Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

Research Article

X-ray crystal structure and specificity of the *Toxoplasma gondii* ME49 *Tg*APN2

Emilia M. Marijanovic[1], Karolina Weronika Swiderska[2], James Andersen[1], Jasmin C. Aschenbrenner[1], Chaille T. Webb[1], Marcin Drag[2], Nyssa Drinkwater[1] and Sheena McGowan[1]

[1]Biomedicine Discovery Institute, Department of Microbiology, Monash University Clayton, Melbourne, VIC 3800, Australia; [2]Department of Chemical Biology and Bioimaging, Wroclaw University of Science and Technology, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland

Correspondence: Sheena McGowan (Sheena.McGowan@monash.edu)

Toxoplasmosis is a parasitic disease caused by infection with *Toxoplasma gondii* that currently has few therapeutic options. The M1 aminopeptidase enzymes have been shown to be attractive targets for anti-parasitic agents and/or vaccine candidates, suggesting potential to re-purpose inhibitors between parasite M1 aminopeptidase targets. The M1 aminopeptidase *Tg*APN2 has been suggested to be a potential new drug target for toxoplasmosis. Here we investigate the structure and function of *Tg*APN2, a homologue of the antimalarial drug target *Pf*A-M1, and evaluate the capacity to use inhibitors that target *Pf*A-M1 against *Tg*APN2. The results show that despite a similar overall fold, the *Tg*APN2 has a unique substrate specificity and inhibition profile. Sequence and structure differences are investigated and show how comparative structure-activity relationships may provide a route to obtaining potent inhibitors of *Tg*APN2.

Introduction

In *Plasmodium, Trypanosome, Eimeria* and *Cryptosporidium* parasites, the clan MA family M1 aminopeptidases have been investigated as potential targets for new anti-parasitics and/or vaccine development [1]. M1 aminopeptidases, also known as aminopeptidase N, are generally membrane-anchored metallo-exopeptidases found throughout all kingdoms of life, and have a wide range of functions including cell maintenance, growth and development, and defence [1]. In the apicomplexan malaria parasite *Plasmodium falciparum,* there is a single M1 aminopeptidase, known as *Pf*A-M1 [2]. *Pf*A-M1 has been shown to be important for the final stages of haemoglobin digestion, an essential metabolic pathway required for blood-stage *Plasmodium* growth and development [3]. Inhibition of *Pf*A-M1 activity in parasites has been shown to control both laboratory (*P. falciparum*) and murine (*Plasmodium chaubaudi chaubaudi*) models of malaria [4].

Toxoplasmosis is caused by infection of *Toxoplasma gondii* which, like *Plasmodium, Eimeria* and *Cryptosporidum* is an apicomplexan parasite. Toxoplasmosis in healthy individuals is generally asymptomatic, however, serious complications can arise in immunocompromised patients and pregnant women [5]. Congenital infection can lead to miscarriage, mental retardation, vision and hearing problems [5]. Preventing toxoplasmosis in immunocompromised patients relies on chemotherapy generally consisting of pyrimethamine with either sulfadiazine or clindamycin [6,7]. Agents to treat active and latent disease remain scarce and new molecular targets for therapeutics are desirable for future drug discovery.

The *T. gondii* ME49 strain contains three canonical M1 aminopeptidases, *Tg*APN1 (TGME49_221310), *Tg*APN2 (TGME49_224350) and *Tg*APN3 (TGME4_224460) [8], and their respective structural genes are located on different chromosomes [9]. *Tg*APN1 has previously been experimentally shown to be active as a metallo-aminopeptidase, as well as being immunogenic — suggesting an involvement in infection response [10]. *Tg*APN3 is also functional as a metallo-exopeptidase with a substrate preference for Ala > Tyr/Cys > Arg but interestingly has a signal

Received: 18 July 2020
Revised: 3 September 2020
Accepted: 14 September 2020

Accepted Manuscript online: 16 September 2020
Version of Record published: 12 October 2020

Biochemical Journal (2020) 477 3819–3832
https://doi.org/10.1042/BCJ20200569

peptide in place of the normal N-terminal transmembrane anchor. Recent studies into the localisation of *Tg*APN3 [9] showed that the protease was co-located with the GRA protein in the organelle and parasitophorous vacuole, and could be secreted with tachyzoites [9]. *Tg*APN2 is predicted to be localised within the parasite cytosol (https://toxodb.org [11]), and is expressed throughout the life cycle of the parasite, including during the tachyzoite stage where the parasite undergoes acute reproduction [11,12] (https://toxodb.org). The biological role(s) of *Tg*APN2 remain unclear but a genome wide CRISPR screen of the *T. gondii* reference strain, showed the loss of *Tg*APN2 resulted in a fitness cost of −1.21 in culture, meaning that it is important but probably not essential to tachyzoite growth [13]. The transcript is, however, up-regulated in bradyzoites, which is a more important stage for drug development [13].

Out of the three *T. gondii* M1 aminopeptidases, *Tg*APN2 shares the highest sequence similarity to that of *Pf*A-M1 (38%) and as such, has been suggested to be target worthy of further investigation as a novel drug target [14]. Biochemical characterisation of recombinant *Tg*APN2, however, showed a substrate preference for a P1 Arg > Ala > Leu, suggesting significant changes in the substrate specificity pockets between *Tg*APN2 and *Pf*A-M1 [14,15]. In this study, we aimed to produce structure activity relationship (SAR) data of a selection of M1 aminopeptidase inhibitors with the aim to map the similarities and differences in SAR between *Pf*A-M1 and *Tg*APN2. The identification of a potent inhibitor of enzymatic activity would provide a valuable tool to probe the biological roles of *Tg*APN2 in *T. gondii*, as well as assess the capacity of the enzyme to act as a drug target. Our findings show that *Tg*APN2 displays a significantly different substrate specificity to that of *Pf*A-M1. Additionally, inhibition studies indicate that there are significant changes in inhibitory kinetics likely due to differences within the S1 substrate specificity pocket of *Tg*APN2, as characterised in the crystal structure.

