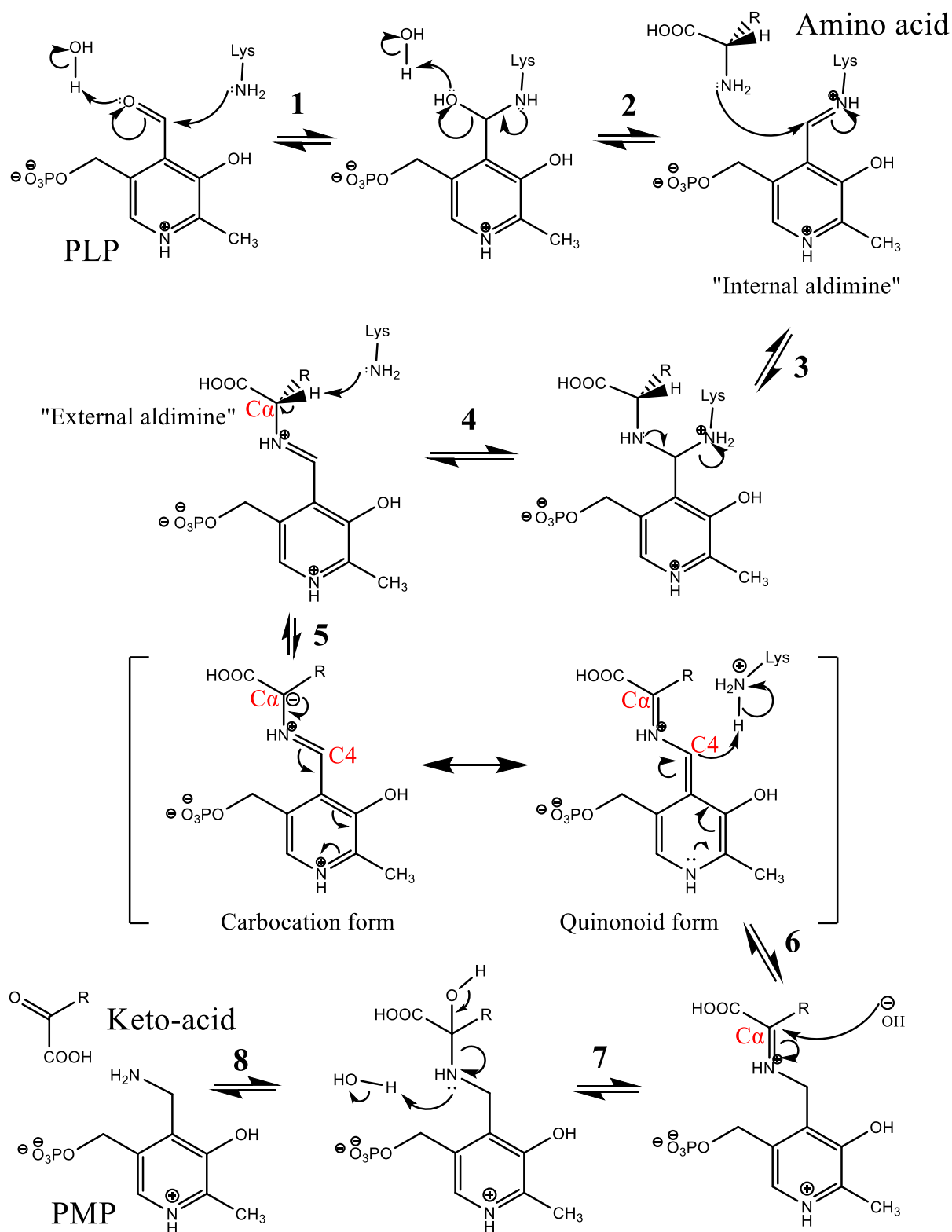
Figure 1.2 Catalytic lysine forms an imine linkage with PLP (**Steps 1+2**) resulting in the internal aldimine state; the lysine is then displaced when the amino acid substrate forms a new imine bond with PLP (**Steps 3+4**) as the external aldimine state. Deprotonation (**Step 5**) yields a carbocation stabilised by a quinonoid resonance form, hydrolysis of which (**Steps 6+7+8**) results in the release of the keto-acid product and formation of PMP.

TAs belong to the wider family of pyridoxal 5'-phosphate (PLP)-dependent enzymes, with PLP being one of vitamin B6's six vitamer forms, the form which is active as an enzyme cofactor[7]. A dominant theme across these PLP-dependent enzymes is the electron-withdrawing chemistry of PLP, being exploited to break a bond within the substrate. The bond which is to be broken must be aligned so as to overlap with the antibonding orbital of the C=N imine moiety in the PLP-substrate covalent adduct (Figure 1.3)[8]. In those TAs which accept amino acid substrates, it is the Cα which must be deprotonated (i.e. the Cα-H bond must be broken); whereas in other PLP-dependent enzymes, the Cα-COOH bond or the Cα-R can be broken instead, leading to different enzyme functionalities.
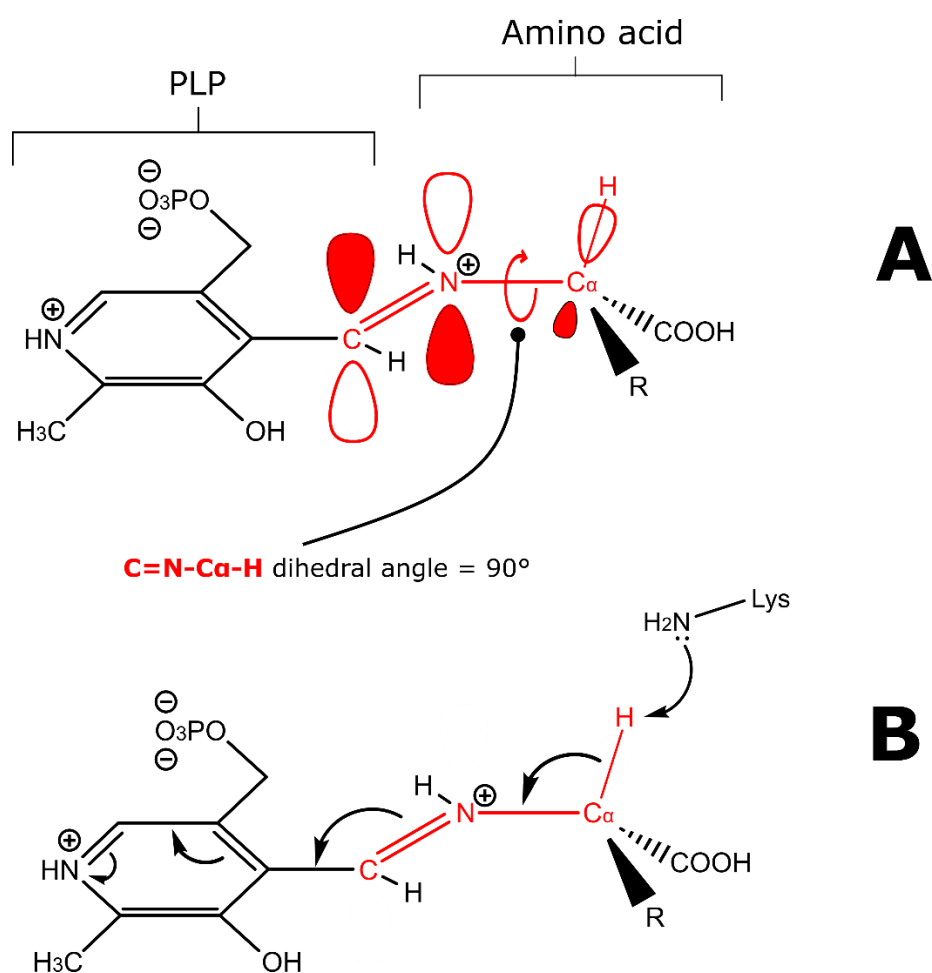
Figure 1.3 **(A)** A dihedral angle of 90° between the C=N imine bond and the Cα-H bond in the PLP-substrate adduct, allows for optimal alignment of the filled Cα-H bonding orbital in the amino acid substrate, with the empty antibonding orbital of the PLP-substrate C=N linkage. **(B)** Since the C=N antibonding orbital is part of a conjugated system with the pyridinyl moiety of PLP, this has the effect of facilitating the deprotonation of Cα by the catalytic lysine residue.

The superfamily of B6-dependent enzymes has been divided into seven structural fold types to date, and enzymes with transamination function belong to either fold-type I or IV (). These fold types however, are not exclusively limited to a single functionality; for example, fold-type I includes both TAs and decarboxylases.

Focusing specifically on the TAs, sequence-based classification further divides them into classes, each covering a smaller range of substrate specificities, as reviewed comprehensively by Steffen-Munsberg *et al* [9]. Most of these TA classes adopt the fold-type I tertiary structure, with the exception of the TA class IV, which adopts the fold-type IV tertiary structure ().
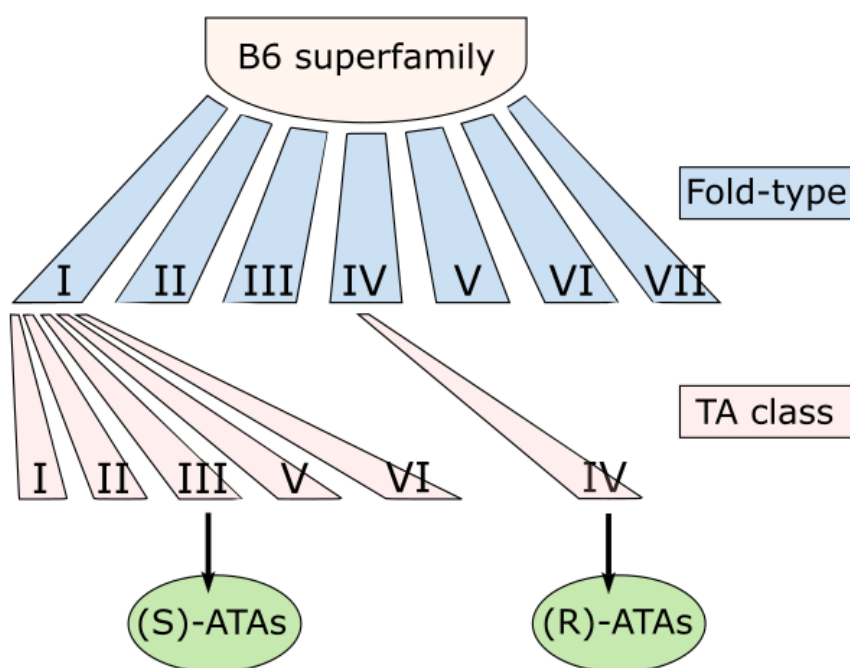
Figure 1.4 Relationship between the six TA class categories and the seven fold types of the superfamily of PLP-dependent enzymes. Class III and class IV contain the (*S*)- and (*R*)- selective ATAs, respectively

Amongst all the classes of TAs, it is the class III TAs (belonging to fold-type I) and class IV TAs (belonging to fold-type IV) which are most interesting from a biocatalysis perspective. The former is limited to only (*S*)- selectivity, whilst the latter includes (*R*)- selectivity in its substrates, but both two classes are unlike any of the other classes, in that for the purpose of substrate recognition, they do not strictly require a geminal -COOH moiety on the carbon atom to be (de-)aminated, which makes them more applicable to a wider range of chemical transformations.

TAs displaying this ability, can recognise both kinds of substrate, i.e. substrates with a geminal -COOH group, and substrates without a geminal -COOH group. To do this, they employ either of two different switching mechanisms (Figure 1.5). The first mechanism uses a "flipping"

6

arginine residue, which can flip into the active site to make an electrostatic interaction with -COOH bearing substrates, and can flip out of the active site for other substrates which lack geminal -COOH. The second mechanism also uses arginine in the same way, but the arginine residue itself does not flip position. Rather, when a substrate without geminal -COOH is bound, a glutamate sidechain flips into the active site, to provide a competing -COOH group to interact with the arginine residue[9].
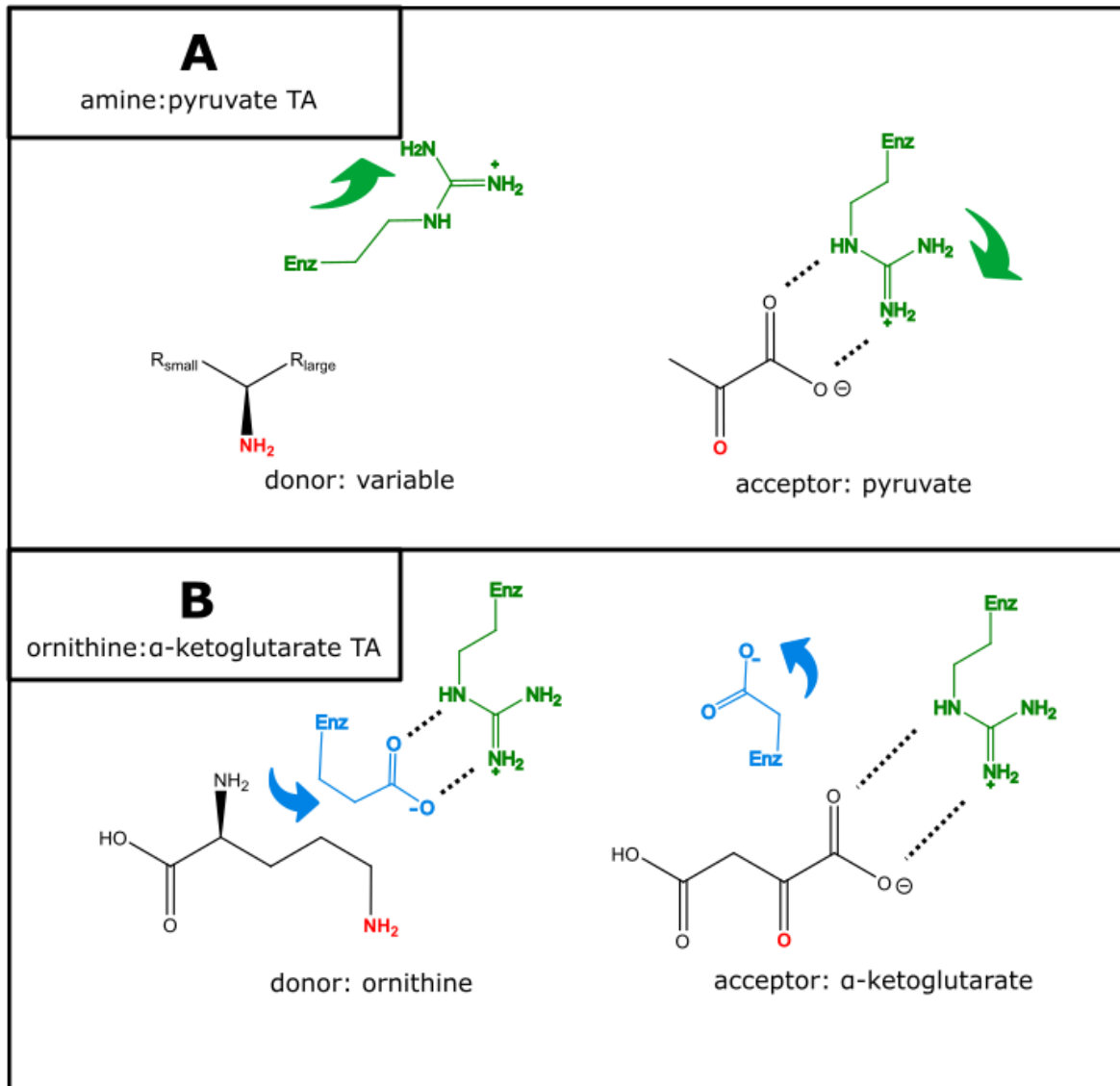
Figure 1.5 Mechanisms for dual substrate recognition in TAs. **(A)** Amine:pyruvate TAs use an arginine residue to recognise pyruvate, whereas a variety of non-specific amine donors may be accommodated in the active site, by first flipping the arginine residue away from the active site. **(B)** Ornithine:α-ketoglutarate TA demonstrates a different mode of dual recognition, since the lack of steric bulk around the amine moiety carbon atom of ornithine (in comparison to the analogous position in α-ketoglutarate) means that a glutamate residue is able to move in to compete for electrostatic interaction with the arginine residue responsible for -COOH recognition.

TAs employing these mechanisms, especially the first mechanism, often accept a wide range of susbtrates, sometimes making the natural substrate harder to identify[10]. Those which accept amino acid substrates, and (de-)aminate positions other than Cα or Cβ (i.e. sidechain positions far from the -COOH group, such as the sidechain amine moiety in lysine or ornithine) are frequently termed "ω-transaminases" (ω-TAs). If an enzyme can accept amine substrates with no -COOH group at all, especially where the natural substrate is not determined, these are sometimes termed simply as "amine TAs" (ATAs).

Amongst these, some have found repeated success as biocatalysts. For example, an (R-) selective (class IV) ATA from *Arthrobacter* KNK168 has been used within synthetic routes for four different (pro-)drugs[11], each of these requiring a different engineered variant of the wild-type enzyme. Likewise, amongst the class III ATAs, those from *Chromobacterium violaceum* and *Vibrio fluvialis* have found multiple implementations in drug syntheses (Figure 1.6)
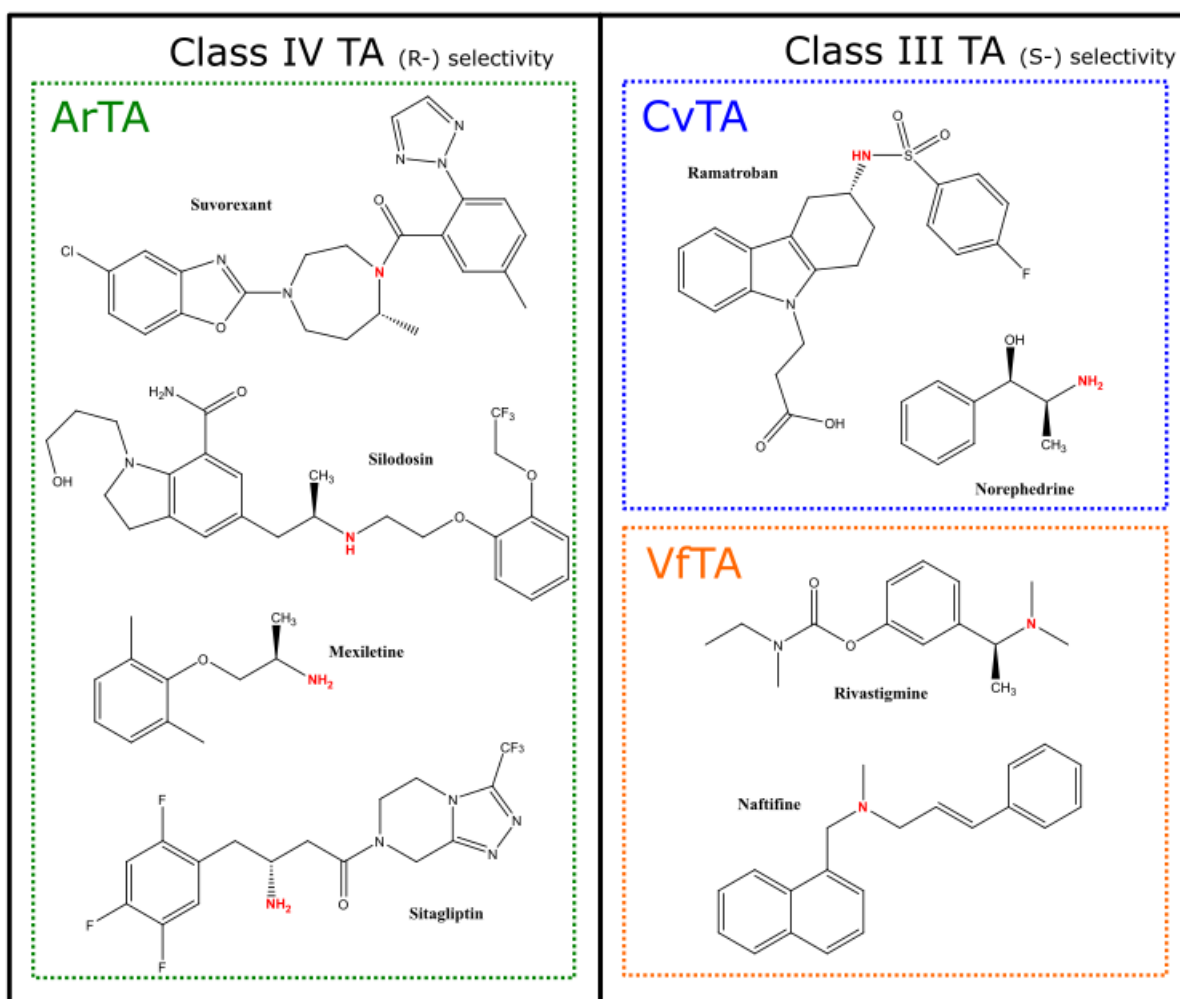
Figure 1.6 Examples of TA implementation in drug syntheses, for *Arthrobacter* KNK168 (ArTA), *Chromobacterium violaceum* (CvTA) and *Vibrio fluvialis* (VfTA) transaminases. The transaminated position is highlighted in red within the final product molecule. Examples selected from Kelly *et al*[11]

1.3   Transaminase stability and oligomerisation

The opportunities for industrial applications of TAs are eminent, with an estimate that 40% of pharmaceuticals/agrochemicals contain chiral amine moieties[12]. However, TAs have several shortcomings which limit their commercial usefulness, especially in the context of inadequate stability under industrial process conditions, such as in the presence of the cosolvents sometimes required to solubilise substrates, and under other challenging conditions as reviewed for example by Guo and Berglund[13]. Engineering the reaction medium, or immobilising the TA, represent approaches to working around this problem, but engineering TAs for resilience under a wider range of conditions would make them more useful as an option in the "toolbox" of synthesis options.

1.3.1 Cofactor binding in transaminases

Taking a rationally guided, structure-based approach towards TA engineering, this project focuses on TAs belonging to the fold-type I, rather than the fold-type IV, since structure-function relationships differ according to the basic fold of the enzyme.

Based on the catalytic mechanism, it is naturally expected that TAs cannot function without binding PLP, but PLP also contributes to TA stability. An overview of the literature establishes that PLP binding in fold-type I TAs cannot occur unless the active site is correctly formed[14], but once PLP is bound in the active site, the PLP helps to stabilise the active site structure, and stabilise the protein against irreversible denaturation[15–17].

This can be considered structurally in terms of the constitution of the active site, which is formed on the homo-dimerization interface where two identical monomers come together, nominally considered as monomer "A" and monomer "B". Monomer "A" forms majority of the interior surface of one active site, surrounding the entirety of PLP's pyridinyl moiety, and part of its phosphate moiety; whilst a long loop belonging to monomer "B" completes the rest of the active site around PLP's phosphate moiety, by forming hydrogen bonds to the phosphate moiety (Figure 1.7). In order to do this, this long loop, referred to as the phosphate-binding loop, must be positioned within a cleft of monomer "A".
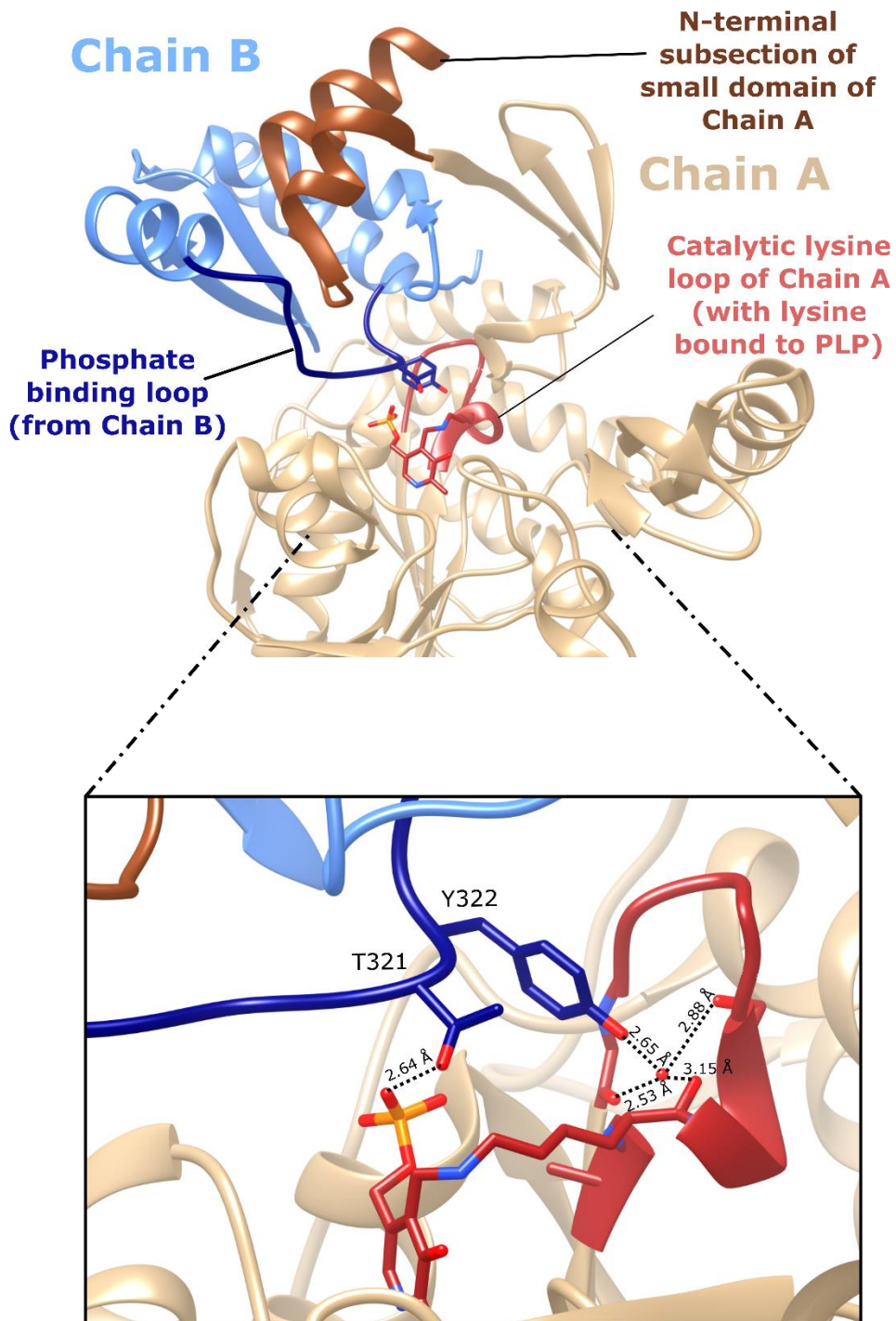
Figure 1.7 Dimeric *Chromobacterium violaceum* TA (CvTA) PDB 4AH3 with PLP covalently linked to catalytic lysine (internal aldimine state). The phosphate binding loop (highlighted dark blue) of the complementary chain (Chain B, pale blue) forms hydrogen bonds to both the PLP phosphate moiety and the catalytic lysine loop (highlighted dark red) of Chain A (pale brown). The phosphate binding loop of Chain B runs along a cleft formed in part from the N-terminal subsection of Chain A small domain (highlighted dark brown).

According to the symmetry of the interface between identical monomers "A" and "B", there is another identical active site, formed mainly from monomer "B", completed by the phosphate-binding loop from monomer "A". Overall, the interface area between "A" and "B" is around

25%-30% of the surface area of a single monomer (Table 4-2 in Results, for comparison of existing PDB structures against this work). This large interface area is typical for an obligate interface, an interface which is strictly necessary to enable protein function and/or stability, where the individual monomers alone are unstable or non-functional[18]. It has been established in general that the presence of PLP is not required for individual monomers to fold, but that PLP will only bind to the dimeric apo-enzyme[19,20]. To the extent of available awareness, only one example is found in literature, with evidence to support TA catalytic activity and PLP binding in a monomeric state[21].

Focusing on how PLP stabilises the active site and dimerization of TAs, from a mechanistic perspective, Humble *et al.*[16] were able to obtain and compare crystal structures for apo- and holo- forms of *Chromobacterium violaceum* TA (CvTA). It was observed, firstly, that the conformation of the catalytic lysine differs, projecting into the active site to form the imine linkage with PLP when PLP is bound, but recoiling away from the active site in the absence of PLP; and that this change is not due to motion of the sidechain of the lysine residue itself, but rather due to a change in the conformation of the loop to which this catalytic lysine belongs. Furthermore, in the holo- state, the phosphate binding loop (from the complementary chain "B") contributes not only a threonine residue to hydrogen bond with the PLP phosphate moiety, but also a tyrosine residue to hydrogen bond with the backbone of the catalytic lysine loop; whereas in the apo state, these interactions are not found, with the whole phosphate binding loop conformation becoming disordered. The phosphate binding loop itself, runs through a cleft formed from both the large domain and N-terminal subsection of the small domain of the complementary chain "A", and this subsection of the small domain is also observed to become disordered. This is significant because the N-terminal subsection of the small domain makes a large contribution to the monomer-monomer interface, wrapping around the large domain of the complementary chain (Figure 1.7).

The relationship between PLP binding and dimer stability inferred crystallographically, was confirmed and more fully elucidated by molecular dynamics simulations on CvTA, which demonstrated that the loss of the PLP-lysine imine linkage, does trigger rearrangement of the phosphate binding loop and the N-terminal subsection of the small domain, culminating in disruption of the monomer-monomer interface[15](Figure 1.8). The molecular dynamics results were validated in part by applying a reducing agent to CvTA to reduce the PLP-enzyme imine linkage (from C=N to C-N). This reduced linkage makes it impossible to complete the catalytic cycle, and effectively traps the CvTA lysine residue in the active conformation, leading to an

increased melting temperature of the enzyme, indicating a stabilising effect. By contrast, the stabilising effect of inorganic phosphate alone, was relatively insignificant.
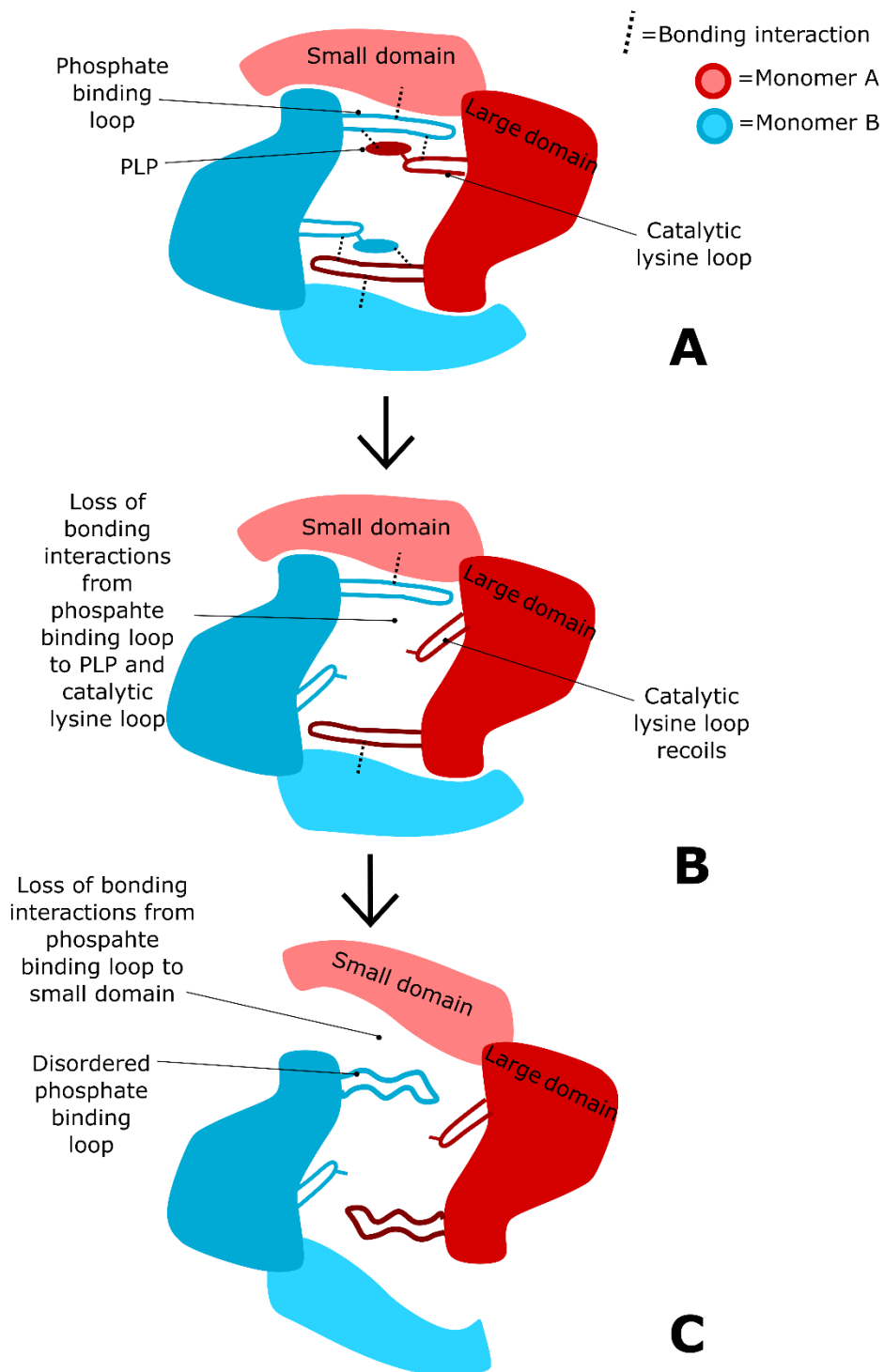
Figure 1.8 Schematic illustration of negative influence of PLP unbinding on active TA holo-dimer stability. (**A**) Active holo-dimer shows inter-chain bonding interactions from the phosphate binding loop to the complementary small domain, catalytic lysine loop, and PLP molecule. (**B**) Loss of PLP causes recoiling of the catalytic lysine loop, with resultant loss of bonding to the phosphate binding loop. (**C**) Phosphate binding loop becomes disordered, with disruption of the interaction between it and the complementary small domain, facilitating dissociation of the dimer.

Despite the apparent importance of PLP in stabilising the internal aldimine state (enzyme-PLP linked state), Humble *et al.* also identified a second apo-structure for CvTA, crystallised under different precipitant conditions, which showed a mixed occupancy of active / inactive conformations of the catalytic lysine residue. This suggests that the presence of PLP within the active site is not strictly required in order to induce a change to the active conformation, but that the active conformation is stabilised by PLP. Adding further complexity to the picture however, it has been observed that in some TA crystal structures, such as the TA from *Bifidobacterium kashiwanohense* (BkTA) (PDB 6JIX)[22], that whilst PLP is present in the active site, the active conformation of the catalytic lysine is not achieved, and the PLP is present in aldehyde form, not linked to the catalytic lysine. Considering the specific example of BkTA, the described crystallisation condition at pH 5.0 would possibly preclude the formation of the imine linkage, since formation of the imine linkage would require the catalytic lysine to be in a neutral (not protonated) state in order to perform the initial nucleophilic attack on PLP.

Considering these observations collectively, it is possible to outline a step-by-step process of PLP binding and enzyme stabilisation (Figure 1.9).
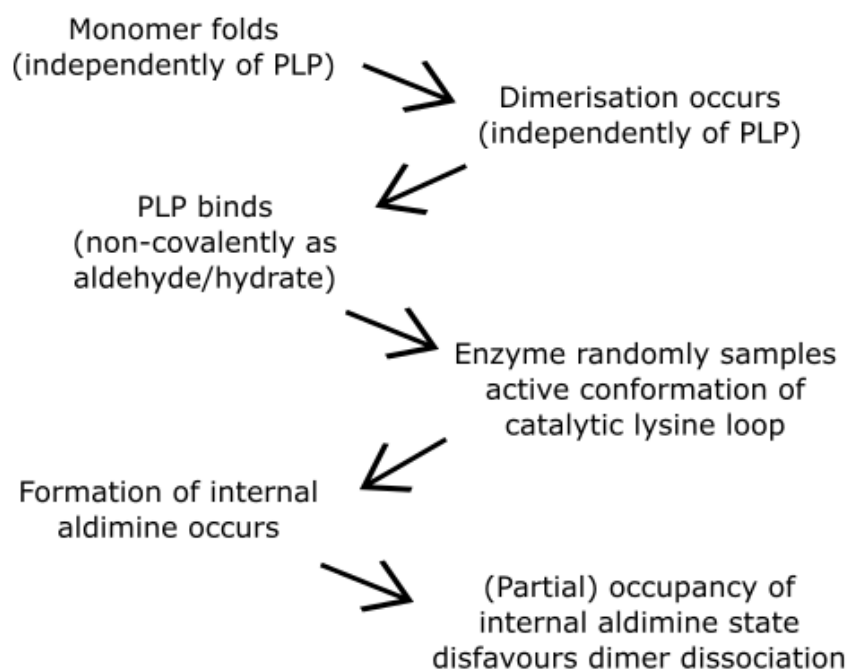
Figure 1.9 Step-by-step process of PLP binding and formation of stable active holo-dimer TA.

Building on these findings, Börner et al.[17] elucidated a substrate-driven inactivation mechanism reflective of the TA catalytic cycle. In the first half-reaction of the TA catalytic cycle, the external aldimine is formed, followed by the tautomerisation and hydrolysis steps, to leave behind the aminated form of PLP (which is pyridoxamine 5' phosphate, PMP) and the deaminated (i.e. carbonyl form) donor substrate (Figure 1.10). In the subsequent second half-reaction, this process is reversed, as the amine acceptor is aminated and PLP is regenerated to its original state. However, this second half-reaction is disfavoured in contexts where a high amine donor loading has been applied to work around unfavourable reactant/product equilibria, causing accumulation of the non-covalent E·PMP complex, from which PMP may dissociate.

Whereas the enzyme-PLP internal aldimine state would stabilise the active conformation and dimeric state of the enzyme, the apo-state would be relatively more prone to dimer dissociation, and the resulting monomers be more prone to unfolding and irreversible aggregation. This aggregation behaviour is not surprising, because it has been previously demonstrated using controlled denaturation, that folded monomeric units can exist in solution but present a larger hydrophobic surface, relative to the dimer, as inferred from the increased binding of hydrophobic dye[23]. For this reason, it is also apparent why the common practice of employing excess PLP in TA-based biocatalysis is helpful, in that it reinforces occupancy of PLP in the active site, despite high loading of the amine donor substrate.
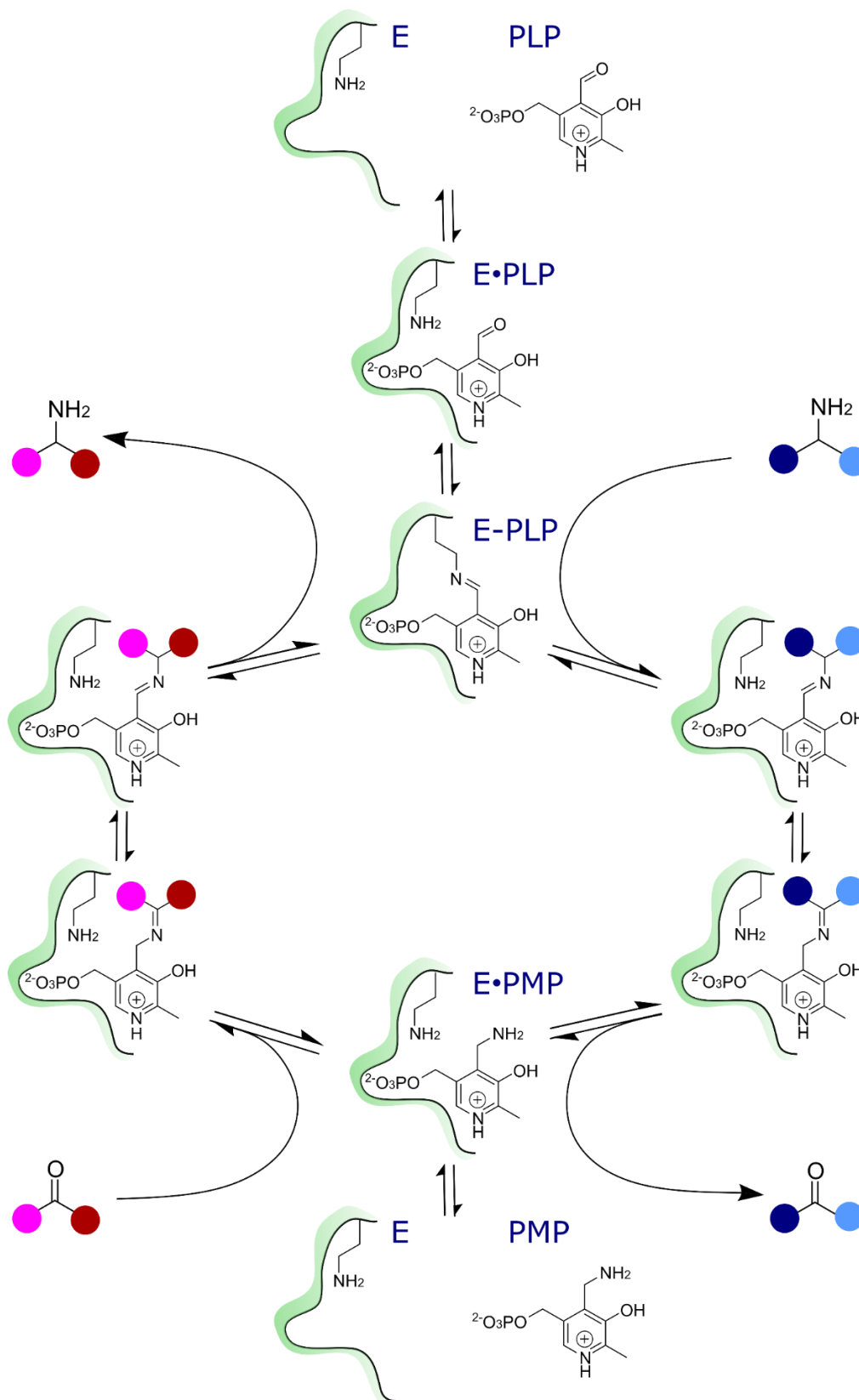
Figure 1.10 Changing states of cofactor PLP/PMP in the active site throughout the catalytic cycle. Excess of amine donor (blue) converts PLP to PMP, driving the reaction towards the E·PMP state, which can accumulate if the lack of amine acceptor (red) prevents the re-formation of PLP.

Börner *et al.*[17], convincingly illustrated this link between substrate loading, PLP dissociation, and enzyme inactivation, using three TAs, two of which were the well-studied dimeric TAs CvTA and VfTA, with the third being a tetrameric TA from *Pseudomonas sp.* (PxTA). Due to the different absorbance maxima of PLP and PMP, it was possible to photometrically follow the rate diminution of PLP and accumulation of PMP under different substrate loadings. It was found that higher loading of the amine donor substrate leads to faster inactivation, reduced T_m, and correlated to increased PMP accumulation.

1.3.2 Engineering transaminases for improved stability

Biocatalysis is hindered by poor enzyme stability, especially in contexts of high substrate loading, use of co-solvents, or elevated temperatures, and therefore stability improvement is a common target in enzyme engineering for biocatalysts. Ways of quantifying stability, for the purpose of comparing wild type against mutated enzymes, commonly include:

T_m the melting temperature of the enzyme, a thermodynamic property reflecting the reversible unfolding of the enzyme,;

$T_{1/2}$ the temperature at which enzyme activity is reduced by one half, when incubated for a given amount of time;

$t_{1/2}$ the incubation time at which enzyme activity is reduced by one half (the half-life), when incubated at a given temperature

Strategies for engineering stability can be broadly divided into directed evolution approaches, and rational (or semi-rational) approaches. In the former, no structural knowledge of the enzyme is strictly required; mutations are introduced randomly; and beneficial mutations are maintained either by selection pressure on the host, or are identified in a high-throughput screen. In the latter, structural information is used to guide the choice of sites for mutation, and often also to guide also the specific choice of mutation for each residue.

In the context of TAs, iterative rounds of library preparation and then HPLC-based screening can be used as the basis for a directed evolution strategy. This has been illustrated in an example from Novick *et al*[24]., whereby residues within 9Å of the active site were subjected to saturation mutagenesis. In a straightforward screening methodology, unpurified cell lysates were mixed with substrate and incubated; the products then solvent extracted, and the conversion assessed by HPLC. Favourable mutations can be incorporated as the basis upon which subsequent

rounds of mutagenesis and screening are based, such that the final enzyme contains multiple accumulated mutations.

In such an approach, no presumptions are made about which mutations will be beneficial, and the results depend on the incubation and selection conditions. For example, Savile *et al.* were unable to identify any TA with activity towards their substrate of interest; therefore, they adopted a "substrate walking" approach, first evolving their enzyme to accept a similar substrate, subsequently evolving it onto their substrate of interest and optimising the conversion for that substrate[4].

Rational design approaches include rigidification of the enzyme structure directly by introducing targeted bonding interactions, and the approaches of consensus mutagenesis, and incorporation of non-canonical amino acids which indirectly achieve the same effect. These are all general approaches which have also been applied to examples of TA biocatalysts.

Rigidification views the protein as a polypeptide chain, of which the conformational state is limited by the rigidifying effects of the bonding present in the folded state, and these bonds compensate for thermal motion, favouring the folded state over the unfolded state. In this view, the most flexible parts of the protein are seen as the limiting factor in the overall stability of the protein, and are targets for mutation. Identifying these sites can be done by using the crystallographic B-factor to infer flexibility. Alternatively, it can be done using molecular dynamics simulations, taking the root-mean-square fluctuation (RMSF) of each residue relative to its reference position, instead of the B-factor. The latter option can be made more efficient by constraining all covalent bond lengths except dihedral angles, for simulation. Introduction of hydrogen bonds, salt bridges, hydrophobic interactions, can be effective to increase rigidity. The introduction of covalent disulphide linkages can be especially effective, and proline substitutions can be used to rigidify protein backbone, due to the limited conformational fleixibility of proline[25].

Huang *et al.* applied this approach to TA engineering, calculating folding free energies for prospective mutations to the identified sites, taking forward the predicted best mutations to be validated, and achieving improvements in $T_{\frac{1}{2}}$ of up to $5°C$[26].

Consensus mutagenesis is an indirect extension of the view of protein folding in terms of rigidity. In this approach, variant versions of the same protein (homologous proteins from different species, for example), represent an ensemble. It is assumed that random mutations amongst the ensemble are more likely to be de-stabilising than stabilising, but that these

destabilising mutations are tolerated under selection pressure, as long as the destabilising effect is not too great. Therefore, for any given position in a sequence alignment, it is considered that the more frequently any one residue appears (as a proportion out of the total number of sequences), the more likely it is to represent a stabilising contribution to the overall folding free energy, to the stability of the folded state over the unfolded state[27]. Using a consensus mutagenesis approach to introduce only two mutations, Xie *et al.* were able to extend by 6-fold the $t_{1/2}$ (at 40 °C) of an *Aspergillus terreus* TA, and by introducing an additional disulphide bond, were able to achieve a 16-fold improvement versus the wild-type [28].

Incorporation of non-canonical amino acids can be achieved by two different methods. One method is to effectively expand the genetic code, by introducing a plasmid conferring the host with the ability to make a novel tRNA carrying a supplied non-canonical amino acid, which is then incorporated site-specifically wherever the corresponding codon is engineered into the overexpressed protein sequence. Alternatively, the host's ability to produce one of the canonical amino acids can be knocked out, and a sterically similar non-canonical substitute supplied, which is then incorporated globally, through the normal tRNA synthesis and normal translation processes[29].

Commonly, the latter method is employed using fluorinated analogues of canonical amino acids. Deepankumar *et al.* applied this approach using global substitution of (4R)-fluoroproline to stabilise a *Sphaerobacter thermophilus* TA and achieve a doubling of the $t_{1/2}$ from 2.2 hours to 4.5 hours, at 70 °C[30]. The stabilising effect of the fluorinated analogues would be considered to arise from the features of the C-F bond. The electronegativity of fluorine leads to it being able to form hydrogen bonds, but the presence of the "σ-hole" on the crown of the fluorine atom, also allows it to participate in halogen bonds[31].

Whilst approaches such as these apply to rigidifying the protein structure itself, it is also possible to target mutations in the protein to stabilise enzyme-cofactor interactions. The same group of Börner *el al.* which identified the relationship between substrate loading, cofactor loss, and instability in their three studied TAs, also exploited their methodology to demonstrate that mutations which reduced cofactor dissociation also conferred improved T_m and $t_{1/2}$ values, with reduced sensitivity of the T_m and $t_{1/2}$ values towards amine donor substrate loading, as compared to the wild-type, and without significant loss of specific activity[32].

1.4 Activity and stability in context of transaminase oligomerisation

1.4.1 Background to higher oligomerisation states in transaminases

Amongst the TAs used in the key study of Börner *et al.* establishing the PLP dissociation-based inactivation mechanisn, were two examples of two dimeric TAs (*Vibrio fluvialis* VfTA and *Chromobacterium violaceum* CvTA) and a tetrameric TA (*Pseudomonas* sp. PxTA), the latter of which showed relatively lower sensitivity to high substrate loading conditions, in terms of cofactor dissociation and inactivation. Though fold-type I PLP enzymes are obligate dimers, they can also exist as assemblies of dimers, including tetramers, and more rarely also hexamers and even dodecamers[33]. Comparing crystal structures of CvTA, VfTA and PxTA to explain their findings, Börner *et al* concluded that inter-dimer contact in PxTA is likely to impede PLP dissociation. Specifically, inter-dimer contact confines the conformational space available to one flexible loop which forms the active site wall next to PLP's pyridinyl moiety, a conformational change (away from the holo-structure conformation) in this loop being necessary in order to allow egress of PLP from the active site.