Experimental

Expression and purification of recombinant ecto-domain of *Tg*APN2

DNA coding for residues 520–1419 of TGME49_224350 from *T. gondii* ME49 (UniProt S8G5K8) with an in-frame C-terminal His6 tag was chemically synthesised by DNA2.0. Potential glycosylation sites at N840 and N1303 were altered to encode glutamine. The gene was provided in the pJ404 vector for expression, which also encodes for ampicillin resistance. The construct was transformed into BL21 (DE3) cells and protein expressed using an auto-induction method based on previously described protocols [16]. A 200 ml culture was grown for 24 h at 28°C followed by a further 24 h at 20°C. Cells were harvested and resuspended in PBS pH 8.0, 0.3 M NaCl, 5% (v/v) glycerol prior to lysis by sonication. Clarified lysates were bound to a Ni-NTA-agarose column in nickel-affinity buffer (PBS pH 8.0, 300 mM NaCl, 5% glycerol, 20 mM imidazole), and eluted in nickel-affinity buffer supplemented with 250 mM imidazole. The protein was further purified by size exclusion chromatography on a Superdex 200 16/60 using an AKTAxpress high throughput chromatography system in 50 mM Hepes pH 8.0, 0.3 M NaCl, 5% (v/v) glycerol. The major peak, containing *Tg*APN2, was pooled and concentrated to 5 mg/ml. The purity of the protein was assessed by SDS–PAGE and protein concentration determined using Pierce™ BCA protein assay.

Enzymatic analysis

Aminopeptidase activity was determined by measuring the hydrolysis of 7-amido-methylcoumarin (Mec) from the fluorogenic substrate L-leucine-7-amido-4-methylcoumarin hydrochloride (Leu-Mec) (Sigma L2145) or L-arginine-7-amido-4-methylcoumarin hydrochloride (Arg-Mec) (Sigma A2027). Reactions were carried out in 96-well microtitre plates (100 µl total volume, 60 min, 37°C) using a spectrofluorimeter (BMG LabTech Optima) with excitation at 355 nm and emission at 460 nm. For determination of kinetic parameters, enzyme was first added to 100 mM Tris–HCl, pH 8.0 for 10 min prior to the addition of substrate. Initial rates were obtained at 37°C over a range of substrate concentrations spanning K_M (0.5–500 µM) and at fixed enzyme concentration (20 nM). Calculations of Michaelis–Menten constants (K_M) and rates (k_{cat}) were performed using GraphPad Prism. Observation of the rate of substrate hydrolysis under chelating and different pH conditions was performed by alteration of the enzyme buffer by either the addition of EDTA (0–200 mM) or alteration of pH (pH 5–9) of the buffer prior to enzyme addition. Assays were performed with fixed enzyme concentration (20 nM) and fixed substrate concentration (15 µM). Statistical significance was calculated using a Mann–Whitney test, comparing the mean of the slope velocity (fluorescence per sec, $n = 6$) between the different pH conditions. $K_i^{(app)}$ values were calculated as described for *Pf*A-M1 [17], where inhibitor (0–1 mM) and enzyme were pre-incubated in 100 mM Tris–HCl, pH 8.0 for 20 min prior to the addition of substrate. Substrate

concentration (15 μM) was selected to allow sensitive detection of enzyme activity while not exceeding the K_M. Enzyme concentration was fixed (20 nM). The $K_i^{(app)}$ values were calculated by plotting the initial rates versus inhibitor concentration, and fitting to the Morrison equation in GraphPad Prism (non-linear regression method).

Substrate-specificity profiling was achieved by the use of a fluorogenic substrate library containing 63 amino acids [15]. For convenience in solid-phase synthesis, we employed a 7-amino-4-carbamoylmethylcoumarin (ACC) fluorogenic leaving group in this library. The final screening of the library was carried out at 10 μM substrate and 0.5 nM enzyme as described above and release of free ACC fluorophore was monitored. Each experiment was repeated at least three times and the average value with standard error was calculated. Concentration of DMSO in the assay was <1% (v/v).

Crystallisation and structure refinement

*Tg*APN2 was concentrated to 5, 10 and 13.5 mg ml^{-1} for initial sparse-matrix screening. The Monash Macromolecular Crystallization Facility was utilised to set up sitting-drop sparse matrix screens (INTELLI-PLATE 96-3, Art Robbins Instruments). Crystals were observed in seven different conditions. Two conditions were selected for optimisation/screening by the hanging drop method: (1) 15–25% PEG 8000, 0.1 M Hepes pH 6.5–8.0 and (2) 20–30% PEG 3350, 0.1 M Bis-Tris pH 5.0–7.0. Each tray was set with 6, 8 and 10 mg ml^{-1}. Crystals were reproduced in both conditions and the best diffraction was observed from screening a crystal grown from 6 mg ml^{-1} in the PEG3350/Bis-Tris screen. Final crystals of *Tg*APN2 were obtained with a protein concentration of 5 mg ml^{-1} in 28% PEG 3350, 0.1 M Bis-Tris pH 5.0. Cryo-protection was achieved by briefly soaking single crystals in the reservoir containing 20% glycerol.

A single *Tg*APN2 crystal was harvested (CryoLoop^{TM}, Hampton Research) and flash cooled in liquid nitrogen. X-ray diffraction data were collected at 100 K using synchrotron radiation on the MX2 beamline (3BM1) at the Australian Synchrotron. Data were indexed and integrated using XDS [18], and scaled and averaged in AIMLESS [19] from the CCP4 suite [20]. Initial phases were solved by molecular replacement using PHASER [21] and 3EBG as a search model [22]. Model building and refinement were performed using Phenix [23] and Coot [24]. Initial refinement made use of non-crystallographic symmetry until model was near completion. Final data collection and refinement statistics are shown in Table 1. The coordinates and structure factors are available from the Protein Data Bank (PDB ID: 6OIU). Analysis of the surface was performed using PDBePISA [25]. All crystal structure figures and analyses were prepared in the Pymol Molecular Graphics System, version 1.3r2 (Schrödinger, LLC). Structure and sequence comparisons were carried out using PDBeFold [26], PDBePISA [25] Stride [27] and Clustal Omega[28,29].

Results

Expression and purification of *Tg*APN2

M1 aminopeptidases are membrane anchored enzymes that can be anchored either intra- or extracellularly [1]. *Tg*APN2 is confidently predicted to be present in the cytosol (Probability MAP/MCMC 0.98/1, https://toxodb.org [11]) and has previously been shown to be a functional aminopeptidase [14]. In contrast with the previous biochemical study that — on the basis of amino acid sequence — suggested *Tg*APN2 is homohexameric [14], our analysis indicated that *Tg*APN2 is likely monomeric. The *TgAPN2* structural gene encodes a 1419 amino acid protein (predicted molecular weight 156.7 kDa) that has an identifiable transmembrane region ∼300 amino acids from the N-terminus (Supplementary Figure S1) as is also observed in the other *T. gondii* M1 aminopeptidases, as well as to *Pf*A-M1 (Supplementary Figure S1). This arrangement suggests an N-terminal domain that would be located on the opposite side of the membrane from the enzymatic domain. Sequence analysis of the first 319 amino acids shows significant regions of low-complexity, and analysis using the D2P2 server [30] confirms regions of intrinsic disorder (Supplementary Figure S2). PSI-Blast, and fold and function assignment failed to identify any homologues of this domain with known structure or function [31,32].

As we were interested in characterising the aminopeptidase function of *Tg*APN2, we chose to produce the soluble ecto-domain, excluding the N-terminal and transmembrane domains from our expression construct. The first residue of our truncated *Tg*APN2 (residues 520–1419) aligned with the beginning of truncated *Pf*A-M1 (Supplementary Figure S1) [4]. Truncated *Tg*APN2 was produced using a bacterial expression system and purified from the soluble fraction using a two-step purification facilitated by the inclusion of an in-frame

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

Table 1 Data collection and refinement

Data collection	
Diffraction source	MX2 beamline, AS
Wavelength (Å)	0.9537
Temperature (K)	100
Detector	CCD Quantum 201r (ADSC)
Crystal-to-detector distance (mm)	250
Rotation range per image (°)	0.5
Total rotation range (°)	0–90
Exposure time per image (s)	1.0
Space group	P 1 21 1
a, *b*, *c* (Å)	92.93, 207.87, 102.87 90, 94.34, 90
Mosaicity (°)	0.17
Resolution range (Å)	38.54–2.20 (2.28–2.20)
Total No. of reflections	741 396 (35 716)
No. of unique reflections	195 801(19 286)
Completeness (%)	99.6 (98.1)
Multiplicity	3.8 (3.8) 5.3 (1.4)
R-merge (%)	0.205 (1.062)
R-pim (%)	0.122 (0.629)
$CC_{1/2}$	0.979 (0.433)
Overall B factor from Wilson plot (Å^2)	23.92
Refinement	
R-work	0.1904 (0.2641)
R-free	0.2357 (0.3160)
non-hydrogen atoms	30 396
Macromolecules	27 586
Ligands	62
Water	2748
Protein residues	3575
RMS(bonds)	0.002
RMS(angles)	0.41
Ramachandran favoured (%)	97.47
Ramachandran outliers (%)	0.11
Clashscore	2.71/100th percentile[1]
Average B-factor	29.26
Macromolecules	28.99
Ligands	39.46
Solvent	31.73
PDB ID	6OlU

Values in parentheses are for the outer shell.
[1]Molprobity $N = 456$, 2.200 Å ± 0.25 Å.

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

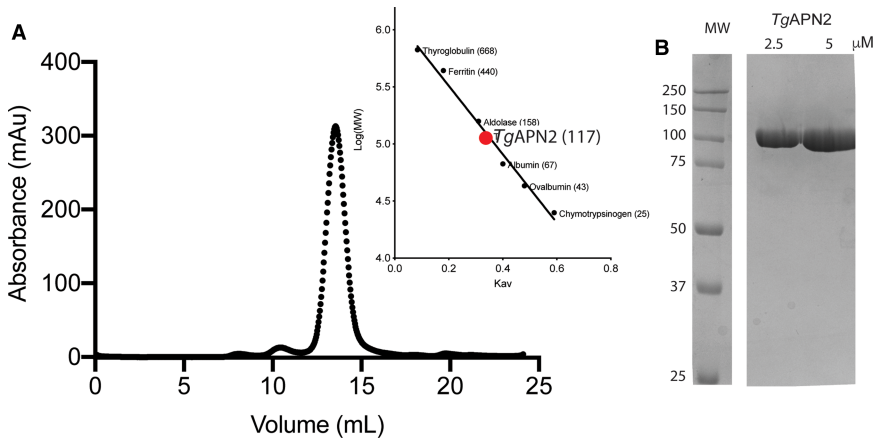


Figure 1. Expression and purification of *Tg*APN2.

(**A**) Analytical size exclusion chromatography of *Tg*APN2 using an S200 10/300 column and the standard curve showing the calculated molecular weight of purified *Tg*APN2 (red point). (**B**) An SDS–PAGE gel of purified *Tg*APN2.

C-terminal hexa-histidine tag. Analytical size exclusion chromatography indicated that the purified, truncated *Tg*APN2 is a monomer with an expected molecular weight of 117 kDa (Figure 1).

Activity and substrate specificity of *Tg*APN2

To confirm if our purified recombinant *Tg*APN2 could function as an aminopeptidase, we used a fluorescence based assay with the fluorogenic substrate, *L*-Leucine-7-amido-4-methylcoumarin hydrochloride (Leu-Mec) [4]. The results show a slow enzyme possessing Michaelis–Menten kinetics with a K_M of 31.8 ± 3.0 μM and a turnover rate of 1.32 s^{-1} (Figure 2A, and Table 2). Given the slow kinetics for Leu-Mec and previous study identifying Arg as the preferred P1 residue [14], we also determined the kinetic parameters toward L-arginine-7-amido-4-methylcoumarin hydrochloride (Arg-Mec) (Table 2 and Figure 2A). In comparison with a P1 Leu (Leu-Mec), the results suggest that *Tg*APN2 has a weaker substrate affinity for Arg (K_M of 148 μM), but a significantly higher turnover rate (k_{cat} of 1017 s^{-1}). To investigate the metal-ion dependency of *Tg*APN2, we monitored aminopeptidase activity of *Tg*APN2 at increasing concentrations of EDTA and observed a reduction in enzyme velocity as EDTA concentration increased (Figure 2B). These results confirmed that, like other M1 aminopeptidases, *Tg*APN2 activity is metal-dependent. The pH dependency of the recombinant enzyme was investigated and found to be most active at pH 8.0, but retained majority of its activity at a neutral pH (Figure 2C). Changing the pH to either 6.0 or 9.0 resulted in a statistically significant decrease in the activity of *Tg*APN2 (*P* value of 0.0087 and 0.015, respectively) (Figure 2C).