These observations suggest that oligomerisation effects arising from multi-dimer assembly in TAs may need be taken into consideration when considering the selection and engineering of TAs suitable for biocatalysis. This would especially be the case if multi-dimer assemblies bind or retain PLP through mechanisms differing from those seen in single-dimer structures. However, these questions are little explored beyond the key work of Börner *et al.*.

As described in the above subsection (1.3.2), the same group has successfully targeted cofactor binding by mutation of cofactor-binding secondary structure elements, but without exploring effects which may be specifically attributable to the formation of higher oligomerisation states. Likewise, whilst another group has directly targeted mutation of the TA intra-dimer interface as a means of attenuating the established monomerization-based inactivation pathway[34], no engineering attempts targeting inter-dimer interfaces have appeared in the TA literature. This group used a mixture of folding free energy calculation and consensus approach to increase T_m of their *Pseudomonas jessenii* TA by 23 °C with a sixfold mutant, the authors also noting that individual mutations made on the subunit interface had a much higher chance of introducing a stabilising effect (56% success) than mutations chosen using the same methodology, but on the non-interface solvent exposed surface (6% success rate).

Accordingly, to inform any investigation into the effect of oligomerisation state on the performance of class III TA biocatalysts, the existing literature for tetrameric fold-type I

structures was reviewed, with interest especially in enzyme-cofactor relationship. Higher oligomerisation states (hexamer, dodecamer) were excluded due to their paucity within the literature.

To help ensure relevance and reliability, only literature for TAs with at least one item of biophysical evidence for tetramerisation was reviewed, thereby excluding some purportedly tetrameric TAs available in the PDB.

From a crystal structure, the oligomerisation state of an enzyme can be predicted using various programs which use different combinations of parameters (e.g. calculated binding energy, or count of inter-residue contacts, interface area etc.) to evaluate interfaces in the lattice[35]. By reference to a training set of structures of known oligomerisation state, a likely oligomerisation state for the query structure is inferred. Though an overall accuracy of ~90% can be achieved, distinguishing biological from crystallographic interfaces is non-trivial, and dimer/tetramer misassignment has been identified in particular[36]. Biophysical evidence has shown that TAs in high concentration can present non-natural oligomerisation states, in pre-crystallisation solution[37] or in the crystal[38]. Additionally, some TAs are also known to exist in dimer-tetramer equilibrium, such as the *E. coli* TA and human ornithine TA discussed below, examples of context-dependent oligomerisation which can be characterised biophysically under different conditions, but not inferred from the crystal structure.

1.4.2 Relevant biophysical techniques for verifying oligomerisation state

The most common technique seen in the literature reviewed was size-exclusion chromatography (SEC). SEC infers solute protein's mass from its hydrodynamic radius, in terms of the protein's ability to enter the small pores of a resin matrix, which depends partially on its shape, and can possibly be affected by non-specific solute-matrix interactions[39]. The protein mass can only be estimated by comparison of its own elution volume against those of other proteins of known mass.

Scattering techniques were also applied for some TAs, either as dynamic light scattering (DLS) which infers oligomerisation in terms of the mobility of the solute under Brownian motion[40], or as small-angle X-ray scattering (SAXS), which infers oligomerisation from the scattering effect of atom pairs within single solute particles, which was used only in one work. Whilst SAXS can be used *ab initio* to estimate a molecular envelope, SAXS data can also be fitted to models derived from a crystal structure to confirm what assembly is most likely.

Analytical ultracentrifugation (AUC) was used on four of the reviewed TAs. In AUC the solute protein is characterised either by the rate of sedimentation (sedimentation velocity) or by the change in sedimentation boundary position at different centrifugal forces. Provided estimates for the density of the protein, solvent density and solvent viscosity, the sedimentation profile depends on the mass of the protein and on its hydrodynamic properties, with heavier and more condensed solutes giving faster sedimentation. The molecular mass is calculated from the sedimentation coefficient (in Svedburgs, $1 \ S = 10^{-13}$ seconds) and a frictional ratio f/f_0, which accounts for the hydrodynamic properties of the solute.

1.4.3   Available literature on PLP binding in tetrameric fold-type I enzymes

One of the most well characterised tetrameric TAs reviewed is human ornithine TA (hOAT), though the quaternary structure of hOAT is dissimilar to other tetrameric fold-type I enzymes, as shown in Figure 4.1 (in comparative structural analysis section 4.1). hOAT exists in concentration dependant self-association of dimers, and the tetramer-dimer dissociation constant is lower for the holo-enzyme (0.070 ± 0.005 µM) than for the apo-enzyme (0.32 ± 0.04 µM), suggesting PLP association promotes tetramer formation[38]. Liu *et al.* also found using SEC, that their studied *Escherichia coli* TA (EcTA) exists as a dimer/tetramer mixture, and that the proportion of tetramer was increased by adding exogenous PLP[41].

A single mutation in hOAT (R217A) vitiated tetramerization, and the dimer showed absorbance, fluorescence and circular dichroism profiles which matched the wild type (WT), suggesting the 2°/3° structure was unaffected by the mutation. PLP affinity was not significantly different from the tetrameric WT, and towards the substrates L-ornithine and 2-oxoglutarate, its kinetic properties were comparable[38]. However, the retention of PMP was not assessed, and the different quaternary structures of hOAT and PxTA make the suggested mechanism of enhanced PMP retention in PxTA inapplicable to hOAT, since the cofactor-binding secondary structure element analogous to that identified in PxTA, does not participate in dimer-dimer interface formation in hOAT as it does in PxTA.

The available literature also calls into question whether the PLP-enzyme covalent linkage is the dominant stabilising factor in all TAs, as it is in the example of CvTA cited above. A tetrameric *Pseudomonas aeruginosa* TA (PaTA) structure solved by Sayer *et al.*[42] (PDB ID: 4B9B), showed a total PLP occupancy <80% and an internal aldimine occupancy <20%, and another tetrameric TA from *Pseudomonas putida* (PpTA, PDB 3A8U) also shows mixed aldehydic and internal aldimine forms in its active sites. Photometrically monitored reduction

of the PLP-enzyme imine bond in PpTA revealed only one equivalent of reactive PLP, suggesting the possibility that three of the four chains are not active[43]. Similarly, titration of the same enzyme with the suicide inhibitor gabaculine showed no more than 1.5 equivalents reacted, even after incubation with excess PLP [44] TAs which retain PLP mostly in the aldehydic form may form more of the internal aldimine at warmer temperatures[43] and this could explain the observation of Roura-Padrosa *et al.* that warm incubation can have an activating effect[45].

This predominantly non-linked PLP in the active site of PaTA occasioned the suggestion from Sayer *et al.* that the occupancy and state of PLP in different TAs might be dependent on the energetic requirements of attaining the whatever conformation may be necessary to accommodate the formation of internal aldimine. Speculatively, it would be possible that tetrameric TAs retain cofactor in non-covalent states (E·PLP and E·PMP) without needing to form the E-PLP linkage to improve PLP retention or stabilise the enzyme. However, some dimeric TAs also accommodate cofactor as E·PLP rather than E-PLP in the crystal structure, such as *Bifidobacterium kashiwanohense* TA (BkTA, PDB ID: 6JIX)[22].

It seems possible that tetramerization might be a feature not always designed to improve catalytic stability or efficiency, but to fulfil some other function. Speculative suggestions would include:

- It might be a way to regulate transamination activity in the presence of excess PLP, either in terms of having a low affinity for PLP as suggested by Sayer *et al*[42], or in terms of exhibiting inhibition by PLP as Börner *et al*[17]. observed for PxTA;
- It might be a way to help store excess PLP, since excess PLP can be toxic in the cell due to the reactive aldehyde moiety[7];
- It might be a way to help reduce de-phosphorylation of PLP in the cell, by having the PLP bound to a protein carrier[46].

Any of these roles would suit a TA which is then able to bind the PLP in all four active sites, but without attaining an active internal aldimine state, whether through a negative cooperativity effect or otherwise.

Overall, whilst individual examples can be found for tetrameric or dimeric transaminases showing more or less occupancy of the internal aldimine state, there is a lack of meaningfully comparable information for biophysically determined tetrameric or dimeric transaminases of

similar substrate specificities with data available for both cofactor occupancy and specific activity.

1.5   Aim of this PhD project

The general aim of this PhD project was to investigate and understand the effects of oligomerisation on transaminase stability and activity, in order that this would inform future TA engineering. The following objectives were formed:

- Identifying a suitable tetrameric TA for study. For this, the *Pseudomonas fluorescens* TA (PfTA) first described by Ito *et al.*[47] and used previously by a collaborating group at Nottingham was initially chosen as a candidate TA, on the basis of previous work by them providing some suggestion that it would be tetrameric.

- Obtaining biophysical evidence to confidently establish the oligomerisation state of the candidate tetrameric TA. Obtaining a crystal structure for the same.

- Use mutagenesis to vitiate tetramerization in candidate TA, generating several different dimeric mutants using as conservative mutation as possible, with computationally guided rational design. Obtain biophysical evidence to confirm changed oligomerisation, and crystal structures for each mutant. Assay WT and mutant TAs to compare resilience towards high substrate loading and specific activity and thus determine contribution of oligomerisation to these.

- Time permitting, according to progress on previous objectives, use mutagenesis to stabilise tetramerization in candidate TA, and/or induce tetramerization in other non-tetrameric TA, by means of mutagenesis or chimeric design, then obtain biophysical, crystallographic data for stabilised/tetramerised TA and compare with original using the substrate loading/specific activity assays.

2 Methods

2.1 Plasmid preparation, protein expression, and purification

2.1.1 Plasmids and site-directed mutagenesis

The N-terminal His-tagged PfTA WT was expressed from a pET-28b construct kindly prepared and provided by Dr. D. Roura-Padrosa. The pET-28b plasmid contains sequence for N- or C-terminal His-tag and for kanamycin resistance. Site-directed mutagenesis was performed according to the method of Naismith and Liu[48], in which the primers overhang each other at the 3' end, enabling each to bridge the nicked site in the synthesised DNA strand which has been extended from the opposite primer, helping to re-circularise amplified plasmid. The primer designs are given in Appendix 8.3. Primers were supplied by Merck. Annealing temperatures were chosen at 5 °C below the average of the calculated melting temperatures for the two primers (calculated for the non-overlapping section, and ignoring contribution from mismatched bases), incrementing ±2°C in cases where the initial PCR attempt failed (judged by failure to produce colonies upon transformation of PCR product).

Mutagenesis PCR was run using 250 ng of each primer and 44 ng of PfTA WT template plasmid, with 1 µL Phusion polymerase (NEB), 1 µL dNTP mix, 10 µL of 10x HF reaction buffer and 0.5 µL DMSO in total volume 50 µL and run according to the following protocol:

1) Initial denaturation, 95 °C, 5 minutes;
2) Denaturation, 95 °C, 30 seconds;
3) Annealing, (temperature as described above), 1 minute;
4) Extension, 72 °C, 7 minutes;
5) Return to step 2, for 35 cycles
6) Final extension, 72 °C, 10 minutes

Mutagenesis PCR products were digested with 1 µL DpnI (NEB) for 3 hours at 37 °C, and 4µL of this was then transformed into 20 µL aliquots of NovaBlue competent cells prepared using the Inoue method. Single colonies grown from LB agar were picked and grown in 5mL LB medium overnight at 37 °C, using kanamycin at 50 µg/mL to select for the plasmid.

MiniPrep kits (Merck) were used for plasmid extraction and purification, and plasmids were sequenced (SourceBioscience) under T7 forward and T7 reverse primers to verify presence of correct mutation.

2.1.2 Expression

Competent *E. coli* BL21-CodonPlus prepared using the calcium chloride method was used as the host, chosen based on previous experience in group of good yield from this host. Host was transformed by addition of 1 µL plasmid, rest period of 20 minutes on ice, heatshock at 42 °C for 45 seconds, rest period of 2 minutes on ice, addition of 500 µL LB broth with 1 hour incubation. 100 µL of resulting culture was applied to LB agar selection plates with 50 µg/mL kanamycin (to select for plasmid kanamycin resistance gene) and 25 µg/mL chloramphenicol (to select for host chromosomal chloramphenicol resistance gene). 30 mL LB broth cultures with the same antibiotic conditions were inoculated with single colonies and grown overnight at 37 °C.

10 mL overnight culture was used to inoculate each 800 mL LB expression culture, under the same antibiotic conditions. At target OD_{600} 0.7~1.0 expression was induced with 0.5 mM 1-thio-β-D-galactopyranoside according to the lac operon-based expression system. Expression was allowed to proceed for either 3 hours at 37 °C, or 18 hours at 20 °C, with the latter generally giving more reliably good expression. Harvested pellet was stored at -20 °C.

Expression from glycerol stock derived from expression-verified overnight culture was also attempted. Overnight culture was pelleted and resuspended in LB with 30% glycerol before storage at -80 °C. Expressions from glycerol stock were found to be less reliable in yield than from freshly transformed host.

2.1.3 Purification (including analytical SEC)

Cells were resuspended in buffer A (150 mM NaCl, 50 mM sodium phosphate pH 7.2, 0.1 mM PLP, 10 mM imidazole) and lysed by sonication. Lysate was filtered (0.45 µm syringe filter) and loaded onto a HisTrap HP 5mL chelating column (GE Healthcare, pre-charged with nickel sulphate). Buffer B (identical to buffer A, but with 500 mM imidazole and pH 7.5) was used to elute the protein, typically over a gradient of 30 column volumes, with a prolonged hold at 5% buffer B being found to improve the elimination of contaminants prior to elution of the desired protein.

Fractions were pooled after evaluation of purity by SDS-PAGE, against a 116kDa reference protein ladder (Thermo Fischer). Samples were held at 95 °C for ten minutes and run at 175 V, then stained with Coomassie brilliant blue G. Pooled fractions were desalted immediately using PD-10 column (Cytiva) if being stored overnight, as sample was found to be sensitive to

precipitation under the presence of imidazole from buffer B. Dialysis was found less successful than desalting, with sample precipitating on inner surface of dialysis membrane.

Concentrated (Vivaspin V20, 30 kDa PES membrane) pooled fractions were subjected to SEC using a pre-equilibrated Superdex 200 HiLoad 16/60 column (GE Healthcare), typically with 50 mM HEPES or sodium phosphate, pH 7.2, 150 mM NaCl, 0.1 mM PLP, at a flow rate 0.8 mL/min. Concentrations of NaCl and PLP were varied according to requirements of individual biophysical experiments. Glycerol at 1% was included in preparations for crystallization trays.

Analytical SEC was performed in like manner, using a pre-equilibrated Superdex 200 Increase10/300 column (GE Healthcare) at a flow of 0.5 mL/min, with a manual injection of sample from a 500 µL capillary loop. Calibration plots from reference proteins were plotted as the decimal logarithm of the molecular weight of the marker versus k_{av}, where:

$$k_{av} = (\text{elution volume} - \text{void volume}) / (\text{column volume} - \text{void volume})$$

This yielded an approximately linear plot from which analyte mass was interpolated.

2.2 Biophysical methods

2.2.1 Analytical ultracentrifugation

Samples were prepared using the usual purification protocol as described above, taken fresh from the preparative SEC step, and concentrated to the maximum concentration desired for data analysis. Twofold dilutions of the sample were made with concentration filtrate, the filtrate also being used as the reference buffer. Two-sector cells were loaded with 400 µL sample in the sample sector, and ~420 µL reference buffer in the reference sector. For data series with variable PLP concentration, both sample and reference were first spiked to the desired PLP concentration using a 10 mM stock of PLP prepared by dissolving PLP into an aliquot of the reference buffer.

A Beckmann-Coulter Optima AUC with a 6-hole rota was used to collect interference data, and absorbance data at 280 nm, at 60,000 RPM. Data was collected in sedimentation velocity mode, with scans in 20s intervals, until sedimentation was completed according to visual evaluation. Buffer viscosity was determined using a rolling-ball viscometer (Anton Paar). Partial specific volume was estimated from protein sequence using UltrascanIII[49].

Data analysis was performed in Sedfit[50]. Initial curation of the data was done to identify and exclude those scans which occurred after the sample had already fully sedimented, since the centrifugation was not stopped immediately after sedimentation was visibly finished. The

measured viscosity and density of the buffer, along with the estimated partial specific volume of the protein, were applied as parameters in data fitting. Initial attempts showed the absorbance data to be more tractable toward fitting than was the interference data, absorbance data being used for all subsequent attempts at fitting.

The fundamental principle of sedimentation underpinning data fitting, are the forces determining the motion of a solute particle[51]:

-The force acting on the particle to the apparent "gravity" of sedimentation, the product of particle mass and the gravitational force at the radius the particle is situated at;

-The buoyancy effect according to the mass of the displaced solvent;

-The frictional force acting against the particle moves in any direction;

-The effect of Brownian motion working against sedimentation

Initial fitting to a $c(s,f/f_0)$ model produced an estimated distribution with f/f_0 values lower than considered realistic by reference to other globular proteins[52]. The dimensionless f/f_0 frictional ratio is the ratio of the true frictional coefficient of the solute protein, f, to the frictional coefficient of a hard spherical particle of equivalent mass, f_0, and would typically be ≥ 1.2 for a globular protein[52]. All subsequent fitting attempts were therefore made using the simpler $c(s)$ model, with a fixed f/f_0 value of 1.2 imposed on the fitting, producing a realistic distribution on the single variable, the sedimentation coefficient, s.

In principle, it is possible to distinguish between heavy particles with large frictional ratio, and light particles with small frictional ratio, and which therefore have identical apparent sedimentation coefficients, according to the higher diffusivity of the latter, manifested in a wider and shallower boundary between the high and low concentration sections of the sedimentation profile. The boundary width and gradient is the means by which diffusivity (and hence also f/f_0) is quantifiable and fitted in the $c(s,f/f_0)$ model; a more diffusive particle having a wider and shallower boundary[51]. Whereas in the simpler $c(s)$ model, only the rate of movement of the midpoint of the sedimentation curve, corresponding to the sedimentation coefficient s, is fitted.

Successful data fitting was refined by further curation of the input data to exclude the very first scans for those samples where these first scans showed a large fitting residual, caused by small amounts of aggregated protein (the aggregated material having a large and non-specific

molecular weight, with very fast sedimentation, and affecting the earliest scans the most). This exclusion resulted in loss of no more than 4% of scans. In some samples, further improvement in fitting was obtained by sampling the sedimentation coefficient, s, on a logarithmic scale. This enabled the fitting model to account for small amounts of aggregate of very high molecular weight (at high values of s) using just a few widely spaced sampling points, whilst sampling more finely at the lower values of s, relevant for non-aggregated sample, to allow better precision on the value of s

2.2.2 Small-angle X-ray scattering

Samples were prepared as concentration series, with an ideal series being nominally a series of four concentrations; 10, 5, 2 and 1 mg/mL each in 50 µL volume. However, not all samples were obtained in sufficient yield to allow this. Minimally, therefore, samples were submitted as a series of three concentrations, formed as successive twofold dilutions. Filtrate from concentration was retained as reference buffer. All samples included glycerol at 5% to minimise radiation damage. All samples were flash-frozen, shipped on dry ice and run by beamline staff of ESRF, Grenoble, due to contemporary travel restrictions. Samples were run in batch mode.

Data reduction (averaging and buffer subtraction) was performed in the ExiSAXS automated pipeline, with Kratky plots showing a bell-like curve indicative of a folded protein. Since a crystal structure was available for PfTA WT, ab initio SAXS modelling was not attempted, but rather the FoXS server[53] was used to generate simulated scattering profiles for 1-,2-,3-,or 4-meric PfTA, and compare the simulated scattering to the observed scattering for each sample.

N-terminal His-tags were neglected in generating the simulated scattering data, though present in the real samples, because the small mass of the tag relative to the protein, makes little contribution to the scattering profile.

The closeness of fit, derived from the residuals $I(q)_{obs}$-$I(q)_{model}$ for all values of the scattering vector q, for each model, was expressed as a fit value denoted X^2. Two other parameters, one representing the excluded volume of the solute molecule (denoted c_1), and the other representing the density of its hydration layer (denoted c_2), were automatically adjusted by the FoXS fitting algorithm to minimise the X^2 fit value for each model, since the hydration layer contributes to scattering but its density may vary within a typical range. Fittings in which the X^2 fitting score could be minimised only by modelling an atypically dense hydration layer (to compensate for lacking protein mass) were rejected.

2.3 Crystallography methods

2.3.1 Crystallisation and data collection

Fresh (not defrosted) protein was used in all screenings and crystallisation experiments, and was supplemented with 20 mM dithiothreitol (DTT) before concentrating to target concentration of 8~12 mg/mL.

Initial screening was done using sitting drop vapour diffusion in MRC 96-well plates (Molecular Dimensions), with each well having sub-wells for reservoir and two sitting drops. Commercially available screens (LMB and PACT Premier, Molecular Dimensions) were dispensed at 75 µL reservoir volume. For initial screening, protein stock at two concentrations (typically 12 and 6, or 8 and 4 mg/mL) was dispensed at 400 nLvolume, and mixed at 1:1 or 1:1.5 volume ratio with reservoir condition, using a Mosquito liquid handling robot (TTP Labtech). Trays were incubated at 20 °C.

For PfTA WT, conditions G1 and G6 from the LMB screen gave the most promising results. Variation of the protein and precipitant concentrations in a sitting drop 24-well format with 800 µL reservoir, showed the optimal conditions to be 6 mg/mL protein stock (in 50 mM HEPES, pH 7.5, 300 mM NaCl, 0.1 mM PLP, 1% glycerol, 20 mM DTT) mixed in equal parts with the G1 condition (0.1 M sodium citrate, pH 5.6, 12% w/v PEG 6000,0.1 M lithium sulfate), in total 4 µL (i.e. 2 µL protein stock with 2 µL reservoir condition). Diffraction-quality single crystals grew within a week. Harvested crystals were soaked in cryoprotectant solution (six parts G1 condition with four parts glycerol) and flash-frozen. Data was collected on beamline I24 at Diamond Light Source at 100K.

For PfTA R215E, following the same general screening approach, sitting drop vapour diffusion plates were set up at 6.4 and 3.2 mg/mL and incubated at 20 °C. Conditions F3 and F4 from the LMB screen (Molecular Dimensions) produced the most promising crystal forms, as a 1:1 mixture with the R215E stock solution. Since the precipitants and pHs in the F3 and F4 conditions were similar (Table 2-1), optimisation was attempted by reproducing the F3 condition but varying the precipitant concentration (15%, 18%, 21%), also varying the protein:condition ratio and protein stock concentration, in 24-well format. Unfortunately, this did not yield any improvement, but original crystals from F3 and F4 conditions gave diffraction data. Data was collected on beamline I24 at Diamond Light Source at 100K, with the F4 condition yielding marginally higher data quality and being taken forward for processing.

Condition	Precipitant 1	Precipitant 2	Buffer	Salt
F3	18% w/v PEG 4000	-	100 mM sodium citrate pH 6.0	200 mM ammonium acetate
F4	14% w/v PEG 4000	6% v/v MPD	100 mM sodium/potassium phosphate pH 6.2	-

Table 2-1 Composition of LMB screen conditions F3 and F4 (Molecular Dimensions)

2.3.2 Data processing

For PfTA WT, integration and indexing were performed in the automated Xia2 dials pipeline. Molecular replacement was performed with the model of an uncultivated transaminase from *Pseudomonas fluorescens* sp. (PDB: 5LHA) in PHASER[54], with the calculated Matthews coefficients indicating that four copies of the PfTA WT chain would be the most likely contents of the asymmetric unit. Iterative model building in Coot[55] and refinement in Refmac5[56] were used to optimise the model, especially to extend missing N-terminal residues, and to model in PLP to active sites. Sulphate ions were also added to unexplained tetrahedral densities appearing at sites on the surface of the protein model. Secondary structure restraints were used throughout the refinement process, and NCS restraints were also used in the earlier cycles. Crystallographic waters were added automatically in the final stages of refinement and edited manually where necessary. Model building and refinement proceeded until improvements in R/R$_{free}$ could not be made, and all Ramachandran outliers, density fit outliers, and other validation outliers had been examined to ensure no improvement could be made.

For PfTA R215E, data integrated and indexed by the Xia2 3dii pipeline was taken for solution by molecular replacement using the deposited structure for PfTA WT (PDB: 6S54) in PHASER[54], finding only two chains in the asymmetric unit. The same iterative model building and refinement procedure was used as for the PfTA WT, except that unexplained tetrahedral densities were modelled as phosphate rather than sulphate, to reflect the content of the crystallisation conditions.

2.4 Molecular dynamics methods

2.4.1 System preparation and production dynamics

Initial structures for the mutants were prepared from the PfTA WT crystal structure in Swiss PDB, selecting for each mutated sidechain a conformation with minimal steric clashing. The

N-termini of PfTA WT and mutants were rebuilt in Coot according to the original sequence to complete missing residues, whilst PLP was deleted from all chains to afford the tetrameric apo-enzyme. These were processed in the H++ server [57] to assign protonation states for titratable residues at pH 7.2. All cysteines were modelled as reduced, since no Cys-Cys distances were close enough for cystine formation to be feasible.

Simulations were performed in the Amber suite using the standard Amber forcefield GAFF with ff14SB. The structures were solvated with TIP3P water, neutralised (requiring between 19 to 35 Na^+) and raised to 150 mM with Na^+ and Cl^-, in a truncated octahedral box with a 10.0 Å boundary between the edge of the model and the edge of the box.

The system was minimised with 500 steps of steepest descent and 500 of conjugate gradient, then heated to 300 K over 80 ps and held at 300 K for a further 20 ps, using the Langevin thermostat at 1 collision ps^{-1}, in NVT ensemble with harmonic restraints of 100 kcal/mol. Equilibration and production dynamics were run in NPT ensemble and without restraints, using the Amber SHAKE algorithm to constrain all bond lengths for bonds to hydrogen atoms, in order to improve simulation efficiency.

The density of the system was plotted to determine when the system was sufficiently equilibrated. The trajectories (each 206.05 ns in total, for the mutants) had all reached an equilibrium density after roughly 50 ns, therefore, the first 56.05 ns of each was discarded, leaving three replicates of 150 ns for each mutant. The same process was followed for the PfTA WT system, leaving a single simulation of 500 ns for PfTA WT. Simulation progress was recorded every 2 ps.

2.4.2 RMSD and RMSF analysis

RMSD was plotted for each trajectory, using Cα positions only, and using the initial model derived from the crystal structure (prior to minimisation) as the reference. Per-residue RMSF was also plotted using Cα positions only, in order to represent fluctuation in the backbone position rather than mobility of the sidechain conformation.

2.5 Stability assays

2.5.1 Differential scanning fluorimetry

PfTA was prepared for DSF by defrosting aliquots and desalting (PD-10 column, Cytiva) into 50 mM HEPES, pH 7.5, 150 mM NaCl. This stock was concentrated to 1 mg/mL and mixed with 4 µL Sypro Orange (x5000 stock, Merck) to give a final dye concentration of x20. This

was dispensed at 10 µL per well into 96-well plates (Applied Biosystems), after which the buffer condition was added to the plate at 40 µL per well, from a 96-well block prepared in advance at x1.25 concentration. The plate was sealed with optically clear film, incubated at room temperature for twenty minutes, and the contents settled by centrifugation at 300 x g for 3 minutes. The DSF assay was then performed in a real-time PCR instrument (Applied Biosystems), with an initial incubation at 25 °C for 5 minutes, followed by a step-and-hold ramp from 25 °C to 95 °C in increments of 1 °C, held for 1 minute each increment.

The fluorescence intensity data was processed in NAMI[58] to identify T_m for each well. Checking fluorescence plots by eye was required to ensure that T_m was identified correctly, because for some data, the fitting had failed to identify real inflection points, whilst for other data the fitting had mis-identified noise as inflection points, leading to the mis-assignment of secondary melting points for the protein. Whilst melting can occur over several phases for proteins with more than one domain, inspection of all the plots did not reveal any secondary melting point for PfTA.

Data was collected for all conditions in triplicate (with the exception of a data collection failure for one single well in the R215E plate, which was therefore only measured in duplicate) and the sample standard deviation was taken to use as the uncertainty value on each T_m. Uncertainties in ΔT_m values were therefore propagated from the independent $T_{m\ (condition\ A)}$ and $T_{m\ (condition\ B)}$ uncertainties according to the formula:

$$\sigma_{AB} = \sqrt{\sigma_A^2 + \sigma_B^2}$$

2.5.2 Activity assay

The standard assay used for measuring transamination activity was adopted from Roura-Padrosa *et al.*[59]. To 500 µL of a solution of freshly mixed (S)-methylbenzylamine (SMBA, 2.5 mM), sodium pyruvate (2.5 mM) and PLP (100 µM) in sodium phosphate buffer (50 mM, pH 8.0), was immediately added 5 µL of enzyme solution. SMBA stock solution was kept tightly sealed when not being dispensed, to avoid evaporative loss of SMBA from solution. Formation of acetophenone was monitored at 245 nm in a quartz cuvette using an Implen NanoPhotometer NP80, for two minutes at five second intervals. Samples were diluted as necessary with self-matching buffer, to bring photometric readings within the linear optical range of the instrument. An extinction coefficient of 12,600 $M^{-1}cm^{-1}$ was used for acetophenone. Freshly mixed assay

solution with no enzyme added was used to establish blank values arising from reaction of SMBA with PLP in the presence of no enzyme. Blank values established in this way were subtracted from the sample values; all measurements were made in triplicate using standard deviation as the attached uncertainty value, and the overall uncertainty propagated as the root of the sum of squares of the uncertainties of the sample and the blank. One unit of activity was defined as production of acetophenone at 1 µmol min^{-1}.

2.5.3   Incubation assays

For the experiment varying incubation temperature, frozen aliquots of PfTA WT or R215E were defrosted and desalted into sodium phosphate buffer pH 7.2, with 150 mM NaCl and 0.1 mM or 1 mM PLP, using a PD10 column (Cytiva). Protein was diluted with buffer to achieve a stock concentration of 1.4~1.5 mg/mL for each stock. Each stock was divided into 50 µL aliquots, kept on ice, the incubation periods for each aliquot to be started at staggered intervals

From the 50 µL aliquots, a control aliquot was reserved for the purpose of checking that sample activity would not be significantly diminished during the time window of the experiment.

The control aliquot was subjected to the activity assay to determine its activity. Following this, the first sample aliquot (60 °C incubation) was taken from the ice bucket and placed into a PCR machine being used for incubation. After half an hour, the second sample aliquot (50 °C incubation) was taken out from the ice box and placed into the other side of the PCR machine sample platform (platform gradient allowing 60 °C on one side and 50 °C on the other side). Likewise the third and fourth samples were started at staggered further half hour intervals, using a second PCR machine to incubate at 40 °C and 30 °C.

After the first sample had been incubated 2 hours, it was removed from the PCR machine, put onto ice for 20 minutes, and centrifuged for 10 minutes at 12,000 g in a refrigerated centrifuge to sediment any potential aggregation. It was then subjected to the activity assay described above, making dilutions of the sample with activity assay buffer where necessary, to obtain a linearity of absorbance change in the activity assay.

The remaining samples were processed in the same way as each staggered 2 hour incubation period came to an end. Once all sample aliquots had been processed in this way, the control aliquot which had remained on ice for the duration of the experiment, was subjected to the

activity assay again. All activity assays were performed in triplicate with the standard deviation used to quantify the measurement uncertainty.

For the experiment with variable substrate concentration, protein stock was prepared in the same way by desalting in buffer A (50 mM sodium phosphate, pH 8.0, 150 mM NaCl, 0.1 mM PLP), to give a stock at a target concentration of 4 mg/mL. Due to yield limitations, a stock at 4.1 mg/mL was obtained for PfTA R215E, but slightly less for PfTA WT, at 3.4 mg/mL.

Stock solutions of 2 M isopropylamine in buffer A, and 100 mM sodium pyruvate in buffer A, were prepared. The pH of stock solutions was adjusted to pH 8.0.

The experiment was performed with a control, and with staggered starts for each sample, as described for the variable temperature experiment above. However, for the variable substrate incubation assay, the initiation of the incubation period was the point at which each sample (40 µL) was supplemented with buffer A, and isopropylamine stock, and sodium pyruvate stock, in the proportions necessary to afford the desired concentration of substrates at a final volume of 100 µL.

Initial work with this variable substrate concentration experiment included a second and complementary supplementing of each sample at the end of the incubation period, to adjust each sample such that the final concentration for all samples would match the maximum substrate condition. The rationale for this was that the concentration of isopropylamine (1M) in the highest concentration sample, would be sufficient to carry over enough isopropylamine when diluted 1:100 in the activity assay, to raise the isopropylamine concentration in the activity assay to 10 mM, and potentially compete with the SMBA activity assay substrate, rendering the activity assay unfair. However, it was found in initial work that inclusion or non-inclusion of this amount of isopropropylamine in such a complementary supplementation made no significant difference to observed activity, and therefore no such complementary supplementation was included in the final form of the incubation assay.

3 Crystallographic and biophysical studies of PfTA

3.1 Introduction

To study structure and oligomerisation of PfTA WT, it was subjected to crystallographic and biophysical analysis. Crystallography provided a detailed view of the inter-dimer interface in terms of the individual atom-to-atom and residue-to-residue interactions, providing the opportunities for informed mutation of interfacial residues, and for structural comparison against homologous B6 enzymes. But attempting to infer oligomerisation from a crystal structure is less accurate than determining the oligomerisation experimentally by biophysical methods. Therefore, PfTA WT was also studied using AUC, SAXS and SEC, to establish more certainly its possible oligomerisation state(s) and whether this would be affected by PLP, inorganic phosphate, or salinity.

3.2 Summary of expression and purification of PfTA WT

Expression followed the description provided under Methods. For PfTA WT, a 3 hour expression at 37 °C was tried, and an overnight (18 hours) expression at 20 °C was tried, with the latter proving to give a better yield more reliably.

Purification followed the two-step IMAC and SEC procedure given under Methods. For crystallography preparation, a minor modification was made to the IMAC purification, by increasing the buffer concentration of NaCl to 1 M, as this was found to afford a small improvement in purity, likely attributable to a reduction in non-specific binding on the IMAC column. For SEC step prior to crystal tray preparation, 1% glycerol was included in buffer.

Whilst no calibration was performed for the preparative SEC column used in these instances of purification, plots obtained previously in the group for the same column model (HiLoad 16/600 Superdex 200 PG), suggested a likely MW range 74-96 kDa (Figure 3.1). The monomer of PfTA WT has an expected MW of 52.5 kDa, calculated from the amino acid sequence including the N-terminal His-tag.
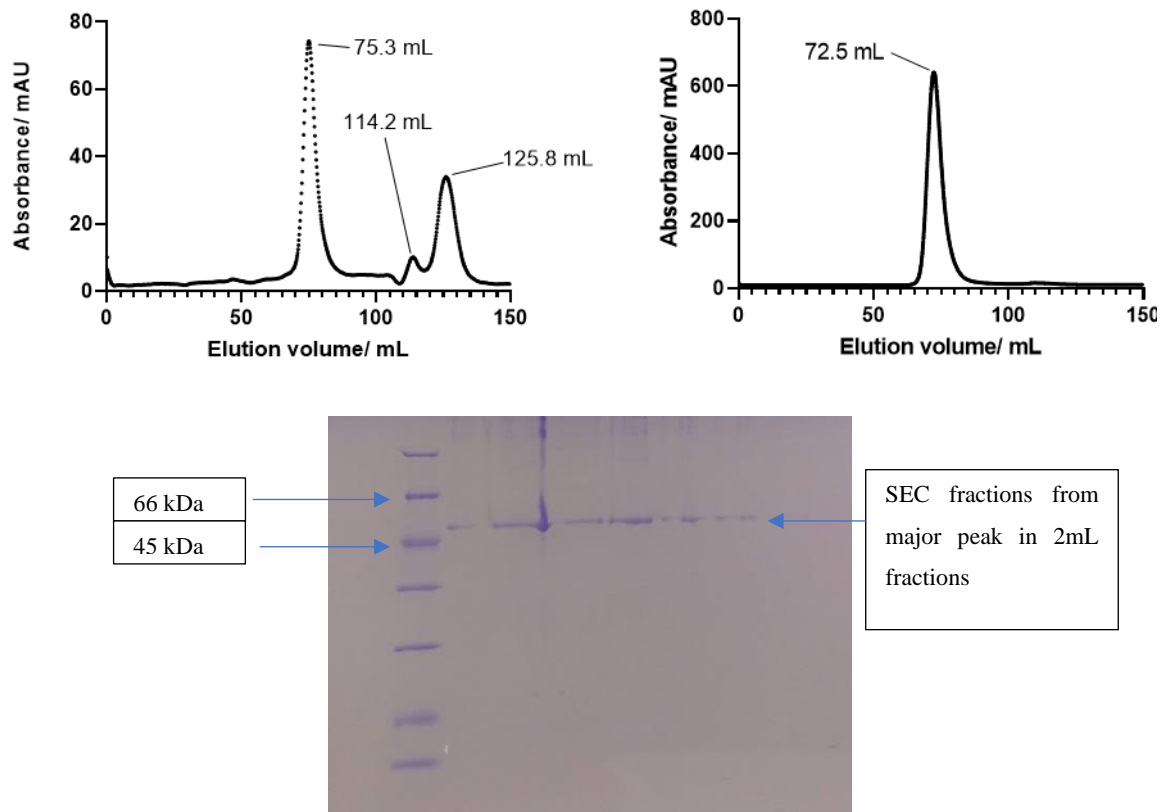
Figure 3.1 Two preparative-scale SEC chromatograms for PfTA WT (the major peak in each chromatogram eluting ~72-76 mL, HEPES 50 mM, NaCl 250 mM, PLP 0.1 mM, pH 7.5) with SDS PAGE showing elution between 66 and 45 kDa bands (expected 52.5 kDa). Protein yield varied significantly, with attempts to express from glycerol stock often producing inferior results; left and right chromatograms represent poor and good yields respectively.

The same method of purification was followed for the biophysical experiments, except where indicated otherwise in the results described. Specifically, where experimental conditions dictated, HEPES was substituted for sodium phosphate; or PLP was omitted, in the buffer conditions; or concentration of NaCl was altered to suit the requirement for final salt concentration in the sample.

3.3   Summary of PfTA WT diffraction data processing

Figure 3.2 Single crystals of PfTA WT formed under optimised conditions, and were identified by their yellow colour, imparted by PLP.

The quality of the map differed markedly across the asymmetric unit. The $2F_o$-F_c map was clear enough to trace almost the entirety of chains A an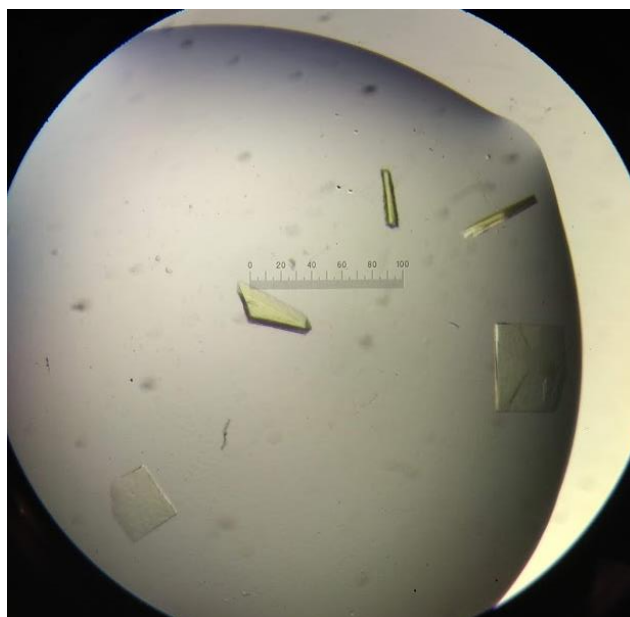d B when contoured to 2σ, whereas the map quality for chains C and D deteriorates further away from the (AB)-(CD) interface, and is not fully continuous at a 1σ contour. Unsurprisingly the B-factors for C and D were higher overall than for A and B. Fortunately, the discontinuities in the map for C and D were few and small enough, that it was possible to trace the backbone in continuity without breaks.

The reason for the dissimilarity in map quality is not entirely clear but is likely to be related to packing effects. The asymmetric unit is packed more closely around chains A and B (2081 atoms from neighbouring units are within a 10 Å cutoff) than chains C and D (1292 atoms within the 10 Å cutoff).

With manual rebuilding, some improvement at the N-termini was achieved, with each chain missing between 1-7 residues of the native sequence, in addition to the whole of the N-terminal His-tag. The model extends as far as the C-terminus for every chain.

	PfTA WT
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	92.02, 94.54, 242.22
Resolution (Å)	2.21
R_{pim}	0.131 (0.723)
$I / \sigma I$	3.9 (1.71)
CC1/2	0.973 (0.589)
Completeness (%)	99.8 (88.07)
Redundancy	6.5 (6.5)
Wilson B-factor (Å^2)	28.7
No. of unique reflections	106548 (7759)
Refinement	
Resolution range (Å)	88.07 – 2.21
R_{work} / R_{free}	0.211 / 0.243
No. atoms	
Protein	13838
PLP	60
Water	328
B-factors (Å^2)	
Protein	65.35
PLP	56.13
Water	38.86
RMS deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.062

Table 3-1 Data collection and refinement statistics for PfTA WT. Statistics for highest resolution shell are in parentheses.

The cofactor PLP was modelled into the active sites of all four chains to fit the positive F_o-F_c density present in the active site after molecular replacement. The F_o-F_c map was clearest for chain A and least clear for chain D, so Polder maps were prepared for both, in order to confirm the presence of PLP. The Polder maps for chains A and D (Figure 3.3) were consistent with

each other and clearly suggestive of the shape of PLP. The positive density in the Polder map extended beyond just the PLP itself, past the C4 atom of PLP, into the substrate pocket, but since no substrate was present in the crystallisation conditions, this density is likely an artefact of water in the substrate pocket. Whilst the $2F_o$-F_c map was clearly suggestive of a Lys-PLP linkage, an attempt was made to model aldehydic PLP anyway. The resulting F_o-F_c map clearly indicated the aldehydic oxygen atom did not fit properly, thus ruling out the aldehydic state and confirming the Lys-PLP linkage.

The model was validated with Molprobity[60] and deposited with PDB code 6S54.

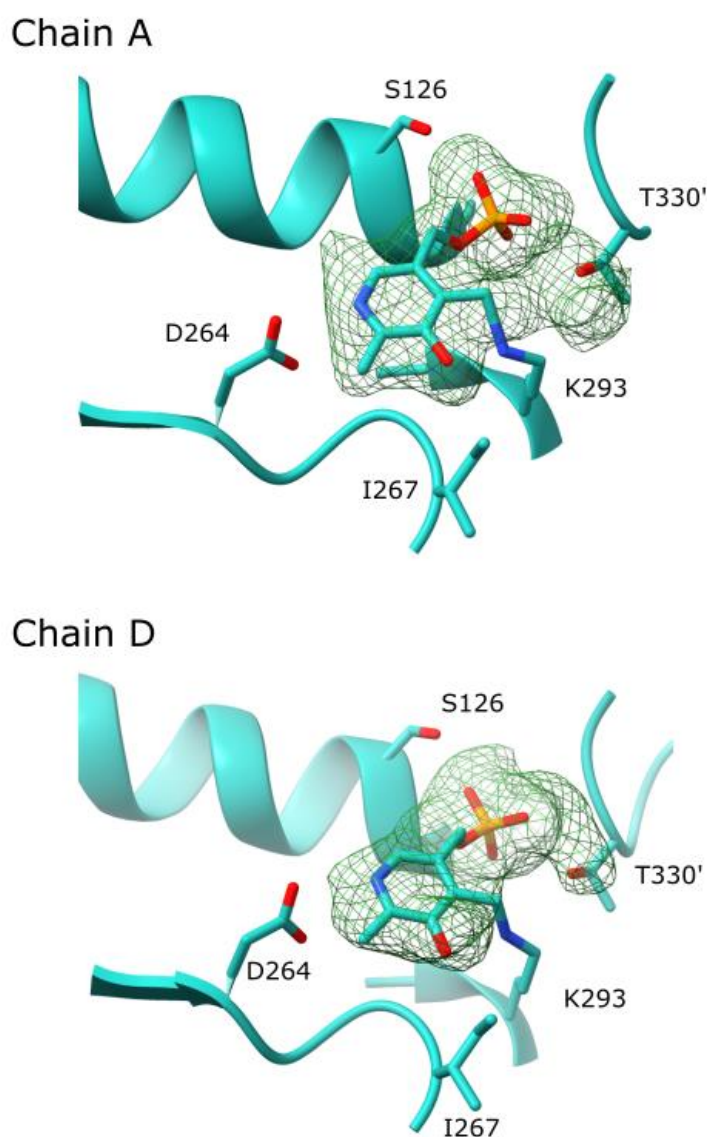Figure 3.3 Polder maps for PfTA WT active sites, prepared by omission of modelled PLP. Whereas a 3σ $2F_o$-F_c map around the chain D was lower quality than around chain A, and had not covered all atoms of PLP, the 3σ Polder maps for chains A and D were more clearly interpretable, consistent and supported all atoms of the PLP unambiguously in both chains other chains likewise showed the same presence.

3.4 Crystal structure description

3.4.1 Overall quaternary structure

The PfTA WT crystal structure presents four chains in the asymmetric unit, corresponding to a dimer-of-dimers structure with D2 symmetry. This can be represented schematically as in Figure 3.4, where dimer AB associates with dimer CD. Analysis of the interfaces using PISA[61] shows that the average interface area for intra-dimer interface AB (or its NCS equivalent interface CD) is 5412 Å^2, whereas the inter-dimer interface AD (and NCS equivalent BC) averages only 1089 Å^2. Inter-dimer interface AC (and BD) are very small, averaging 183 Å^2. The large extent of the AB interface, approximately 35% of the monomer total surface area, is reflective of an obligatory, permanent interface which is critical to enzyme function.
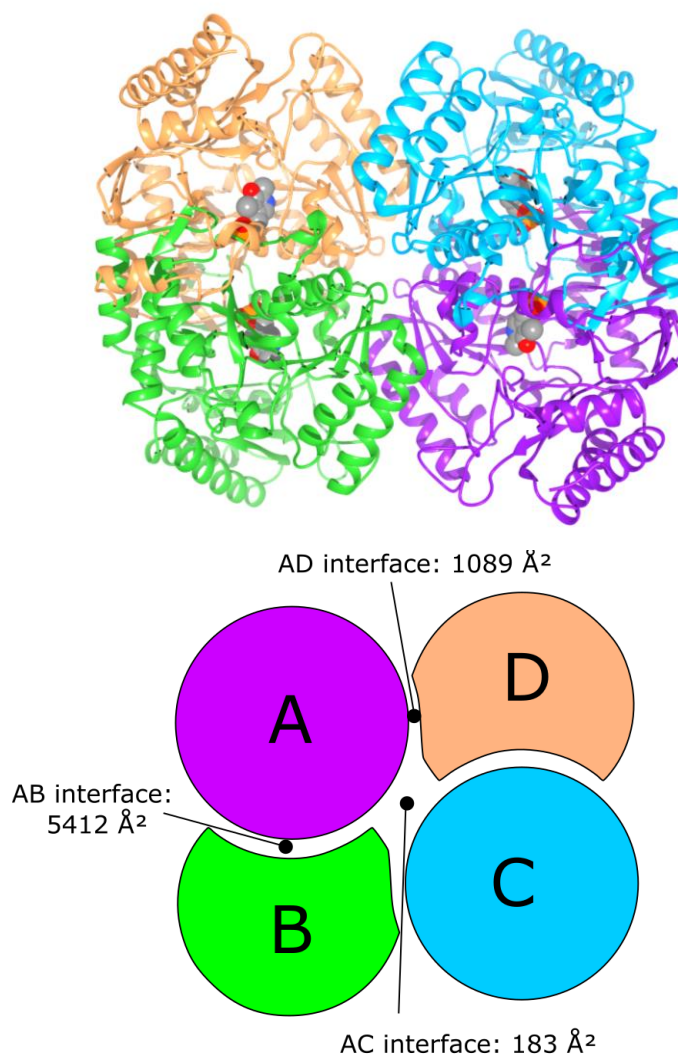

Figure 3.4 PfTA WT and its schematic representation as a dimer of dimers (AB) and (CD), the inter-dimer interfaces being significantly smaller in area than the intra-dimer interfaces.