To determine the P1 substrate specificity, we screened the activity of *Tg*APN2 against a fluorogenic peptide substrate library that incorporates 19 natural and 44 unnatural amino acids [15]. A total of 63 substrates were monitored for fluorescence post-cleavage, and the results confirmed that *Tg*APN2 strongly prefers a P1 Arg (Figure 2D). Homo-arginine (hArg), containing four carbons in the side chain, was the most active and the data was normalised by defining the activity against hArg as 100%. A natural Arg residue showed 81% activity, whilst activity toward Lys was 52% (Figure 2D). These results also correlate with the observed activity of *Tg*APN2 toward the P1 Leu as in our library screen the enzyme was only able to achieve 26% activity (Figure 2D). Alanine was able to be hydrolysed (~20%) however other small hydrophobic residues such as valine, serine, proline or threonine showed no measurable activity (Figure 2D). Interestingly, P1 phenylalanine and tyrosine were digested poorly (<10% activity) but non-natural substitutes with longer aliphatic groups were effectively digested, (hTyr 54% and hPhe 18%), suggesting the extra carbon may allow the residues to occupy a more favourable position within the enzyme (Figure 2D).

X-ray crystal structure of *Tg*APN2 aminopeptidase domain

To investigate the structural basis for the substrate specificity, we solved the X-ray crystal structure of truncated *Tg*APN2 to 2.2 Å. A total of four *Tg*APN2 monomers were placed in the asymmetric unit, however, alignment

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

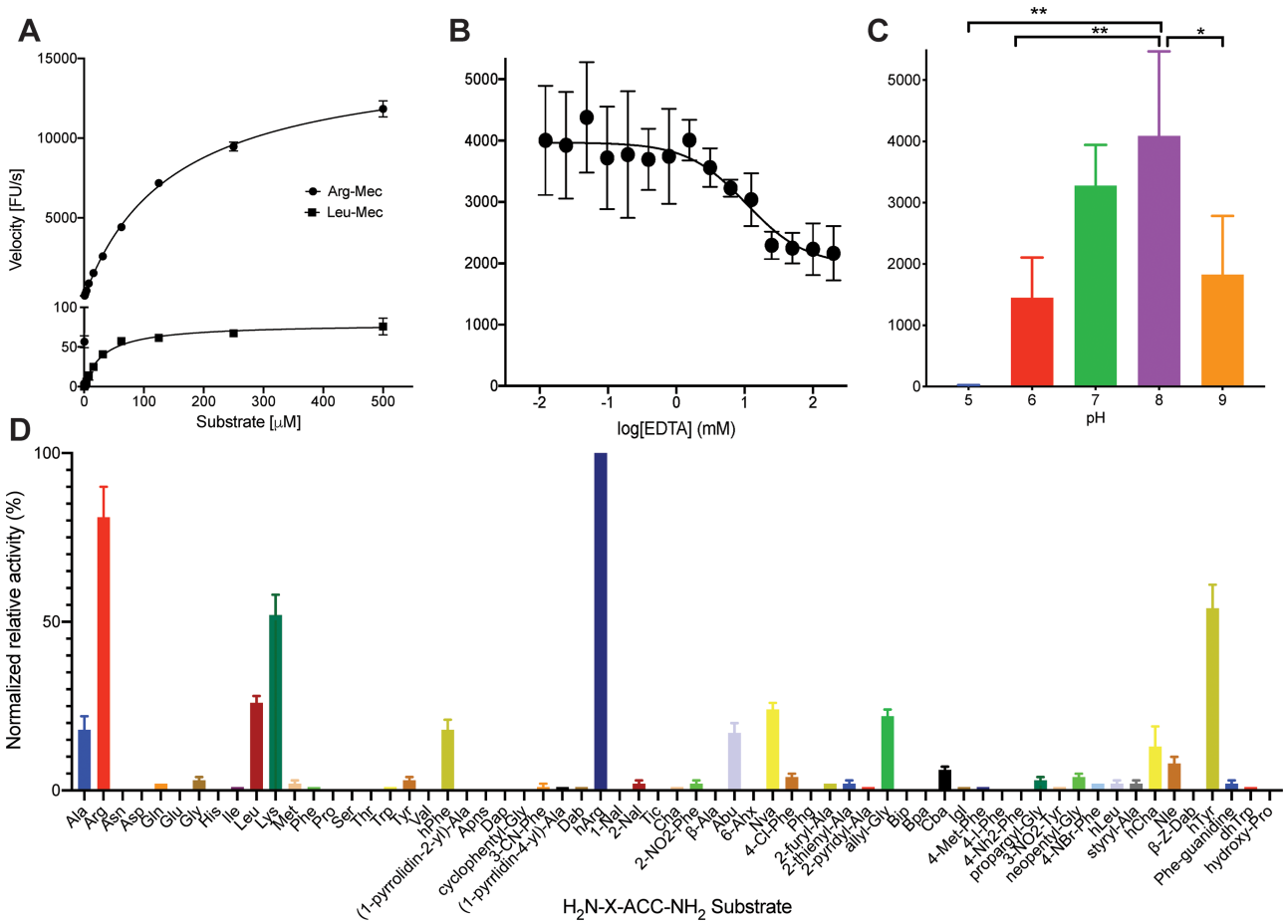


Figure 2. Aminopeptidase activity and P1 substrate specificity of *Tg*APN2.

(**A**) Michaelis–Menten plot for *Tg*APN2 against substrates Leu-Mec and Arg-Mec. (**B**) The effect of increasing concentration of EDTA (x-axis) on the velocity of *Tg*APN2 (*y*-axis). (**C**) Enzyme velocity of *Tg*APN2 at different pH conditions. Statistical significance was calculated using a Mann–Whitney test and comparing the mean ($n = 6$) of the velocity (fluorescence units per sec) at different pH conditions to pH 8.0. (**D**) Normalised activities of *Tg*APN2 against the extended substrate library where the *x*-axis represents the abbreviated amino acid names and *y*-axis represents the normalised relative activity relative to unnatural amino acid substrate hArg.

of the four chains indicates that all are essentially identical (RMSD over 901 Cα atoms is 0.18–0.34 Å between chains A–D). Chain C in the asymmetric unit was the most complete, with only two residues not resolved in the density (1208–1209). Chain D was missing residues 520–521, and chains A and D also lacked continuous density for residues 1204–1210. The structural description and figures in this paper relate to chain C.

The final model confirmed that *Tg*APN2 is a monomer with a standard M1 aminopeptidase ecto-domain fold of four domains with a total of seven α-sheets and 24 α-helices (Figure 3A). The four domains consist of an N-terminal domain I (residues 520–731), the catalytic domain II (residues 732–980), a small Ig-like fold for domain III (residues 981–1080) and an entirely helical C-terminal domain IV (residues 1081–1419)

Table 2 Kinetic parameters of *Tg*APN2

Substrate	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} μM^{-1})
L-Leu-Mec	31.8 ± 3.0	1.32 ± 0.03	0.0415
L-Arg-Mec	148 ± 5.1	1017 ± 15.2	6.87

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

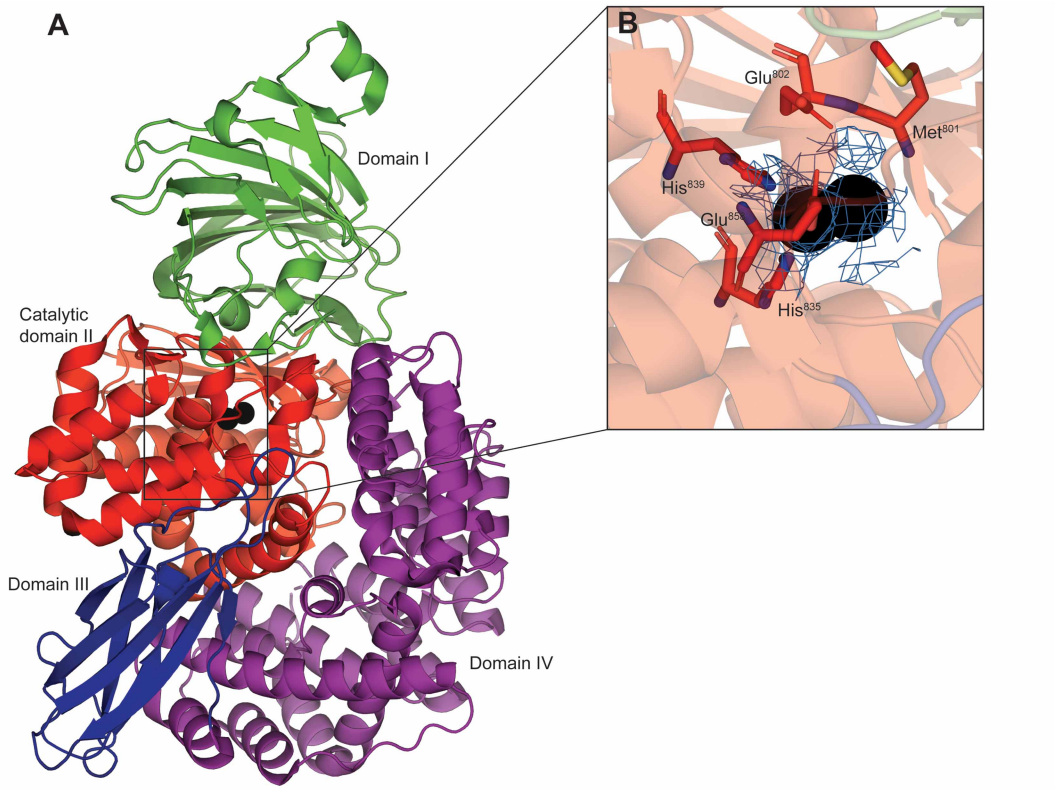


Figure 3. The X-ray crystal structure of *Tg*APN2 (PDB ID: 6OIU).

(**A**) Cartoon representation of chain C of *Tg*APN2, showing each of the four domains, domain I (residues 520–731 in green), catalytic domain II (residues 732–980 in red), domain III (residues 981–1080 in blue) and domain IV (residues 1081–1419 in purple), and the zinc ions (grey spheres). (**B**) The electron density $2F_o–F_c$ maps contoured at one sigma showing density over the two zinc ions within the active site. Zinc coordinating residues (His^{835}, His^{839} and Glu^{858}) and Met^{801} and Glu^{802} of the conserved G^{799}AMEN motif are shown as sticks with red carbon atoms.

(Figure 3A). The overall structure was in a closed and compact confirmation with the helical domain IV packed against the catalytic domain II (Figure 3A). Analysis of the structure shows that the tertiary structure of *Tg*APN2 is highly consistent with that of the clan MA, family M1 aminopeptidases (Supplementary Figure S3). The bacterial and parasite homologues showed the highest structural similarity, with an overall RMSD (Cα atoms) of 1.2 Å to the *E. coli* aminopeptidase N (PDB ID: 2DQ6 [33]) and 1.6 Å to *Pf*A-M1 (PDB ID 3EBG [4]) (Supplementary Figure S3A). Comparison to the structures of human aminopeptidase N (APN, PDB ID: 5LHD [34]) and endoplasmic reticulum aminopeptidase 1 (ERAP1, PDB ID: 2YD0 [35]) when in a similar conformation to that of *Tg*APN2 share an overall RMSD (Cα atoms) of 3.5 Å and 2.3 Å, respectively (Supplementary Figure S3B). In particular, the N-terminal domain shows significant differences between *Tg*APN2 and human APN and ERAP1. The most noticeable difference is a small insertion that forms a helical loop in *Tg*APN2 domain I (residues 590–601) that is not present in the homologues (Supplementary Figure S3).

The active site of *Tg*APN2 in domain II contains the two conserved M1 aminopeptidase motifs, the zinc-binding motif ($H^{835}EYFHX_{17}KE^{858}$) and substrate binding G^{799}AMEN motif (Figure 3B). Similar to other M1 aminopeptidases, a zinc ion is coordinated by a catalytic triad made up of the side-chains of a His^{835}, His^{839}, Glu^{858} (Figure 3B). Generally in unbound structures of M1 aminopeptidases, the tetrahedral zinc coordination state is completed by a water molecule in the active site cavity [36,37]. However, during the refinement of our structural model, it became evident that a water molecule would not satisfy the density observed in this position (Supplementary Figure S4). Individual refinement of a chloride, magnesium, nickel and zinc ion in this position indicated that the zinc and nickel were able to satisfy the density (Supplementary Figure S4). The second ion