Each chain adopts the expected B6-dependent fold-type I, with a small domain (corresponding to the N- and C- terminal sections of the sequence) and a large domain (corresponding to the middle of the sequence). The large domain is an α/β/α sandwich, the individual β-strands covering residues M116-G122 (β5), K150-L154 (β6), F185-V188 (β7), V225-A230 (β8), L260-D264 (β9), I288-L291 (β10), and S303-N308 (β11). The β sheet is arranged as β5(p), β11(a), β10(p), β9(p), β8(p), β6(p), β7(p), which gives the characteristic topology (+5x, -1, -1x, -1x, +1x, -2x) for the B6 fold-type I.
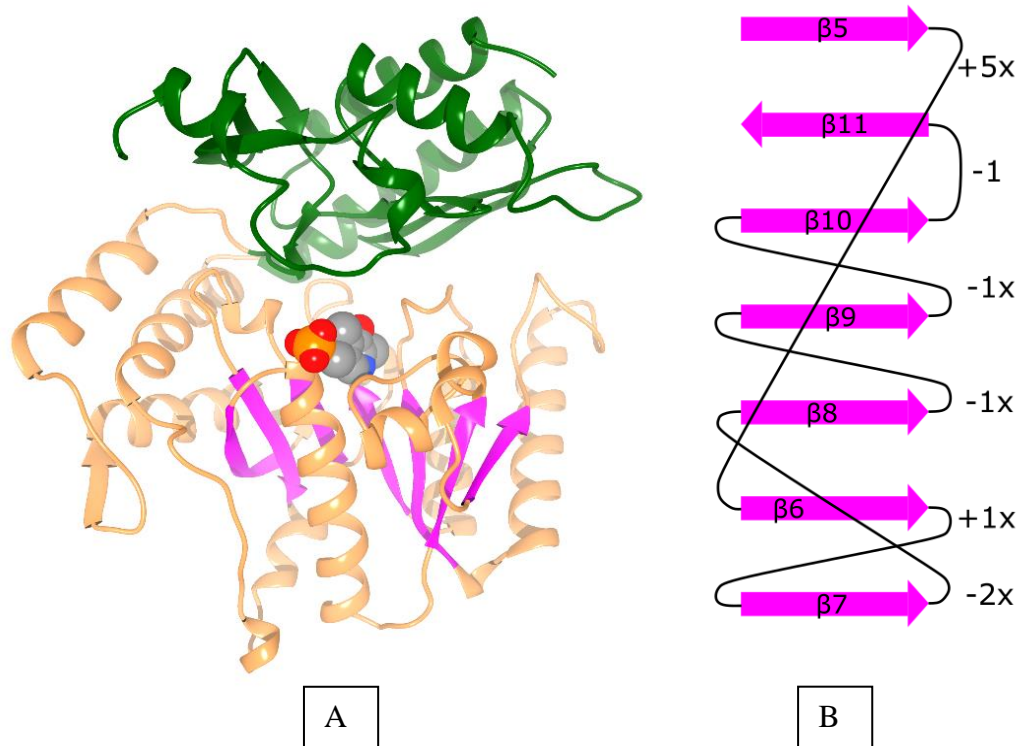
Figure 3.5 (A) Cofactor PLP (grey) sits in the active site which is positioned between the large (orange) and small (green) domains of a monomeric unit of PfTA WT. The core β-sheet is highlighted pink. (B) The topology for the core β-sheet is characteristic of the fold-type I B6 enzymes.

3.4.2   Active site

The active site sits between the large and small domains, however, the PLP forms contacts and hydrogen bonds with the large domain only. Its phosphate moiety bonds with the backbone N atoms of G125 and S126, which belong to helix α7. S126 sidechain Oγ is also positioned to bond to the phosphate moiety, and S126 is conserved as either serine or threonine amongst 16 members of a sample of 21 homologous fold-type I structures (Table 4-1). The binding pocket for the phosphate moiety of PLP is completed by T330' from the other intra-dimer chain (i.e.

the pocket of chain A is completed by T330' of chain B), which hydrogen bonds at atoms N and Oγ (Figure 3.6, Table 3-2).

Chain	Distance to nearest O atom of phosphate moiety, from by donor atom (Å)				
	G125 (N)	S126 (N)	S126 (Oγ)	T330' (N)	T330' (Oγ)
A	2.91	3.01	2.61	2.96	2.57
B	3.04	2.98	2.57	3.10	2.63
C	3.58*	3.01	3.17	2.90	2.72
D	2.58	2.52	2.67	2.94	2.69
Average	**3.01**	**2.88**	**2.76**	**2.98**	**2.65**
*Too distant to constitute hydrogen bond					

Table 3-2 Interatomic distances between PLP phosphate moiety and the protein residues which are close enough to form direct hydrogen bonds.
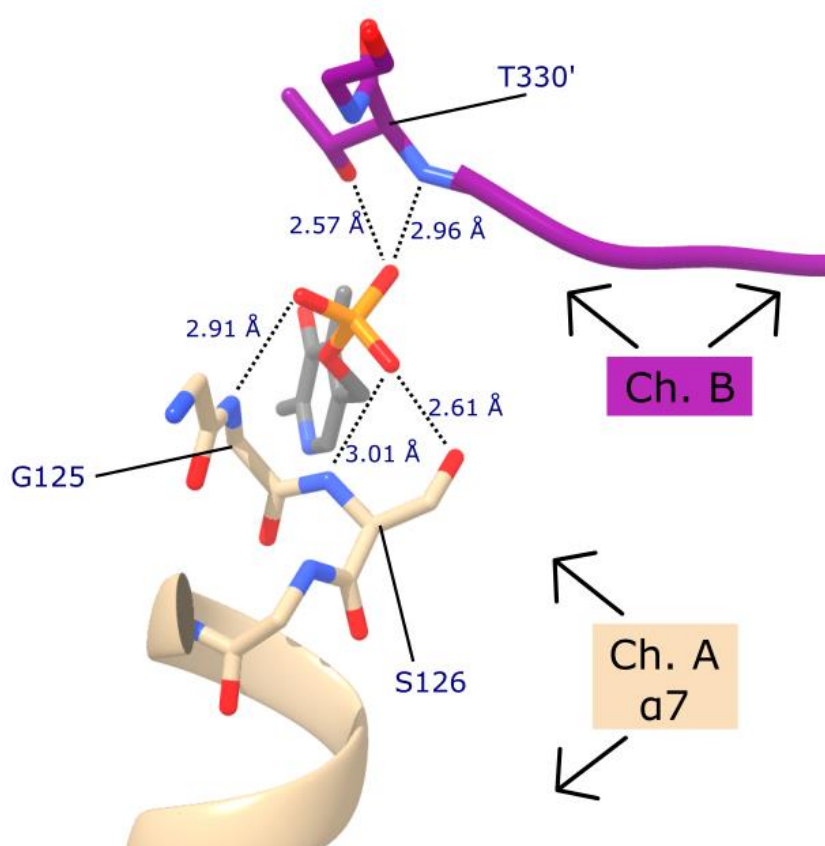
Figure 3.6 The active site of PfTA WT comprises residues from both members of the intra-dimer chains A and B (or equivalently C and D) in order to complete the phosphate binding cup.

The part of the PLP binding pocket occupied by the PLP pyridinyl moiety, is defined on one face by Y158, which is heavily conserved as an aromatic residue (as Tyr [17/21], Phe [3/21] or Trp [1/21], amongst 21 structurally similar B6 enzymes). The residue Y158 is part of a knotty sequence of three short loop-interspersed α-helices (α8, α9, α10) which here is defined as the "ring-binding motif" (RBM) after Börner *et al.*[17], because it has a conserved structural role in defining this part of the PLP pocket, adjacent to the PLP pyridine ring. The opposite

face of the pyridine ring is abutted by V266 (Figure 3.7). The RBM also includes a fully conserved His residue (H159), which together with D264 forms a relay which controls the protonation state at the PLP pyridine N atom, since the enzyme functionality depends on the electron-withdrawing properties of the pyridine moiety of PLP[59,62].
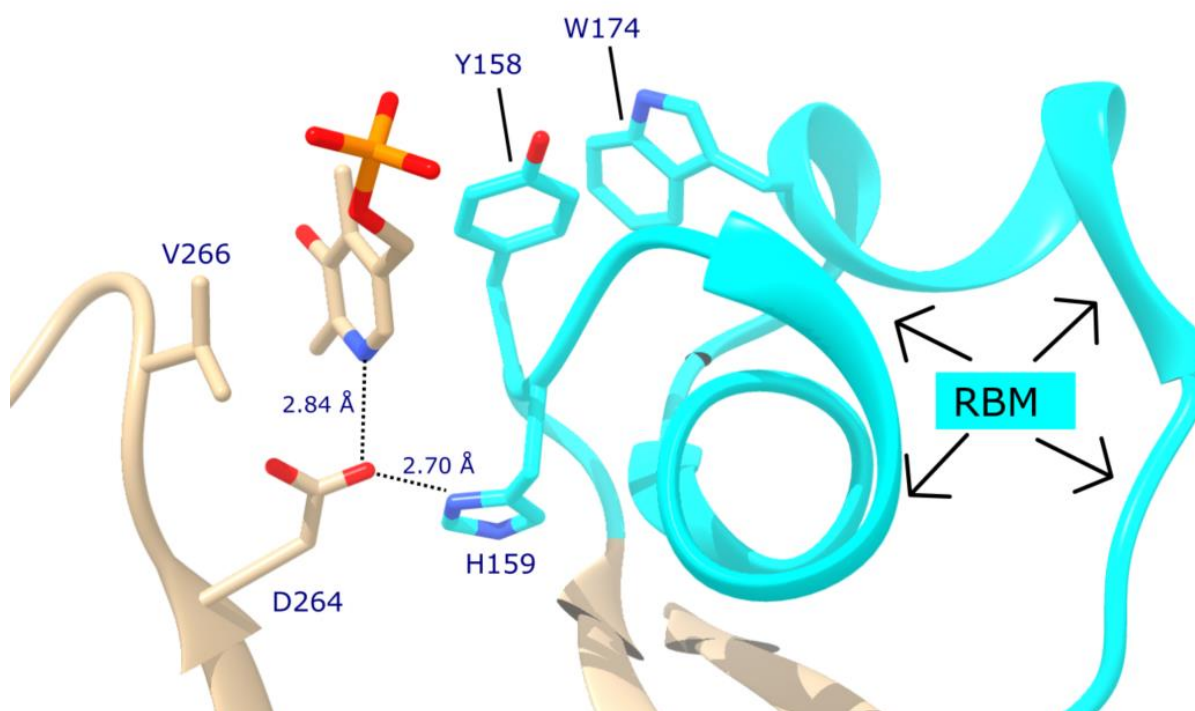
Figure 3.7 The PfTA WT active site features around the pyridine ring moiety of PLP. The ring is sandwiched between V266 and Y158, the latter's position suggesting the possibility π-π stacking interactions with PLP and with W174. The ring-binding motif (RBM) also features H159, in a relay with D264 to the pyridine atom N.

3.4.3 Tetramerisation interface

The program PISA[61] was used to identify the interfacial residues in PfTA WT which form the tetramer (i.e. inter-dimer) interface. An interfacial residue is defined as a residue which has a reduced solvent-exposed area because of the presence of another protein chain(s) in its vicinity, which is not the same protein chain to which the residue itself belongs. By this definition, it is not necessary for a residue to make intermolecular bonds or have contact with the other chain.

The interfacial residues are all within the large domain, specifically a contiguous section of chain between residues 125 to 230 (out of a total of 454 residues in each monomer, ignoring the His-tag). Figure 3.8 shows the distribution of the interfacial residues on one monomer of PfTA. In the PfTA secondary structure, this section spans helices α7-α11 and strands β6-β8. The section encompassing the small helices α8, α9 and α10 and the interspersed loops, is the ring-binding motif as defined above.
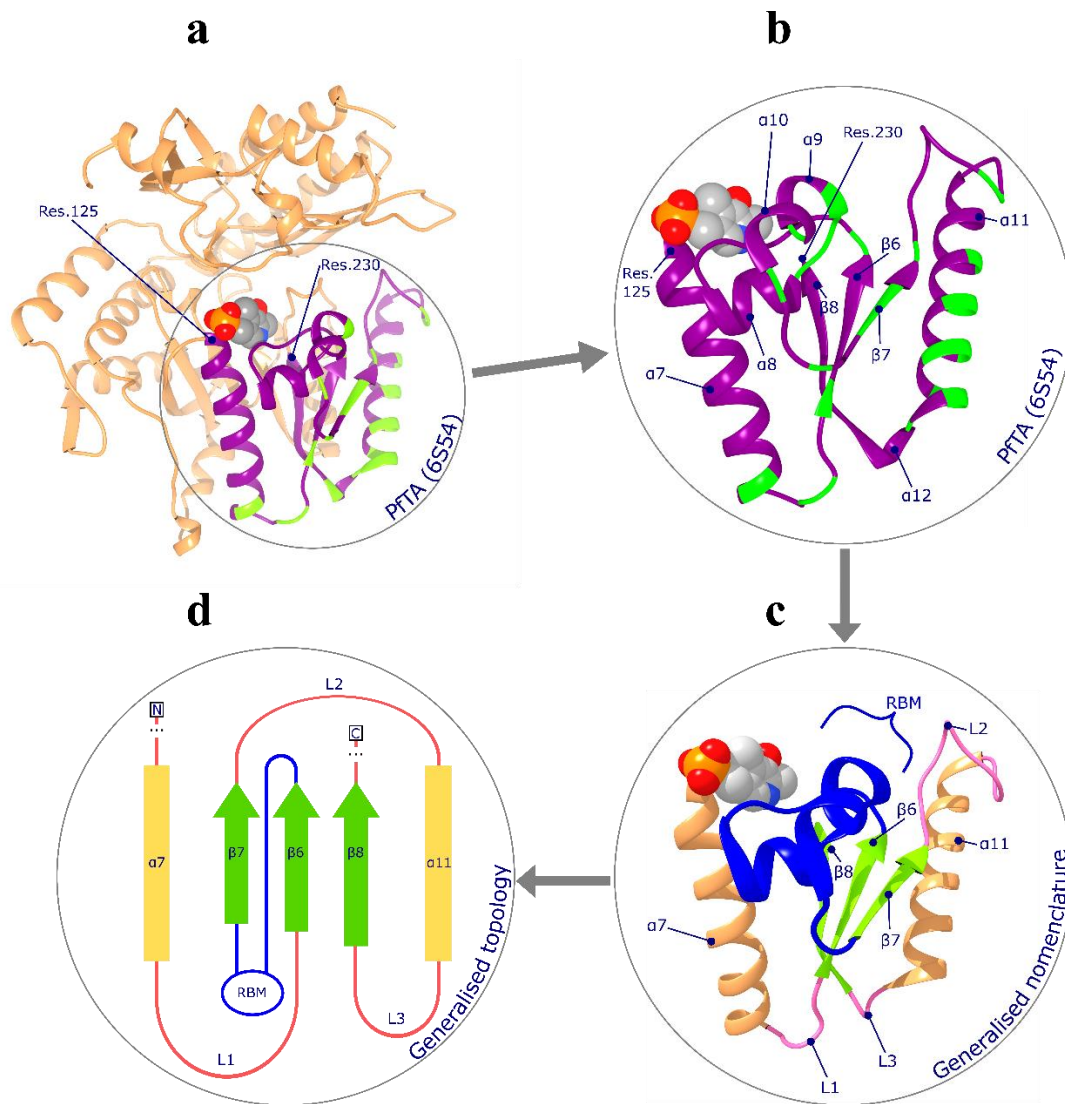
Figure 3.8 **(A)** The residues in one monomer of PfTA WT which form the dimer-dimer interface are highlighted in green, these all fall within a continuous section of chain from residue 125 to 230 (purple section), the secondary structure elements of which are labelled in the next panel **(B). (C)** Since the exact secondary structure of consecutive helices α8-10 is not conserved among homologous B6 enzymes, they are grouped together and defined as the ring-binding motif (RBM) according to custom in literature. The remaining loops L1-3 also labelled, and presented in a schematic view in the next panel **(D).**

The inter-dimer hydrogen bonds and salt bridges are shown in Figure 3.9, and all involve either the helix α11 or the RBM, which suggested these areas as likely targets for mutations to vitiate inter-dimer interface formation, especially noticeable being the R215-E189 salt bridge. The residues P197, F198, I208 and I212 (which are at the N-terminal part of helix α11 and the preceding loop L2) interface with the identical residues P197', F198', I208' and I212' in a hydrophobic interaction, another likely target (Figure 3.10).
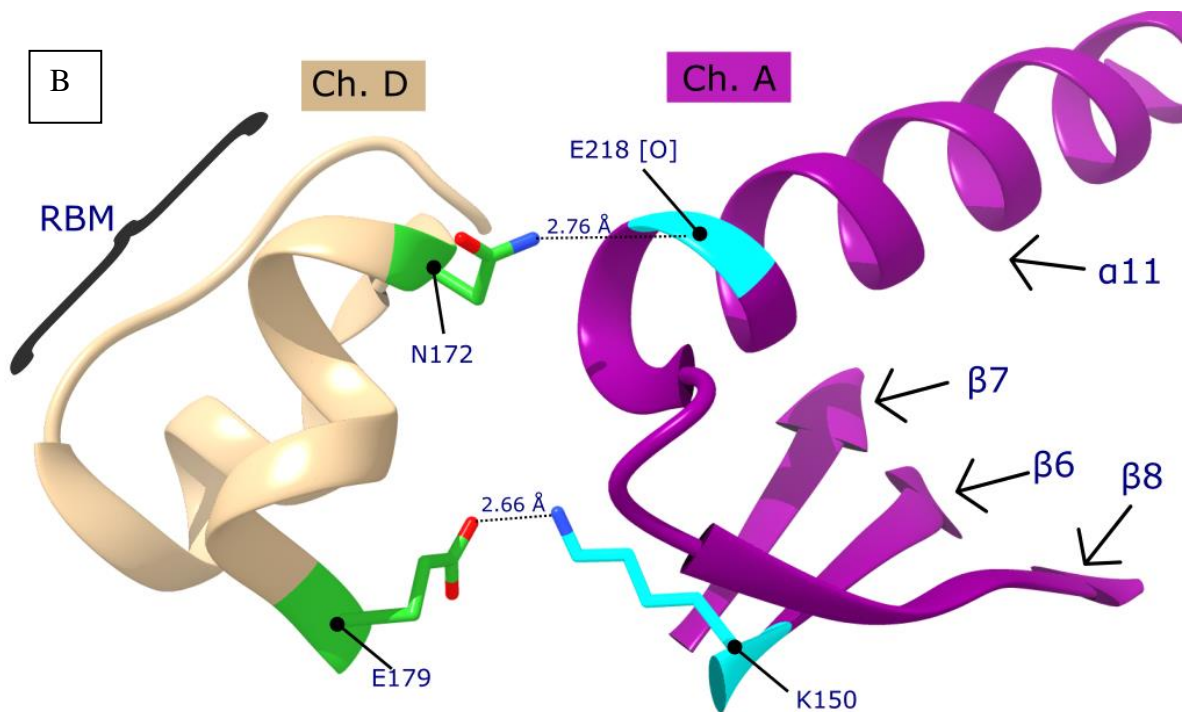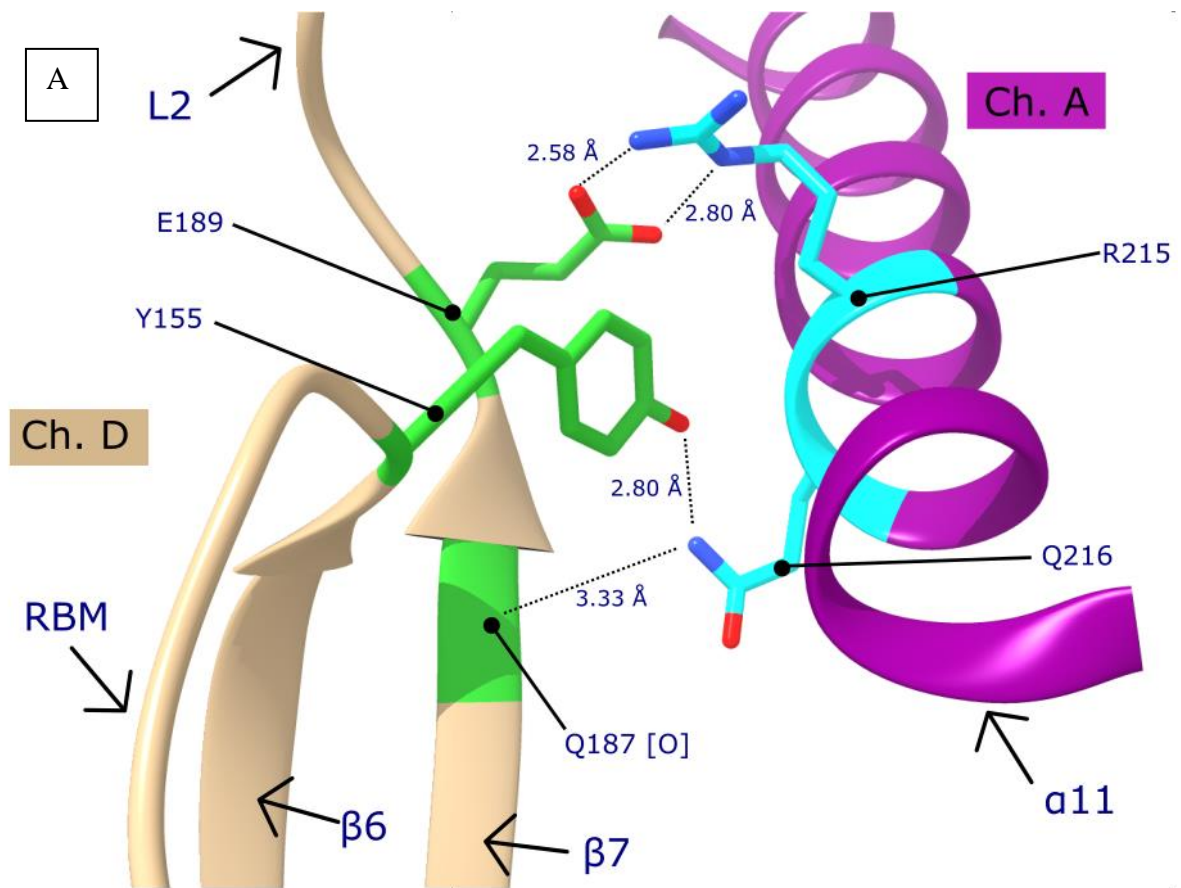
Figure 3.9 The nine individual residues forming inter-dimer hydrogen bonds and salt bridges in PfTA WT are distributed across several of the secondary structure elements, but all the bonds/bridges involve at least one residue from either **(A)** the α11 helix or **(B)** the RBM, or both.
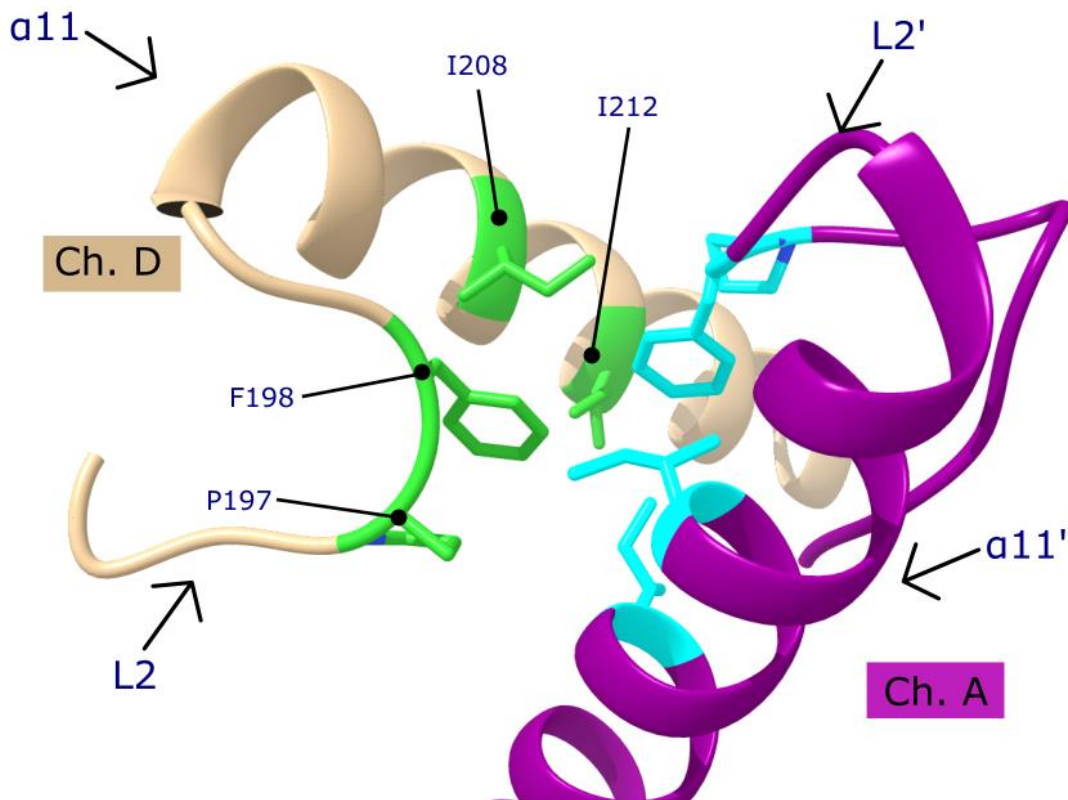
Figure 3.10 Amongst the several hydrophobic interactions contributing to the inter-dimer interface in PfTA WT, the largest single cluster is centred around L2 and the N-terminal end of α11, the residues of which interact with their non-crystallographic symmetry equivalents in the other dimer.

3.5 Biophysical characterisation

3.5.1 Size exclusion chromatography (SEC)

To obtain an estimate for the oligomerisation state of PfTA WT in solution, it was subjected to analytical SEC (using a Superdex 200 10/300 GL). This was repeated for different buffering conditions, to determine whether the presence or absence of PLP, or inorganic phosphate, would have any impact on the observed oligomerisation state.

It has been described previously for transaminases that PLP dissociation from within subunits can induce dimer-to-monomer dissociation of the subunits[15]. Conversely, it has been shown that PLP association induces tetramerization in SHMT2, a PLP-dependant transferase which shares with PfTA the same fold-type I structure, and which exists in a dimer-tetramer equilibrium[63,64]. There is also computational evidence that inorganic phosphate can partially mimic the stabilising effects of PLP, by binding in the same phosphate-binding pocket[65].

An initial attempt at analytical SEC, with phosphate buffer (50 mM) and PLP (0.1 mM), was dissatisfactory, showing multiple peaks (Figure 3.11). To calculate MWs, a calibration plot was prepared using standard proteins of known MW (Table 8-1), and this was checked for

consistency against plots produced by the column manufacturer. Based on this, the two major peaks had inferred MWs of 191 kDa, and 75 kDa. The former would correspond to a PfTA WT tetramer (expected at 210kDa), the latter would be unclear, since it falls between the expected MWs for the monomer (52.5 kDa) or dimer (105 kDa). This result stood in contrast to the previous preparative SEC results which had shown only a single major peak for PfTA WT.
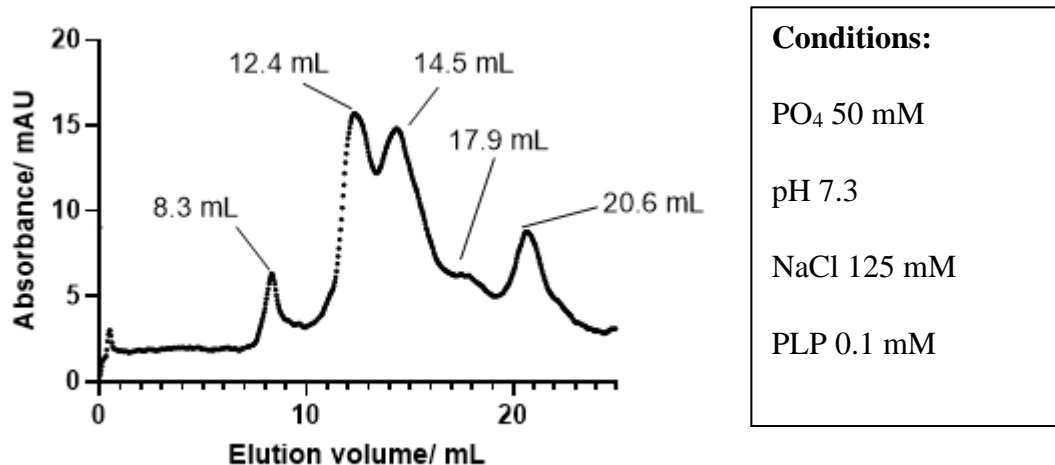
Figure 3.11 First (failed) analytical SEC of PfTA WT, with PLP and phosphate

Another analytical SEC, from a fresh purification, and under the same conditions (i.e. in the presence of both PLP and inorganic phosphate), produced a chromatogram (Figure 3.12) with a single major peak with a calculated MW of 167 kDa. This MW more likely corresponds to the tetrameric PfTA (expected MW 210 kDa) rather than the dimer (expected MW 105 kDa) because the discrepancy in mass would be smaller for 210 kDa than 105 kDa. The trimer (expected MW 157.5 kDa) is considered unlikely, even though it would seemingly be the closest match, since a dimer is the minimal functional unit, and therefore TAs exist as multiples of the dimer: as dimers, tetramers, or hexamers. Trimeric TAs are not reported in the literature. A trace peak, corresponding to a calculated MW of 449 kDa, was also present, but was disregarded because of its negligible size.
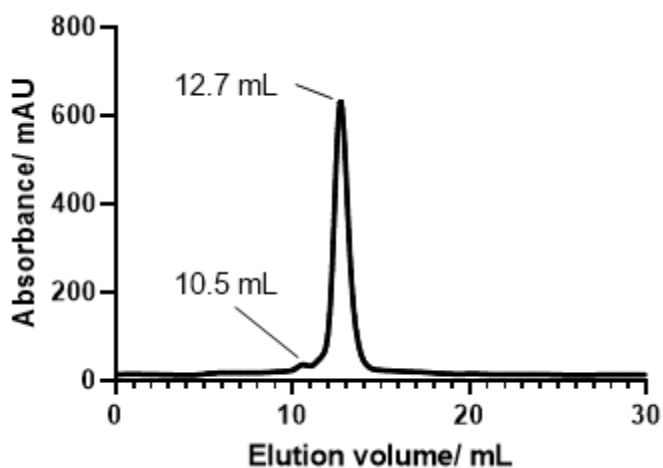
Figure 3.12 Second analytical SEC of PfTA WT, with PLP and phosphate

A A second aliquot of the same sample of PfTA WT was spiked to raise the salt (as NaCl) concentration to 1 M, to assess the possibility that oligomerisation might be influenced by the ionic strength of the solvent. The SEC chromatogram for this salt-spiked condition showed no significant change, with the major peak calculated to an MW of 153 kDa (Figure 3.13).
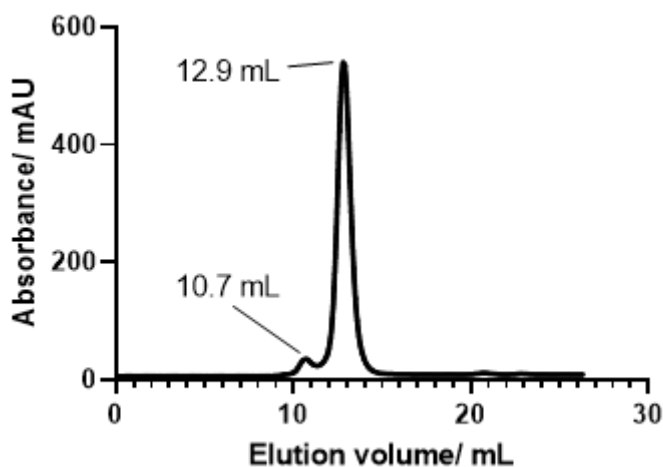
Figure 3.13 Analytical SEC of PfTA WT, with high-salt conditions

Another sample of PfTA WT was subjected to dialysis to remove the PLP introduced through the purification process, before the analytical SEC was performed. The chromatogram produced in this way, again showed no significant difference in elution profile (Figure 3.14).
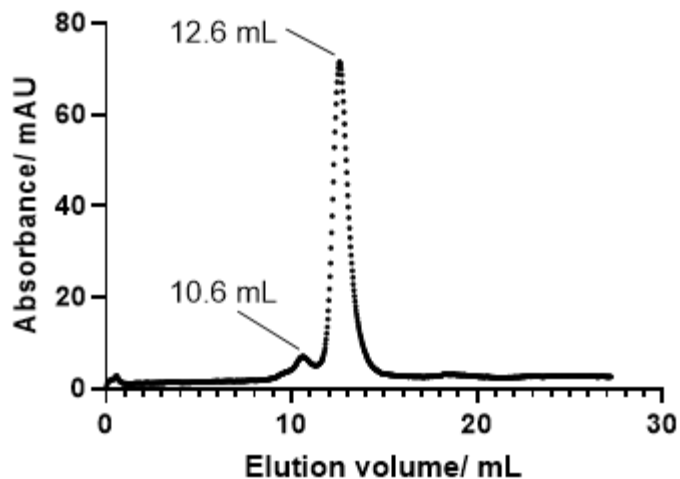
Conditions:

PO$_4$ 50 mM

pH 7.5

NaCl 125 mM

No PLP

Figure 3.14 Analytical SEC for PfTA WT following dialysis to remove PLP

Finally, PfTA WT was purified and subjected to analytical SEC in buffer devoid of both inorganic phosphate and PLP; throughout the whole purification process they were not used. The chromatogram for this again showed no significant difference in the elution volume, although it did show some tailing on the major peak (Figure 3.15). An aliquot of the same sample of PfTA WT was spiked with PLP, but this also produced no significant change (Figure 3.16).
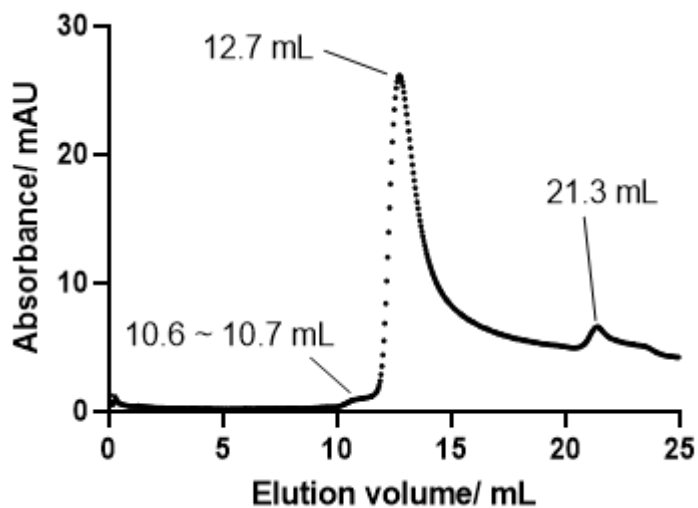
Conditions:

HEPES 50 mM

pH 7.2

NaCl 125 mM

No PLP

Figure 3.15 Analytical SEC for PfTA WT with no PLP and no phosphate
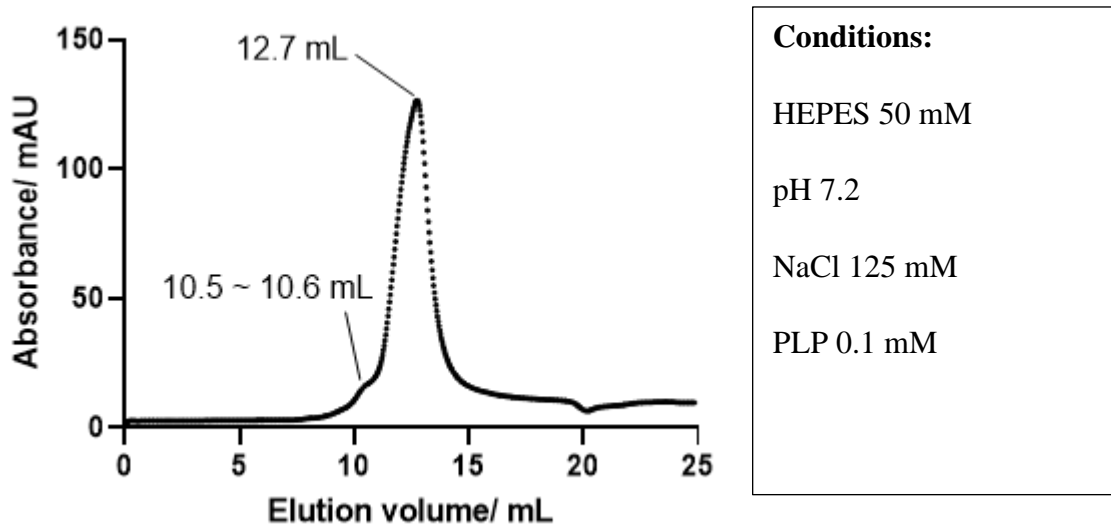
Figure 3.16 Analytical SEC for PfTA WT, purified without PLP or phosphate, but spiked with PLP just prior to sample application.

The SEC results taken together suggested that PfTA WT oligomerisation is not significantly affected by the presence of absence of PLP or inorganic phosphate, nor by the ionic strength of the solvent (within the timeframe of the sample preparation and SEC experiment). However, the MW values inferred from the elution volumes, were often more than 40 kDa below the expected mass of 210 kDa for a tetramer. Because of this, additional evidence of oligomerisation behaviour was sought from alternative methods, specifically AUC and SAXS.

3.5.2 Analytical Ultracentrifugation (AUC)

The first sedimentation velocity experiment (Figure 3.17) examined the oligomerisation behaviour of PfTA WT in the absence of both PLP and inorganic phosphate, but varied the concentration of the PfTA WT enzyme. The sample was prepared entirely in the absence of phosphate buffer and exogenous PLP.

A concentration series was used for the PfTA WT enzyme (8 µM, 4 µM and 2 µM), to assess for non-ideal solute behaviour. This refers to a situation in which the apparent MW of the protein varies with the concentration of the protein, because at higher concentrations, the protein self-associates increasingly, leading to an apparent increase in the sedimentation coefficient (and mass) of the protein. In such a case, it could be indicative of a concentration-dependant equilibrium between two (or more) oligomeric states. The modest protein concentrations used here (<10 mg/mL) were considered unlikely to introduce artefactual self-association arising from sheer quantity of protein, based on literature guidance[66].

The experiment found no evidence of non-ideality. The c(s) distribution plot for each sample showed a single major peak at around 8.7 to 8.8 S, for all three concentrations. When corrected for the solvent viscosity and density, the average value corresponds to an MW estimate of 178 kDa, with a minimal MW of 134 kDa, which reasonably rules out the possibility of a dimeric state (which would be expected at 105 kDa).

The c(s) model was used, with the frictional ratio *f/f₀* fixed at a value of 1.2, a typical value for globular proteins[52], and the values for sedimentation coefficient were sampled along a log scale, in order to allow the model to better account for small amounts of aggregated material without distortion of the fitting. Whilst this choice of model and analysis parameters worked best, as described under Methods, it is likely that inaccuracy in the value of f/f_0 is a significant factor in the observed discrepancy between the calculated MW (178 kDa) and the expected value (210 kDa for tetramer).
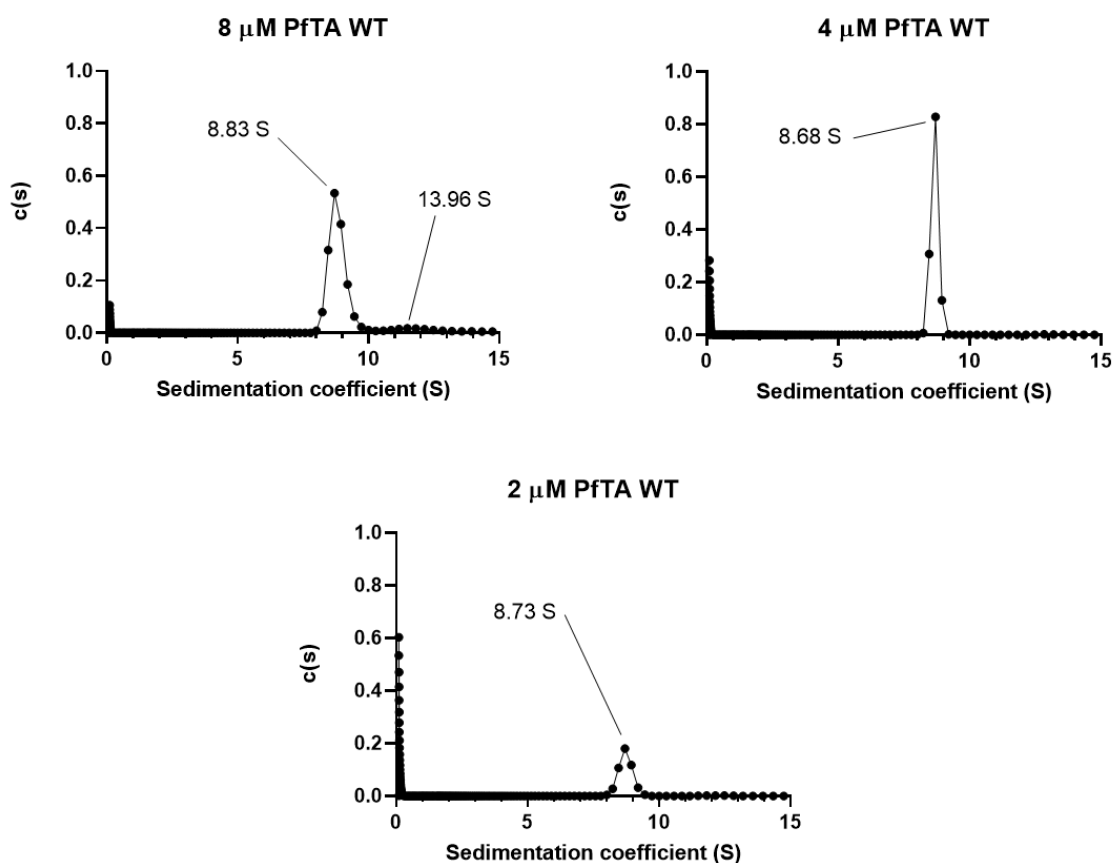
PfTA WT concentration (μM)	Sedimentation coefficient (S)	Sedimentation coefficient (20 °C, water) (S)	MW estimate (for $f/f_0 = 1.2$) (kDa)	Minimum MW (for $f/f_0 = 1.0$) (kDa)
8	8.83	9.41	181	138
4	8.68	9.24	177	134
2	8.73	9.30	178	136

Figure 3.17 AUC results for PfTA WT at three concentrations, in the absence of PLP and phosphate, producing an estimate of 178 kDa, and a minimal mass of 134 kDa, with no evidence of non-ideal behaviour. Values were produced from fitting to c(s) distribution in Sedfit.

For the second AUC experiment (Figure 3.18), the concentration of PfTA WT was kept constant at 4 μM, but PLP was added to each sample, to give a concentration series for PLP from 8 μM to 64 μM, a twofold to sixteen-fold excess of PLP.

The model and method of analysis were the same as for the first experiment. The sedimentation coefficient did not differ significantly across the different concentrations of PLP (average

estimated MW of 181 kDa and minimal MW of 136 kDa). These estimated and minimal MW values do not differ significantly from those produced in the absence of PLP altogether.

PLP concentration (μM)	Sedimentation coefficient (S)	Sedimentation coefficient (20 °C, water) (S)	MW estimate (for f/f_0 = 1.2) (kDa)	Minimum MW (for f/f_0 = 1.0) (kDa)
8	8.76	9.33	179	136
16	8.86	9.44	182	139
32	8.87	9.45	183	139
64	8.82	9.39	181	138

Figure 3.18 AUC results for PfTA WT under several PLP concentrations, producing an estimate of 181 kDa and minimal mass of 136 kDa, with no evidence of non-ideality. Data analysed in Sedfit

To confirm the AUC results, a repetition was made for both the PLP-absent condition and the PLP-present condition. A third condition was also tested, with inorganic phosphate but without PLP. All data for PfTA WT were collected in person, with assistance from Dr. Gemma Harris.

For each condition, the buffering was kept consistent throughout the whole purification process. This entailed three separate purifications of PfTA WT, but through optimisation of the IMAC, a single step purification was found to yield sufficiently pure PfTA WT. The model and method of data analysis were as before.

The PLP-absent condition (Figure 3.19, Figure 3.20) again showed no evidence of non-ideality. The estimated MW was 168 kDa after correction to standard conditions; the apparent sedimentation coefficient in the distribution here appears lower than in the former distribution, due to additional glycerol in the solvent and hence higher viscosity. Additional glycerol was necessary to protect the samples, which had to be sent frozen due to travel restrictions.

The PLP-present condition (Figure 3.21) was performed with a fixed concentration of 100 µM PLP, rather than the 8 µM - 64 µM concentration series used previously. This produced an estimate of 173 kDa.

The condition with 50 mM inorganic phosphate, but no PLP, unfortunately produced unusable data, after suffering precipitation on thawing. The residual PfTA WT concentration (0.04 mg/mL) fell below the threshold of the absorbance optics. An attempt to use the data from the interference optics was unsuccessful.