Biochemical Journal (2020) 477 3819–3832
https://doi.org/10.1042/BCJ20200569

forms a bond with the active zinc ion (distance 2.3 Å, Supplementary Figure S4B) whilst the remaining coordination of the second ion results from longer bonds (2.9–3.7 Å) with residues Glu^{802}, Glu^{836}, Tyr^{919} and a water molecule (Supplementary Figure S4). A glycerol molecule is also present in this area (Supplementary Figure S4). Whilst the identification of a potential second metal site was unexpected, subsequent alignment of the coordinates of *Tg*APN2 with the unbound structure of *Pf*A-M1 (PDB ID 3EBG [4]) showed that there is a Mg^{2+} ion in the same position of *Pf*A-M1. The Mg^{2+} ion is suggested to be a result of high concentrations of Mg^{2+} in the crystallisation conditions of 3EBG [4]. For *Tg*APN2, there were no supplemental ions in the crystallisation buffers suggesting that the metal ion may have come from the nickel resin used to purify the protein or is a zinc ion that has likely been scavenged during purification. The final coordinates of *Tg*APN2 (6OIU.pdb) have been modelled with a zinc ion.

The substrate specificity pockets possess a similar architecture to that of *Pf*A-M1 but sequence changes dramatically alter the local environment (Supplementary Figure S5A). The main wall of the S1 pocket is formed by the G^{799}AMEN^{803} motif (Supplementary Figure S5A) which is highly conserved in M1 aminopeptidases, forming part of the S1 pocket and stabilising substrate prior to hydrolysis. Immediately preceding Gly^{799}, *Tg*APN2 has a Met^{798} residue that provides the pocket with a more amphipathic nature than the equivalent hydrophobic valine in *Pf*A-M1 (Supplementary Figure S5A). At the top of the S1 pocket, further sequence variation may account for substrate differences between *Pf*A-M1 and *Tg*APN2, where *Tg*APN2 has a Gln^{1362} in place of the Met^{1034} in *Pf*A-M1, and an Asp^{911} instead of the flexible Glu^{572} in *Pf*A-M1 (Supplementary Figure S5B). The remaining residues that form the S1 pocket — Glu^{657} and Tyr^{914} — are conserved across the two proteins. The S1' pocket is large and stretches into the internal cavity formed by the domain IV helices with the main contributing residues consisting of Arg^{828}, Asn^{831}, Val^{832} and Val^{862} (Supplementary Figure S5B).

Structure-activity relationships of P1 arginine mimetics

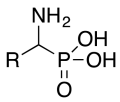
Previous studies have mapped the P1 substrate specificity for *Pf*A-M1 and shown that this enzyme had a strong preference for a P1 Leu or Met [15]. Given that *Tg*APN2 prefers a P1 Arg and/or a longer substituent, we selected three arginine mimetics from a previous study [38] to probe the differences in the structure-activity relationships of the S1 pockets of *Pf*A-M1 and *Tg*APN2. The compounds all contained a phosphonic acid zinc-binding group as well as various P1 arginine mimetics (Table 3). The inhibitory activity of the three selected compounds tested was moderate, but far greater than was observed for *Pf*A-M1 (Table 3). The most potent inhibitor was compound **1** ($K_i^{(app)}$ 1.2 μM, Table 3) which resembles *Tg*APN2's preferred hArg substrate containing four carbons in the side chain. Compound **2** was shorter with three carbons in its side chain and had a slightly higher $K_i^{(app)}$ of 5.5 μM, matching our finding that longer, charged substrates may be preferred. The third compound tested (**3**) differs from **1** and **2** in that there is a phenyl ring within the side chain (Table 3). The addition of this phenyl ring was detrimental to inhibitory activity, increasing the $K_i^{(app)}$ to 11 μM.

We attempted to solve the co-crystal structure of *Tg*APN2 bound to **1** and/or **2**, however, were unsuccessful at obtaining any bound structures. Therefore, we superimposed the coordinates of *Tg*APN2 onto the structures of *Pf*A-M1 bound to **1** (PDB ID 4K5L) and **2** (PDB ID 4K5 M) [38], aligning the catalytic domains and assuming a displacement of the second, non-catalytic zinc ion from the active site (Figure 4). The phosphonic acid is a strong zinc-binding group and we assume that the coordination of the H_3O_3P group is similar in both enzymes. For *Pf*A-M1-**1**, the guanidino moiety of the hArg forms a hydrogen bond with the backbone carbonyl of Glu^{572} (Figure 4A). A similar interaction would be possible for *Tg*APN1 with equivalent residue Asp^{911}; however, it also seems likely that a second H-bond would be possible with the Oε atoms of Asp^{911} as well (Figure 4A). In addition, *Tg*APN2 has Gln^{1362} in H-bonding distance to the guanidino group, possibly providing a third coordinating bond (Figure 4A). This extra bonding capacity likely accounts for the increase in potency observed for *Tg*APN2 in comparison with *Pf*A-M1, by locking the Arg mimetic into a more rigid position. The slightly shorter **2** was found to occupy two different conformations when bound to *Pf*A-M1, the first was similar to that of **1** and the second saw the guanidino group form interactions with Val^{459}, the residue that immediately precedes the GAMEN motif in *Pf*A-M1 (Figure 4A). For *Tg*APN2, we suggest that this second conformation may not occur due to the ability of the compound to form hydrogen bonds with the side-chains of Asp^{911} and Gln^{1362}. The carbonyl of Asp^{911} is beyond regular hydrogen bonding distance (4.5 Å) and this may account for the loss of potency observed for **2** against *Tg*APN2 compared with **1** — the loss of a hydrogen bond to the backbone may result in less stable binding due to the movement of the side-chains.

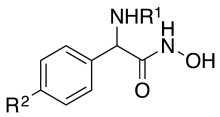
The second conformation of **2** observed for *Pf*A-M1 also highlights the known flexibility of the S1 pockets of M1 aminopeptidases. In both our superpositions, *Tg*APN2 Met^{798} is too close to the compound and would

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

Table 3 Inhibition of *Tg*APN2 by selected phosphonic (MD 1, 2, 3) and hydroxamic acid compounds 9b, 10o and 10q



Compound No[1]	R	K_i (μM)	
		*Tg*APN2	*Pf*A-M1[1]
1	1-amino-5-guanidinopentyl	1.2	11
2	1-amino-4-guanidinobutyl	5.5	193
3	1-amino-2-(4-guanidinophenyl)ethyl	11	>1000



Compound No[2]	R^1	R^2	K_i (μM)	
			*Tg*APN2	*Pf*A-M1[2]
9b	Boc	Br	2.9	0.027
10o	C(O)^{t}Bu	3,4,5-trifluorophenyl	44.9	0.078
10q	C(O)^{t}Bu	thiophen-3-yl	2.4	0.64

K_i values are the mean of three independent experiments.
[1]Sivaraman et al. [38].
[2]Drinkwater et al. [40].

need to move to accommodate the inhibitor. Such a movement was observed in the crystal structure of *Pf*A-M1-**2** where the equivalent residue Val[459] shifted to allow the compound to occupy the second site [38]. A similar movement was also observed with the equivalent *Ec*APN1 residue, with Met[260] being labelled a gatekeeper residue that can move to allow substrate or inhibitor binding [33]. Therefore, it is possible that **2** could form interactions with this region, however, given the ∼17-fold increase in potency observed between *Pf*A-M1 and *Tg*APN2, we feel it is likely that the compound is coordinated in one position in *Tg*APN2 (Figure 4a).

Compound **3**, containing a 1-amino-2-(4-guanidinophenyl)ethyl core, was non-inhibitory toward *Pf*A-M1 but did show moderate activity toward *Tg*APN2 (Table 3). The conclusions of the original study on the SAR of *Pf*A-M1 showed that an amino(4-guanidinophenyl)methyl was capable of inhibition but the addition of the extra CH2 prior to the guanidinophenyl was what was detrimental to activity [38], placing the guanidino nitrogens too close to the end of the S1 pocket. However, *Tg*APN2 appears to like slightly longer charged P1 residues (Figure 2D), suggesting more flexibility of the residues that interact with the inhibitor.

Structure-activity relationships of hydroxamic acid containing *Pf*A-M1 inhibitors

Previous work by our group has generated a library of inhibitors that combine a hydroxamic acid zinc-binding group linked various bi-aryl moieties that show inhibitory activity toward *Pf*A-M1 [17,39,40]. To assess the capacity to re-purpose this scaffold to target *Tg*APN2, we selected three compounds of varying functional groups where R1 is the S1' anchor position and R2 explores the S1 pocket (Table 3). Compound **9b** is the least complex of the selected compounds, providing a Boc group in the S1′ of *Pf*A-M1 and bromo-substituent in R2. The compound showed a near 100-fold loss in activity toward *Tg*APN2 in comparison with *Pf*A-M1 (Table 3). The other two compounds, each had a slightly longer S1′ anchor, probed the capacity of *Tg*APN2 S1 pocket to accommodate either 3,4,5-trifluorophenyl (**10o**) or a thiophen-3-yl substituent (**10q**). The 3,4,5-trifluorophenyl analogue was a poor inhibitor of *Tg*APN2, whereas the thiophen-3-yl analogue had a similar potency to that of

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

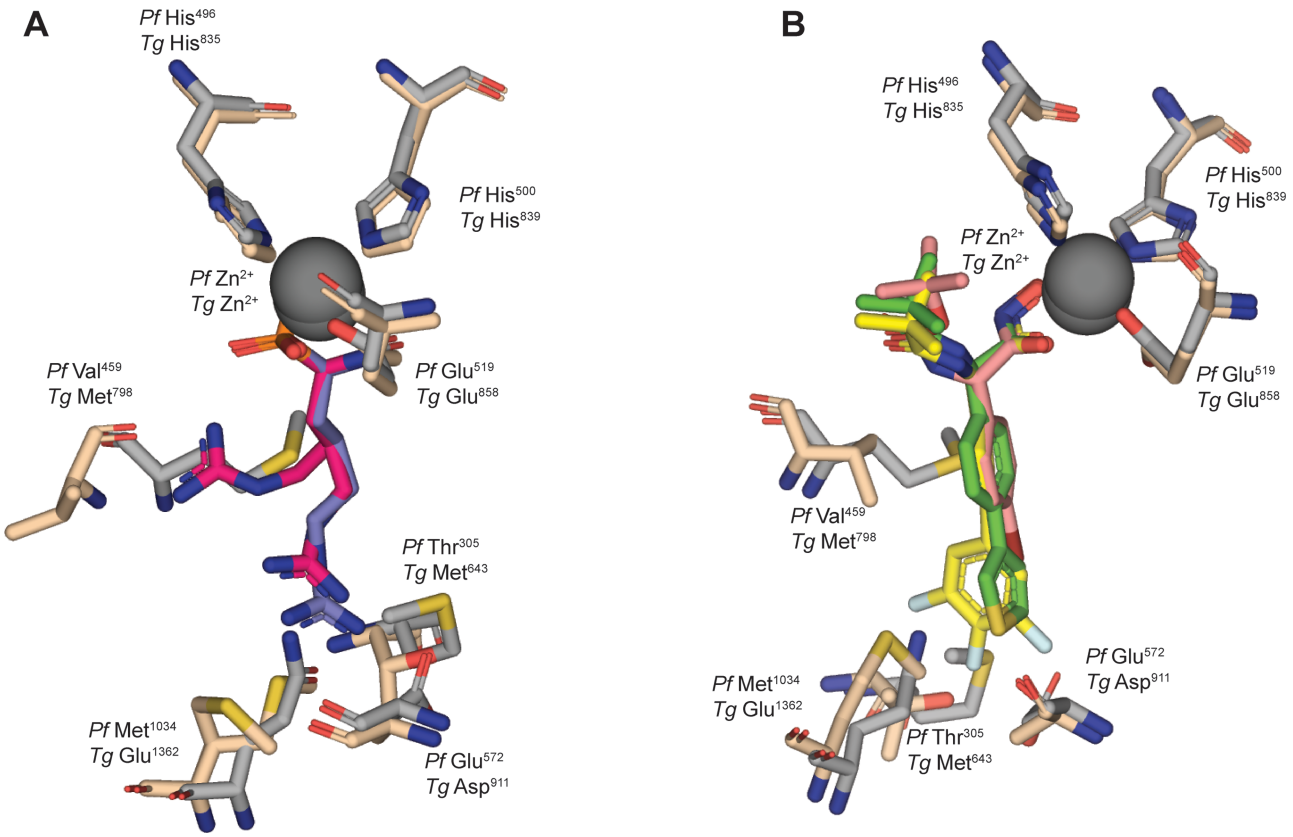


Figure 4. *Pf*A-M1 structures with bound inhibitors showing the superposition of *Tg*APNs.