0.6 mg/mL

**PLP absent
PO$_4$ absent
(1st replicate)**

0.3 mg/mL

0.15 mg/mL

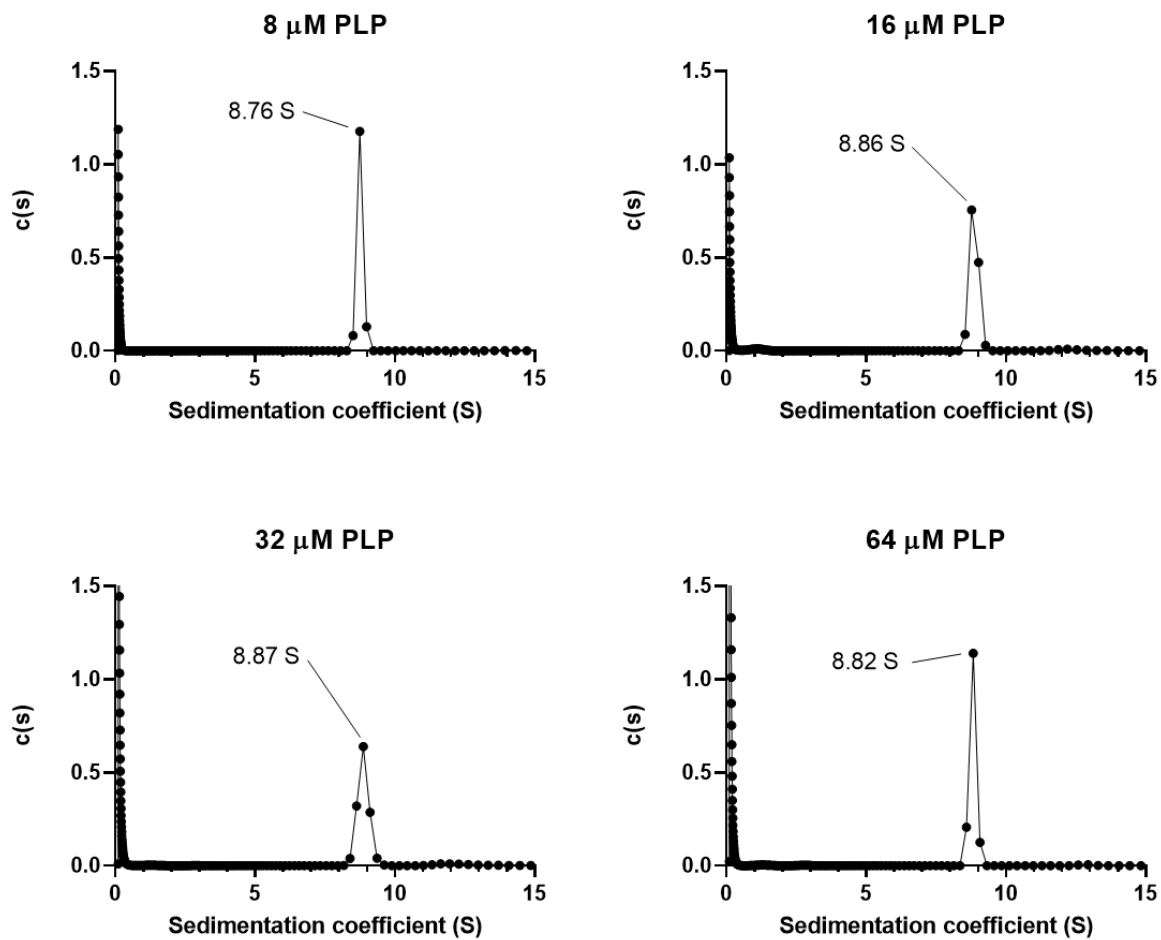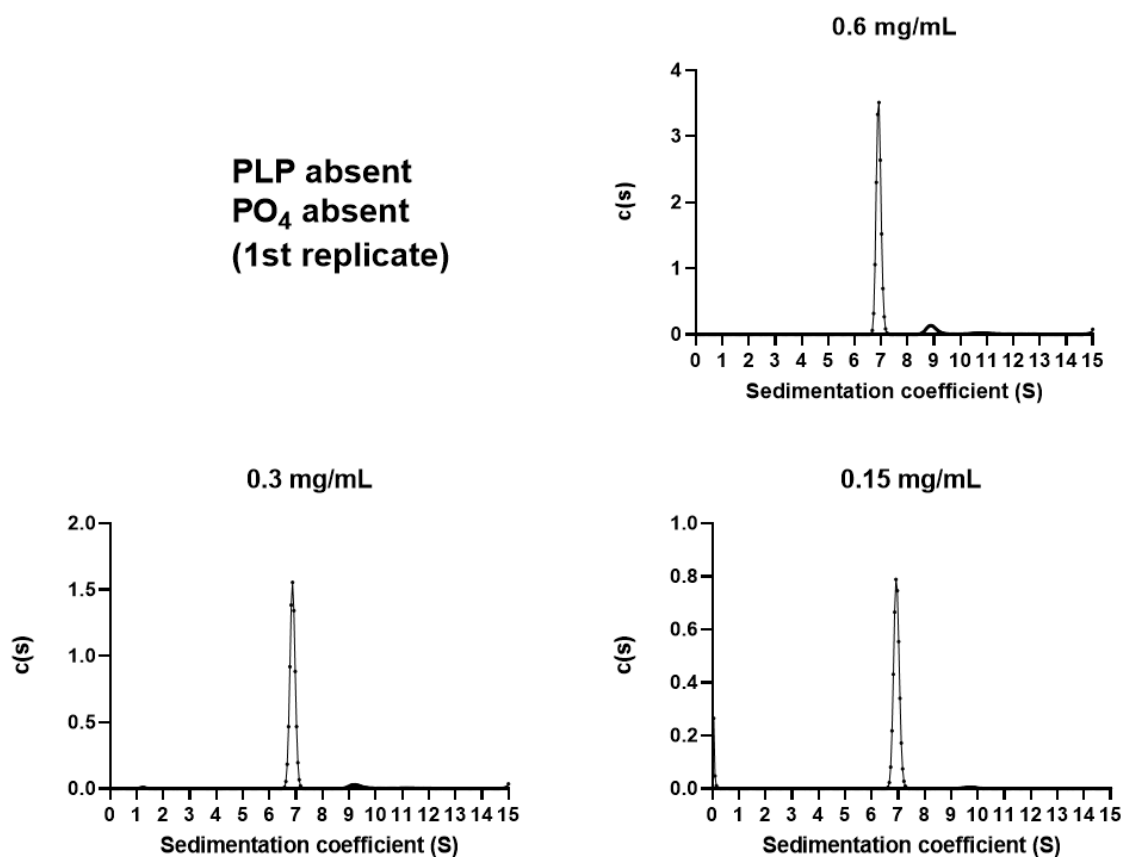| PLP absent; PO$_4$ absent: 1st replicate				
PfTA WT concentration (mg/mL)	Sedimentation coefficient (S)	Sedimentation coefficient (20 °C, water) (S)	MW estimate (for *f/f$_0$* = 1.2) (kDa)	Minimum MW (for *f/f$_0$*= 1.0) (kDa)
0.6	6.91	9.15	168	128
0.3	6.87	9.09	167	127
0.15	6.94	9.19	169	129

Figure 3.19 1st replicate results for PfTA WT derived from single-step purification, showing good purity, and MW estimate of 168 kDa consistent with previous results under the absence of PLP and phosphate. Data analysed in Sedfit

0.14 mg/mL

PLP absent
PO$_4$ absent
(second replicate)

0.07 mg/mL

0.035 mg/mL
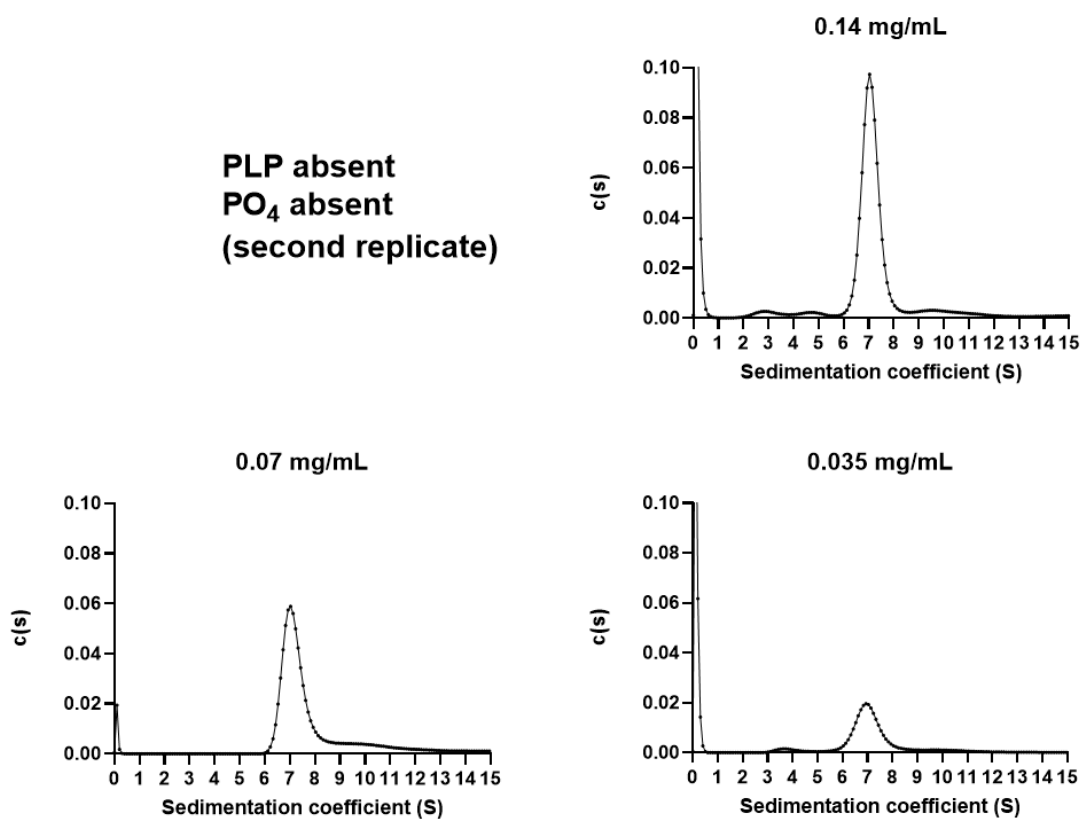
PLP absent; PO$_4$ absent: 2nd replicate				
PfTA WT concentration (mg/mL)	Sedimentation coefficient (S)	Sedimentation coefficient (20 °C, water) (S)	MW estimate (for $f/f_0 = 1.2$) (kDa)	Minimum MW (for $f/f_0 = 1.0$) (kDa)
0.14	7.09	9.28	172	131
0.07	7.12	9.33	174	133
0.035	7.06	9.24	170	130

Figure 3.20 2nd replicate results for PfTA WT derived from single-step purification, showing MW estimate of 172 kDa in consistency with previous results, despite slightly inferior data quality. Data analysed in Sedfit

PLP 100 μM
PO₄ absent
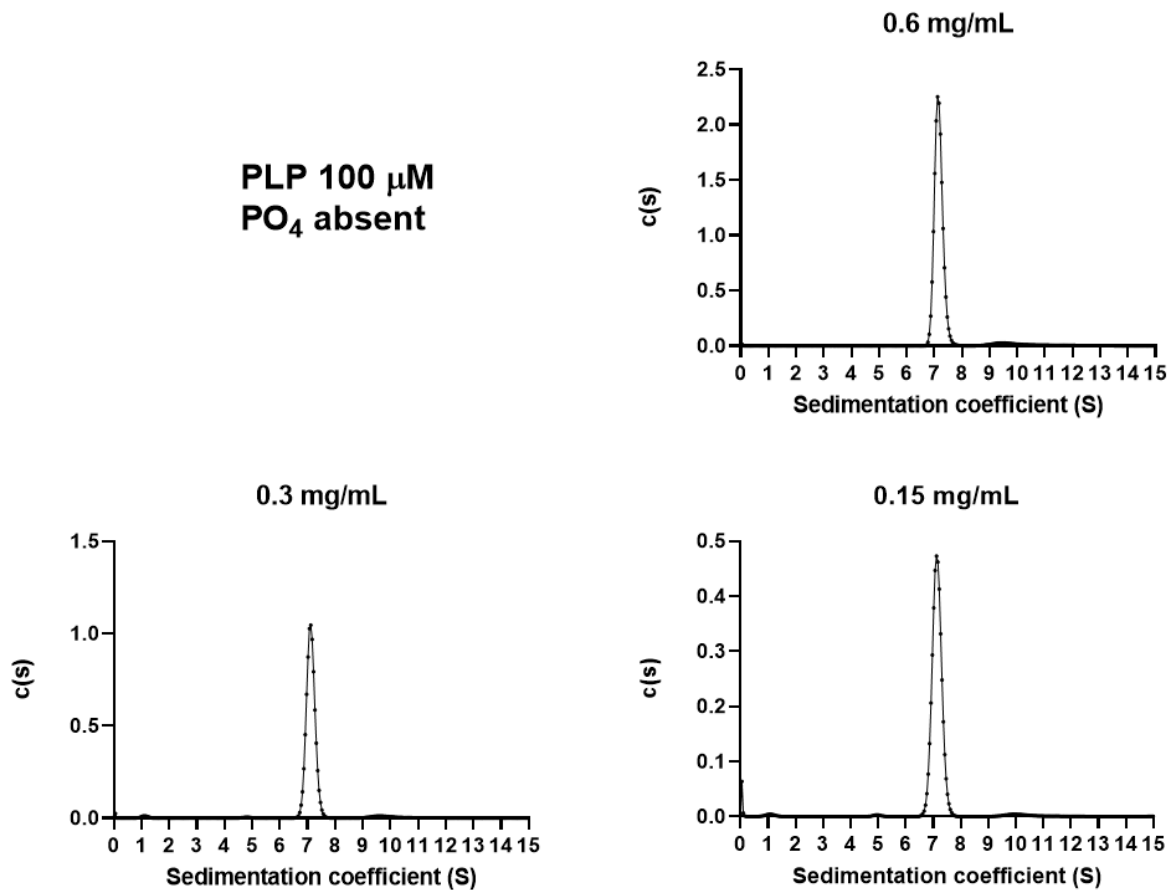
PLP 100 µM; PO₄ absent:				
PfTA WT concentration (mg/mL)	Sedimentation coefficient (S)	Sedimentation coefficient (20 °C, water) (S)	MW estimate (for f/f_0 = 1.2) (kDa)	Minimum MW (for f/f_0 = 1.0) (kDa)
0.6	7.16	9.37	174	133
0.3	7.12	9.32	173	131
0.15	7.14	9.34	173	132

Figure 3.21 PfTA WT derived from single-step purification with PLP, showing MW estimate of 173 kDa in consistency with previous results including non-PLP conditions. Data analysed in Sedfit

3.5.3 Small angle X-ray scattering (SAXS)

Concentration series of PfTA were prepared by dilution from the stock of each different condition using matched buffers. Data was collected by Dr. Anton Popov of the European Synchrotron Radiation Facility. The experimentally determined scattering profiles were compared to simulated scattering profiles, for monomeric, dimeric, trimeric and tetrameric PfTA WT structures, derived from the PfTA WT crystal structure (PDB: 6S54), using the FoXS server[53], as described under Methods.

Three or four concentrations of PfTA WT were tested under each buffer condition: with neither PLP nor phosphate (Table 3-3), with both (Table 3-4), with just phosphate (Table 3-5) or just PLP (Table 3-6). The results for every buffer and every concentration of PfTA WT, clearly showed that the monomeric and dimeric models were the poorest fitting to the experimental data. For half of the samples (7/14), the tetrameric model directly provided the closest fitting, whilst for all other samples, the trimeric model seemed to provide the best fit.

However, it is unlikely that the trimeric model really represents the most accurate fit for those samples for two reasons, in addition to the intuitive unlikelihood of the trimer. Firstly, for all scattering profiles where the trimeric model nominally provided the closest fit, the X^2 fit values for the trimeric and tetrameric models differed very little, especially in comparison to the significantly poorer X^2 fit values of the monomeric and dimeric models. Secondly, all trimeric models showed significant evidence of overfitting, in the hydration layer density values (c_2 values) as described under Methods.

Table 3-3 SAXS scattering profiles for PfTA WT in absence of PLP and phosphate

PfTA WT 4.66 mg/mL				PfTA WT 2.33 mg/mL			
Model	X^2	c_1	c_2	Model	X^2	c_1	c_2
Tetramer	8.17	1.04	0.48	Tetramer	2.48	1.04	0.41
Trimer	16.03	1.05	4.00!	Trimer	4.09	1.05	4.00!
Dimer	1122.69	1.05	4.00!	Dimer	378.27	1.05	4.00!
Monomer	3487.64	1.05	4.00!	Monomer	1170.50	1.05	4.00!

PfTA WT 1.16 mg/mL							
Model	X^2	c_1	c_2				
Trimer	2.18	1.05	3.79				
Tetramer	2.88	1.05	-0.03				
Dimer	154.69	1.05	4.00!				
Monomer	472.86	1.05	4.00!				
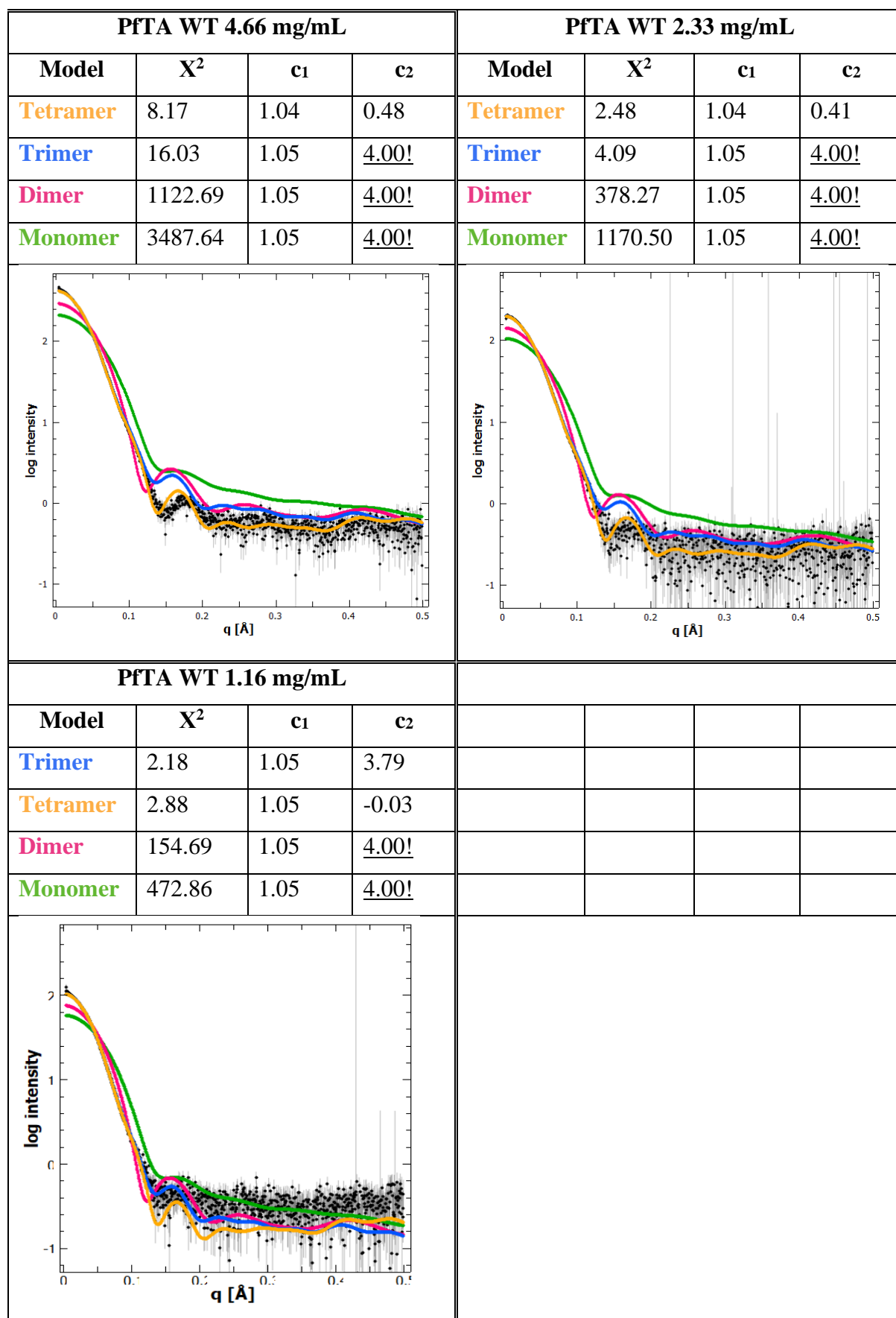
Table 3-4 SAXS scattering profiles for PfTA WT in 100 µM PLP and 50 mM phosphate

PfTA WT 6.42 mg/mL				PfTA WT 3.21 mg/mL			
Model	X^2	c_1	c_2	Model	X^2	c_1	c_2
Tetramer	3.17	1.04	0.48	Trimer	2.45	1.05	3.89
Trimer	3.42	1.05	4.00!	Tetramer	2.76	1.05	-0.00
Dimer	120.93	1.05	4.00!	Dimer	65.06	1.05	4.00!
Monomer	360.02	1.05	4.00!	Monomer	191.72	1.05	4.00!

PfTA WT 1.61 mg/mL							
Model	X^2	c_1	c_2				
Trimer	4.62	1.05	4.00!				
Tetramer	4.76	1.05	0.11				
Dimer	14.80	1.05	4.00!				
Monomer	33.42	1.05	4.00!				
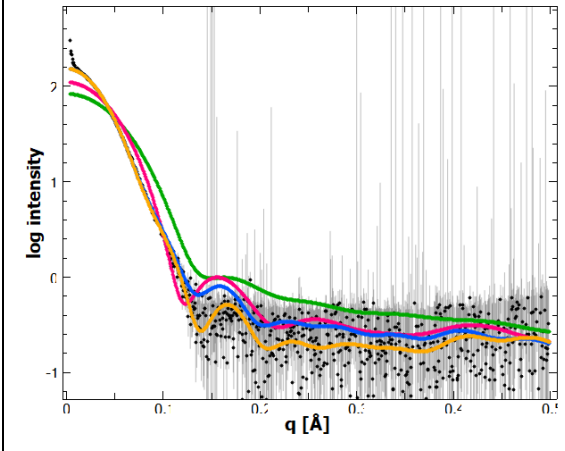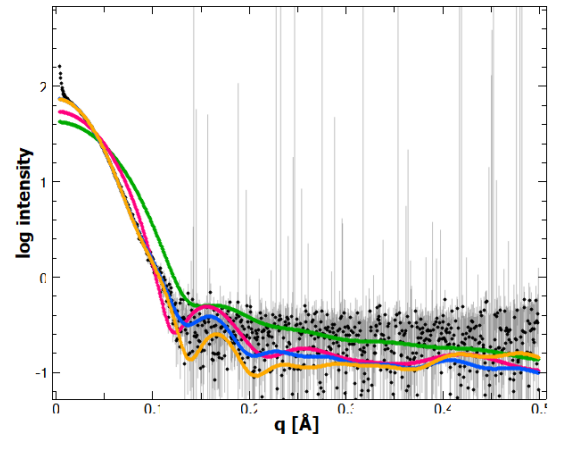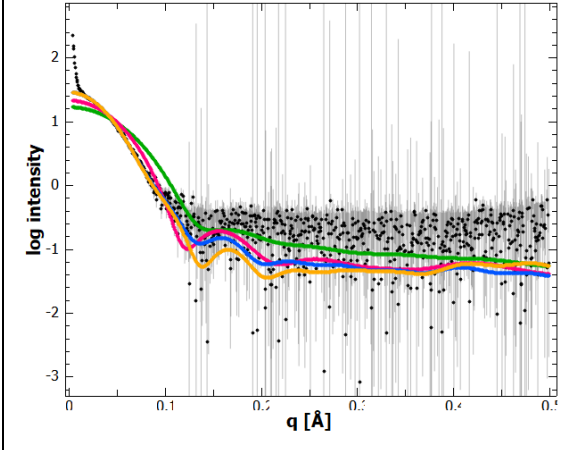
Table 3-5 SAXS scattering profiles for PfTA WT in absence of PLP and presence of 50 mM phosphate

PLP absent; PO$_4$ 50 mM							
PfTA WT 7.82 mg/mL				**PfTA WT 4.9 mg/mL**			
Model	**X^2**	**c$_1$**	**c$_2$**	**Model**	**X^2**	**c$_1$**	**c$_2$**
Trimer	26.65	1.05	<u>4.00!</u>	**Trimer**	7.04	1.05	2.73
Tetramer	26.72	1.05	0.76	**Tetramer**	8.70	1.05	-0.61
Dimer	69.11	1.05	<u>4.00!</u>	**Dimer**	103.74	1.05	<u>4.00!</u>
Monomer	145.10	1.05	<u>4.00!</u>	**Monomer**	309.47	1.05	<u>4.00!</u>

PfTA WT 2.45 mg/mL				**PfTA WT 1.23 mg/mL**			
Model	**X^2**	**c$_1$**	**c$_2$**				
Trimer	1.12	1.05	3.23	**Trimer**	2.38	1.05	1.31
Tetramer	1.23	1.05	-0.33	**Tetramer**	2.58	1.05	-1.40
Dimer	9.94	1.05	<u>4.00!</u>	**Dimer**	8.74	1.05	<u>4.00!</u>
Monomer	28.05	1.05	<u>4.00!</u>	**Monomer**	22.16	1.05	<u>4.00!</u>

Table 3-6 SAXS scattering profiles for PfTA WT in 100 µM PLP and absence of phosphate

PfTA WT 10.65 mg/mL				PfTA WT 5.36 mg/mL			
Model	X^2	c_1	c_2	**Model**	X^2	c_1	c_2
Tetramer	51.96	1.05	0.18	**Tetramer**	11.84	1.05	0.18
Trimer	67.72	1.05	4.00!	**Trimer**	15.03	1.05	4.00!
Dimer	3194.78	1.05	4.00!	**Dimer**	1223.44	1.05	4.00!
Monomer	9996.52	1.05	4.00!	**Monomer**	3793.41	1.05	4.00!

PfTA WT 2.66 mg/mL				PfTA WT 1.33 mg/mL			
Model	X^2	c_1	c_2	**Model**	X^2	c_1	c_2
Tetramer	4.93	1.05	0.15	**Tetramer**	2.34	1.05	0.20
Trimer	5.08	1.05	4.00!	**Trimer**	2.46	1.05	4.00!
Dimer	495.13	1.05	4.00!	**Dimer**	201.26	1.05	4.00!
Monomer	1516.73	1.05	4.00!	**Monomer**	606.14	1.05	4.00!

3.6 Chapter conclusion

A crystal structure for PfTA WT was obtained, showing four chains in the asymmetric unit, arranged in a tetrameric (dimer-of-dimers) structure. Despite limited map quality in some regions, it was possible to model all but the N-terminal regions in every chain, including the sections of chain which form the dimer-dimer interface. Inspection revealed that interfacial interactions (as hydrogen bonds, salt bridges or hydrophobic interactions) are especially clustered around two secondary structural elements, specifically the 21-residue helix α11 and the 30-residue ring-binding motif, suggesting these areas as possible targets for mutation to vitiate tetramerization, if indeed the tetramer in the asymmetric unit is representative of the oligomerisation state in solution, for which biophysical confirmation was sought.

Initial evidence with analytical SEC suggested that PfTA WT oligomerisation is not affected by PLP, nor by inorganic phosphate nor salinity. However, a large apparent discrepancy between the expected tetrameric mass of 210 kDa and the mass calculated from the elution volume (>170 kDa) gave cause for doubt concerning the oligomerisation state, even though the elution volumes were consistent between experiments.

Recourse to AUC showed that, over the concentration ranges for PfTA WT and PLP which were tested, there was no evidence of non-ideal behaviour, thus confirming that oligomerisation of PfTA WT is unlikely to exist in an equilibrium between oligomerisation states, which might otherwise have explained the apparent discrepancy observed by SEC. AUC also allowed the dimeric state to be excluded with good confidence, according to the minimal masses calculated for the observed PfTA WT sedimentation coefficients.

An attempt to extend the AUC results to cover all four possible combinations of PLP/phosphate presence/absence, was partially hindered by contemporary restrictions on sample submission, yet the useable portion of the data obtained did confirm the previous AUC results, and also confirmed that the single-step IMAC purification was able to yield PfTA WT in sufficient purity for biophysical analysis.

It was finally possible to test all four possible combinations of PLP/phosphate presence/absence, as the opportunity arose to use SAXS. The SAXS data analysis was made easier by the already obtained crystal structure for PfTA WT, which allowed fitting of the data against models for the monomer, dimer, trimer and tetramer, and showing that the tetrameric structure as contained in the asymmetric unit is the most reliable fit. Taken together with the

other biophysical results, this gives confidence that PfTA WT is designed to adopt solely the tetrameric (dimer-of-dimers) state.

4 Comparative structural analysis and mutation of PfTA WT with molecular dynamics simulations

4.1 Introduction

To provide insight for the rational choice of mutations which would be likely to vitiate PfTA tetramerization without deleterious effect on function or solubility, PfTA WT was compared to structurally similar tetrameric enzymes in order to identify conserved residues and secondary structure features. Having chosen mutations for further investigation, these were then subjected to molecular dynamics simulation to explore to what extent would it be possible to infer the interface-disrupting effects from changes in interfacial residue-residue interactions. Finally, the mutants were prepared by site-directed mutagenesis and tested for expression, to take forward for crystallography, biophysical characterisation, and stability assays against PfTA WT.

4.2 Comparative structural analysis and selection of PfTA mutations

B6-dependant enzymes were identified from the literature, for which there is biophysical evidence of tetramerization, and for which a crystal structure is available, listed in Table 4-1. The PDB was searched using PfTA WT (ID: 6S54) as a query input to find the most similar structures (searching by density distribution), which are listed as entries 2 to 10. Additionally, the crystal structures for these entries were checked to ensure that they are consistent in the orientation of subunits within the tetrameric assembly, and hence to ensure that the interfacial helices/strands/loops are equivalent for these entries, to enable a meaningful comparison. In contrast, are examples such as human ornithine TA (entry 13, ID: 1OAT), which is fold-type I, but which forms a tetrameric structure where the subunits are oriented differently and hence the interfacial residues are not part of the equivalent helices/strands/loops (Figure 4.1).

Entry	Reference(s)	Source organism, Name, (Code)	Summary of biophysical method(s) employed	Structural / functional classification notes
1	Ito *et al.* 2011 [47]; Mutti *et al.* 2012 [67]; Roura-Padrosa *et al.* 2019 [59]	*Pseudomonas fluorescens* **PfTA** (6S54)	(Unpublished, this work, sections 3.5.2 and 3.5.3): **AUC.** Protein 2 to 8 µM (~0.1 to 0.41 mg/mL) tested with/without PLP, pH 7.2; **SAXS** same conditions but 1.3 to 10.7 mg/mL. (Ito *et al.* 2011): **SEC.** Found ~120 kDa (dimer) with 10 mM potassium phosphate, 150 mM NaCl, 1.3 mM 2-ME, 0.1 mM PMSF, 20 µM PLP, pH 8.0. [SDS PAGE found 53 kDa] SAXS:	Fold-type I, TA class III Accepts SMBA but rejects β-alanine and GABA. 2.6.1.X
2	Borner *et al.* 2017a[17], 2017b[32]	*Pseudomonas sp.* **PxTA** (5LH9)	(Börner et al. 2017a): **SEC** and **DLS.** Several SEC experiments described. **Either: 1)** 10 mg/mL protein, 0.5 mM PLP, 40 mM sodium phosphate, pH 8, substrates present at various concentrations; **or 2)** 100 mM sodium phosphate, 300 mM NaCl, pH 6.8	Fold-type I, TA class III 2.6.1.X

3	Yohana *et al.* 1976 [68], 1977 [69], 1983 [43]; Burnett *et al.* 1980 [44]; Watanabe *et al.* 1989 [70]	*Pseudomonas putida* **PpTA** (3A8U)	(Yonaha *et al.* 1976): **SEC.** Found single, tetrameric species, in 20 mM potassium phosphate pH 7.4 (Yonaha *et al.* 1977): **SEC.** Found ~172 kDa (tetramer) in 20 mM potassium phosphate pH 8; **AUC** found 167 ±3 kDa (tetramer), with conditions not described. [SDS PAGE found ~45 kDa]	Fold-type I , TA class III Able to accept a range of ω-amino acids, but specific to pyruvate as amine acceptor (2.6.1.X)
4	Sayer *et al.* 2013 [42]. **NB** Citation attached to deposition 4B9B, the same enzyme.	*Pseudomonas aeruginosa* **PaTA** (4BQ0)	**SEC** and **DLS.** Found ~200 kDa (tetramer) using 50 mM Tris HCl, 100 mM NaCl, 50 µM PLP, pH 7.5	Fold-type I, TA class III β-alanine:pyruvate TA, also accepts MBA as a donor. (2.6.1.18)
5	Lamartiniere, Itoh and Dempsey 1971 [71]; Toney *et al.* 1993 [72], 1995 [73]	*Burkholderia cepacia* **BcDD** (2DKB)	(Lamartiniere, Itoh and Dempsey 1971): **AUC.** Found 188 kDa (tetramer) for protein at 0.05 to 0.3 mg/mL, with other details of condition not described. [AUC under denaturing conditions found 46-48 kDa]	Fold-type I, TA class III ***Dialkylglycine decarboxylase (4.1.1.64)***
6	Mutaguchi *et al.* 2013 [74]; Hayashi *et al.* 2017 [75]	*Lactobacillus buchneri* JCM 1115 **LbIE** (4YSN)	(Mutaguchi *et al.* 2013): **SEC** found ~200 kDa (tetramer) using 20 mM Tris HCl, 200 mM NaCl, pH 7.5. [SDS PAGE 49 kDa]	Fold-type I, TA class III ***Isoleucine-2-epimerase (5.1.1.21)***
7	Oosterwijk *et al.* 2016 [76]	*Arthrobacter sp.* **AxTA** (5G2P)	**SAXS.** Found 210 kDa (tetramer), R_g= 4.0 nm, D_{max}= 10 nm; against expected values of R_g= 3.6 nm, D_{max}= 11.7 nm from crystal structure. Protein 1.8 to 5.5 mg/mL, 50 mM Tris HCl, 150 mM NaCl, 50 µM PLP, pH 8.0 **SEC.** Found 150 kDa in 50 mM Tris HCl, 200 mM NaCl, 50 µM PLP, pH 8.0	Fold-type I, TA class III Accepts ω-amino acids (2.6.1.X) Sequence identity 95% to 5G0A (entry 8)
8	Oosterwijk *et al.* 2016 [76]	*Bacillus megaterium* **BmTA** (5G0A)	**SEC.** Found 150 kDa in 50 mM Tris HCl, 200 mM NaCl, 50 µM PLP, pH 8.0	Fold-type I, TA class III Accepts ω-amino acids (2.6.1.X) Sequence identity 95% to 5G2P (entry 7)
9	Liu *et al.* 2004 [41], 2005 [77]	*Escherica coli* **EcTA** (1SF2)	(Liu *et al.* 2004) **SEC.** No details supplied. Stated to be dimer/tetramer mixture with proportion of tetramer increased by addition of more PLP.	Fold-type I, TA class III GABA:2-oxoglutarate TA (2.6.1.19)
10	Hillier, Altermark and Leiros 2020 [78]	*Chromohalobacter salexigens* DSM 3043 **CsTA** (6RL5)	**SEC.** Found 172 kDa under 50 mM HEPES, 100 mM NaCl, 0.4 mM TCEP, pH 7.0. **Native PAGE.** Present between 146-242 kDa reference bands. 6 µg protein, under BisTris- and tricine-based buffering, pH 7.0, no exogenous PLP. [Monomer 47 kDa]	Fold-type I, TA class III DABA:2-oxoglutarate TA in ectoine synthesis pathway (2.6.1.76)
11	Sakuraba *et al.* 2004 [79], 2008 [80]	*Thermococcus litoralis* **TlTA** (2ZC0)	(Sakuraba et al. 2004): **Native PAGE.** Found 170 kDa, conditions not described. [SDS PAGE found 42 kDa]	Fold-type I. Archaeal enzyme, and does not classify clearly in current class system. Alanine:glyoxylate TA (2.6.1.44)
12	Yoshikane *et al.* 2006 [81], 2008 [82]	*Mesorhizobium loti* MAFF303099 **MlTA** (2Z9U)	(Yoshikane *et al.* 2006): **SEC.** Found 154 ±7.2 kDa (tetramer), conditions not described. [SDS PAGE found 41 kDa]	Fold-type I, TA class V. PLP independent; not catalytic in PLP.

				Pyridoxamine:pyruvate TA (2.6.1.30)
13	Montioli *et al.* 2017 [38]	*Homo sapiens* **hOAT** (1OAT)	**AUC.** 6 µM protein in HEPES 50 mM, NaCl 500 mM, pH 7.4. Single peak corresponding to tetramer. **SEC.** protein from 0.3 to 75 µM, HEPES 50 mM, NaCl 500 mM, pH 7.4. Elution volume varying with protein concentration, indicating dimer/tetramer in rapid equilibrium.	Fold type I, TA class III Ornithine δ-transaminase (2.6.1.13)
14	Heckmann *et al.* 2020 [83]	*Aspergillus terreus* **AtRTA** (4CE5)	**SEC.** Protein at ~2 mg/mL in potassium phosphate 50 mM, NaCl 100 mM, 0.1 mM PLP, pH 8.0. **MALDI-TOF MS.** Used on mutants only. Potassium phospate 50 mM, PLP 0.1 mM, pH 8 for sample preparation, followed by desalting into 80% acetonitrile, 0.1% TFA, with 10 mg/mL sinapic acid.	Fold-type IV, TA class IV. SEC indicates tetramer main solution species (97%). MS used to analyse mutants.
15	Heckmann *et al.* 2020 [83]	*Thermomyces stellatus* **TsRTA** (6XWB)	**SEC.** Protein at ~2 mg/mL in potassium phosphate 50 mM, NaCl 100 mM, 0.1 mM PLP, pH 8.0. **MALDI-TOF MS.** Used on mutants only. Potassium phospate 50 mM, PLP 0.1 mM, pH 8 for sample preparation, followed by desalting into 80% acetonitrile, 0.1% TFA, with 10 mg/mL sinapic acid.	Fold-type IV, TA class IV. SEC indicates tetramer 90% in solution with 10% dimer. MS used to analyse mutants.

Table 4-1 List of enzymes of the B6-dependant family which meet the two following criteria: i) a crystal structure is available and ii) biophysical evidence has been produced to confirm the tetrameric oligomerisation state. For some, the tetrameric state is not exclusive but depends on the buffering conditions used for the biophysical experiment. Whilst most are transaminases, entries 5 and 6 have non-transamination functionalities.
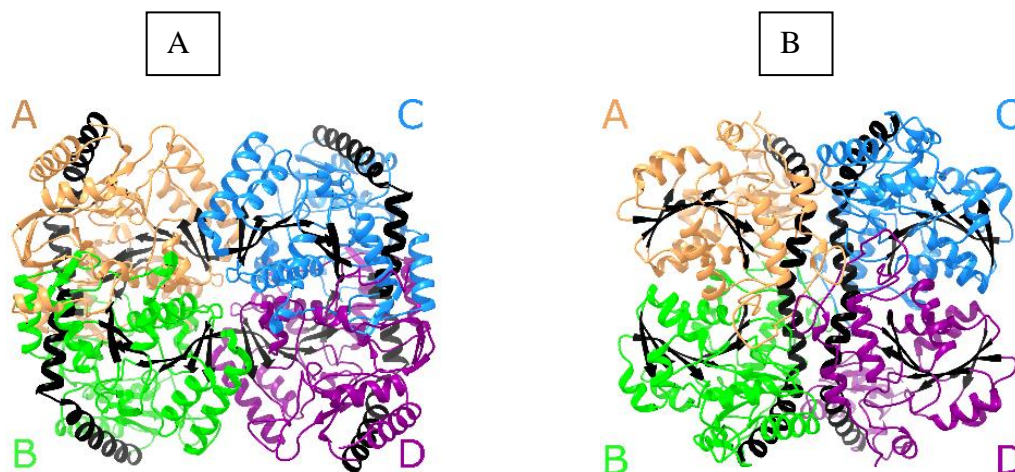
Figure 4.1 Tetrameric fold-type I enzymes presenting dissimilar subunit arrangements in their crystal structures. The core seven-stranded β-sheets and long inter-domain helices are highlighted in black for all chains in **(A)** PfTA WT (6S54) and **(B)** human ornithine TA (1OAT). In PfTA WT, the ends of the β-sheet form part of the inter-dimer interface, whereas the orientation of subunits is opposite in human ornithine TA.

Data for the interfacial surface areas and predicted oligomerisation states of PfTA WT and its close homologs (i.e. entries 2-10) was obtained from the server PISA (Proteins, Interfaces,

Structures and Assemblies)[61], as pre-computed results from the deposited crystal structures. Areas for the inter-dimer interface A-D ranged from 800Å^2 to $1460\ \text{Å}^2$, whilst PfTA WT's three *Pseudomonas*-derived homologs (PxTA, PpTA and PaTA) have the smallest interfacial areas, leading PISA to predict the dimeric oligomerisation state for them. The fact that the dimeric state was predicted, but the tetrameric state was biophysically confirmed, suggested that the Gibb's energy favouring tetramerization may be marginal, and since the PfTA inter-dimer interfacial area is only slightly larger than in PxTA, PpTA and PaTA, it also suggested that a single well-placed mutation in PfTA WT could likely be sufficient to prevent the inter-dimer interface forming.

Name	PDB code	Average area / Å^2				"CSS" score estimated significance of Chain A-Chain D interface*	Predicted oligomerisation state
		Intra-dimer interface-Chain A to Chain B (% of monomer surface)	Inter-dimer interface-Chain A to Chain D (% of monomer surface)	Inter-dimer interface-Chain A to Chain C (% of monomer surface)	Single chain gross surface area		
PfTA	6S54	5410 (28 %)	1090 (5.5 %)	183 (1 %)	19660	0.515	Tetramer
PxTA	5LH9	5280 (28 %)	940 (4.9 %)	110 (1 %)	19180	0.000	Dimer
PpTA	3A8U	6010 (30 %)	800 (4.0 %)	247 (1 %)	19840	0.000	Dimer
PaTA	4BQ0	5830 (30 %)	810 (4.2 %)	282 (1 %)	19380	0.000	Dimer
BcDD	2DKB	4670 (24 %)	1320 (6.9 %)	66 (0 %)	19260	0.000	Tetramer
LbIE	4YSN	5000 (25 %)	1220 (6.2 %)	71 (0 %)	19650	0.955	Tetramer
AxTA	5G2P	5520 (26 %)	1210 (5.7 %)	0 (0 %)	21120	0.080	Tetramer
BmTA	5G0A	5710 (26 %)	1440 (6.5 %)	90 (0 %)	22000	1.000	Tetramer
EcTA	1SF2	4520 (25 %)	1080 (5.9 %)	0 (0 %)	18390	0.182	Tetramer
CsTA	6RL5	4760 (26 %)	1460 (7.9 %)	111 (1 %)	18400	1.000	Tetramer

*PISA estimates the likely oligomerisation state based on Gibb's free energy calculation and assigns a score for the probable biological significance of each interface in the structure as a "CSS" score from 0 (not biologically relevant) to 1 (essential to biological assembly).

Table 4-2 Inter-chain interface areas for PfTA WT and closest tetrameric homologs, chains labelled schematically(A,B,C,D) as illustrated in Figure 3.4, with the oligomerisation state predicted by PISA from the crystal structure.

In an effort to understand the general nature of the inter-dimer interface, an analysis of the distribution of interfacial residues by their physicochemical properties was made (Table 4-3). PISA data was used to group the residues:

i. Core residues- those having no solvent-exposed area;

ii. Non-interfacial surface residues- those having a non-zero solvent-exposed area. The solvent-exposed area is not reduced by the formation of any interface;

iii. Inter-dimer interfacial residues- those where the solvent-exposed area is reduced by formation of the inter-dimer interface;

iv. Intra-dimer interfacial residues- those where the solvent-exposed area is reduced by formation of the intra-dimer interface.

The latter two groups are non-exclusive, since a small number of residues contribute to both inter- and intra-dimer interfaces, and such residues are counted in both groups. For each of the latter three groups (non-interfacial, inter-dimer, intra-dimer), the residues were categorised by charge, polarity, aromaticity and size, according to the scheme of Zvelebil *et al.*[84]

The results showed a general pattern that the inter-dimer interface features more polar (and fewer hydrophobic) residues than either the intra-dimer interface or the non-interface residues, with EcTA and CsTA being exceptions. That the inter-dimer interface is more hydrophilic than the non-interface surface, is not what would be expected for obligate interfaces which are critical to enzyme function, and is more typical of transient interfaces[85], adding weight to the likelihood that PfTA's tetrameric state would not be critical for its function.
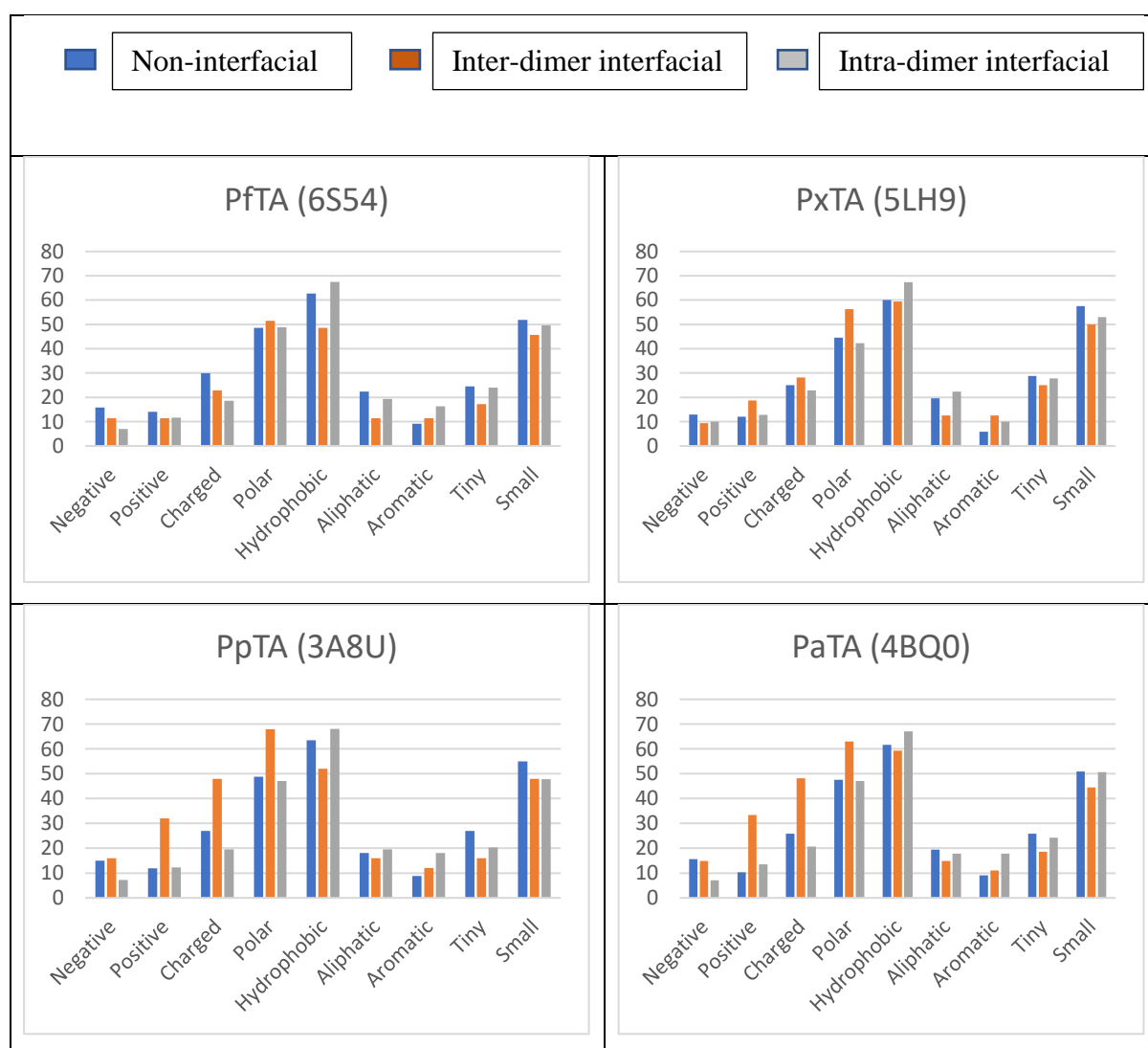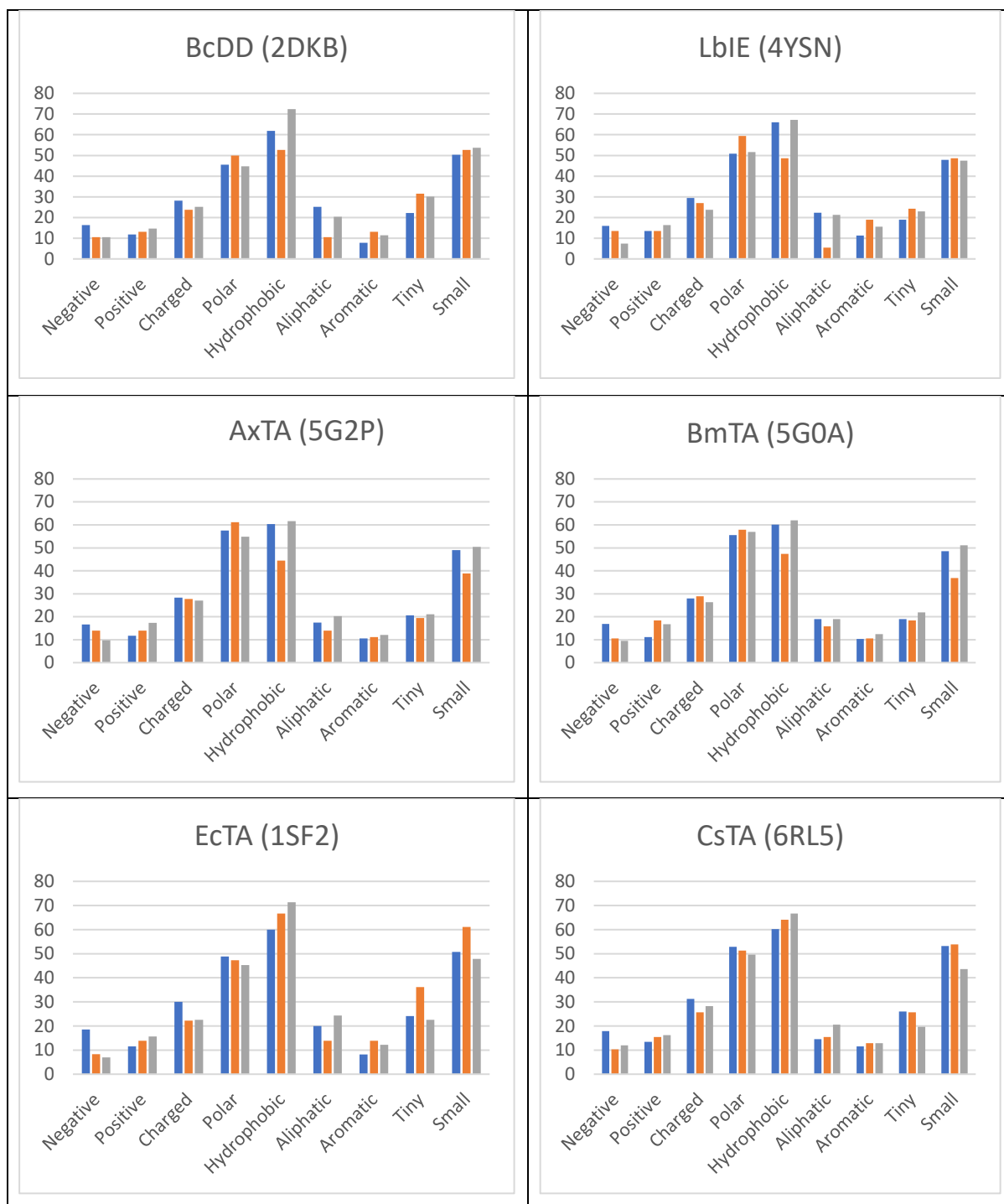
Table 4-3 Plots showing the distribution of intra-dimer, inter-dimer, and non-interfacial residues, by physicochemical category (charge, polarity, size, etc), for PfTA WT and its tetrameric homologs.

Like PfTA WT, most inter-dimer interfacial interactions in the PfTA WT homologs involve either the ring-binding motif (RBM) and/or the section of chain comprising the α11 helix with its N- and C-terminal loops (L2 and L3). Inspection of the RBM/α11/L2/L3 in those homologs with the least inter-dimer interfacial area (PxTA, PpTA and PaTA) unsurprisingly showed fewer residues contributing to interfacial hydrophobic interactions, hydrogen bonds and salt bridges, than in the other homologous TAs. Despite this, and even within this small sub-group,

no general pattern was found for the type or distribution of interfacial residues within the RBM/α11/L2/L3, with the exception of the pair PaTA and PpTA, which are noticeably alike in both the type and distribution of residues (Figure 4.2 and Figure 4.3). Within the wider group of homologs, AxTA and BmTA are also very similar, which is attributable to their high sequence identity (95%), and they share an extended loop structure within the RBM that is not shared by any of the other homologs.Figure 4.2 Ring-binding motif (RBM) in PfTA WT and tetrameric homologs, with inter-dimer interfacial residues highlighted pale blue. Bond labels indicate secondary structure element to which residue is bonding.

Focusing specifically on PfTA WT, the cluster of hydrophobic residues situated on the N-terminal end of α11 and the preceding loop L2 (described in section 3.4.3) is not generally conserved among its homologs, due to the variable conformation of L2 among the homologs (Figure 4.3). Likewise, the position and number of the interfacial hydrogen bond-forming residues on α11 and the RBM is not closely matched with any other homolog.

Therefore, since comparison against its homologs did not reveal that tetramerisation demands any specific residues in specific positions, by which to guide the choice of specific mutation sites in PfTA WT, instead residues were prioritised for mutation based on intuitive criteria:

i.   The number of inter-dimer bonds they contribute to the inter-dimer interface;
ii.   Their contribution to hydrophobic effect of the inter-dimer interface (assessed from PISA's calculated desolvation energy);
iii. They are not one of the small number of residues contributing to both the inter- and intra- dimer interface, since mutation on the intra-dimer interface could have unintended effects.
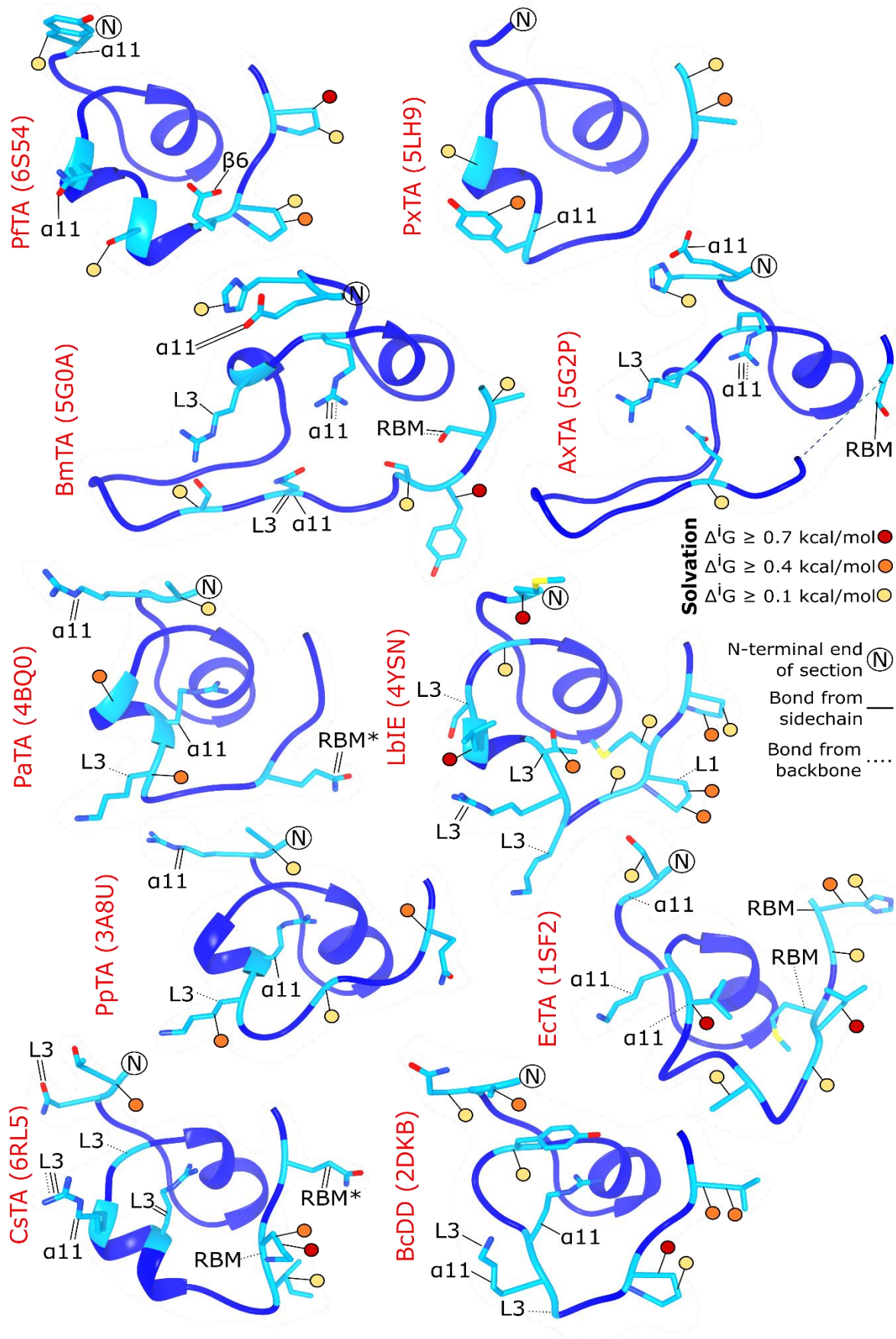
Figure 4.2 Ring-binding motif (RBM) in PfTA WT and tetrameric homologs, with inter-dimer interfacial residues highlighted pale blue. Bond labels indicate secondary structure element to which residue is bonding.

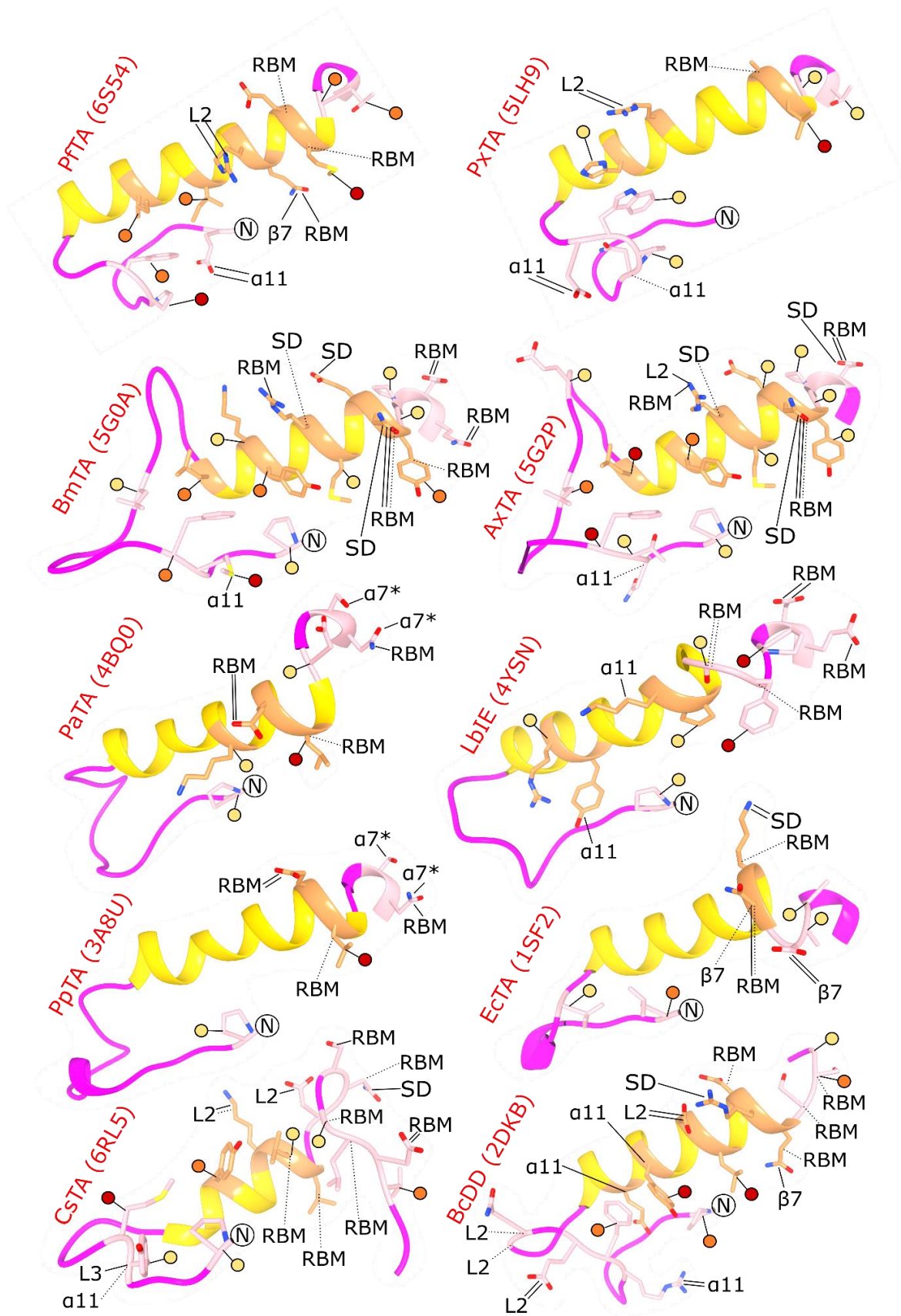Figure 4.3 α11 with its N- and C- terminal loops (L2 and L3) in PfTA and tetrameric homologs, inter-dimer interfacial residues highlighted pale pink and dark orange. Bond labels indicate secondary structure element to which residue is bonding.

Although the sites targeted for mutation were chosen without reference to the homologs, a multiple structural alignment of PfTA WT and the homologs was created in order to inform the decision of which residue to use as substitute at each site (Figure 4.6), to minimise the chance of destabilising the fold. This type of alignment aligns residues in structurally equivalent positions, and was produced in Chimera by iteratively least-squares fitting Cα atoms in the homologs to the equivalent Cα atoms of PfTA WT, followed by elimination of poorly-matched Cα pairs before repeat fitting. For this alignment, both tetrameric and dimeric PfTA WT homologues taken from the PDB were included.

The first mutation target in PfTA WT was the E189-R215' salt bridge (Figure 3.9Figure 3.9 The nine individual residues forming inter-dimer hydrogen bonds and salt bridges in PfTA WT are distributed across several of the secondary structure elements, but all the bonds/bridges involve at least one residue from either **(A)** the α11 helix or **(B)** the RBM, or both.. R215E was selected to make the electrostatic interaction unfavourable for interface formation. Additionally, whilst five of the tetrameric homologs featured basic residues in positions structurally equivalent to R215, amongst the dimeric homologs the acidic residues were more common in this position (7/12), suggesting that Glu would be a tolerable substitute (Figure 4.4). E183P was also selected to target the other side of the salt bridge, because Pro is the most common residue at the equivalent position in the homologs, despite the fact that the conformational constraint of Pro would normally make it an unlikely choice of substitute.

The next target in PfTA WT was the hydrophobic interaction involving α11 and the preceding loop L2, which interact with the symmetry-related α11' and L2'. The residues involved are P197, F198, I208, I212 (and hence also P197', F198', I208', I212' as shown in Figure 3.10 above). Because of the symmetry, substitution of either F198 or I212 with a charged residue would introduce repulsion towards the corresponding F198' or I212'. Since several homologs have Glu in equivalent position to I212, the mutation I212E was chosen, whereas finding structural equivalents for F198 was not possible because it exists as part of the loop L2 with non-conserved conformation between homologs. Therefore, F198E was chosen arbitrarily out of all the possible charged residues.

The final targets were Y155 and M219, being close enough to suggest the possibility of an inter-dimer sulphur-π interaction (Figure 4.5), when making allowance for the fact that the sidechain conformations may deviate from the those observed in the crystal structure. Additionally, M219 is likely to make a significant contribution to inter-dimer hydrophobic

effect (Figure 4.3) and Y155 forms an inter-dimer hydrogen bond to Q216' (Figure 3.9). The homologs show aromatic and charged residues are both common in position Y155; however, Y155A was chosen since Ala would eliminate all potential for inter-chain hydrogen bonding or sulphur-π interaction. M219R was chosen since Arg is common in this position amongst the dimeric homologs (5/12), and the larger size and conformational freedom of Arg has potential to make interface formation unfavourable from steric and entropic perspectives.
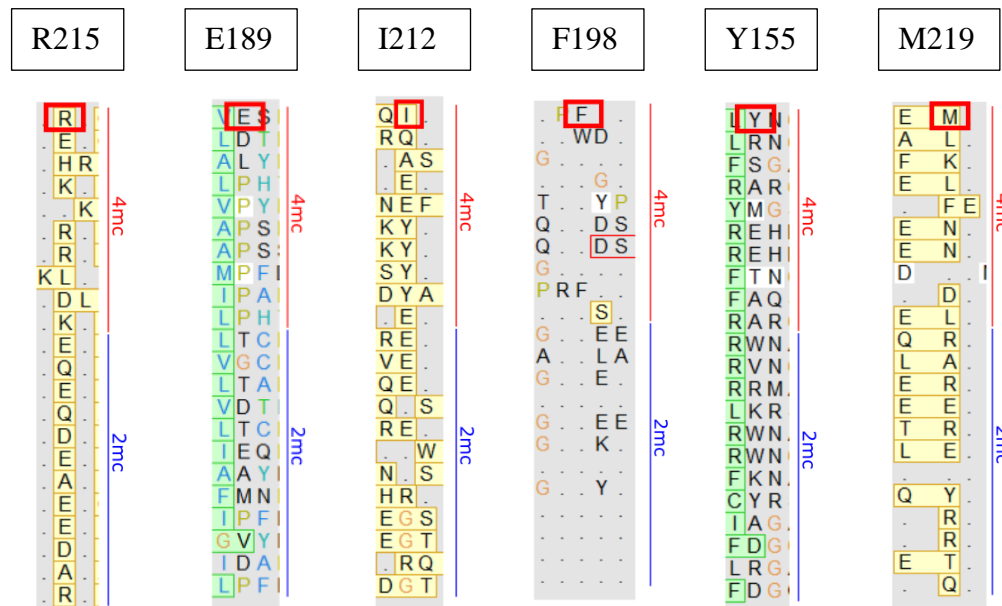

Figure 4.4 Detail from the multiple structural alignment (Figure 4.6), around the residues selected for mutation.
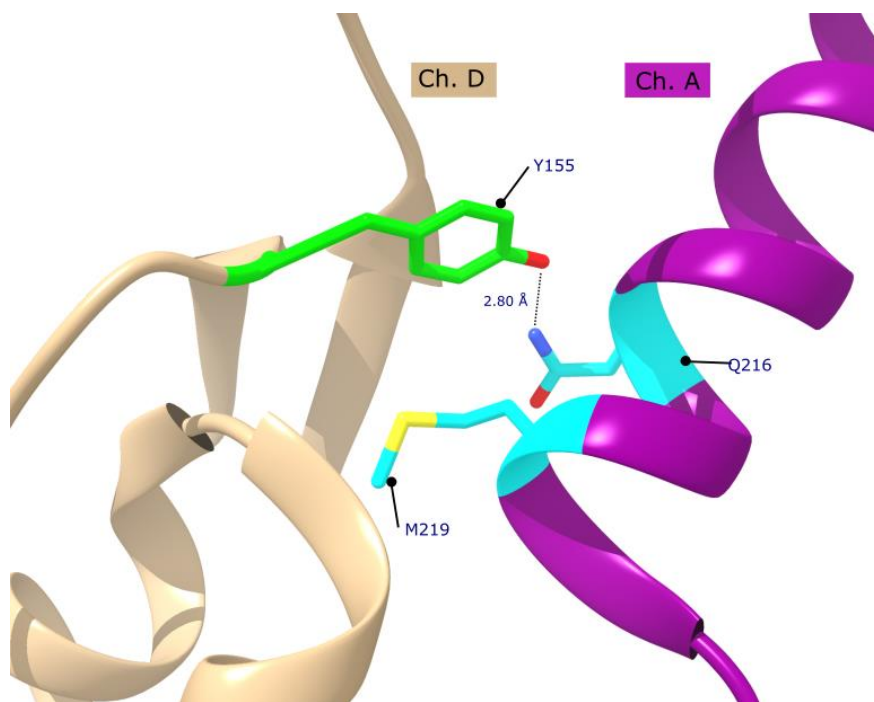
Figure 4.5 PfTA WT residues Y155 and M213. The Y155 forms inter-dimer hydrogen bond with Q216, whilst the displacement of solvent around M219 gives a calculated solvation energy change (averaged over four chains) of 2.53 kcal/mol
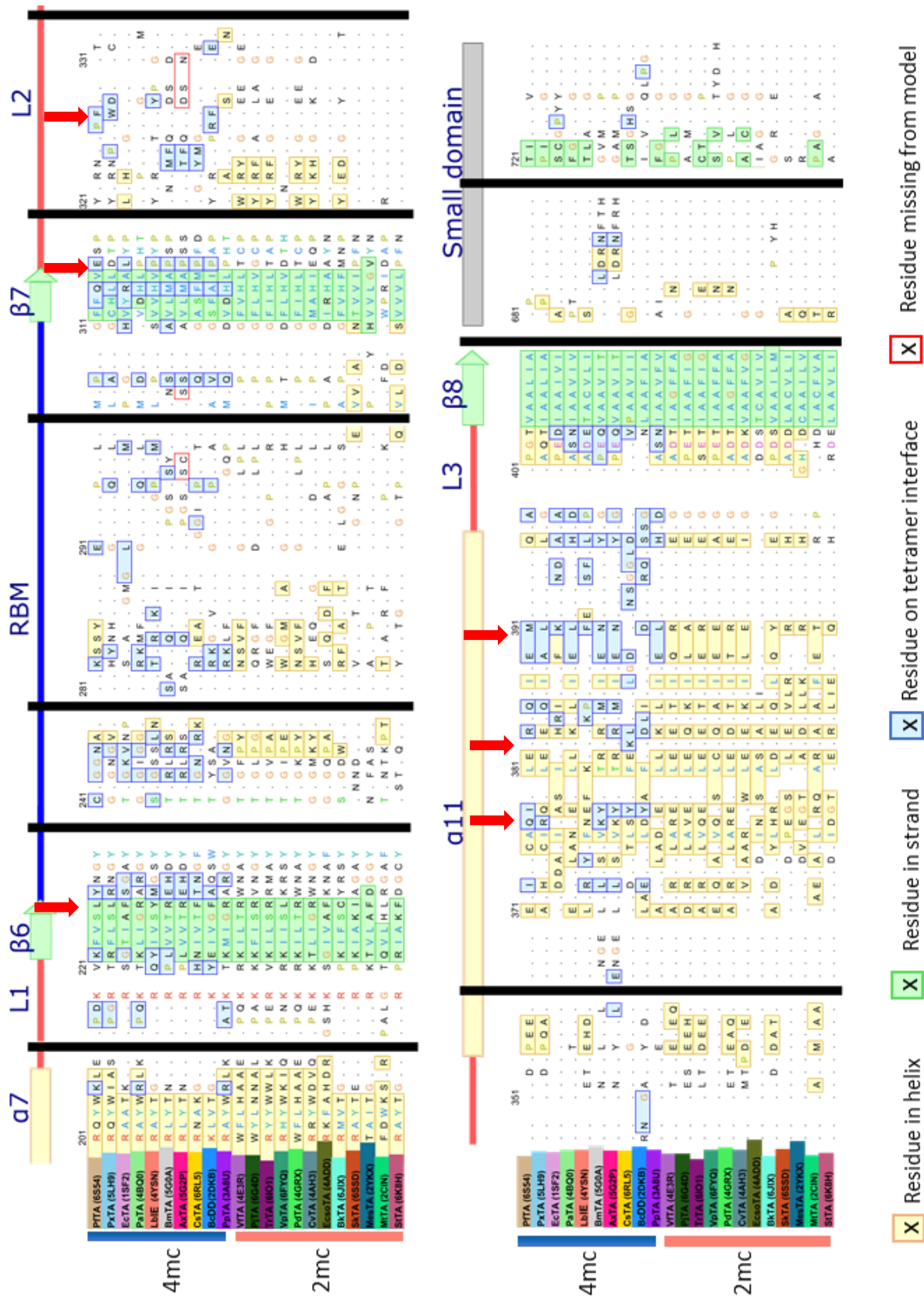
Figure 4.6 The full multiple structural alignment produced in Chimera, for PfTA WT against tetrameric (labelled '4mc') and dimeric ('2mc') homologs, covering all sections of chain in which interfacial residues occur. PfTA mutation sites are marked by red arrows. The RBM and the loops L1, L2, L3 are dissimilar in conformation amongst the homologs, meaning that their residues often cannot be aligned by meaningful structural equivalence.

In total therefore, six single mutants were planned: Y155A, E189P, F198E, I212E, R215E, M213R, and the double mutant I212E/R209E was also planned, in consideration of the possibility that a single mutation might be insufficient.

4.3   Molecular dynamics simulations

The seven planned PfTA mutants and the PfTA WT were simulated to provide more information on the interface-disrupting effects of each mutation towards predicting the likelihood of inducing a change of oligomerisation state. Since the timescale of the simulation was too short to observe whole subunit dissociation, it was attempted to infer the disruptive effects by focusing on individual residue-residue pairs in the interface. The effect of mutation was therefore quantified, firstly, in terms of the residue-residue total fractional contact time, using the "native contacts" function in the Amber suite. The "total fractional contact" is defined as the fraction of total frames in the trajectory for which the two residues are in contact with each other (contact for this analysis is defined by a 3.0 Å cutoff), as the sum over atom-atom pairs within the residue-residue pair, which means that the fraction may be greater than 1 if several atom-atom pairs exist within the residue-residue pair. It was quantified secondly by the residue-residue electrostatic potential using the "linear interaction energy" (LIE) function in Amber.

Adopting these tools was done in hope of finding a convenient heuristic approach, and differs from more rigorous and costly methods of assessing impact of point mutation on protein:protein interfaces, such as described by Zhang *et al.*[86], which require the simulation of WT-mutant intermediate states not reflective of physical reality.

The PfTA WT simulation was run as a single production of 500 ns (excluding equilibration). Since preparing the mutants' starting structures created the possibility of creating clashes/voids due to substitution of sterically larger/smaller residues, and the possibility of unrealistic simulation behaviour caused by that, therefore each mutant was simulated in triplicate (150 ns each) with resetting of the random seed used for initial velocity assignment.

The simulation period required to achieve full equilibration varied between 10 ns to 50 ns, being judged by the system density (i.e. the total system mass divided by the volume of the truncated octahedral solvent box). During this initial period, the system density gradually increased, corresponding to ingress of water molecules into various small solvent-accessible clefts, which had been left empty in the initial assignment of water positions. The part of each

trajectory which was subsequently used for data analysis (500 ns for PfTA WT, 150 ns for each PfTA mutant) excluded this initial equilibration period.

RMSD was calculated for the trajectory of each equilibrated system. The RMSD is used as a tool to measure the extent to which the overall protein structure deviates from its initial conformation at subsequent time points. In general, the fully equilibrated systems showed a stable RMSD, with no overall trend of increase or decrease in RMSD being apparent above or beyond the random variation in RMSD (Figure 4.7, Figure 4.8). A minority of the systems (such as the Y155A Replicate 2 simulation (Figure 4.7)) did seem to show change in RMSD beyond the scope of random variation, however manual inspection of the trajectory, did not reveal significant change in the active site environment or overall structure. Accordingly, the systems were deemed sufficiently free from artefacts to use them for analysis.
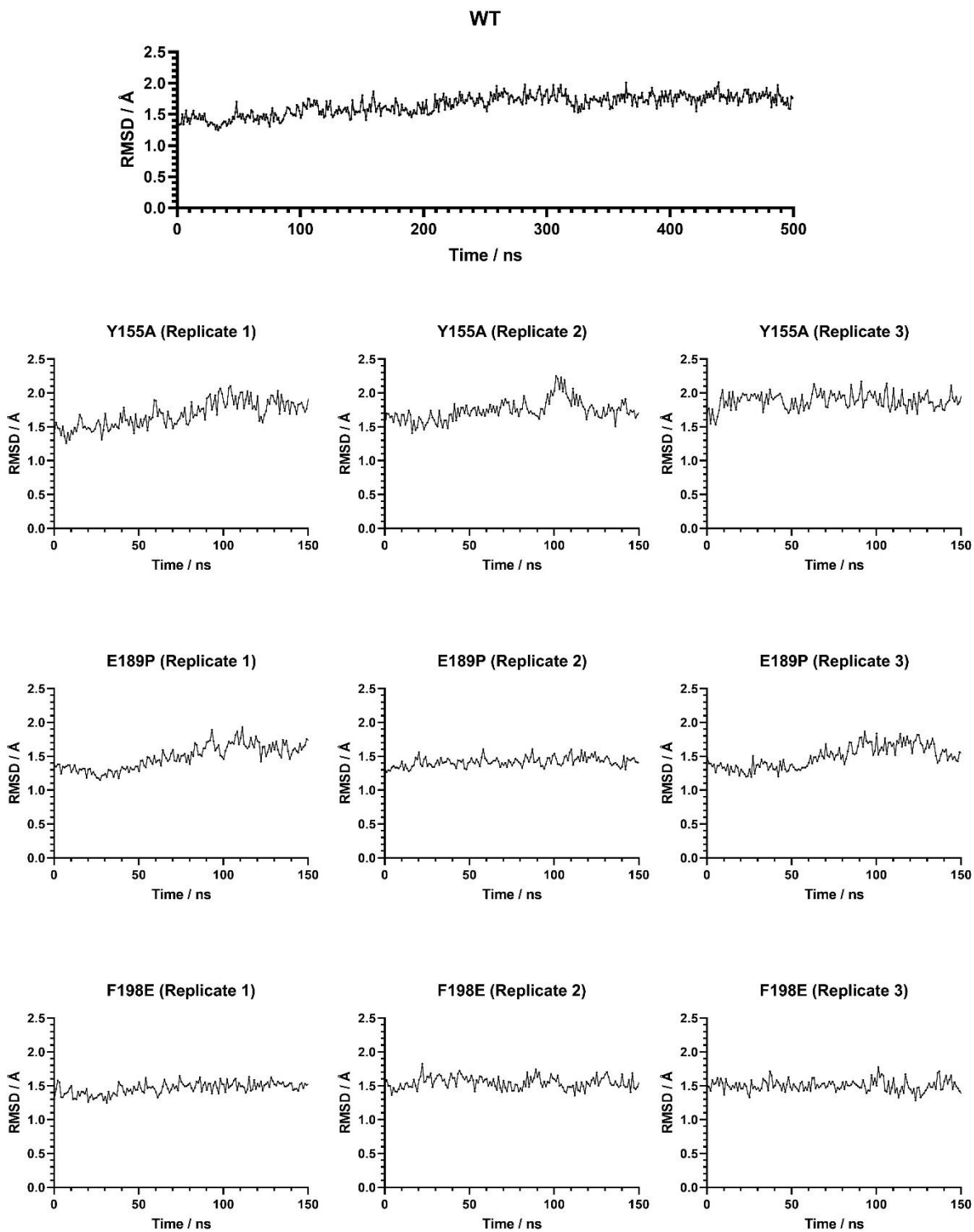
Figure 4.7 RMSD plots for each MD simulation of PfTA WT, Y155A, E189P and F198E
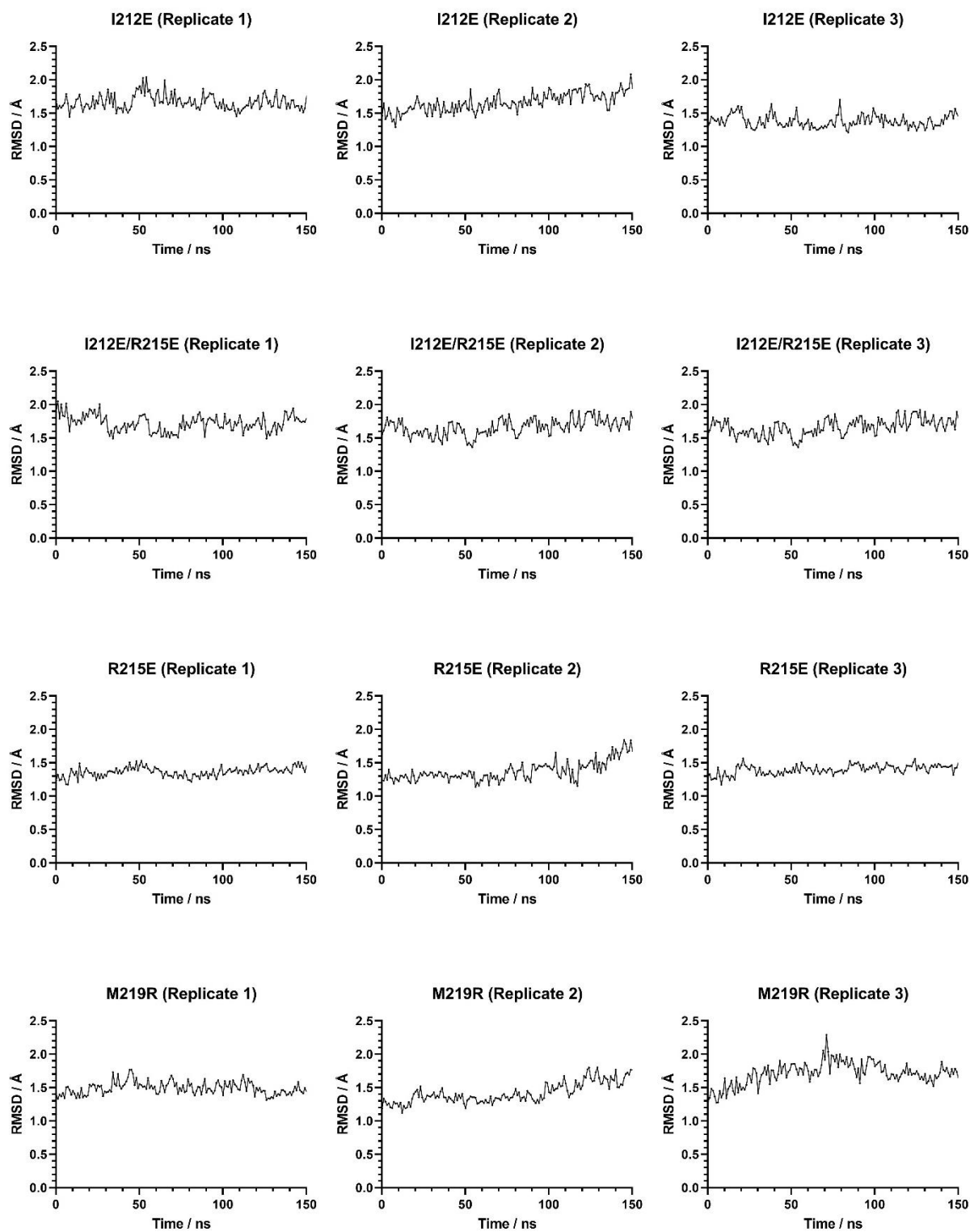
Figure 4.8 RMSD plots for each MD simulation of PfTA I212E, I212E/R215E, R215E, and M219R

The per-residue RMSF was also calculated. The RMSF is a single value for each residue, indicating the extent of fluctuation around its average position. In this work, RMSF was calculated for Cα positions only, to reflect the fluctuation of the backbone, whereas including sidechain atoms in RMSF calculation would conflate this with sidechain conformation fluctuation, since longer and solvent-exposed sidechains would be expected to fluctuate more than shorter and buried sidechains.

Considering PfTA WT, its RMSF values can be rationalised with reference to the overall structure and the active site (Figure 4.9). For example, the C-terminal subsection of the small domain shows higher than average RMSF. Due to its location within the overall structure, and relatively small area of interface formation (Figure 4.9A), it seems likely that rigid-body motion of this subsection may be occurring, which in addition to local fluctuations, would cause RMSF values to be higher for this subsection. (Overall rigid body motion as rotation or translation of the entire tetramer assembly is removed by fitting structure to the original crystal structure prior to RMSF calculation).

Some residues and secondary structure features may exhibit relatively high RMSF values in relation to the vacant active site, and their role in formation of the active site structure. These include D264, to which the pyridine nitrogen atom of PLP is hydrogen bonded in the holo-state, and the α15 helix, which occurs immediately prior to the phosphate group binding loop (Figure 4.9B), and upon the conformation of which depends the positioning of this binding loop within the cleft of the neighbouring chain, in order to form the completed active site. Fulfilling this role puts α15 in a position which is relatively less well-packed, compared to the other helices of the large domain, which pack against the central β-sheet and against each other.

With reference to the secondary structure features involved in tetramer interface formation, the RMSFs for the whole of the L2 loop and for part of the RBM, are also relatively high. The former may simply reflect the length and lack of regular structure in the L2 loop. The latter may reflect the discussion of Börner *et al.* [17] concerning conformational flexibility required of the RBM in order to allow PLP to enter the active site. Likewise, the suggestion made by the same authors, that interface formation limits conformational freedom in the RBM, is in keeping with the result seen here, that those residues within the RBM showing the highest RMSF, are ones which do not participate in interface formation (Figure 4.9C).

A small number of other areas show relatively high RMSF values for reasons which are less clear, such as the β5 strand which forms one end of the core β-sheet, and the C-terminal loop

of the β8 strand, which is in proximity to the active site, but do not directly participate in bonding to PLP.
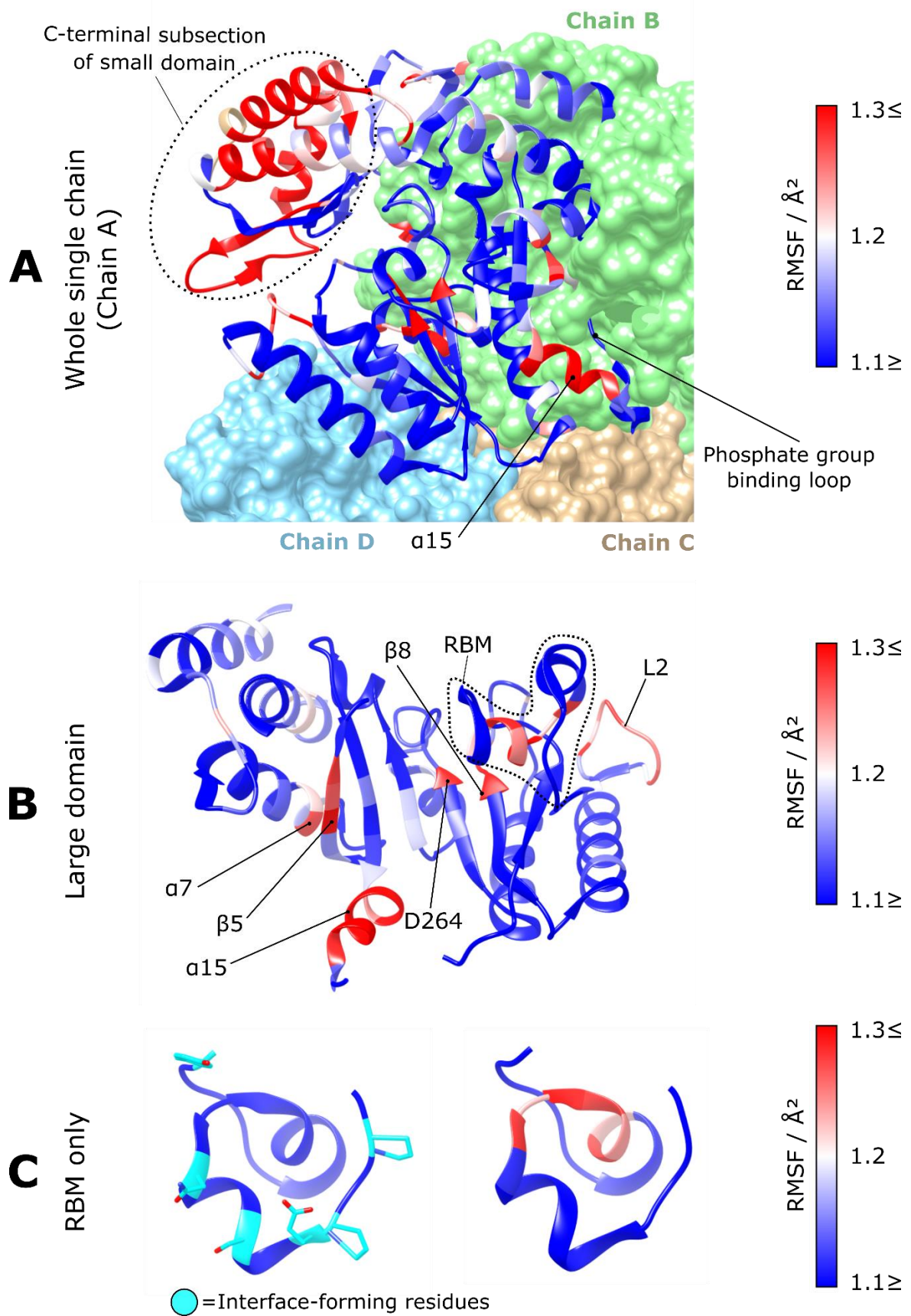
Figure 4.9 PfTA WT coloured by per-residue Cα RMSF values (A) as a whole single chain, the other chains represented as surfaces; (B) as a focused view of the large domain only; (C) as a focused view of the RMB only.
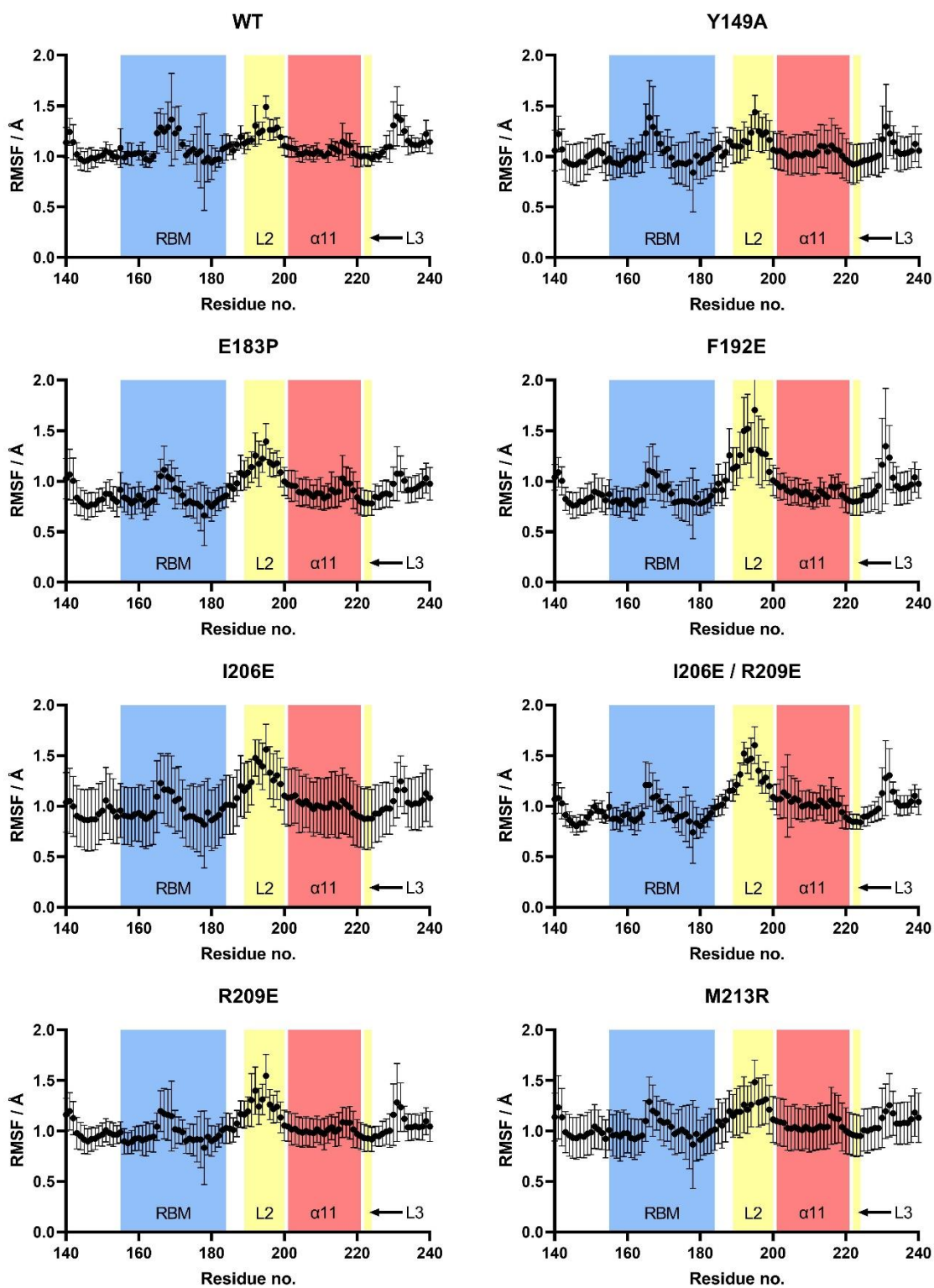
Figure 4.10 RMSF plots for the interface-forming region in PfTA WT and PfTA mutants. In accordance with the homomeric nature of PfTA, RMSFs for each residue are represented as the mean average and standard deviation over all equivalent chains (A,B,C,D) and replicate simulations (3 replicates for mutants, 1 replicate for wild-type).

The RMSF results do not convincingly show any difference in the RMSF between the PfTA WT and PfTA mutants, though there is noticeable variation in the consistency of results for each mutant, in terms of the associated uncertainty values, and this remains true when focusing on the region responsible for tetramer interface formation, into which the mutations were introduced (Figure 4.10). Rather, the RMSF results taken altogether, consistently show a relatively higher RMSF in the central part of the RBM, and in the loop L2, the former of which would suggest agreement with the findings from Börner *et al.* that the RBM is a suitable target for stabilising mutations.[32]

For the "native contacts" analysis, all contacts were counted, not only "native" residue-residue pairs found in contact in the reference (crystal) structure. The data for each residue-residue pair were averaged over all four equivalent chains, and over all replicates. For simplicity, the variability between chains/replicates was assessed by taking the standard deviation, though it was not tested whether the variability follows a normal distribution.

Figure 4.11 to Figure 4.24 shows for each mutant, firstly a plot for the residue-residue pairs with the greatest difference in total fractional contact (between WT and mutant) and secondly, a illustration of the location of those pairs on the inter-dimer interface. For ease of comparison, a threshold of absolute change in total fractional contact of 1.5 is used, to avoid overcrowding the illustrations. The results reveal a contrast between mutants in terms of how localised the effects of mutation are, and in whether the effect is predominantly increased or decreased contact.

The most localised effect is seen in M219R, in which only three residue-residue pairs have significantly changed fractional contact. Specifically, R219 loses contact with Q187, and gains contact with N156, whilst Q187 gains contact with Q220. These changes could be rationalised by a slight twisting of α11 about its axis (Figure 4.24) to accommodate the larger Arg residue and/or permit hydrogen bond formation at the guanidyl moiety.

Similarly, in Y155A, the change from Tyr to the smaller Ala directly accounts for the two significant losses of contact (A155-M219 and A155-Q216), and it could be the same steric effect enables the pairs I212-P197 and Q187-Q220 to achieve increased contact.

The Glu introduced in mutation F198E shows a large increase in contact with R215, though Figure 4.15indicates that this is not uniform for all equivalent chains. Increased contacts for

the pairs P180-M182 (RBM-RBM) and F186-Q216 (β7-α11), which in PfTA WT have almost no contact, are harder to rationalise, except by a larger scale of structural change, with the secondary structure elements (α11, RBM etc) needing to change position relative to one another, or the RBM needing to change conformation, in order to bring the residues into contact.

The four mutants E189P, I212E, R215E and I212E/R215E show the largest scale of change. In both the single mutant I212E and the double mutant I212E/R215E, the change is predominantly a loss of contact between α11 and L2, with very few pairs anywhere across the inter-dimer interface having increased contact. By contrast, in E189P and R215E there is widespread increase of contact concentrated around the RBM (RBM-α11, RBM-β6). L1, which is even further away from the mutation sites than the RBM is, also shows an increase in contact (P146-D147).
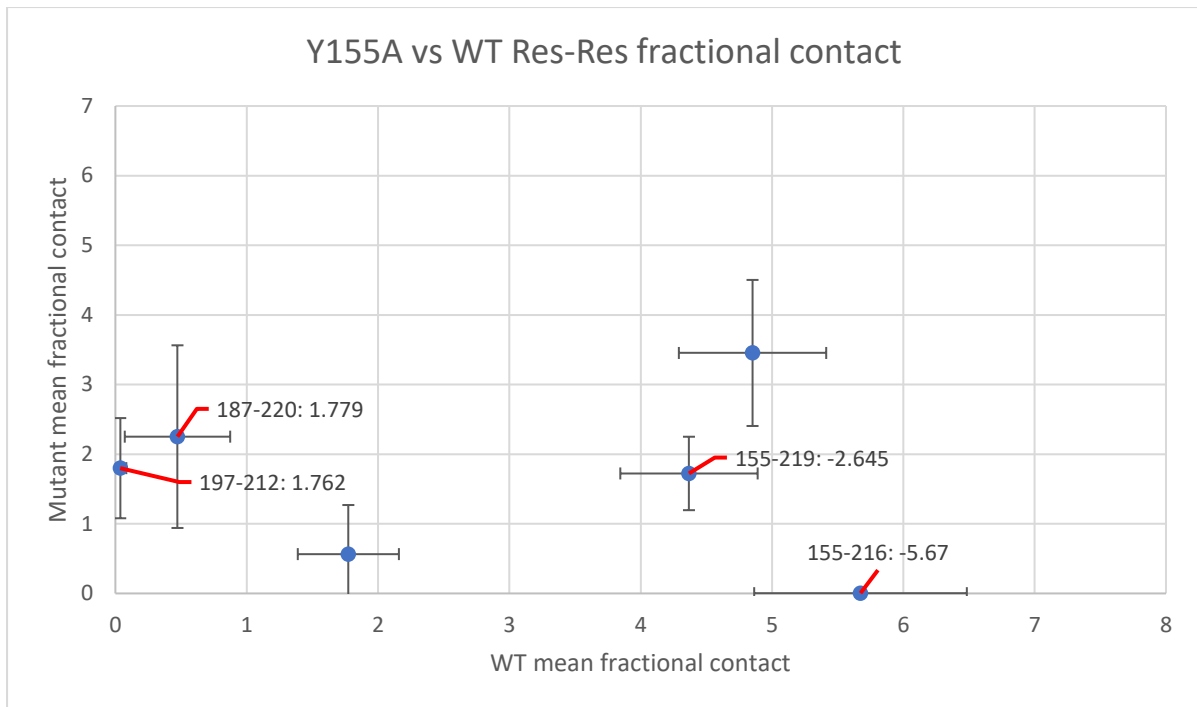
Figure 4.11 Residue-residue pairs plotted by mean fractional contact time in the Y155A mutant (y-axis) versus wild-type (x-axis) simulations for PfTA. Error bars represent the standard deviation over all replicates over all chains, which are equivalent by NCS (n=12 for mutant, n=4 for WT). Pairs with an absolute mutant-to-WT difference greater than 1.5 are labelled, but those with a difference smaller than 1.0 are not plotted, for visual clarity.

Figure 4.12 PfTA WT inter-dimer interface. Residues with the greatest loss (orange) or gain (green) in residue-residue contact are shown, for the calculated effect of the **Y155A** mutation, with the value of the change (and difference in total fractional contact) labelled for each residue-residue pair.

90

Figure 4.13 Residue-residue fractional contact for E189P vs WT.

Figure 4.14 PfTA WT inter-dimer interface. Residues with the greatest loss (orange) or gain (green) in residue-residue contact are shown, for the calculated effect of the **E189P** mutation, with the value of the change (and difference in total fractional contact) indicated by the labelled dashed lines.
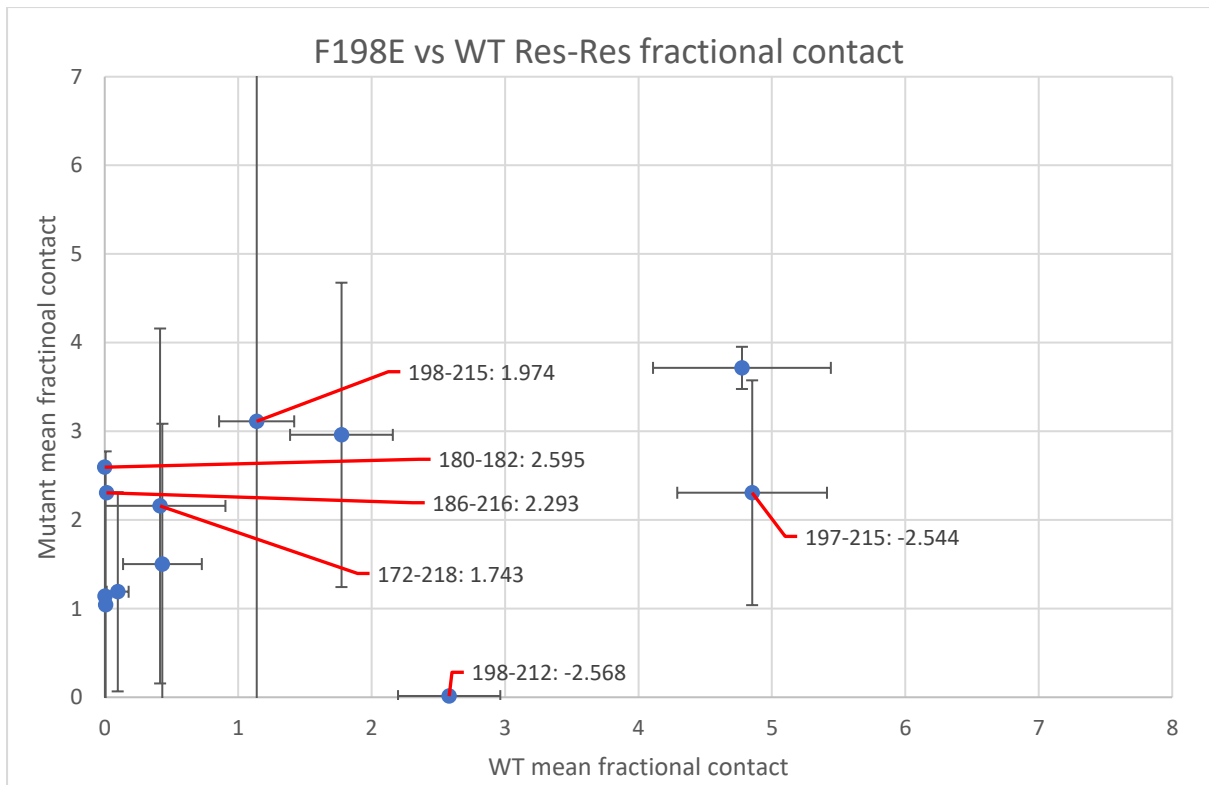
Figure 4.15 Residue-residue fractional contact for F198E vs WT.
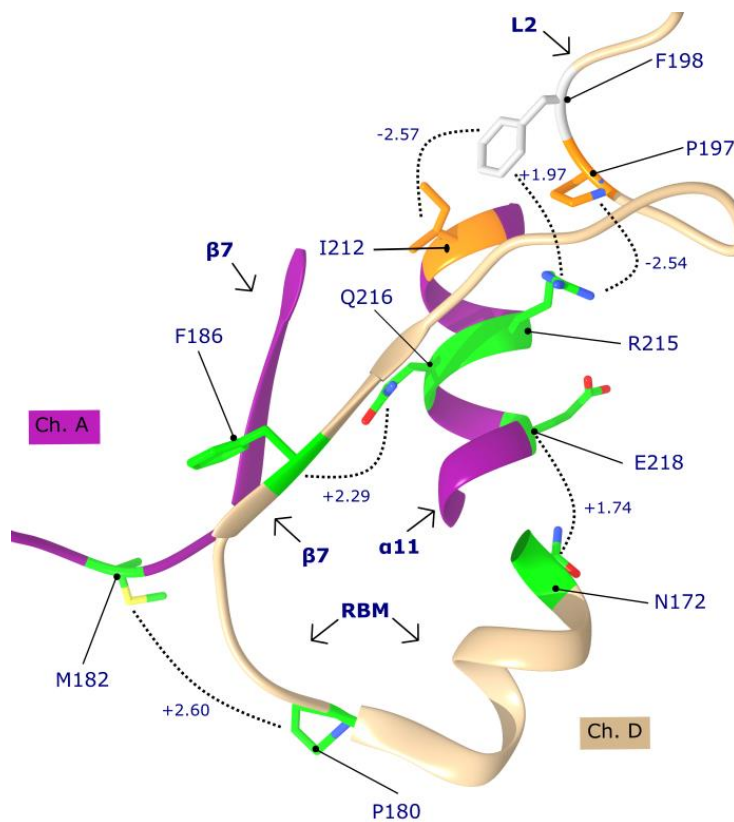
Figure 4.16 PfTA WT inter-dimer interface. Residues with the greatest loss (orange) or gain (green) in residue-residue contact are shown, for the calculated effect of the **F198E** mutation, with the value of the change (and difference in total fractional contact) indicated by the labelled dashed lines.
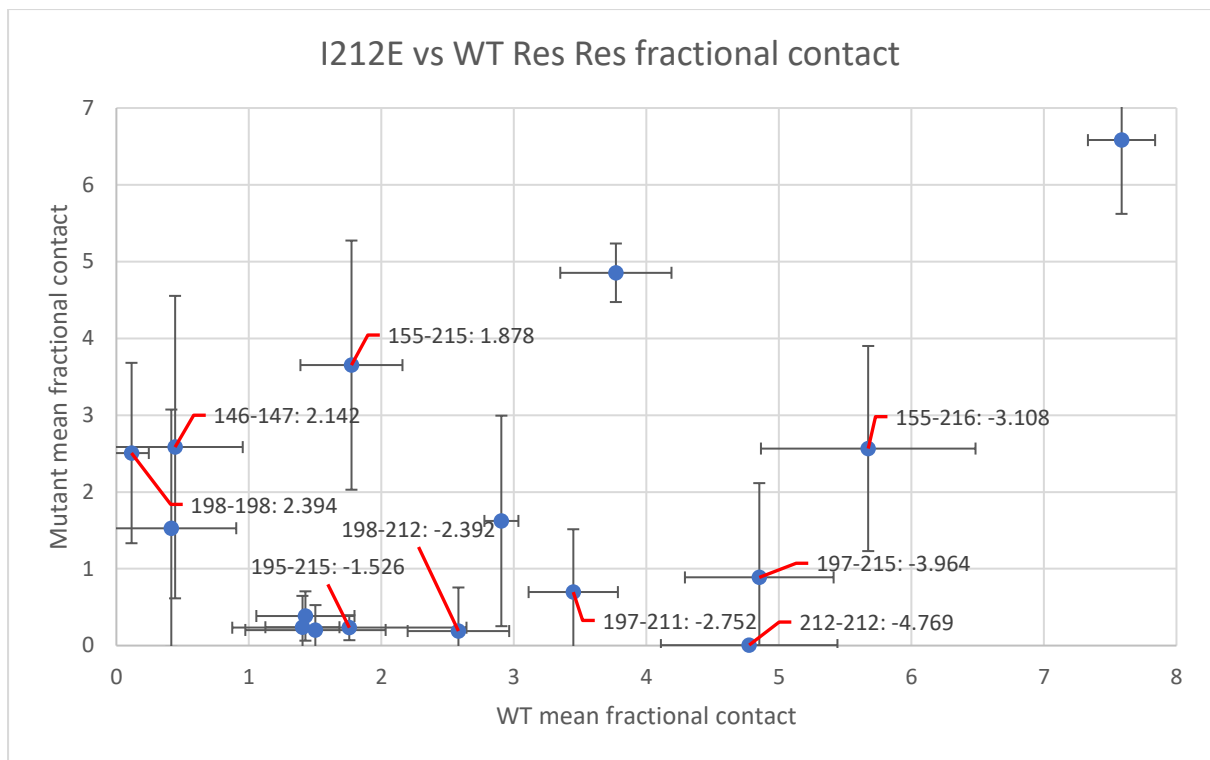
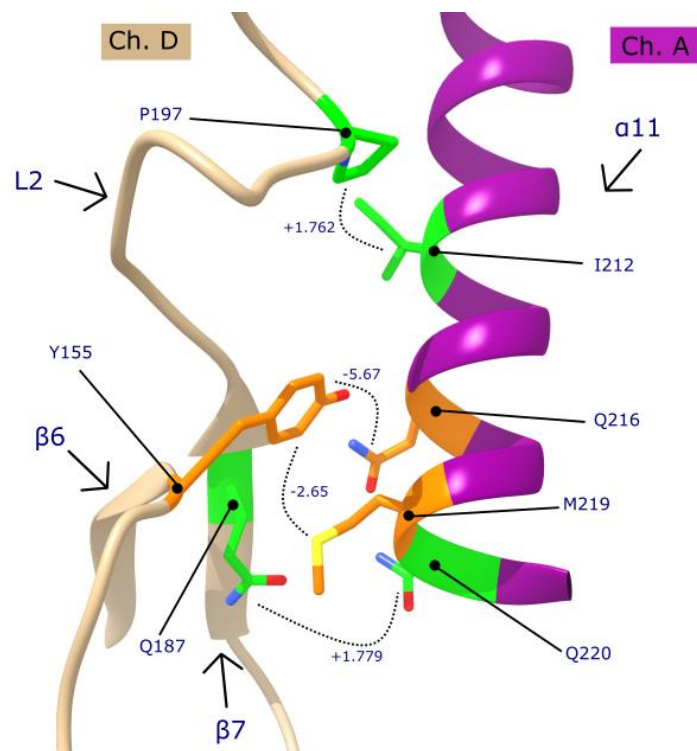Figure 4.17 Residue-residue fractional contact for I212E vs WT.
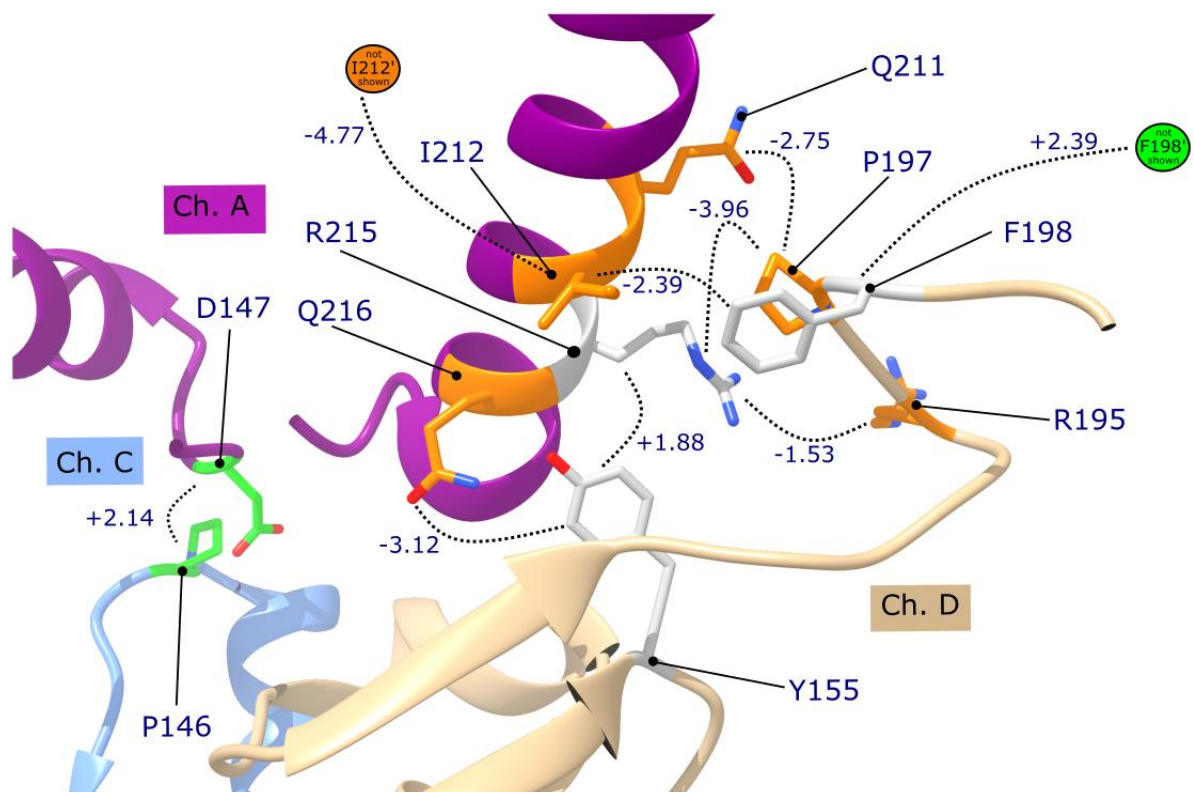
Figure 4.18 PfTA WT inter-dimer interface. Residues with the greatest loss (orange) or gain (green) in residue-residue contact are shown, for the calculated effect of the **I212E** mutation, with the value of the change (and difference in total fractional contact) indicated by the labelled dashed lines.
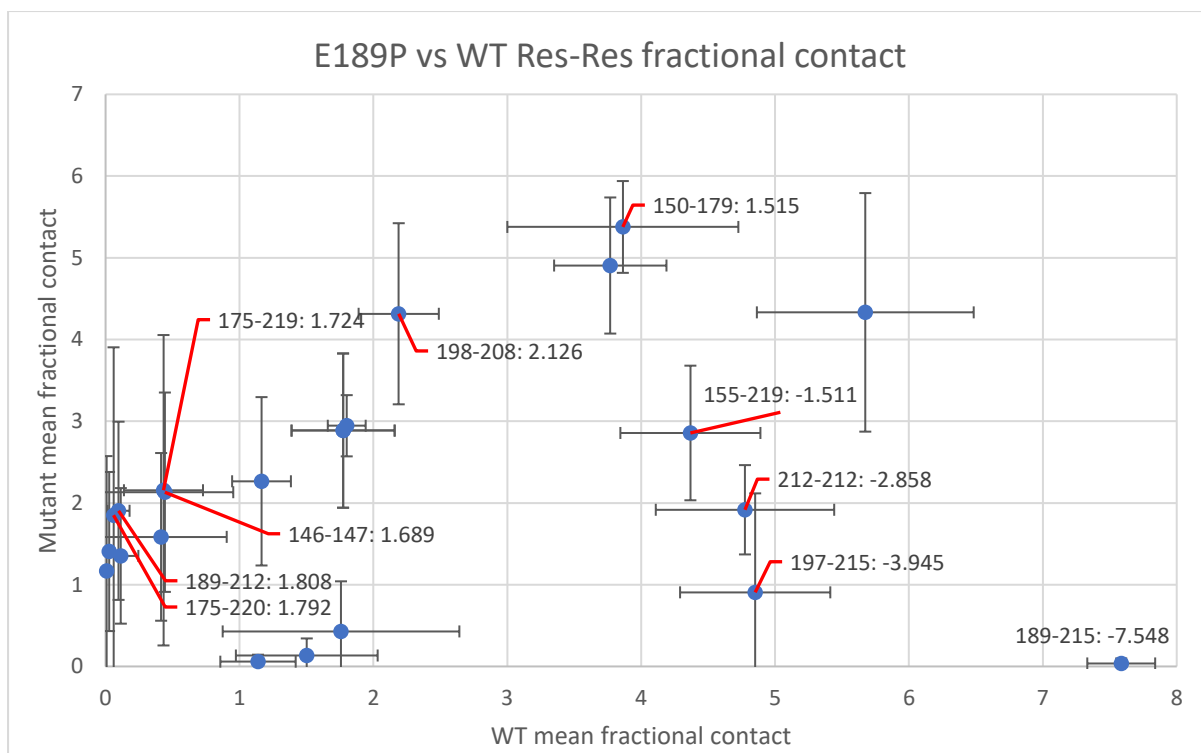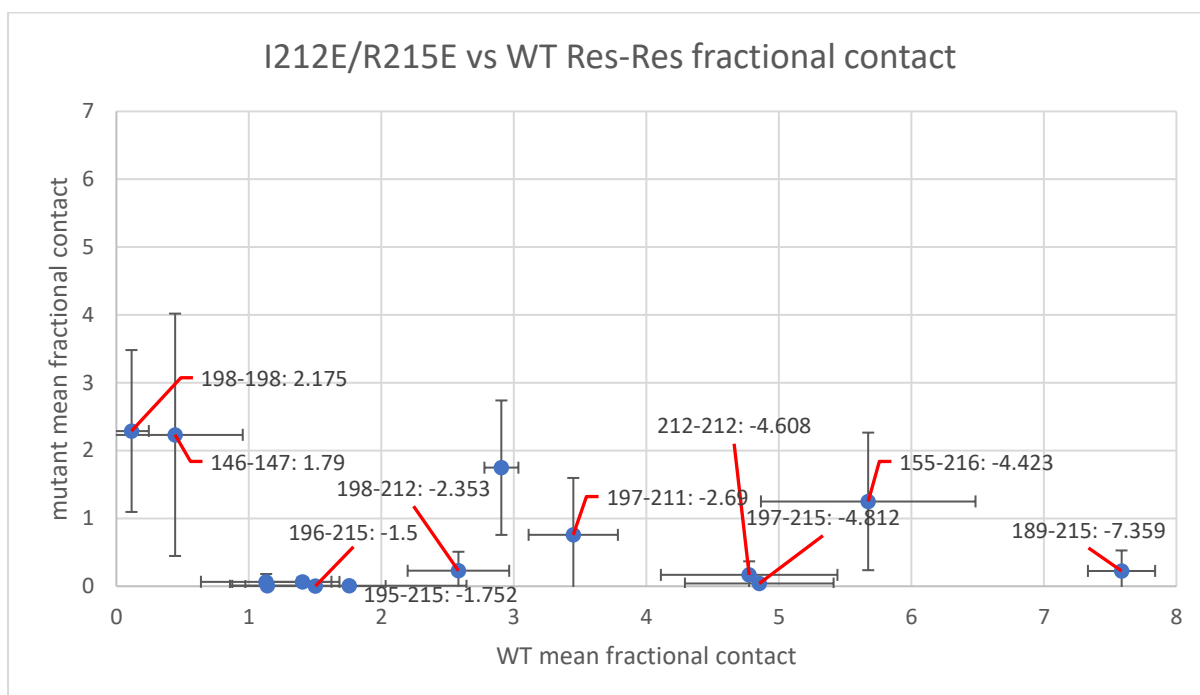
Figure 4.19 Residue-residue fractional contact for I212E/R215E vs WT.
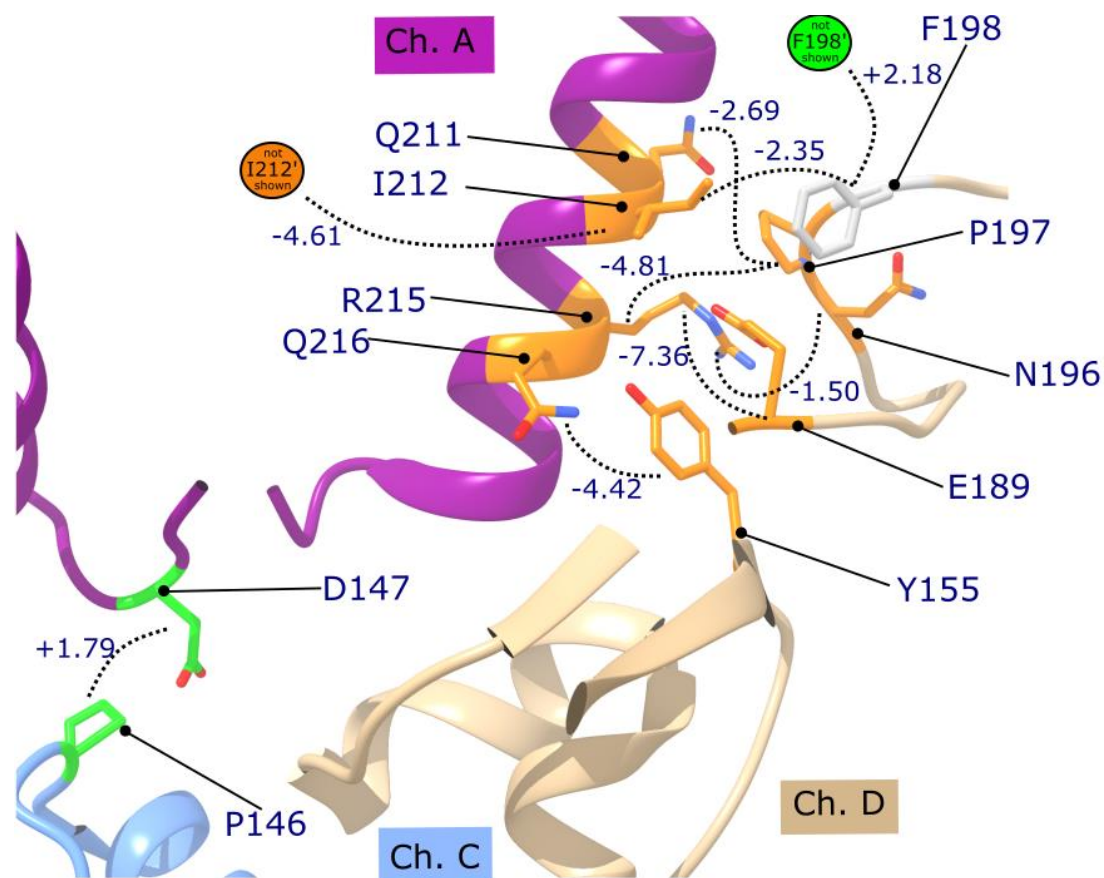
Figure 4.20 PfTA WT inter-dimer interface. Residues with the greatest loss (orange) or gain (green) in residue-residue contact are shown, for the calculated effect of the **I212E/R215E** double mutation, with the value of the change (and difference in total fractional contact) indicated by the labelled dashed lines.
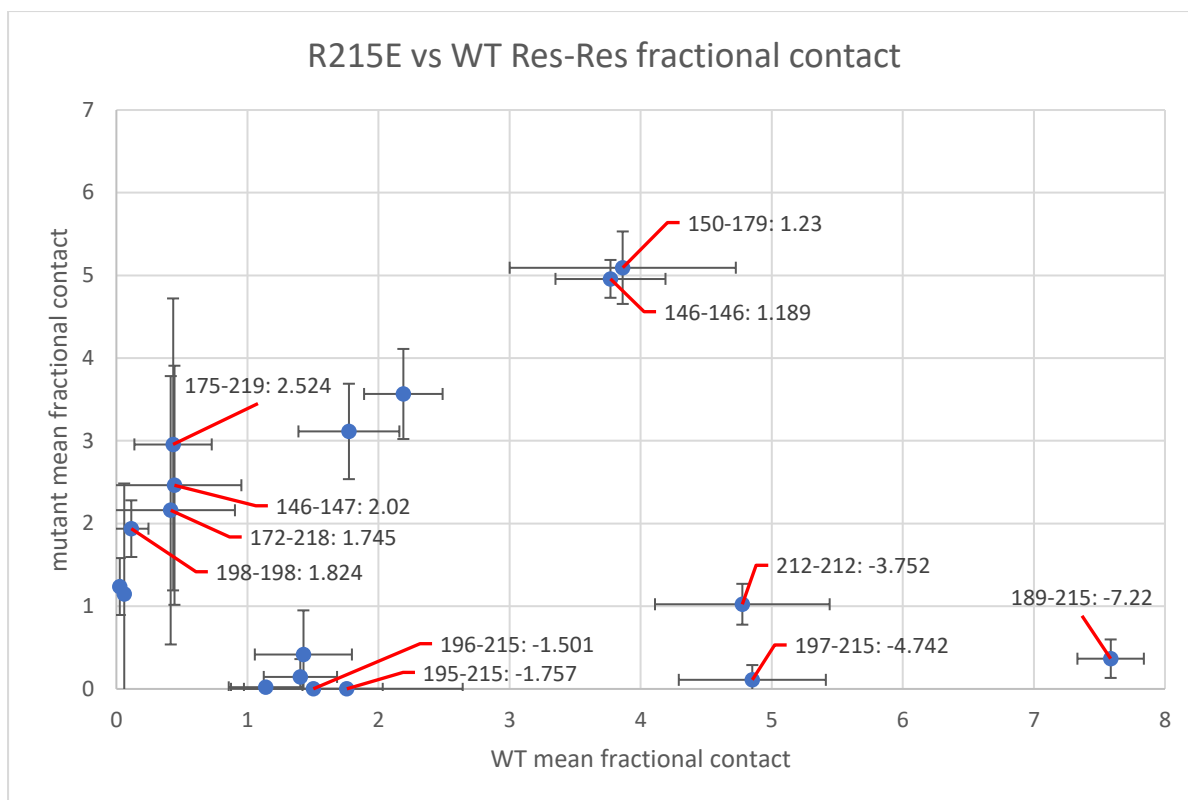
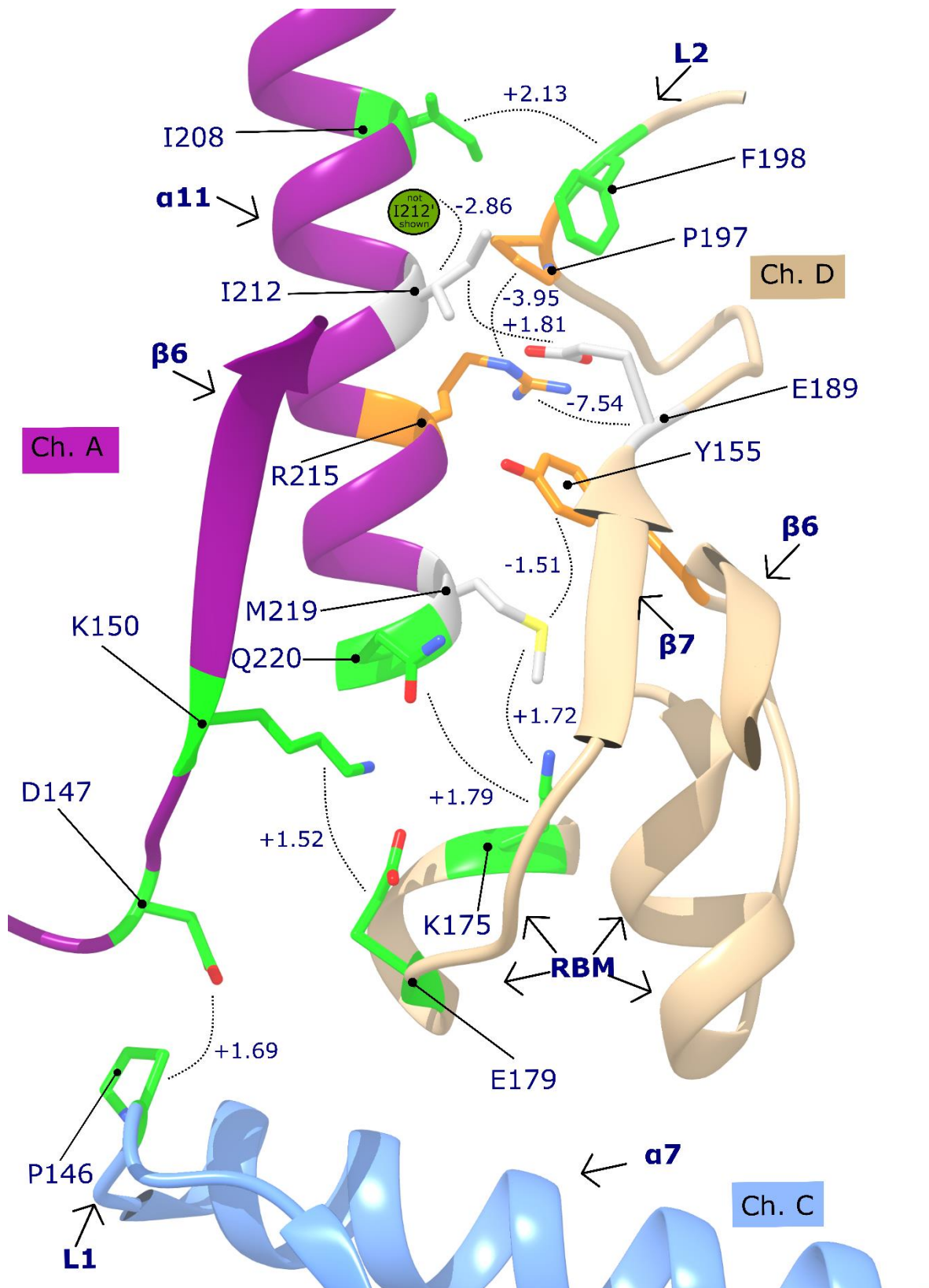Figure 4.21 Residue-residue fractional contact for R215E vs WT.
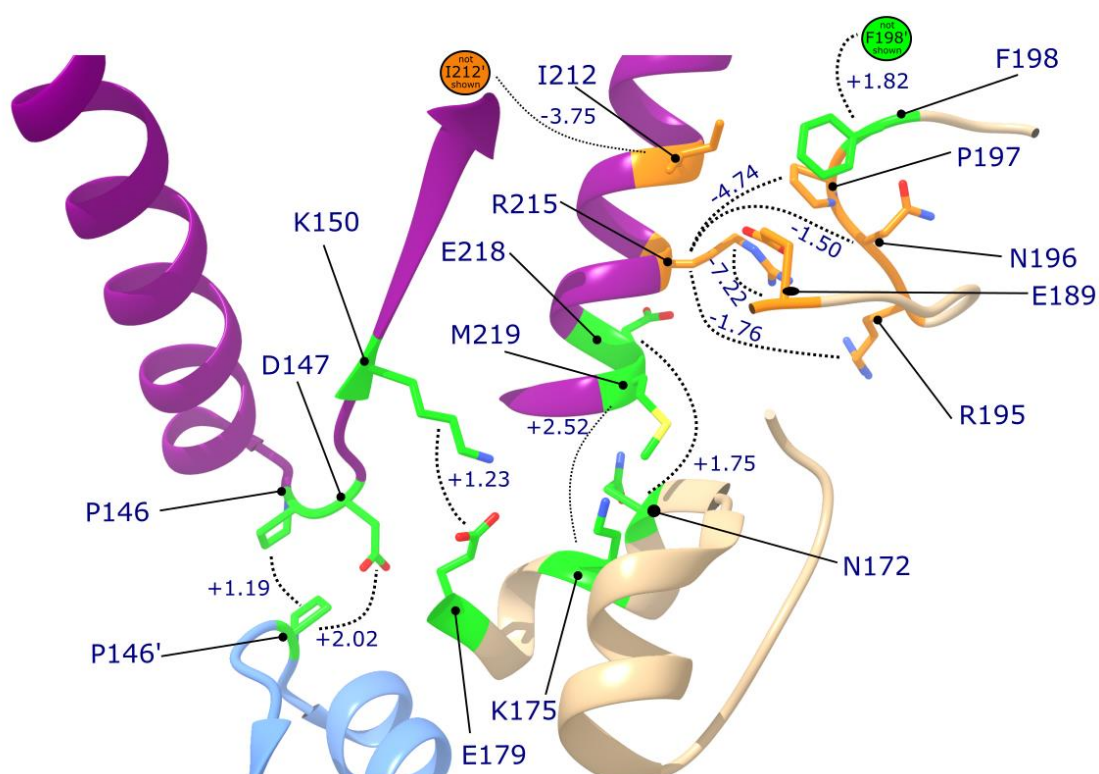
Figure 4.22 PfTA WT inter-dimer interface. Residues with the greatest loss (orange) or gain (green) in residue-residue contact are shown, for the calculated effect of the **R215E** mutation, with the value of the change (and difference in total fractional contact) indicated by the labelled dashed lines.
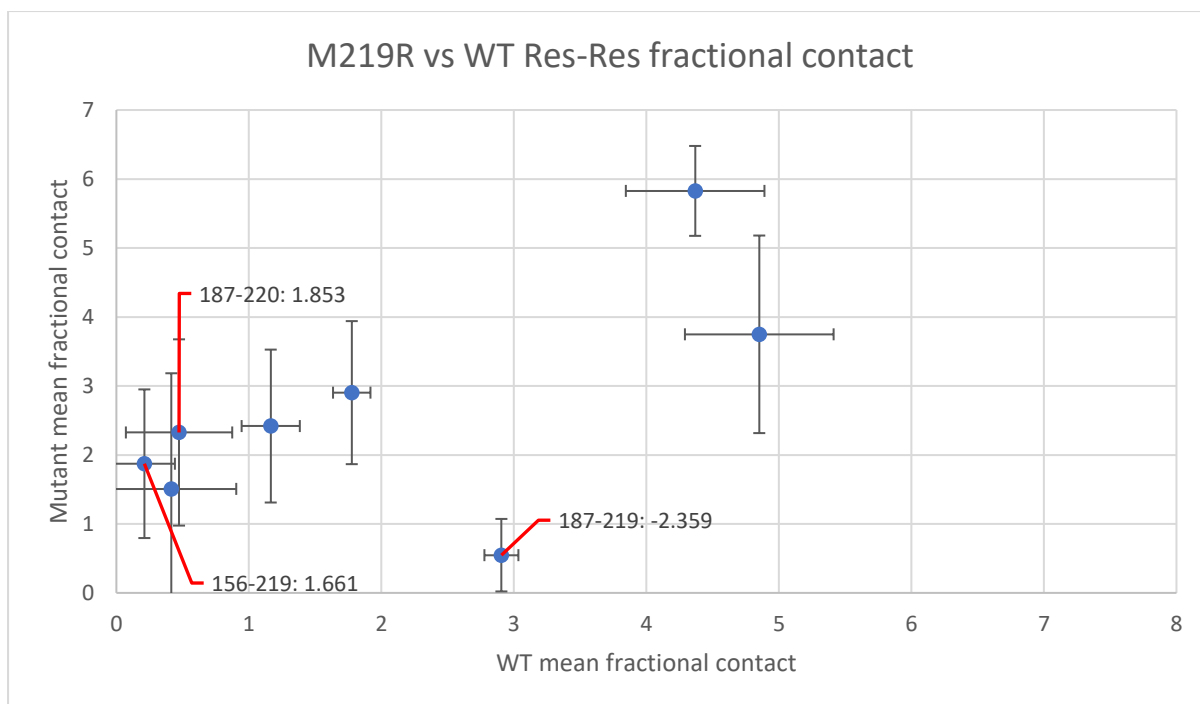
Figure 4.23 Residue-residue fractional contact for M219R vs WT.
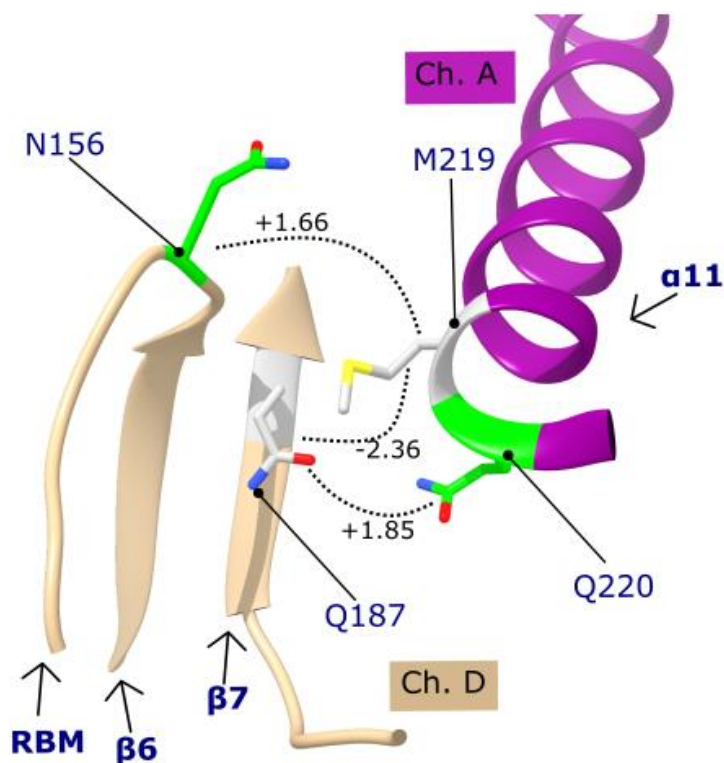
Figure 4.24 PfTA WT inter-dimer interface. Residues with the greatest loss (orange) or gain (green) in residue-residue contact are shown, for the calculated effect of the **M219R** mutation, with the value of the change (and difference in total fractional contact) indicated by the labelled dashed lines.

Change in the electrostatic component of linear interaction energy (LIE$_{elec}$) was also calculated for interfacial residue-residue pairs (Table 4-4), since several of the mutations aimed to introduce charge-charge repulsion. This calculation does not account for interactions with the

rest of the protein or the solvent, but a larger change in LIE_{elec} would suggest a mutation which is more disruptive within the original PfTA WT interface structure. The results suggest that E189P, R215E and the double mutation I212E/R215E would be the most disruptive, through the intended disruption of the R215-E189 salt-bridge pair. Interestingly, in the F198E mutation, the largest electrostatic repulsion introduced is E198-E204 (rather than the intended E198-E198), and in I212E the repulsive effect is comparable for both pairs E189-E212 and E212-E212.

	Res1	Res2	Calculated WT LIE_{elec} / kcal mol^{-1}		Calculated Mutant LIE_{elec} / kcal mol^{-1}		Calculated ΔLIE_{elec} (Mutant-WT) / kcal mol^{-1}
			Mean	S.D.	Mean	S.D.	
Y155A	150	179	-73.78	8.98	-81.68	2.92	-7.90
	189	215	-83.11	1.21	-76.90	8.37	6.21
	155	215	7.11	0.20	2.11	1.40	-5.01
	150	172	-7.19	4.05	-3.13	5.11	4.06
E189P	189	215	-83.11	1.21	2.25	0.61	85.36
	150	179	-73.78	8.98	-83.77	3.59	-10.00
	155	215	7.11	0.20	1.69	1.16	-5.42
	195	215	-2.43	1.83	2.59	3.12	5.02
	196	215	5.86	1.32	0.89	2.75	-4.97
	189	211	4.68	0.58	-0.24	0.07	-4.92
	179	220	10.40	1.59	15.07	3.77	4.67
	147	147	4.95	0.01	9.01	1.99	4.06
	141	147	-72.46	7.04	-68.42	12.74	4.05
F198E	198	204	1.53	0.04	26.44	9.32	24.91
	198	215	2.18	0.17	-7.56	21.60	-9.74
	198	198	0.17	0.02	5.92	4.18	5.75
	198	207	-0.32	0.02	5.37	2.50	5.69
	175	219	-9.02	2.75	-14.03	7.48	-5.02
	172	218	-5.07	1.25	-9.15	4.27	-4.09
	218	172	-5.07	1.25	-9.15	4.27	-4.09
I212E	189	212	2.46	0.33	27.71	11.07	25.25
	212	212	0.09	0.01	24.19	20.23	24.10
	212	215	0.86	0.14	-7.40	2.60	-8.27
	150	175	32.05	7.99	23.96	8.66	-8.09
	141	147	-72.46	7.04	-64.96	16.41	7.50
	189	215	-83.11	1.21	-75.85	9.10	7.27

	150	179	-73.78	8.98	-79.81	7.48	-6.03
	147	147	4.95	0.01	10.16	1.88	5.21
I212E/R215E	189	215	-83.11	1.21	35.07	15.89	118.18
	189	212	2.46	0.33	47.74	14.84	45.28
	212	212	0.09	0.01	39.61	24.22	39.52
	155	215	7.11	0.20	-4.99	2.39	-12.11
	189	189	0.08	0.07	10.28	8.99	10.20
	141	147	-72.46	7.04	-63.49	13.93	8.97
	196	215	5.86	1.32	-2.64	0.70	-8.50
	150	175	32.05	7.99	23.87	5.11	-8.18
	212	215	0.86	0.14	8.90	13.39	8.03
	190	215	-5.14	0.21	1.92	0.99	7.06
	197	215	4.36	0.54	-1.83	0.54	-6.18
	150	179	-73.78	8.98	-77.85	5.10	-4.07
R215E	189	215	-83.11	1.21	40.89	8.66	124.01
	155	215	7.11	0.20	-4.24	0.96	-11.36
	150	179	-73.78	8.98	-83.26	1.92	-9.49
	196	215	5.86	1.32	-3.01	0.80	-8.87
	175	219	-9.02	2.75	-16.94	7.14	-7.92
	190	215	-5.14	0.21	1.97	0.74	7.11
	197	215	4.36	0.54	-1.75	0.42	-6.11
	150	172	-7.19	4.05	-1.38	3.52	5.81
	179	220	10.40	1.59	16.07	2.31	5.68
	172	218	-5.07	1.25	-10.75	4.89	-5.68
	141	147	-72.46	7.04	-68.12	7.11	4.34
	147	147	4.95	0.01	9.23	1.65	4.28
M219R	175	219	-9.02	2.75	-0.04	1.88	8.97
	189	219	-2.39	0.17	-11.34	3.32	-8.95
	170	219	-1.23	0.54	-6.52	6.65	-5.30
	172	219	-1.92	1.74	3.13	2.88	5.05
	155	219	-1.09	0.07	3.66	0.60	4.75
	171	219	-1.97	0.68	2.61	1.44	4.58
	150	175	32.05	7.99	27.51	5.21	-4.55
	172	218	-5.07	1.25	-9.24	5.65	-4.18

Table 4-4 Change in residue-residue pair electrostatic linear interaction energy for seven PfTA WT mutants, ordered by largest-to-smallest ΔE_{elec}, (cutoff ± 4 kcal/mol)

Therefore, whilst native contacts analysis would suggest E189P, R212E, R215E and I212E/R215E to be the most disruptive mutations, the LIE$_{elec}$ would suggest that among these four, the latter two would be most effective in terms of introducing an electrostatic contribution. Limitations to the accuracy of these predictions are partially due to the fact that such interfaces,

unlike protein-ligand interfaces, involve much larger networks of interaction which cannot quickly reorganise when the mutation is introduced into the simulation, which leads to inaccurate sampling of conformational space[86]. Additionally, these analyses do not attempt to provide any information on whether the mutations would have any deleterious effect on the protein stability.

The seven PfTA mutants were subsequently prepared by mutagenesis, in the hope that biophyscial results would identify suitable dimeric PfTA variants for crystallography, and that these biophysical results would also provide validation for the simulation-based analyses, to establish whether these methods would be useful for screening mutations *in silico*.

4.4 Preparation and expression of PfTA mutants

The PfTA single mutants Y155A, E189P, F198E, I212E, R215E, M219R and double mutant I212E/R215E were prepared by site-directed mutagenesis described above (section 0). All mutants except F198E were successfully obtained using the first set of designed primers, though in some cases it was necessary to repeat the transformation of Novablue and isolation of plasmid DNA, where sequencing indicated that the plasmid did not bear the designed mutation, presumably as a result of incomplete DpnI digestion. However, the F198E mutation was not obtained, despite trying three different primer designs (Appendix 8.3) and 6 different annealing temperatures (54 °C to 64 °C in 2 °C intervals).

Expression of the six remaining mutants followed the same method as for PfTA WT (section 3.2). Because initial purification attempts had revealed that most mutants expressed poorly, the level of expression was therefore tested by chemical lysis of harvested samples of pre- and post- induction culture, followed by SDS-PAGE of the soluble and insoluble fractions, before making further attempts at purification. This expression testing showed no evidence of any expression for I212E nor the double mutant I212E/R215E, despite being careful to BL21 from a competent batch with known ability to express PfTA WT.

Y155A and M219R showed overexpression, but at unexpected molecular masses, being expected at 52.5 kDa but presenting ~25 kDa for Y155A and ~45 kDa for M219R. BL21 was transformed with the sequenced plasmids, and two single colonies each of Y155A, M219R and WT were cultured and induced, to provide duplicate results, with the WT included for reference. The results (Figure 4.25) were alike for both duplicates. Reasons for an unexpectedly low molecular mass might include an unintended mutation, or an abundance of rare codons[87].

However, both Y155A and M219R are single mutations, and the sequencing indicated no mutation other than the designed mutation.

The purification of the M219R mutant was attempted despite the unpromising result from the overexpression test. Surprisingly, SDS PAGE of the IMAC fractions did reveal a band in the expected position for PfTA, but unfortunately no fraction contained solely this band. The bands with the least impurity were pooled for analytical SEC, which yielded a poorly-resolved, multiple-peak chromatogram. Despite this, SDS-PAGE of the fractions showed a band compatible with the expected mass of PfTA, which corresponded to an elution volume of 9.1 mL in the SEC chromatogram. Even if this band were to be PfTA M219R, this suggested the it would be a higher, not lower, oligomerisation state than PfTA WT, since this elution volume is below the expected elution volume of the tetrameric PfTA WT (~12.5-12.7 mL), and therefore still of less interest to pursue.
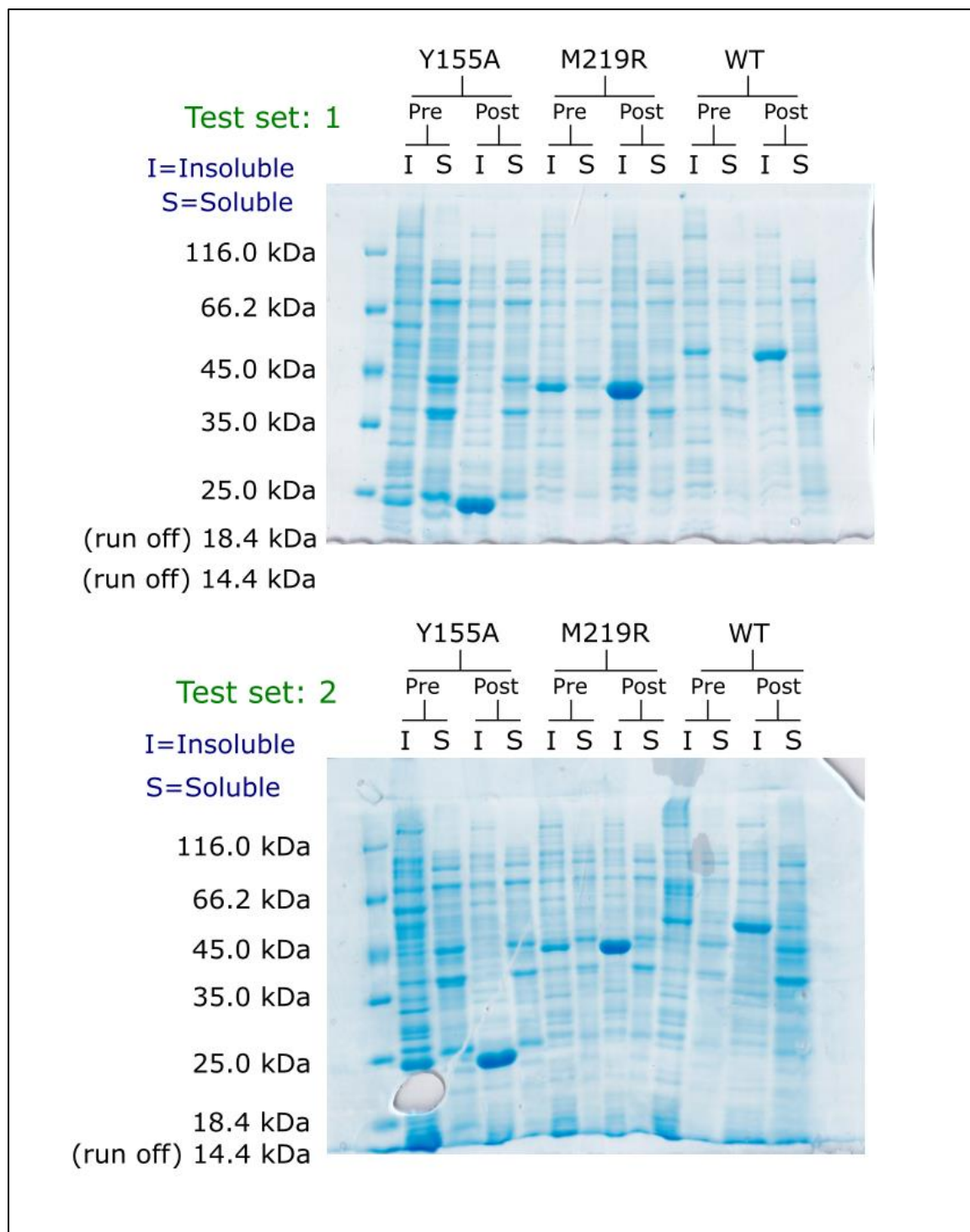
Figure 4.25 SDS-PAGE gels for PfTA Y155A and M219R with PfTA WT as a reference. Duplicate sets were derived from separate individual colonies from plates of BL21 freshly transformed with the respective plasmid. In both duplicates overexpression was indicated for Y155A and M219R at an unexpected molecular mass, whereas PfTA WT band appeared in the expected position (expected mass 52.5 kDa).
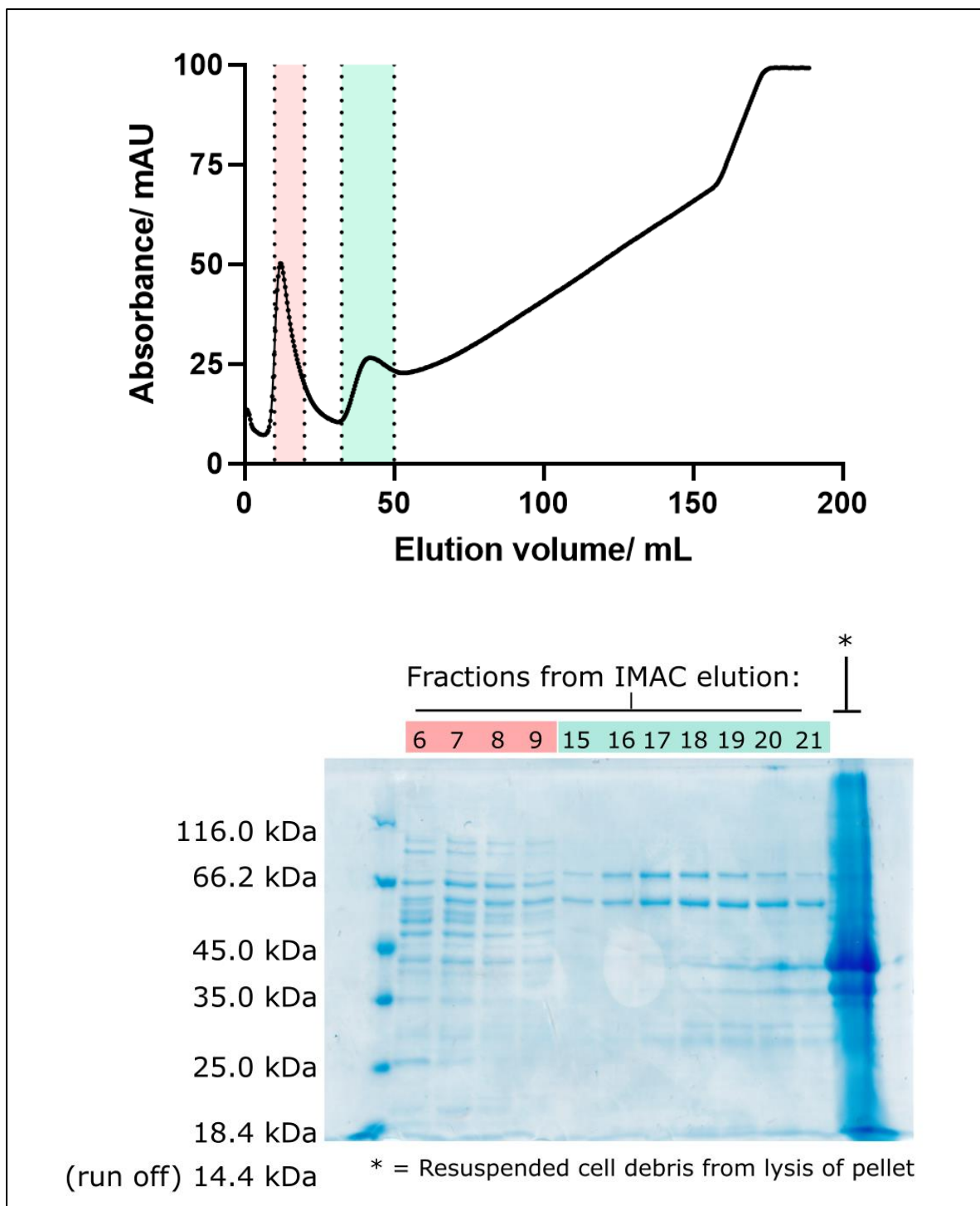
Figure 4.26 IMAC elution profile for PfTA M219R with SDS PAGE of eluted fractions. Fractions highlighted pink and green correspond to pink and green peaks in chromatogram. Expected mass for M219R is 52.5 kDa.
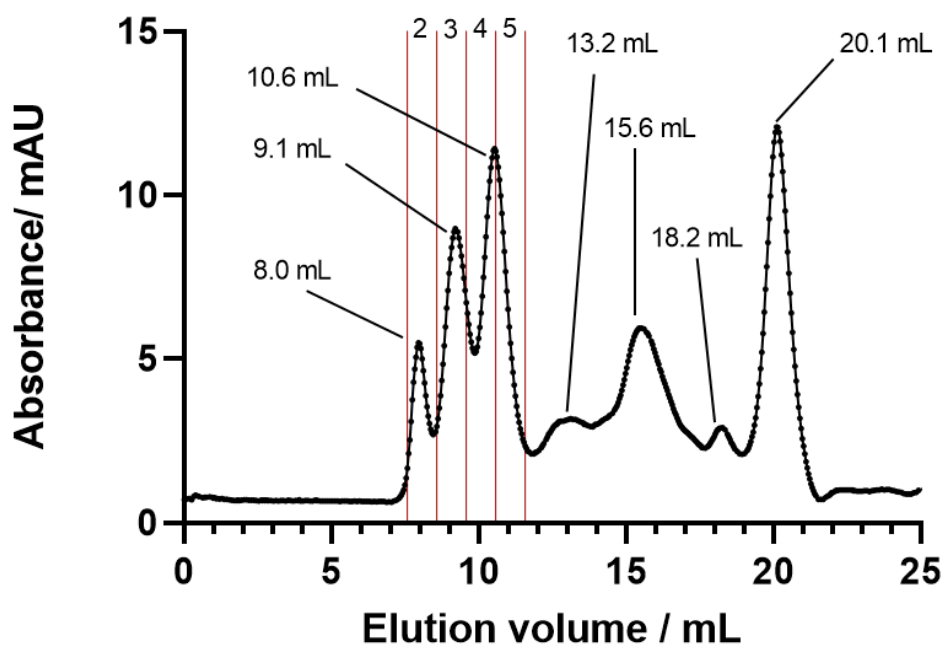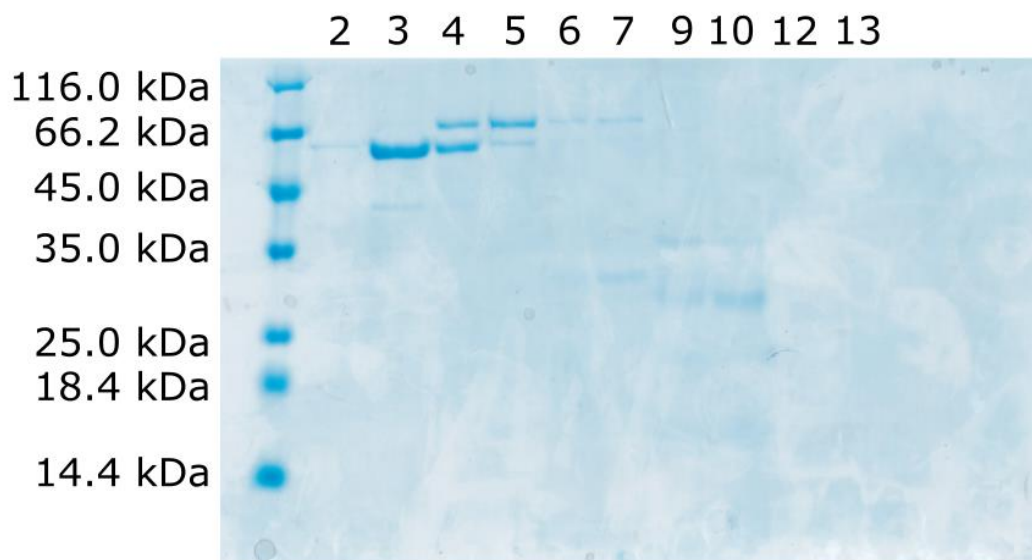
Figure 4.27 SEC profile for PfTA M219R with SDS-PAGE of eluted fractions.

An initial attempt at IMAC purification of E189P using a wash step at 5% buffer B (corresponding to 34.5 mM imidazole), indicated that it would likely be possible to obtain E183P in good purity if the imidazole level in the wash step was slightly increased (Figure 4.28). However, analytical SEC showed a peak elution volume of 9.1 mL for those fractions which presented a SDS PAGE band compatible with the expected mass of PfTA (Figure 4.29). Therefore, just like M219R, even if this SDS PAGE band does represent E198P, it is unlikely that E189P exists in the dimeric state, making it of less interest to pursue.
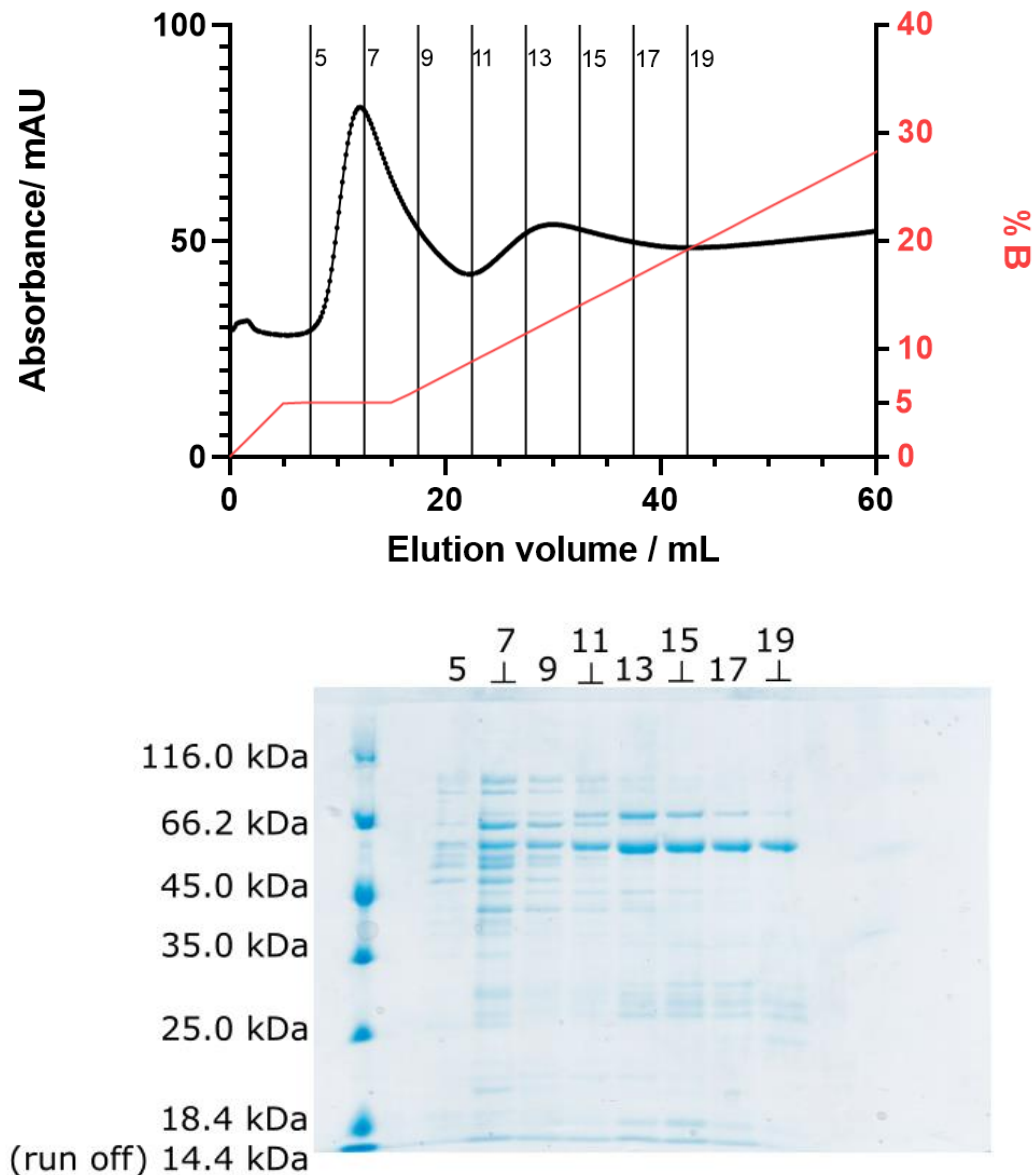
Figure 4.28 IMAC elution profile of PfTA E189P with SDS PAGE of eluted fractions.
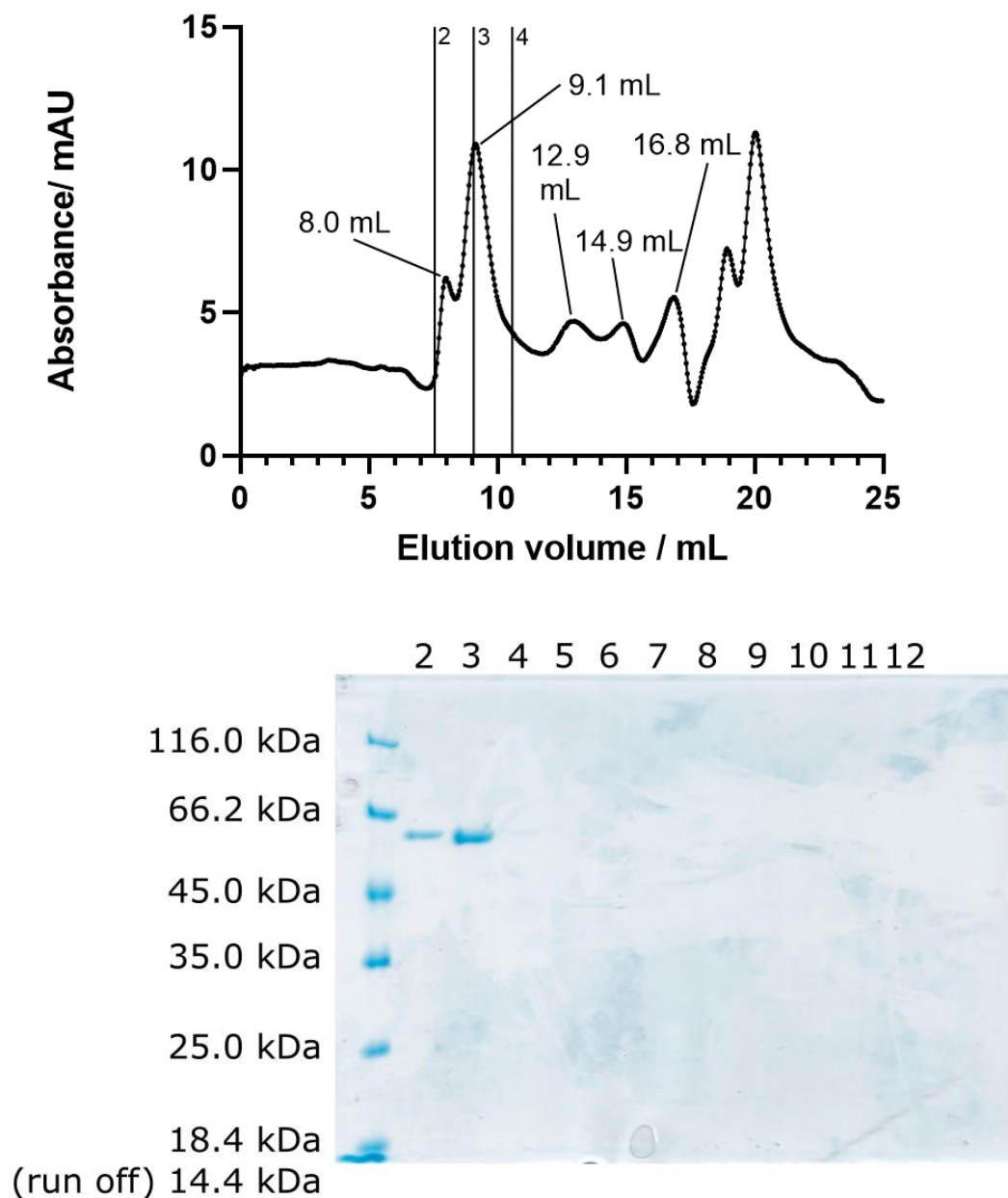
Figure 4.29 Analytical SEC elution profile for PfTA E189P with SDS PAGE of eluted fractions.

R215E was by far the most well-expressed mutant, and was successfully purified by IMAC (Figure 4.30). The fractions under the single peak shown in the analytical SEC chromatogram produced an SDS-PAGE band compatible with the expected mass of 52.5 kDa, whilst the elution volume of 13.4 mL corresponds to a calculated mass of 122 kDa (expected mass of dimer 105 kDa). PLP was included in the SEC buffer (HEPES 50 mM, PLP 0.1 mM, NaCl 150 mM) for the avoidance of doubt that the dimeric state would exist only under low-PLP conditions as in the *E. coli* TA reported by Liu *et al.*[41]
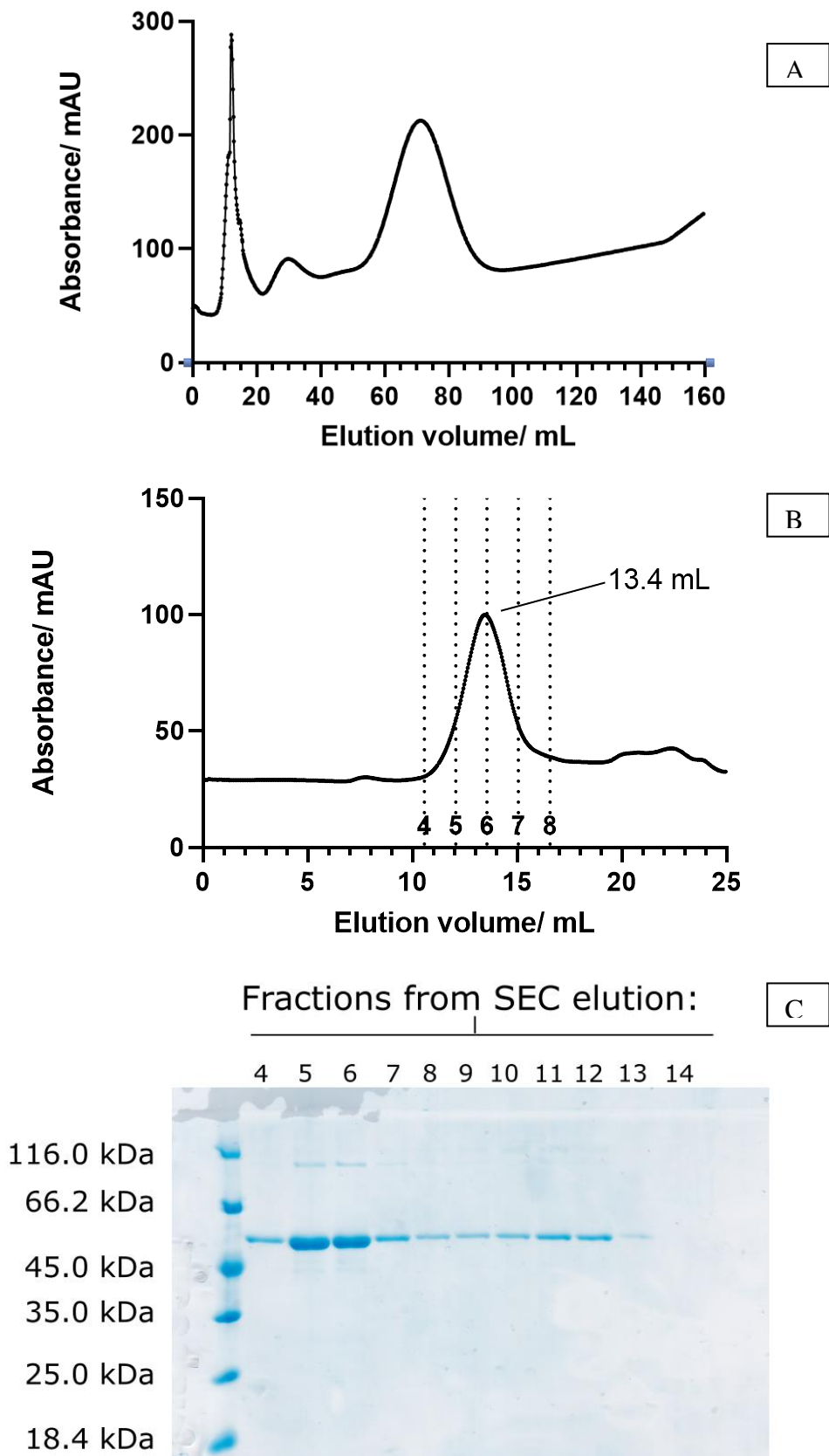
Figure 4.30 **(A)** IMAC elution profile showing peak elution of R215E at 65 mL. **(B)** Analytical SEC elution profile showing R215E peak elution at 13.4 mL. The fractions marked by abscissae correspond to the fractions in the SDS-PAGE gel **(C)**.

4.5　Chapter conclusion

PfTA WT was compared to a set of its biophysically evidenced tetrameric and dimeric homologs to guide the mutation of PfTA WT to change its oligomerisation state from tetrameric to dimeric. Whilst this the structural comparison did not yield information on which residues to target, it did inform the choice of substitute residue to use for each mutation. Seven mutants were designed (six single and one double).

The seven PfTA mutants and the PfTA WT were simulated, and the Amber suite tools used to identify changes in interfacial residue-residue pair contacts and electrostatic linear interaction energies, with the aim of providing a method to predict how likely would be each mutation to cause the disruption of the inter-dimer interface to yield a dimeric PfTA. From the simulation analyses, the following relative likelihood of disrupting the inter-dimer interface was hypothesised:

$$R215E, I212E/R215E > I212E, E189P > Y155A, F198E, M219R$$

Whilst a greater diversity of expressible dimeric mutants would have been desirable for the purposes of validation, nevertheless, the later verification of R215E as an entirely dimeric species regardless of protein or PLP concentration (described in Chapter 5) did accord with the hypothesis that R215E or I212E/R215E were the likeliest candidates to achieve dimerization in a single mutation.

Mutant	Summary result
F198E	Not obtained. Mutagenesis unsuccessful despite using multiple primers and annealing temperatures
I212E	Not obtained. Mutagenesis successful, but expression not detected
I212E/R215E	Not obtained. Mutagenesis successful, but expression not detected
Y155A	Overexpressed band present at unexpected molecular mass despite expected plasmid sequencing result
M219R	Expression poor; oligomerisation not certainly determined
E189P	Expression poor; oligomerisation not certainly determined
R215E	Expression level good, and oligomerisation state likely to be dimeric

Table 4-5 Summary of results for seven PfTA mutants

5 Crystallographic and biophysical characterisation of PfTA R215E and stability assays in comparison to PfTA WT

5.1 Introduction

The PfTA R215E mutant identified in section 0 was subjected to AUC to confirm the assignment of the dimeric oligomerisation state which had been initially made by SEC, with twice attempted SAXS unfortunately failing to produce usable data, due to sample transit problems which could not be mitigated. The R215E crystal structure was then determined, to establish whether the overall fold and intra-dimer interface remained as expected for PfTA. Once this was established, PfTA R215E and WT were subjected to residual activity and DSF assays to compare their stabilities and dependencies on PLP.

5.2 AUC characterisation

Sedimentation velocity characterisation was performed for PfTA R215E in the same way as previously used for PfTA WT, but using only a single experiment, varying the concentration of R215E but keeping the concentration of PLP constant at 100 µM. All data for PfTA R215E were collected by Dr. Gemma Harris, due to contemporary travel restrictions.

Like PfTA WT, the experiment found no evidence of non-ideality for R215E. The c(s) distribution plot for each sample in the concentration series used for R215E (17.1 µM, 8.57 µM, and 4.29 µM) showed a single major peak at around 4.15 to 4.29 S, except for the 4.29 µM sample, which additionally showed a very broad peak around 12 S of unknown but presumably artefactual origin, since all the samples were prepared from a common stock. With solvent correction, the average sedimentation coefficient corresponds to an MW estimate of 90 kDa, and a minimal MW of 68 kDa, suggesting a dimeric state (dimer expected mass 105 kDa). The estimated MW and minimal MW were calculated using fixed values of $f/f_0 = 1.2$ and $f/f_0 = 1.0$, respectively, in the same way used for analysis of the PfTA WT data.

R215E 17.1 μM

4.15 S
[5.92 S (20,w)]
87.5 kDa

5.62 S
[8.03 S (20,w)]
138 kDa

R215E 8.57 μM

4.22 S
[6.03 S (20,w)]
89.8 kDa

6.70 S
[9.56 S (20,w)]
180 kDa

R215E 4.29 μM

4.29 S
[6.13 S (20,w)]
92.2 kDa

12.75 S
[18.21 S (20,w)]
472 kDa

PfTA R215E Concentration (μM)	Sedimentation coefficient (S)	Sedimentation coefficient (20 °C, water) (S)	MW estimate (for f/f_0= 1.2) (kDa)	Minimum MW (for f/f_0= 1.0) (kDa)
17.1	4.15	5.92	87.5	66.6
8.57	4.22	6.03	89.8	68.3
4.29	4.29	6.13	92.2	70.2

Figure 5.1 AUC results for PfTA R215E concentration series, with 100 μM PLP, with estimated MW of 90 kDa, and a minimal mass of 68 kDa. Minimal variation in sedimentation coefficient for different R215E concentrations suggests no evidence of non-ideal behaviour.

5.3 Summary of diffraction data processing
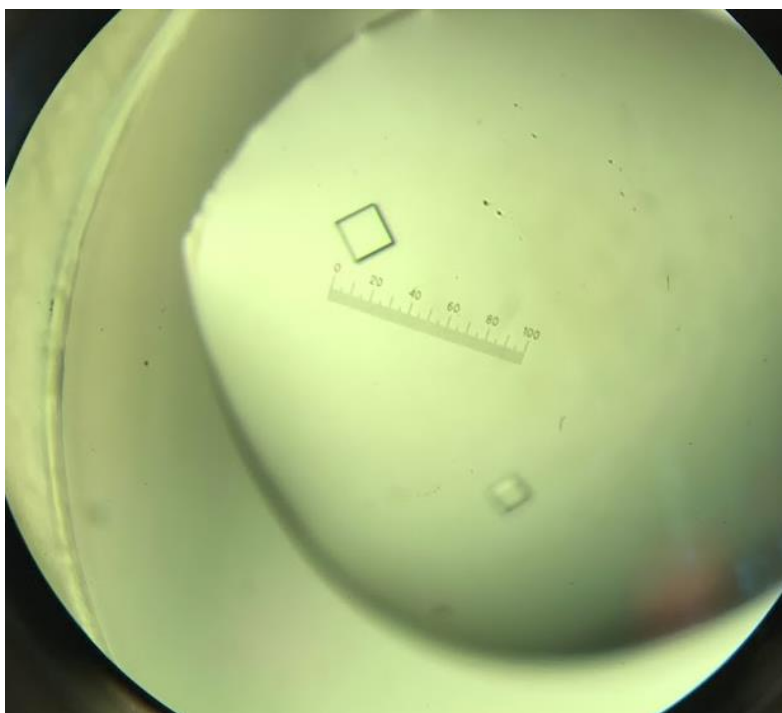
Figure 5.2 Diffraction quality PfTA R215E crystals formed within 12 days.

The quality of the R215E map was markedly more consistent than that of the WT map, with the backbone almost wholly continuous at a 1.5σ contour in the $2F_o$-F_c map, for the whole of both chains except the N-termini. Accordingly, both chains were modelled in full apart from the N-termini, each of which was missing 4 or 5 residues of the original sequence, and the N-terminal His-tag. As expected, the F_o-F_c map supported the modification of the model to reflect the R215E mutation.

The F_o-F_c map clearly indicated the presence of PLP in both active sites (Figure 5.3), with continuous density suggesting a covalent linkage to the catalytic Lysine (K293). Initially, the linkage was modelled by treating Lys and PLP as individual components, by modifying the PLP and generating a link using AceDRG, according to current best practice advice[88], which has the advantage of preserving the identities of Lys and PLP. However, it was found that modelling Lys-PLP as a single composite component (PDB dictionary identifier: LLP) yielded better geometry around the covalent linkage and better R_{free} upon refinement. Ramachandran validation indicated outlaying φ/ψ values for K293 in both chains, but it was not possible to improve these whilst maintaining the density fitting of the modified Lys-PLP composite component.

	PfTA R215E
Data collection	
Space group	$P2_12_12$
Cell dimensions	
a, b, c (Å)	93.75, 115.56, 92.63
Resolution (Å)	2.47
R_{pim}	0.122 (1.084)
$I / \sigma I$	6.07 (0.77)
CC1/2	0.99 (0.326)
Completeness (%)	99.84 (99.83)
Redundancy	6.78 (6.77)
Wilson B-factor (Å^2)	46.08
No. of unique reflections	36768 (3609)
Refinement	
Resolution range (Å)	34.14 – 2.47
R_{work} / R_{free}	0.2082 / 0.2473
No. atoms	
Protein	6868
PLP	30
Water	116
B-factors (Å^2)	
Protein	51.67
PLP	40.07
Water	49.23
RMS deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.70

Table 5-1 Data collection and refinement statistics for PfTA R215E. Statistics for highest resolution shell are in parentheses.

The F_o-F_c map for R215E showed tetrahedral volumes of density located between H42, K44 and H447 which were interpreted as inorganic phosphate ions, since the crystallisation conditions contained phosphate buffer. Interestingly, the WT structure shows similar volumes

in the same locations, but these were interpreted as sulphate ions since the WT crystallisation condition contained lithium sulphate, but did not contain any kind of inorganic phosphate.
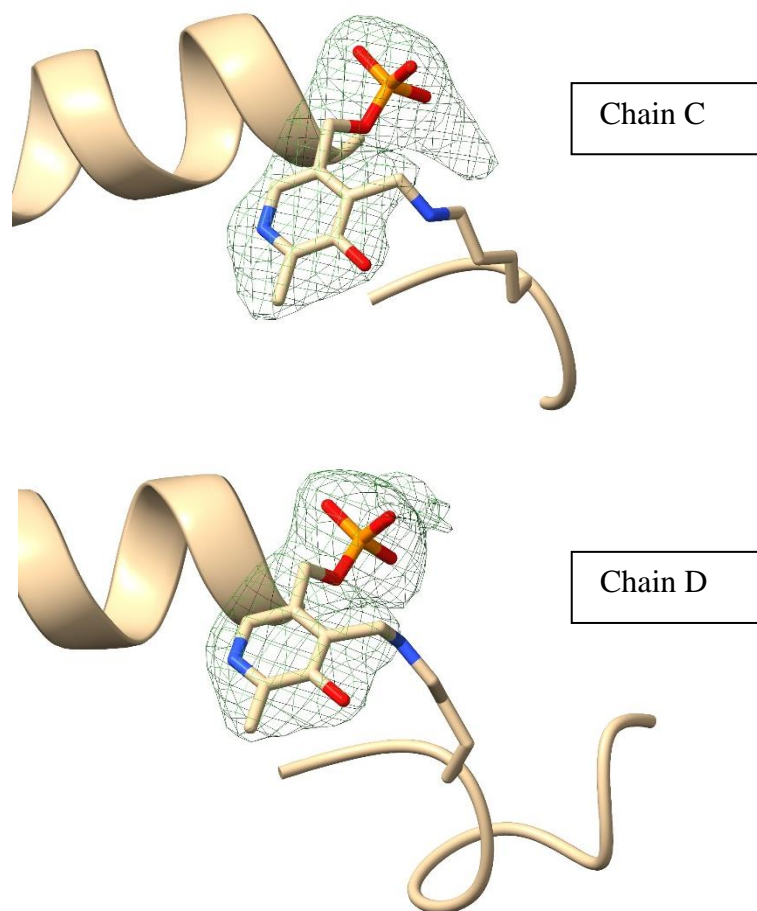
Figure 5.3 Polder maps for both chains (chains C and D) of PfTA R215E, prepared by omission of PLP, contoured at 5σ.

5.4   Crystal structure description

PfTA R215E shows only two chains in its asymmetric unit, presenting a single dimer rather than dimer-dimer pair as seen for PfTA WT (Figure 5.4). Despite this overall change in oligomerisation state, the intra-dimer interface does not show evidence of significant change, and the interfacial area of 5458 Å^2 (calculated by PISA) is not significantly different from the 5412 Å^2 for the equivalent intra-dimer interface in the WT. Similarly, overlaying single chains from WT and R215E shows negligible change in the folding (Figure 5.5).
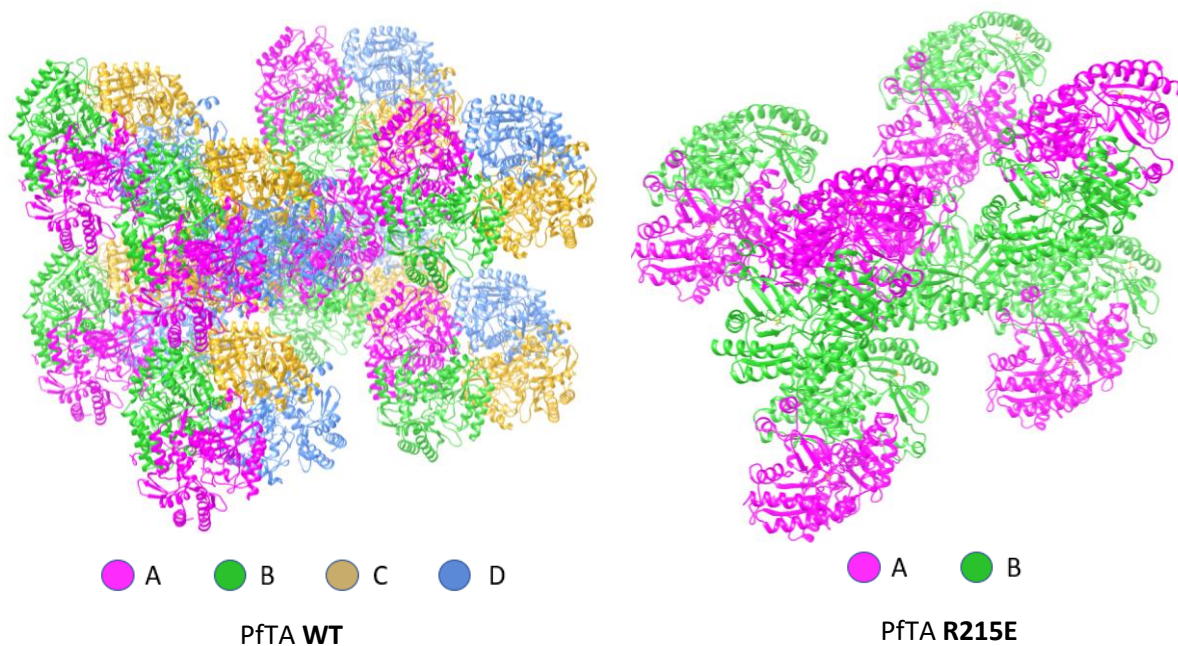
Figure 5.4 Crystal packing in PfTA R215E and WT. The R215E mutant shows only two chains in the asymmetric unit, rather than four, and the inter-dimer interface found in the WT crystal structure is not found in R215E.
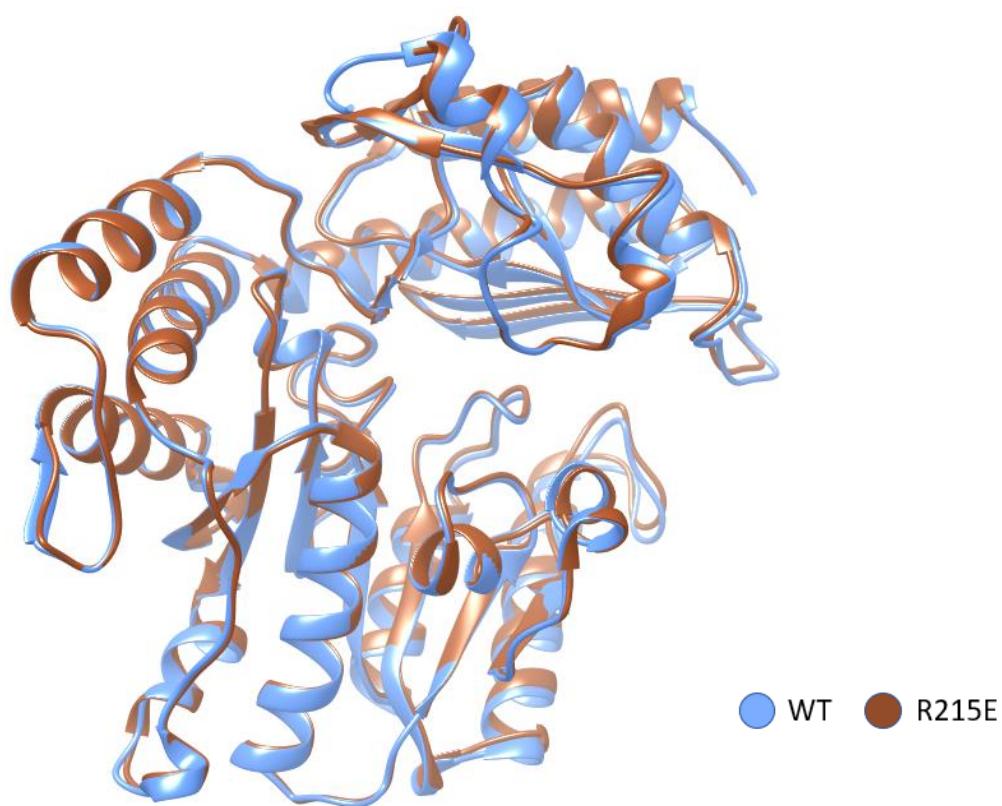
Figure 5.5 Single chains of PfTA WT and R215E superimposed show negligible difference in folding.
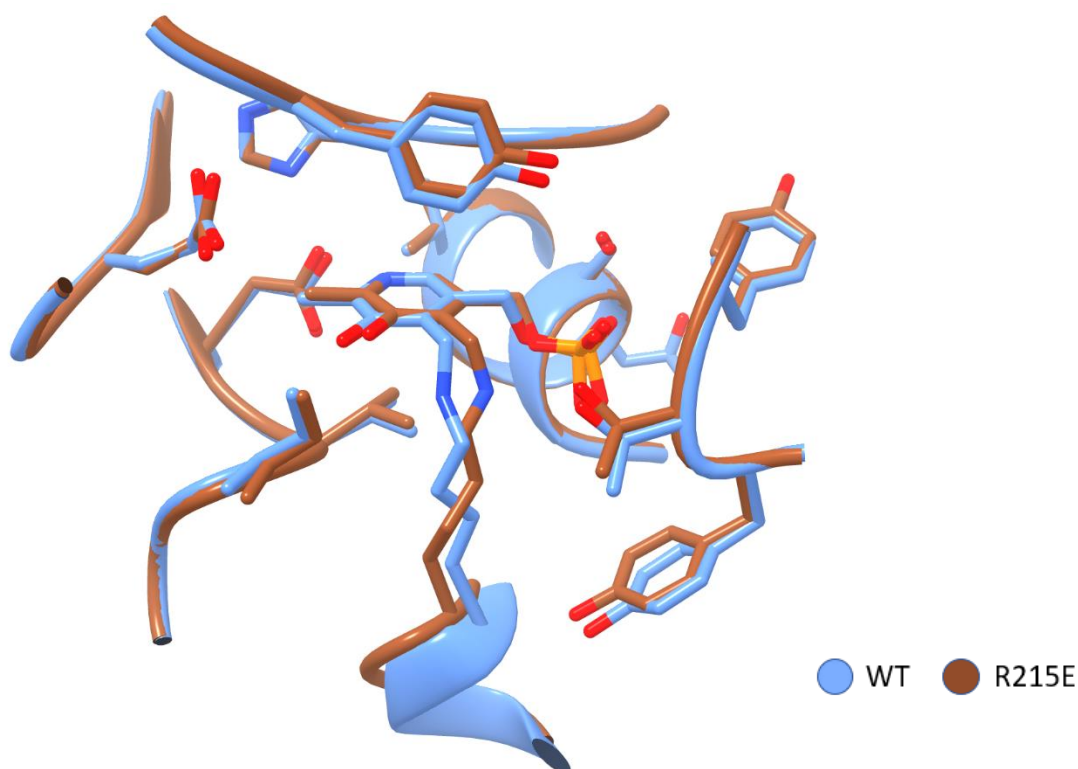
Figure 5.6 Single chains of PfTA WT and R215E superimposed show negligible difference in in conformation of active site residues.

5.5   Stability assays against PfTA WT

The stabilities of PfTA R215E and WT were compared, using differential scanning fluorimetry (DSF) to assess how much the stability of WT/R215E would be affected by the presence of PLP and of amine donors, inferred from the extent to which these would cause a change the melting point (T_m). Melting points were determined under resting conditions (with no substrate), or under working conditions (using either benzylamine or SMBA at 10 or 5 mM)(Figure 5.7). Anticipating that the presence of amine donor substrates would reduce the T_m, as found by Börner *et al.* for PxTA, according to their deduced substrate-driven inactivation mechanism, the working conditions were performed either as the amine donor alone, or with 1 equivalent of sodium pyruvate as the amine acceptor, to assess to what extent the presence of the amine acceptor might mitigate the anticipated destabilising effects of the amine donor. All conditions, both resting and working, were tested with or without addition of 1mM exogenous PLP, in contrast to the work of Börner *et al.* in which all conditions included either 0.1 mM PLP or 1 mM PLP. This was done according to the hypothesis that the stabilising effect of PLP would be most marked under working conditions, for both WT and mutant, but that any

difference in relative stabilising effect (comparing WT with mutant) would be seen most clearly by having a larger difference in PLP concentration, and in consideration that the lack of exogenous PLP does not guarantee a PLP-free or apo-enzyme condition, due to bound PLP carried over through the desalting process[59].

For purposes of control, resting conditions were designed to include 1% or 2% DMSO, because 1% DMSO was required in the working conditions to solubilise SMBA. The results showed no significant difference between using 1% DMSO and 2% DMSO.
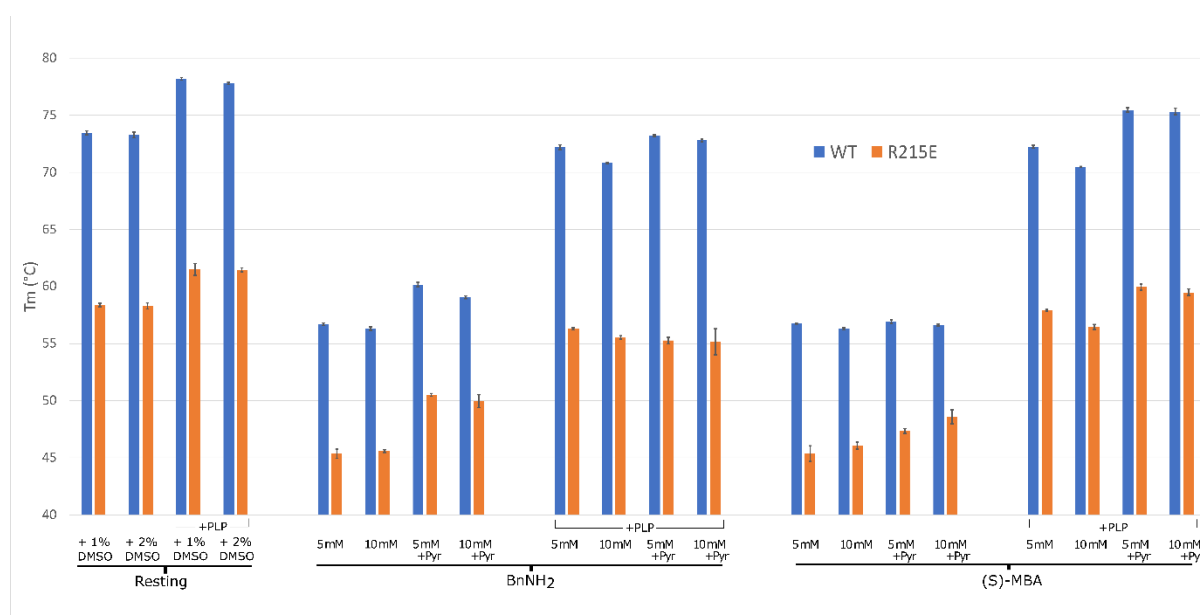
Figure 5.7 Absolute melting temperatures for PfTA WT and R215E under resting (defined here as being without substrate) and working conditions. Working conditions included amine donor substrate with or without equimolar sodium pyruvate, and with or without 1 mM PLP. The series for L-Alanine is omitted here for space constraint.
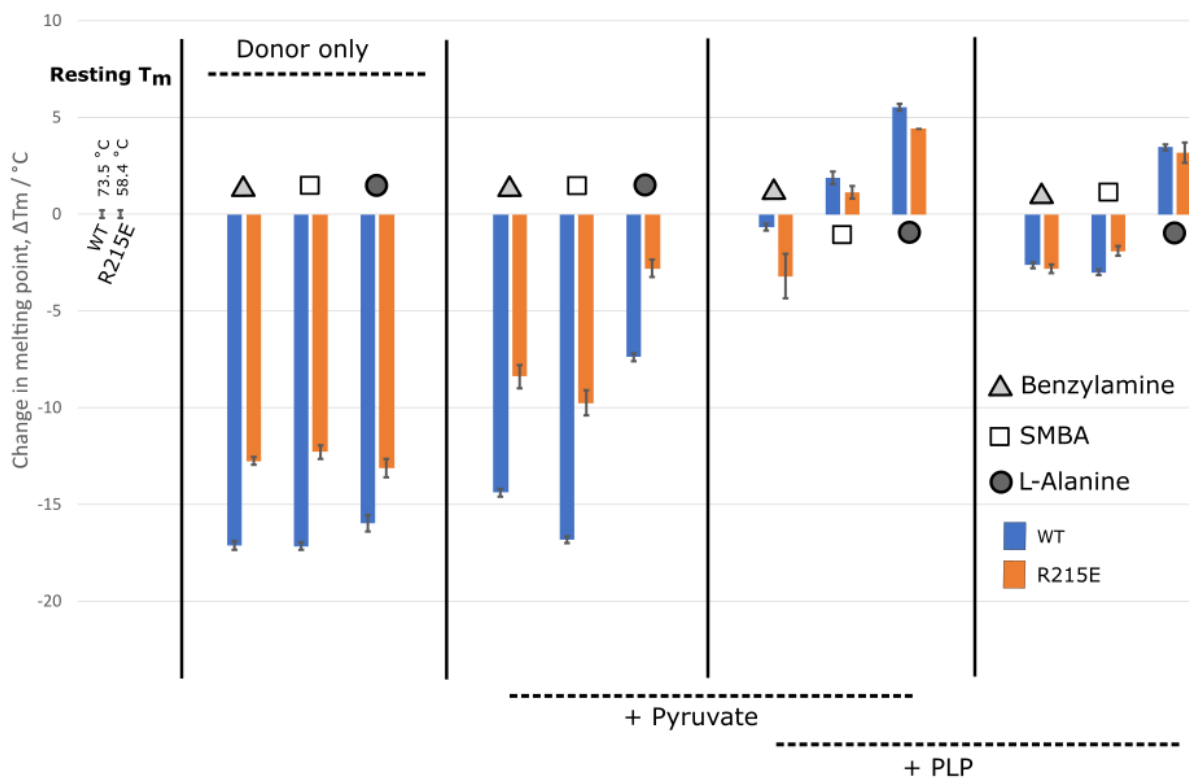
Figure 5.8 Comparing relative ΔT_m for PfTA WT and R215E, as the reduction of thermal stability upon addition of 10 mM donor substrate, and the extent of recovery of thermal stability, with the addition of equimolar acceptor substrate and/or excess PLP. The baseline represents the T_m with neither substrate nor exogenous PLP, defined here as the resting condition. The series for 5 mM donor substrate follows the same pattern as for 10 mM donor substrate, and is omitted for space constraint.

The results showed the R215E mutant to have lower T_m than the WT, for any condition tested. Under the defined resting conditions and without added PLP, the R215E T_m was 15.1 °C lower than WT, whilst with added PLP, it was 16.7 °C lower than WT (Figure 5.7). Based on this alone, and given the solvent-exposed position of R215E, it would seem reasonably likely that oligomerisation itself does play a non-trivial role in bolstering the stability of the protein, since it would seem relatively unlikely that packing, solvation or entropy effects would be the sole or dominant factor arising in this position and with this specific Arg-Glu mutation, and leading to a T_m change of this degree.

Unsurprisingly, it was observed that the introduction of donor substrate always reduced T_m (Figure 5.8). It was found that the identity of the donor (benzylamine, SMBA, or L-Alanine) made no significant difference, when allowing for error; consistently, the reduction for WT was 17 °C, and for R215E 13 °C, to the nearest whole degree. The lesser degree of reduction for R215E, may be seen in a context in which the resting T_m was already much lower for R215E than WT.

Another constant applying to both WT and R215E, was that additional PLP was able to restore T_m (or even take the T_m slightly above its original value, in the example of L-Alanine donor substrate). Whilst, interestingly, this effect did vary according to the identity of the donor substrate, nevertheless the effect applied equally to WT and R215E.

However, a difference between WT and R215E was observable under the condition of having donor and acceptor substrates, but with no PLP. Under this condition, R215E recovered a greater amount of thermal stability, as a proportion of the thermal stability it had lost due to donor substrate alone. This applied for all three donors, though not to a consistent degree (Table 5-2).

Donor (10 mM)	WT			R215E		
	Stability lost to donor alone ΔT_m /°C	Stability regained with equimolar pyruvate $\Delta\Delta T_m$ /°C	Regain as percentage, $\Delta\Delta T_m/\Delta T_m$	Stability lost to donor alone ΔT_m /°C	Stability regained with equimolar pyruvate $\Delta\Delta T_m$ /°C	Regain as percentage, $\Delta\Delta T_m/\Delta T_m$ %
Benzylamine	-17.1 ± 0.2	+2.7 ± 0.3	**16±2 %**	-12.8 ± 0.2	+4.4 ± 0.6	**34±5 %**
SMBA	-17.2 ± 0.2	+0.3 ± 0.3	**2±2 %**	-12.3 ± 0.3	+2.5 ± 0.7	**20±6 %**
L-Alanine	-16.0 ± 0.4	+8.6 ± 0.5	**54±3 %**	-13.1 ± 0.5	+10.3 ± 0.7	**79±5 %**

Table 5-2 Thermal stability loss due to donor substrate, and percentage regain with acceptor substrate, for WT and R215E

Taking these observations together, it may be tentatively suggested that R215E is more sensitive to the effects of donor/acceptor substrate driving the forward/reverse half-reactions, according to the associated PLP-PMP interconversion, but that this effect becomes less visible in the presence of added excess PLP.

To compare the rates of inactivation of PfTA WT and R215E under reaction-like conditions, incubation solutions were used to expose them to varying temperature and substrate concentrations.

Depending on the exact stock defrosted, PfTA WT had a specific activity from 25~33 µmol/min/mg, and PfTA R215E a specific activity of 20~25 µmol/min/mg (specific activity was defined as µmol of acetophenone produced per minute per milligram of enzyme in the activity assay).

In agreement with DSF data, R215E showed markedly lower stability than WT at higher temperatures (Figure 5.10). In PfTA WT, incubation for 2 hours at temperatures up to 60°C

showed no evidence of reducing enzyme activity, regardless PLP concentration (0.1 mM or 1 mM PLP), and for the temperatures 40°C and 50°C, residual activity was even higher than for the control sample. This accords with the observation of Roura *et al.*[59] , in which PfTA WT was found to become marginally more active after incubation at moderate temperatures, possibly because these temperatures might lead to formation of Lys-PLP linkages in a greater proportion of active sites. There was no significant evidence for PfTA WT that the higher PLP concentration lead to a greater residual activity, except at the highest tested temperature of 60°C, whereas R215E clearly showed lower residual activity in the absence of PLP, for all temperatures. However, it should be noted that the data series for R215E with 0.1 mM PLP, was affected by a control problem, in that the control kept on ice lost ~25% activity over the total duration (around 5 hours) of the experiment (Figure 5.9 A). Since the staggered incubation periods started for the 60 ° sample first, and the 30 °C sample last, this means that the assay is likely to have slightly underestimated the true 2-hour residual activity of those samples run last, i.e. the lower temperature samples which were waiting longer in the ice box to be processed.
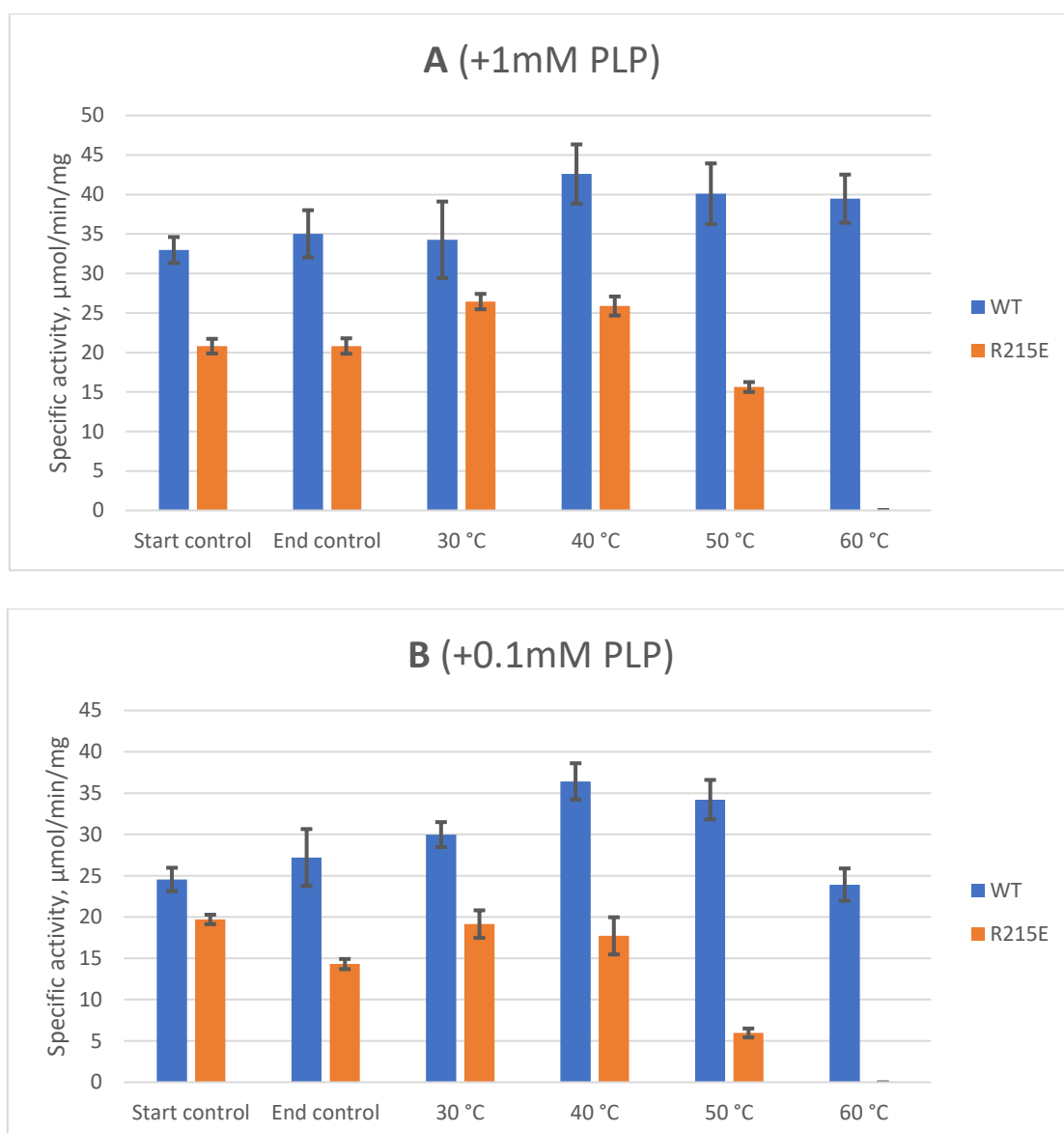
Figure 5.9 Residual specific activities for PfTA WT and R215E (**A**) with 1mM PLP or (**B**) 0.1mM PLP, incubated for 2 hours at the indicated temperatures, in the absence of any substrate.
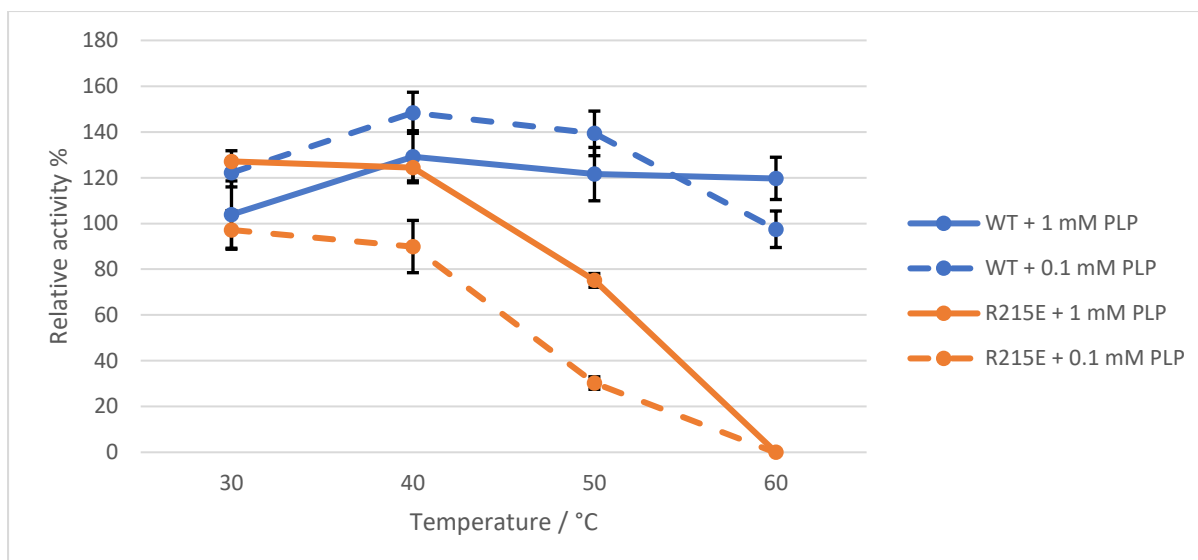
Figure 5.10 Residual activity after 2 hours incubation at different temperatures for PfTA WT and R215E, with or without exogenous PLP. Values are normalised to activity of freshly defrosted sample not subjected to incubation. Control for loss of activity of aliquots whilst stored on ice revealed no loss of activity for any sample during ice storage up to 4.5 hours, except in the case of R215E stored without PLP, which lost 25-30% activity in 4.5 hours.

R215E also showed markedly more sensitivity to higher amine donor loading than PfTA WT, when incubated with 10 mM sodium pyruvate and varying concentrations of iPrNH$_2$ from 2 mM to 1 M (Figure 5.11). Normalised to resting conditions, a low concentration of SMBA and pyruvate (2 and 10 mM respectively) had a mildly activating effect on R215E, whilst for 1 M SMBA there was no detectable activity, whereas the inactivation observed for WT did not differ greatly across the different donor/acceptor ratios tested, with residual activity always around 75-85%.

This provides complementary information to the DSF findings, in which the amine donor concentration was no greater than 10 mM, in suggesting that R215E might indeed be more sensitive to donor substrate concentration, whereas DSF provided only a modest suggestion that acceptor substrate might help "rescue" the stability of R215E more than the stability of WT. However, it should also be considered that the T_m and the inactivation rate are dissimilar properties, in that the former is thermodynamic and melting is reversible, whereas the latter is kinetic, inactivation being essentially irreversible, and the samples with the least residual activity showing significant precipitation.
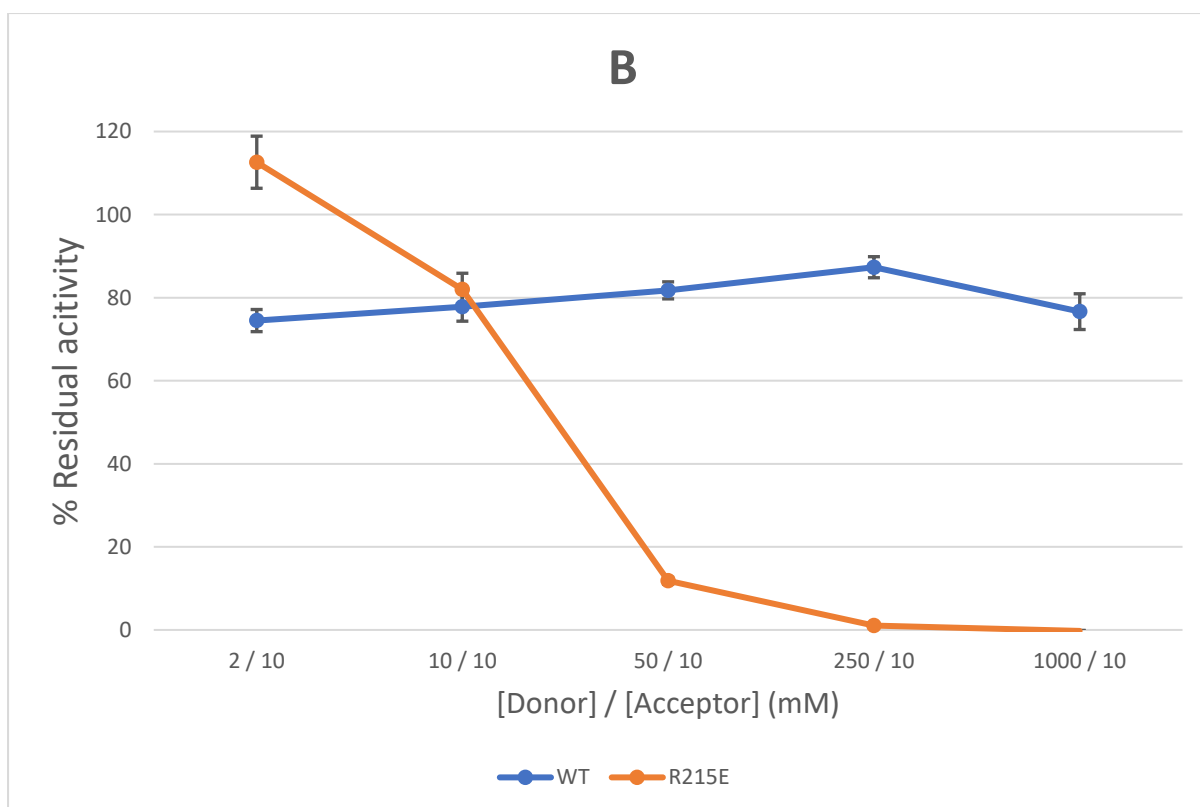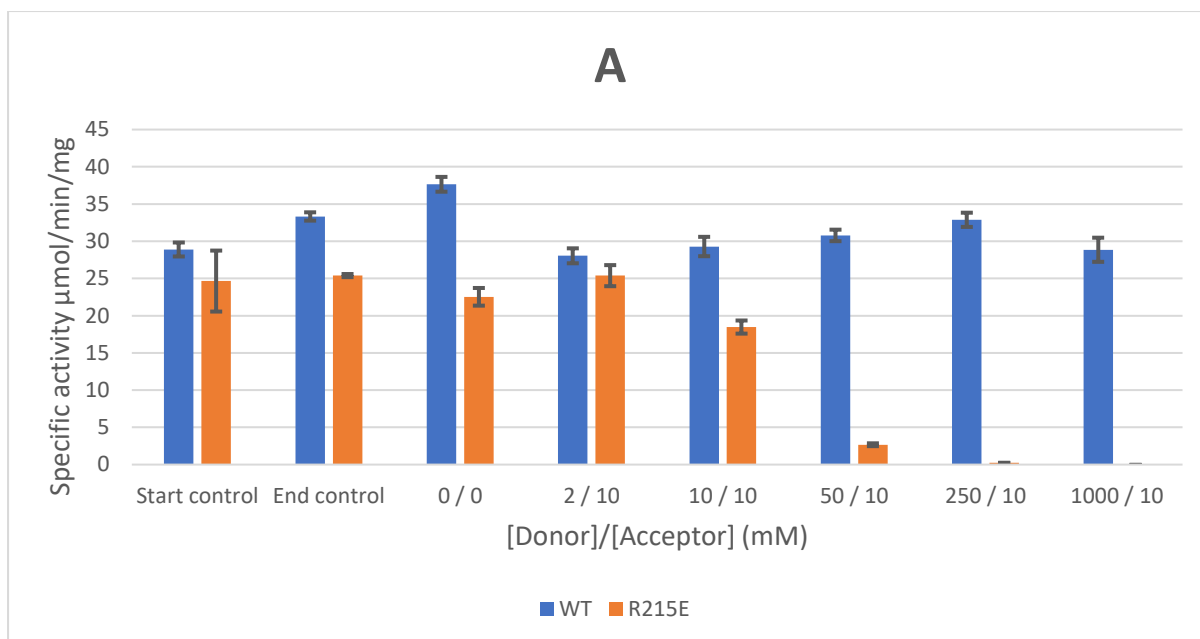
Figure 5.11Residual activities for PfTA WT and R215E incubated 2 hours with variable ratio of amine donor/acceptor concentrations (**A**) as specific activities, or (**B**) normalised to a sample having no substrate.

5.6   Chapter conclusion

AUC evidence for PfTA R215E showed a minimal mass (for $f/f_0 = 0$) greater than that of a monomer, and a likely mass (for $f/f_0 = 1.2$) significantly lower than that of the tetrameric WT,

and in agreement with a dimeric state (observed 90 kDa, expected 105 kDa). Like PfTA WT, R215E also showed no evidence of non-ideal behaviour over the concentrations analysed.

A crystal structure for R215E was obtained, showing two chains in the asymmetric unit, arranged as a dimer. Comparison with the WT crystal structure revealed that despite the changed oligomerisation state, that the intra-dimer interface, the presence of PLP in the active site, the active site residue conformations, and the overall folding all remained unchanged. Specific activity for R215E was comparable to, or only marginally lower than, the WT; especially when taking into account the variability seen amongst different batches of each.

R215E had lower T_m under all tested conditions, with a 15 °C difference under conditions with no substrate and no added PLP. It would be difficult to directly apportion causality for this destabilisation, between oligomerisation-related effects and other effects, albeit that the solvent-exposed position of this single residue substitution may suggest the former to seem more likely. However, DSF also suggested that acceptor substrate may help to "rescue" the donor-induced destabilisation more in R215E than the WT.

The greater loss of activity seen for R215E under high donor substrate loading may be still more informative, since it relates directly to the effect of amine-driven PMP accumulation. However, whilst the pH was controlled for in the substrate loading experiment, the ionic strength was not (since higher concentrations of $iPrNH_2$ are present mainly as the chloride salt under the pH 8.0 buffered conditions), and therefore it cannot be certainly asserted that the faster precipitation of R215E (in comparison to WT) at higher substrate loadings was not due to salting out effects.

6   General discussion, conclusion, and future work

Exploring the effects of higher oligomerisation states in TAs potentially offers a novel approach to engineer their performance as biocatalysts, and whilst existing works have engineered the intra-dimer interface this has not been extended to inter-dimer interfaces.

The PfTA chosen as candidate enzyme herein proved useful, yielding a novel crystal structure. In combination with biophysical evidence, it was possible to determine it forms a tetrameric structure specifically of the same arrangement of chains as found in its homolog PxTA for which Börner *et al.* first found a possible mechanism of improved PLP/PMP retention, and specifically that its oligomerisation is very unlikely to be concentration or PLP-dependant as it is for some homologs.

Mutation of tetrameric PfTA WT to yield a dimer showed a low success rate, with most mutants expressing poorly, and attempts to predict the effect of individual mutations using convenient molecular dynamics analysis tools also proved unreliable. Future attempts would likely benefit from more sophisticated recently emergent methods[86].

Nevertheless, the obtained R215E mutant was amenable to purification, with biophysical evidence demonstrating a dimeric state even in the presence of excess PLP, and yielding a crystal structure revealing no change in the folding, active site environment, or presence of PLP in the active site.

R215E showed a lower thermal stability than PfTA WT in all contexts. Its demonstrated a greater sensitivity to amine donor concentration in residual activity assay, and somewhat higher sensitivity to the "rescue" effect of acceptor substrate in the DSF assay, which could be tentatively interpreted in the context of the inactivation scheme of Börner *et al.*, to mean that the R215E mutant is more sensitive to substrate driven inactivation. However, the robustness of this finding would be complemented by determining the effects of ionic strength/salt concentration on inactivation rate, and controlling for this in the inactivation assay.

These results for WT and R215E could be further complemented by exploring in more depth the enzyme behaviour towards PLP/PMP binding, especially in terms of a PMP release assay to accompany inactivation assay data. Titration data to quantify internal aldimine occupancy would be helpful to complement the occupancy inferred from crystal data. Exploring PLP inhibition effects for WT and R215E would also enable more detailed comparison, using the same Cornish-Bowden analysis as applied to PxTA.

In focusing on enzyme-PLP relationship more deeply, MD approaches could also be applied, following a similar approach to the simulations of Ruggieri *et al*[15] on the dimer dissociation process. Modifying this to explore whether a significant energy barrier exists between the active site secondary structure conformations belonging to the apo- state, the aldehydic PLP state, and the internal aldimine state, and whether this would differ between dimer and tetramer, would be especially helpful. MD has the advantage that PfTA WT can be simulated as a dimer without the need to introduce any mutation.

Moving beyond PfTA, it would be advantageous to be able to generalise these findings to other tetrameric TAs. For example, the tetrameric *Bacillus megaterium* TA (BmTA) used by Oosterwijk *et al*[89] shows an unusual extended loop in the RBM, suggesting different dynamic behaviour of the RBM may be involved. It would be informative to prepare a dimeric mutant for this, and other TAs, just as done in this work for PfTA, to investigate how for the findings would be supported and generalised. Generation of both single and multiple mutants may help to refine the extent to which non-oligomerisation based effects, attached to specific individual mutations, can be discerned from generalisable oligomerisation-based effects In preparing such mutants, it may be possible to take advantage of cryo-EM as a method which would allow simultaneous structure and oligomerisation state determination, without need for crystallisation. Whereas previously reserved for larger proteins and complexes, recent improvement in achievable contrast has made structures as small as 65 kDa tractable[90].

A very interesting possibility for modifying oligomerisation state upwards rather than downwards, would be the generation of TA chimeras. For example, the overall tertiary structure of VfTA is extremely similar to PfTA, despite the former being dimeric and the latter tetrameric. Therefore, if dimer-dimer interface formation is considered in terms of a shape complementarity problem, then it may be possible to make VfTA tetrameric by simply taking the ~50 residue section of sequence from PfTA which covers all dimer-dimer interface-forming residues, and substituting this into VfTA. The chance of success of such a technique is difficult to estimate, but if successful could be very rewarding. Applying an interface stabilisation technique such as that employed by Meng *et al* [91] , to stabilise a dimer-dimer interface in a naturally tetrameric TA, may represent a more reliable option.

Ultimately, in reference to biocatalytic applications, it would be helpful to explore the optimal catalytic temperature for tetrameric transaminases (as a balance of activity and stability, to achieve maximum substrate conversion). Because of the existing, though limited, evidence to

suggest that higher temperatures may promote greater occupancy of the internal aldimine state, conversely therefore at lower temperatures the activity per mass of enzyme for tetrameric transaminases may be inferior.

7 References

1.   Alonso, D. & Mondragón, A. Mechanisms of catalytic RNA molecules. *Biochemical Society Transactions* vol. 49 1529–1535 (2021).

2.   Sheldon, R. A. & Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chemical Reviews* vol. 118 801–838 (2018).

3.   France, S. P., Lewis, R. D. & Martinez, C. A. The Evolving Nature of Biocatalysis in Pharmaceutical Research and Development. *JACS Au* vol. 3 715–735 (2023).

4.   Savile, C. K. *et al.* Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **329**, 303–305 (2010).

5.   Jansen, R. S. *et al.* Aspartate aminotransferase Rv3722c governs aspartate-dependent nitrogen metabolism in Mycobacterium tuberculosis. *Nat. Commun.* **11**, (2020).

6.   Kelly, S. A., Mix, S., Moody, T. S. & Gilmore, B. F. Transaminases for industrial biocatalysis: novel enzyme discovery. *Appl. Microbiol. Biotechnol.* **104**, 4781–4794 (2020).

7.   Richts, B., Rosenberg, J. & Commichau, F. M. A survey of pyridoxal 5'-phosphate-dependent proteins in the gram-positive model bacterium Bacillus subtilis. *Front. Mol. Biosci.* **6**, (2019).

8.   Dunathan, H. Conformation and reaction specificity in pyridoxal phosphate enzymes. *PNAS* **55**, 712 (1966).

9.   Steffen-Munsberg, F. *et al.* Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. *Biotechnol. Adv.* **33**, 566–604 (2015).

10.   Coscolín, C. *et al.* Bioprospecting reveals class ω-transaminases converting bulky ketones and environmentally relevant polyamines. *Appl. Environ. Microbiol.* **85**, (2019).

11.   Kelly, S. A. *et al.* Application of $ω$-Transaminases in the Pharmaceutical Industry. *Chemical Reviews* vol. 118 349–367 (2018).

12.   Patil, M. D., Grogan, G., Bommarius, A. & Yun, H. Oxidoreductase-Catalyzed Synthesis of Chiral Amines. *ACS Catal.* **8**, 10985–11015 (2018).

13.   Guo, F. & Berglund, P. Transaminase biocatalysis: optimization and application. *Green*

Chem. **19**, 333 (2017).

14. Lee, J., Gokey, T., Ting, D., He, Z. H. & Guliaev, A. B. Dimerization misalignment in human glutamate-oxaloacetate transaminase variants is the primary factor for PLP release. *PLoS One* **13**, (2018).

15. Ruggieri, F. *et al.* Insight into the dimer dissociation process of the Chromobacterium violaceum (S)-selective amine transaminase. *Sci. Rep.* **9**, (2019).

16. Humble, M. S. *et al.* Crystal structures of the Chromobacterium violaceumω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. *FEBS J.* **279**, 779–792 (2012).

17. Börner, T. *et al.* Explaining Operational Instability of Amine Transaminases: Substrate-Induced Inactivation Mechanism and Influence of Quaternary Structure on Enzyme−Cofactor Intermediate Stability. *ACS Catal.* **7**, 1259–1269 (2017).

18. Ali, M. H. & Imperiali, B. Protein oligomerization: How and why. *Bioorganic and Medicinal Chemistry* vol. 13 5013–5020 (2005).

19. Cai, K., Schirch, D. & Schirch, V. The affinity of pyridoxal 5′-phosphate for folding intermediates of Escherichia coli serine hydroxymethyltransferase. *J. Biol. Chem.* **270**, 19294–19299 (1995).

20. Malerba, F., Bellelli, A., Giorgi, A., Bossa, F. & Contestabile, R. The mechanism of addition of pyridoxal 5′-phosphate to Escherichia coli apo-serine hydroxymethyltransferase. *Biochem. J.* **404**, 477–485 (2007).

21. Stránská, J., Tylichová, M., Kopečný, D., Snégaroff, J. & Šebela, M. Biochemical characterization of pea ornithine-δ-aminotransferase: Substrate specificity and inhibition by di- and polyamines. *Biochimie* **92**, 940–948 (2010).

22. Li, M. *et al.* Biochemical and structural investigation of taurine:2-oxoglutarate aminotransferase from Bifidobacterium kashiwanohense. *Biochem. J.* **476**, 1605–1619 (2019).

23. Bertoldi, M., Cellini, B., Laurents, D. V & Borri Voltattorni, C. *Folding pathway of the pyridoxal 5-phosphate C-S lyase MalY from Escherichia coli. Biochem. J* vol. 389 (2005).

24. Novick, S. J. *et al.* Engineering an amine transaminase for the efficient production of a chiral sacubitril precursor. *ACS Catal.* **11**, 3762–3770 (2021).

25. Yu, H. & Huang, H. Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnology Advances* vol. 32 308–315 (2014).

26. Huang, J., Xie, D. F. & Feng, Y. Engineering thermostable (R)-selective amine transaminase from Aspergillus terreus through in silico design employing B-factor and folding free energy calculations. *Biochem. Biophys. Res. Commun.* **483**, 397–402 (2017).

27. Steipe, B., Schiller, B., Pluckthun, A. & Steinbecker, S. Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain. *J. Mol. Biol.* **240**, 188–192 (1994).

28. Xie, D. F. *et al.* Construction of stabilized (R)-selective amine transaminase from Aspergillus terreus by consensus mutagenesis. *J. Biotechnol.* **293**, 8–16 (2019).

29. Lugtenburg, T., Gran-Scheuch, A. & Drienovská, I. Non-canonical amino acids as a tool for the thermal stabilization of enzymes. *Protein Engineering, Design and Selection* vol. 36 (2023).

30. Deepankumar, K. *et al.* Engineering transaminase for stability enhancement and site-specific immobilization through multiple noncanonical amino acids incorporation. *ChemCatChem* **7**, 417–421 (2015).

31. Politzer, P. & Murray, J. S. Halogen bonding: An interim discussion. *ChemPhysChem* vol. 14 278–294 (2013).

32. Börner, T. *et al.* Three in One: Temperature, Solvent and Catalytic Stability by Engineering the Cofactor-Binding Element of Amine Transaminase. *ChemBioChem* **18**, 1482–1486 (2017).

33. Chen, H. J., Ko, T. P., Lee, C. Y., Wang, N. C. & Wang, A. H. J. Structure, Assembly, and Mechanism of a PLP-Dependent Dodecameric l-Aspartate β-Decarboxylase. *Structure* **17**, 517–529 (2009).

34. Meng, Q. *et al.* Robust ω-Transaminases by Computational Stabilization of the Subunit Interface. *ACS Catal.* **10**, 2915–2928 (2020).

35. Elez, K., Bonvin, A. M. J. J. & Vangone, A. Biological vs. Crystallographic protein

interfaces: An overview of computational approaches for their classification. *Crystals* **10**, (2020).

36. Ponstingl, H., Kabir, T. & Thornton, J. M. Applied Crystallography Automatic inference of protein quaternary structure from crystals. *1116±1122 J.* **36**, (2003).

37. Marchenkova, M. *et al.* Dodecamers derived from the crystal structure were found in the pre-crystallization solution of the transaminase from the thermophilic bacterium Thermobaculum terrenum by small-angle X-ray scattering. *J. Biomol. Struct. Dyn.* **38**, (2020).

38. Montioli, R., Zamparelli, C., Borri Voltattorni, C. & Cellini, B. Oligomeric State and Thermal Stability of Apo- and Holo- Human Ornithine δ-Aminotransferase. *Protein J.* **36**, 174–185 (2017).

39. Hong, P., Koza, S. & Bouvier, E. S. P. A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *J. Liq. Chromatogr. Relat. Technol.* **35**, 2923–2950 (2012).

40. Lorber, B., Fischer, F., Bailly, M., Roy, H. & Kern, D. Protein analysis by dynamic light scattering: Methods and techniques for students. *Biochem. Mol. Biol. Educ.* **40**, 372–382 (2012).

41. Liu, W. *et al.* Crystal structures of unbound and aminooxyacetate-bound Escherichia coli γ-aminobutyrate aminotransferase. *Biochemistry* **43**, 10896–10905 (2004).

42. Sayer, C., Isupov, M. N., Westlake, A. & Littlechild, J. A. Structural studies of Pseudomonas and Chromobacterium ω- aminotransferases provide insights into their differing substrate specificity. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 564–576 (2013).

43. Yonaha, K., Toyama, S. & Kagamiyamagy, H. Properties of the Bound Coenzyme and Subunit Structure of ω-Amino Acid:Pyruvate Aminotransferase. *J. Biol. Chem.* **258**, 2260–2265 (1983).

44. Burnett, G., Yonaha, K., Toyama, S., Soda, K. & Walsh, C. Studies on the Kinetics and Stoichiometry of Inactivation of Pseudomonas w-Amino Acid:Pyruvate Transaminase by Gabaculine. *J. Biol. Chem.* **255**, 428–432 (1980).

45. Roura Padrosa, D. *et al.* Enhancing PLP-Binding Capacity of Class-III ω-

Transaminase by Single Residue Substitution. *Front. Bioeng. Biotechnol.* **7**, (2019).

46. Parra, M., Stahl, S. & Hellmann, H. Vitamin B6 and Its Role in Cell Metabolism and Physiology. *Cells* **7**, 84 (2018).

47. Ito, N., Kawano, S., Hasegawa, J. & Yasohara, Y. Purification and characterization of a novel (S)-enantioselective transaminase from Pseudomonas fluorescens KNK08-18 for the synthesis of optically active amines. *Biosci. Biotechnol. Biochem.* **75**, 2093–2098 (2011).

48. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).

49. Demeler, B. *UltraScan A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments: Modern Analytical UltracentrifugationModern Analytical Ultracentrifugation: Techniques and Methods.* (RSC, 2005).

50. Schuck, P. *Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling.* (2000).

51. Cole, J. L., Lary, J. W., P. Moody, T. & Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods in Cell Biology* vol. 84 143–179 (2008).

52. Smith, C. A. Estimation of Sedimentation Coefficients and Frictional Ratios of Globular Proteins. *Biochem. Educ.* **16**, 104–106 (1988).

53. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424--W429 (2016).

54. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

56. Winn, M. D. *et al.* Overview of the *CCP* 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).

57. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, (2012).

58. Grøftehauge, M. K., Hajizadeh, N. R., Swann, M. J. & Pohl, E. Protein-ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 36–44 (2015).

59. Roura Padrosa, D. *et al.* Enhancing PLP-Binding Capacity of Class-III ω-Transaminase by Single Residue Substitution. *Front. Bioeng. Biotechnol.* **7**, (2019).

60. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).

61. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

62. Richard, J. P., Amyes, T. L., Crugeiras, J. & Rios, A. The PLP cofactor: Lessons from studies on model reactions. *Biochimica et Biophysica Acta - Proteins and Proteomics* vol. 1814 1419–1425 (2011).

63. Walden, M. *et al.* Metabolic control of BRISC–SHMT2 assembly regulates immune signalling. *Nature* **570**, 194–199 (2019).

64. Giardina, G. *et al.* How pyridoxal 5′-phosphate differentially regulates human cytosolic and mitochondrial serine hydroxymethyltransferase oligomeric state. *FEBS J.* **282**, 1225–1241 (2015).

65. Chen, S. *Stability and inactivation mechanisms of two transaminases*. (2017).

66. Edwards, G. B., Muthurajan, U. M., Bowerman, S. & Luger, K. Analytical Ultracentrifugation (AUC): An Overview of the Application of Fluorescence and Absorbance AUC to the Study of Biological Macromolecules. *Curr. Protoc. Mol. Biol.* **133**, (2020).

67. Mutti, F. G. *et al.* Amination of ketones by employing two new (S)-selective ω-transaminases and the his-tagged ω-TA from Vibrio fluvialis. *European J. Org. Chem.* 1003–1007 (2012) doi:10.1002/ejoc.201101476.

68. Yonaha, K., Toyama, S., Yasuda, M. & Soda, K. Purification and crystallization of

bacterial ω-amino acid-pyruvate aminotransferase. *FEBS Lett.* **71**, 21–24 (1976).

69. Yonaha, K., Toyama, S., Yasuda, M. & Soda, K. Properties of crystalline ω-amino acid: pyruvate aminotransferase of Pseudomonas sp. F-126. *Agric. Biol. Chem.* **41**, 1701–1706 (1977).

70. Watanabe, N. *et al.* Crystal Structure Analysis of $ω$-Amino Acid: Pyruvate Aminotransferase with a Newly Developed Weissenberg Camera and an Imaging Plate Using Synchrotron Radiation. *J. Biochem* **105**, 3 (1989).

71. Lamartiniere, C. A., Itoh, H. & Dempsey, W. B. Acid Transaminase from Pseudomonas cepacia. Purification, Crystallization, Physical, and Kinetic Properties. *Biochemistry* **10**, 4783–4788 (1971).

72. Toney, M. D., Hohenester, E., Cowan, S. W. & Jansonius, J. N. Dialkylglycine decarboxylsase structure: bifunctional active site and alkali metal sites. *Science* **261**, 756 (1993).

73. Toney, M. D., Hohenester, E., Keller, J. W. & Jansonius, J. N. Structural and Mechanistic Analysis of Two Refined Crystal Structures of the Pyridoxal Phosphate-dependent Enzyme Dialkylglycine Decarboxylase. *J. Mol. Biol* **245**, 151–179 (1995).

74. Mutaguchi, Y., Ohmori, T., Wakamatsu, T., Doi, K. & Ohshima, T. Identification, purification, and characterization of a novel amino acid racemase, isoleucine 2-epimerase, from Lactobacillus Species. *J. Bacteriol.* **195**, 5207–5215 (2013).

75. Hayashi, J. *et al.* Crystal structure of the novel amino-acid racemase isoleucine 2-epimerase from Lactobacillus buchneri. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 428–437 (2017).

76. Van Oosterwijk, N. *et al.* Structural Basis of the Substrate Range and Enantioselectivity of Two (S)-Selective ω-Transaminases. *Biochemistry* **55**, 4422–4431 (2016).

77. Liu, W. *et al.* Kinetic and crystallographic analysis of active site mutants of Escherichia coli $γ$-aminobutyrate aminotransferase. *Biochemistry* **44**, 2982–2992 (2005).

78. Hillier, H. T., Altermark, B. & Leiros, I. The crystal structure of the tetrameric DABA-aminotransferase EctB, a rate-limiting enzyme in the ectoine biosynthesis pathway. *FEBS J.* **287**, 4641–4658 (2020).

79. Sakuraba, H., Kawakami, R., Takahashi, H. & Ohshima, T. Novel archaeal alanine:glyoxylate aminotransferase from Thermococcus litoralis. *J. Bacteriol.* **186**, 5513–5518 (2004).

80. Sakuraba, H. *et al.* Structure of an archaeal alanine:glyoxylate aminotransferase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 696–699 (2008).

81. Yoshikane, Y., Yokochi, N., Ohnishi, K., Hayashi, H. & Yagi, T. Molecular cloning, expression and characterization of pyridoxamine-pyruvate aminotransferase. *Biochem. J.* **396**, 499–507 (2006).

82. Yoshikane, Y. *et al.* Crystal structure of pyridoxamine-pyruvate aminotransferase from Mesorhizobium loti MAFF303099. *J. Biol. Chem.* **283**, 1120–1127 (2008).

83. Heckmann, C. M., Gourlay, L. J., Dominguez, B. & Paradisi, F. An (R)-Selective Transaminase From Thermomyces stellatus: Stabilizing the Tetrameric Form. *Front. Bioeng. Biotechnol.* **8**, (2020).

84. Zvelebil, M. J., Barton, G. J., Taylor, W. R. & Sternberg, M. J. E. c. *J. Mol. Biol.* **195**, 957–961 (1987).

85. Yan, C., Wu, F., Jernigan, R. L., Dobbs, D. & Honavar, V. Characterization of protein-protein interfaces. *Protein J.* **27**, 59–70 (2008).

86. Zhang, I. Identifying and overcoming the sampling challenges in relative binding free energy calculations of a model protein:protein complex. doi:10.1101/2023.03.07.530278.

87. Swan, P. Troubleshooting: Protein Expression. *https://goldbio.com/articles/article/troubleshooting-protein-expression*.

88. Nicholls, R. A. *et al.* Modelling covalent linkages in CCP4. *Acta Crystallogr. Sect. D Struct. Biol.* **77**, 712–726 (2021).

89. Van Oosterwijk, N. *et al.* Structural Basis of the Substrate Range and Enantioselectivity of Two (S)-Selective ω-Transaminases. *Biochemistry* **55**, 4422–4431 (2016).

90. Renaud, J. P. *et al.* Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nat. Rev. Drug Discov.* **17**, 471–492 (2018).

91. Meng, Q. *et al.* Robust ω-Transaminases by Computational Stabilization of the Subunit

Interface. *ACS Catal.* **10**, 2915–2928 (2020).

92. Midelfort, K. S. *et al.* Redesigning and characterizing the substrate specificity and activity of Vibrio fluvialis aminotransferase for the synthesis of imagabalin. **26**, 25–33 (2013).

93. Kwon, S. *et al.* Structural basis of substrate recognition by a novel thermostable (S)-enantioselective ω-transaminase from Thermomicrobium roseum. *Sci. Rep.* **9**, (2019).

94. Mathew, S. *et al.* Identification of novel thermostable $ω$-transaminase and its application for enzymatic synthesis of chiral amines at high temperature. *RSC Adv.* **6**, 69257–69260 (2016).

95. Palacio, C. M. *et al.* Biochemical properties of a Pseudomonas aminotransferase involved in caprolactam metabolism. *FEBS J.* **286**, 4086–4102 (2019).

96. Guidi, B. *et al.* Strategic single point mutation yields a solvent- and salt-stable transaminase from Virgibacillus sp. in soluble form. *Sci. Rep.* **8**, (2018).

97. Rausch, C., Lerchner, A., Schiefner, A. & Skerra, A. Crystal structure of the $ω$-aminotransferase from Paracoccus denitrificans and its phylogenetic relationship with other class III amino- transferases that have biotechnological potential. *Proteins Struct. Funct. Bioinforma.* **81**, 774–787 (2013).

98. Shin, J.-S., Yun, ·H, Jang, J.-W., Park, ·I & Kim, B.-G. Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from Vibrio fluvialis JS17. *Appl Microbiol Biotechnol* **61**, 463–471 (2003).

99. Newman, J. *et al.* Determination of the Structure of the Catabolic N-Succinylornithine Transaminase (AstC) from Escherichia coli. *PLoS One* **8**, (2013).

100. Mani Tripathi, S. & Ramachandran, R. Direct Evidence for a Glutamate Switch Necessary for Substrate Recognition: Crystal Structures of Lysine ε-Aminotransferase (Rv3290c) from Mycobacterium tuberculosis H37Rv. *J. Mol. Biol.* **362**, 877–886 (2006).

101. Gao, S. *et al.* PMP-diketopiperazine adducts form at the active site of a PLP dependent enzyme involved in formycin biosynthesis. *Chem. Commun.* **55**, 14502–14505 (2019).

102. Kwon, S. *et al.* Structural insights into the enzyme specificity of a novel ω-transaminase

from the thermophilic bacterium Sphaerobacter thermophilus. *J. Struct. Biol.* **208**, (2019).

103. Wybenga, G. G., Crismaru, C. G., Janssen, D. B. & Dijkstra, B. W. Structural determinants of the β-selectivity of a bacterial. *J. Biol. Chem.* **287**, 28495–28502 (2012).

104. Kim, J. *et al.* Cloning and characterization of a novel β-transaminase from Mesorhizobium sp. strain LUK: A new biocatalyst for the synthesis of enantiomerically pure β-amino acids. *Appl. Environ. Microbiol.* **73**, 1772–1782 (2007).

8 Appendices

8.1 Calibration of analytical SEC column

Calibration plot for Superdex 200 10/300 GL, 2019-09-10

Elution vol. /mL	MW /kDa	logMW	Kav	$V_C \sim 24$ mL
8.5	[void volume]	[aggregates]	-	Dextran blue
9.3	669	2.825426	0.051613	thyroglobulin
10.8	440	2.643453	0.148387	ferritin
13.15	158	2.198657	0.3	aldolase
14.5	75	1.875061	0.387097	conalbumin
15.4	44	1.643453	0.445161	ovalbumin

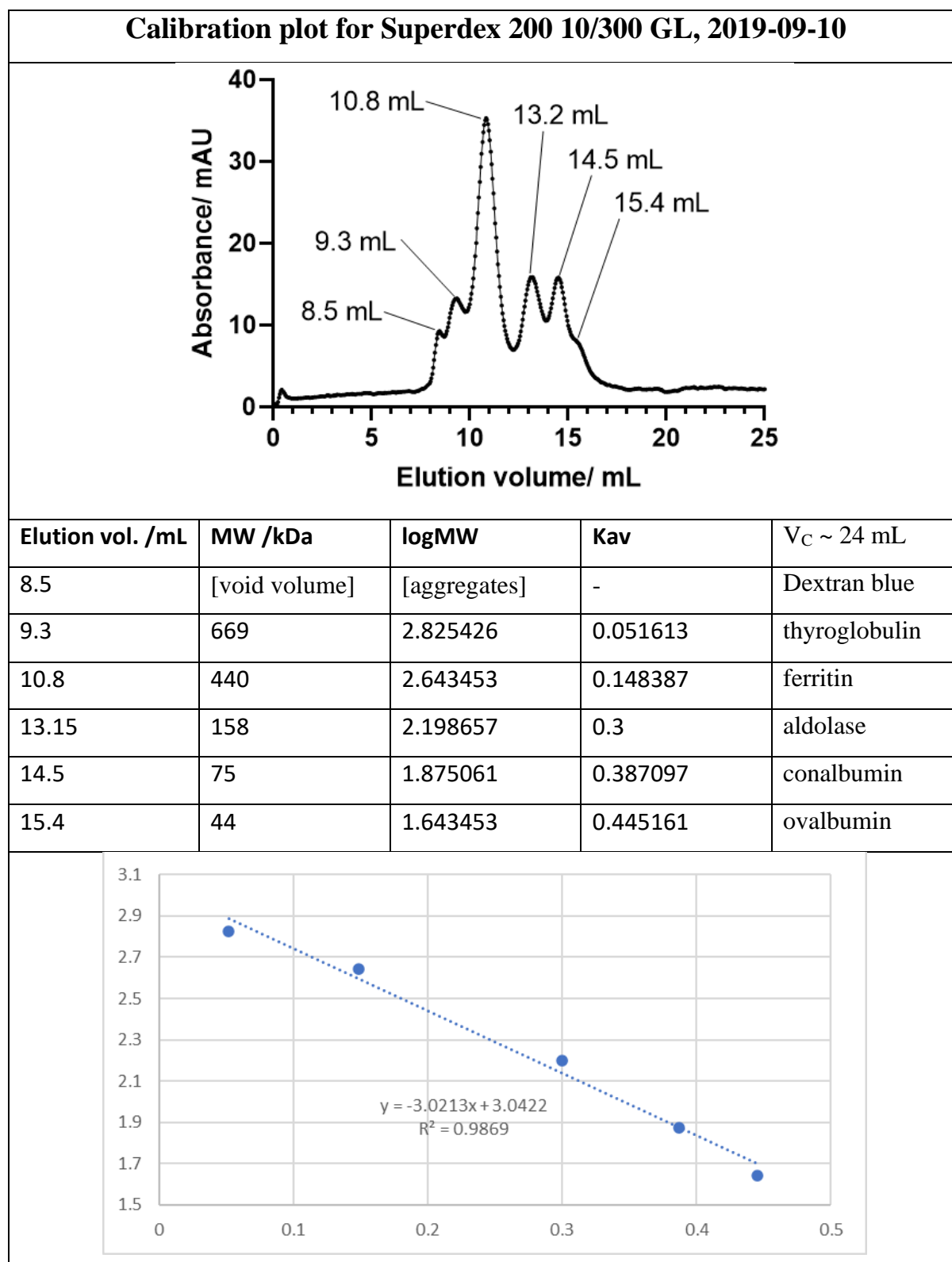$y = -3.0213x + 3.0422$
$R^2 = 0.9869$

Table 8-1: Calibration chromatogram and plot for analytical SEC

8.2 Table of dimeric transaminases used in structural comparative analysis

Dimeric transaminases with closest structural similarity to PfTA were identified from the PDB. Only those with an oligomerisation assignment based on biophysical evidence were tabulated. Since the tetrameric transaminases used for the structural comparative analysis were holo-form structures, the dimeric transaminases tabulated here were also restricted to only holo-form structures, for the sake of fair comparison, since structural differences have previously been observed for apo- and holo- form transaminases[42]. Twelve structures (Table 8-2) were thus identified, including the well-researched 2mcTAs VfTA (*Vibrio fluvialis*)[92] and CvTA (*Chromobacterium violaceum*)[16].

Entry	Reference(s)	Source organism, Name, (Code)	Summary of biophysical method(s) employed
1	Kwon *et al.* 2019 [93] Mathew *et al.* 2016 [94]	*Thermomicrobium roseum* **TrTA** (6IO1)	**SEC-MALS**. 20 mM Tris-HCl, pH 8.0, 500 mM NaCl
2	Palacio *et al.* 2019 [95]	*Pseudomonas jessenii* **PjTA** (6G4D)	PDB entry cites light scattering; details not supplied. Crystal presents dimeric structure.
3	Li *et al.* 2019 [22]	*Bifidobacterium kashiwanohense* PV20-2 **BkTA** (6JIX)	**SEC.** BkTA injected at 14 mg/mL; 10 mM HEPES, pH 7.4, 150 mM KCl.
4	Guidi *et al.* 2018 [96]	*Virgibacillus pantothenicus* 21D **VpTA** (6FYQ)	**SEC.** Found 104 kDa (exp. 106 kDa). 20 mM phosphate pH 8.0, 150 mM NaCl.
5	Rausch *et al.* 2013 [97]	*Paracoccus denitrificans* **PdTA** (4GRX)	**SEC.** Found 86 kDa (exp. 103 kDa, based on sequence). 25 mM HEPES / NaOH, pH 8.3
6	Midelfort *et al.* 2013 [92] Shin *et al.* 2003 [98]	*Vibrio fluvialis* JS17 **VfTA** (4E3R)	(Shin *et al.*) **SEC.** 50 mM potassium phosphate, pH 7.0, 150 mM NaCl.
7	Sayer *et al.* 2013 [42]	*Chromobacterium violaceum* **CvTA** (4AH3)	**SEC.** Found 100 kDa (exp. 102 kDa). Details not supplied.
8	Newman *et al.* 2013 [99]	*Escherichia coli* **EcsoTA** (4ADD)	**DLS.** Details not supplied.
9	Tripathi and Ramachandran 2006 [100]	*Mycobacterium tuberculosis* H37Rv **MtTA** (2CIN)	**SEC.** Details not supplied.
10	Gao *et al.* 2019 [101]	*Streptomyces kaniharaensis* **SkTA** (6SSD)	**SEC-MALS.** 10 mM HEPES, 150 mM NaCl, 1 mM TCEP
11	Kwon *et al.* 2019 [102]	*Sphaerobacter thermophilus* **StTA** (6K8H)	**SEC-MALS.** 20 mM Tris-HCl, pH 8.0, 500 mM NaCl
12	Wybenga *et al.* 2012 [103] Kim *et al.* 2007 [104]	*Mesorhizobium* sp. LUK MesTA (2YKX)	(Kim *et al.*) **SEC.** Details not supplied.

Table 8-2 Biophysically-evidence dimeric holo-form transaminases

8.3 Mutagenesis primers for PfTA mutants

All primers are given 5'→3' with mutated codon highlighted red, forward-reverse primer overlapping section highlighted blue, and overhanging section highlighted green.

<u>Y155A</u>

For: 5' GTT AGC CTG GCG AAC GGT TAT CAC GGC CTG AAC TTT GGT GG 3'
Rev: 5' ATA ACC GTT CGC CAG GCT AAC GAA TTT CAC CTT ATC CGG TTG ACC 3'

<u>E189P</u>

For: 5' C TTC TTT CAG GTT CCG AGC CCG CAC CTG TAT CGT AAC CCG TTC ACC 3'
Rev: 5' CGG GCT CGG AAC CTG AAA GAA GCC CGG CAT CAG CGG CTC G 3'

<u>F198E</u>

1st attempt

For: 5' AT CGT AAC CCG GAA ACC AAC GAC CCG GAG GAA CTG GCG G 3'
Rev: 5' CGT TGG TTT CCG GGT TAC GAT ACA GGT GCG GGC TTT CAA CC 3'

2nd attempt

For: 5' G TAT CGT AAC CCG GAA ACC AAC GAC CCG GAG GAA CTG GC 3'
Rev: 5' T GGT TTC CGG GTT ACG ATA CAG GTG CGG GCT TTC AAC CT 3'

3rd attempt

For: 5' GAA ACC AAC GAC CCG GAG GAA CTG GCG GAA ATC TGC GCG 3'
Rev: 5' C CTC CGG GTC GTT GGT TTC CGG GTT ACG ATA CAG GTG CGG 3'

<u>I212E</u>

For: 5' TGC GCG CAA GAA CTG GAG CGT CAG ATC GAA ATG CAA GCG CCG GG 3'
Rev: 5' ACG CTC CAG TTC TTG CGC GCA GAT TTC CGC CAG TTC CTC CGG G 3'

<u>R215E</u>

For: 5' G CAA ATT CTG GAG GAA CAG ATC GAA ATG CAA GCG CCG GGT ACC G 3'
Rev: 5' C GAT CTG TTC CTC CAG AAT TTG CGC GCA GAT TTC CGC CAG TTC C 3'

<u>I212E/R215E</u>

For: 5' C TGC GCG CAA GAA CTG GAG GAA CAG ATC GAA ATG CAA GCG CCG GG 3'
Rev: 5' CTG TTC CTC CAG TTC TTG CGC GCA GAT TTC CGC CAG TTC CTC CGG 3'

<u>M219R</u>

For: 5' AG CGT CAG ATC GAA CGT CAA GCG CCG GGT ACC GTG GCG GCG C 3'
Rv: 5' G CGC TTG ACG TTC GAT CTG ACG CTC CAG AAT TTG CGC GCA GAT TTC CGC 3'

8.4   List of publications

Roura Padrosa D, Alaux R, Smith P, Dreveny I, López-Gallego F, Paradisi F. Enhancing PLP-Binding Capacity of Class-III ω-Transaminase by Single Residue Substitution. Front Bioeng Biotechnol. 2019 Oct 18; 7:282. doi: 10.3389/fbioe.2019.00282. PMID: 31681755; PMCID: PMC6813460.

8.5   PIP placement review

According to BBSRC funding attached requirements, a PIP (professional internships for PhD students) placement was undertaken, June-August 2021. The host organisation was Tao China Zentrum Schweiz. The business of the host organisation is facilitating relations between Chinese and European enterprises, associations or other bodies.

Initial plans for the project were made in anticipation of being on-site in company offices, and had to be modified due to the impact of Covid-19, to allow for remote working.

The project, as modified, was framed in the context of an agreement recently formed between the psychosomatic research committee of the China Life Care Association (CLCA, a non-profit social welfare organisation) and the World Association for Sustainable Digital Transformation (WASDT), with the cooperating partners evaluating the possible benefits of organising a trade exhibition. The project was an open-ended exploration of the implementation of blockchain technology in healthcare applications.

The project was challenging in the limited information available at the intersection of these two topics (blockchain technology and healthcare systems). Specific research in the area was identifiable, though sparse. A greater amount was available when considering the two topics individually and separately. In essence, blockchain constitutes a distributed ledger system in which any kind of record or set of records can be kept, without the need for a trusted and authoritative body to keep or to validate the recorded information. As such, it can notionally be applied to any problem involving record keeping, especially where a governance structure is poorly developed, for example in the logistics of drug distribution in developing healthcare

systems, to minimize the risk of counterfeiting. The output of the project was a summary report of the information researched and the themes identifiable therein.