The residues of *Pf*A-M1 (denoted *Pf*) is shown with wheat carbon atoms and TgAPN2 (denoted *Tg*) with grey carbon atoms. Protein sidechain numbers are indicated according to each structure. The inhibitors are also shown in stick with compounds **1** (magenta carbon atoms) and **2** (purple carbon) in (**A**) and the hydroxamic acid scaffolds in (**B**). Compounds **9b** (salmon carbon in stick) **10q** (green carbon in sticks) **10o** (yellow carbon in sticks) are shown.

9b. Interrogation of the superpositions of the structures onto PDB ID 4ZX3 provide some structural explanation as to the reason for the loss of potency (Figure 4B). Assuming a movement of the Met^{798}, **9b** can easily fit into the active site but appears to form very few interactions beyond the hydroxamic acid core. A similar situation can be seen with **10q**, however, the thiophenyl group is within bonding distance of Asp^{911} and Gln^{1362} (Figure 4B). The similar potency of **9b** and **10o** suggests there may be a water network that assists the positioning of **9b** that cannot be captured in a superposition, hence stabilising the inhibitor to provide an equivalent potency. The trifluorophenyl of **10o** has a likely clash with Asp^{911} that may not have the flexibility that is the equivalent Glu^{572} has in *Pf*A-M1 [41]. This likely accounts for the extremely poor inhibition by this compound (Figure 4B).

Discussion

Toxoplasmosis is a zoonotic infection caused by the apicomplexan parasite *T. gondii*. Infections are often asymptomatic and only present a risk to health for immunocompromised individuals and through congenital infection of a foetus. Current therapeutics require prolonged courses of drugs from weeks to more than a year and commonly have toxic side effects [42]. New medicines with enhanced efficacy as well as features that address the unique aspects of toxoplasmosis (e.g. eliminating tissue cysts, achieving therapeutic concentrations in the eye and brain) would greatly improve therapy. Conservation amongst apicomplexan protozoan parasites makes drug re-purposing an attractive option for screening for new anti-toxoplasmosis agents. In this study, we have investigated *Tg*APN2 — homologue of the malarial drug target, *Pf*A-M1 — to assess the capacity of current aminopeptidase inhibitors to target *T. gondii* homologues.

*Tg*APN2 is a member of the M1 aminopeptidase family of metallo-exopeptidases that has a membrane anchor, as well as large and potentially disordered N-terminal domain that precedes both the transmembrane anchor and the enzymatic domain. It appears from bioinformatic analysis available on ToxoDB, that *Tg*APN2 may be the only typical intra-cellular M1 aminopeptidase and given its sequence identity to *Pf*A-M1 (38% overall, 55% in the catalytic domain) may be a potential target for inhibition. A previous study has shown that recombinant *Tg*APN2 behaves as a typical metallo-exopeptidase with a substrate preference for a P1 Ala and Arg, and that it can be inhibited by generic aminopeptidases inhibitors bestatin and amastatin [14]. The authors of the biochemical study stated that *Tg*APN2 would share a similar homohexameric arrangement to that of other M1 aminopeptidases [14], which is in contradiction to the field as the M1 aminopeptidases or aminopeptidase N family are generally monomeric or dimeric in structure and function [1]. In this study, we produced the *Tg*APN2 ecto-enzyme and characterised its structure and function, producing comparative structure-activity relationships of *Pf*A-M1 purposed inhibitors.

We produced the recombinant ecto-domain of *Tg*APN2 and confirmed that it acted as a typical M1 aminopeptidase in both structure and function. It exhibited metal-dependent aminopeptidase activity with optimal activity at near neutral pH and formed the typical bacterial M1 aminopeptidase fold. One unexpected finding in the resolution of the structure was the presence of unexplained density coordinated to the catalytic zinc. We have modelled this as a second zinc ion coordinated to the active site zinc. The position of a possible second metal ion is similar to that of magnesium ions in the unbound structures of *Pf*A-M1, but the lack of significant concentration of metal ions in the crystallisation buffers suggests that the ion was scavenged. We attribute this finding to a crystallisation artefact and therefore do not assign a functional role however do suggest that the ability to coordinate a second metal ion to the active site ion is a possible reason that some M1 aminopeptidases are inhibited in the presence of excess zinc. Interrogating the role of the second metal-binding site with molecular biology techniques is challenging as the main bond positioning the ion is to the active site zinc, and residues that contribute to its stabilisation would be required for the normal architecture of the active site. Enzymatic analysis of natively purified *Tg*APN2 would help answer the question as to the number of zinc ions found in the catalytic domain of this protease.

The M1 aminopeptidase family is large and diverse, and often shows a variety of substrate preferences. Comparison of the extended substrate specificity of *Tg*APN2 with *Pf*A-M1 shows a significant difference in what P1 residue the enzymes prefer. *Pf*A-M1 has a much broader substrate specificity but a clear preference for hydrophobic natural amino acids (Met > Leu > Ala) [15]. This P1 preference is similar to the *Eimeria tenalla* APN1 [8] and human aminopeptidase N [43]. Similarly, extensive characterisation of the human ERAP1 family [44] has shown exclusive activity toward a P1 Leu and Met, but also differential activity depending on the length of the peptide substrate[45–48]. Longer peptide substrates can interact with an allosteric site that affects both specificity and enzymatic rates [47–49]. *Tg*APN2 prefers basic branched residues (Arg > Lys) and has a reduced activity toward Leu and Ala. The kinetic parameters (K_M, k_{cat}) confirm this preference but interestingly show that a P1 Arg has a reduced affinity for binding but a significantly increased turn-over rate. The bacterial homologues from *E. coli* (Arg > Ala > Lys [33]) and *Neisseria meningitidis* (Arg > Ala > Lys/Leu/Met [50]) share a preference for a P1 Arg, as does the human insulin receptor aminopeptidase (IRAP, Arg >>> Lys/Met/Leu, [44]) and leukotriene A4 hydrolase (LTA4H, Arg >>> Ala/Lys > Phe/Pro/Leu, [51]). Therefore, it can be hard to understand the capacity to re-purpose inhibitors from substrate profiling alone. The addition of comprehensive structure-activity relationship data to correlate the specificity profiles adds to our ability to predict cross-species homologue inhibitors, and potential off-target effects.

The structure of *Tg*APN2 provides clear insight to explain why compounds that potently inhibit PfA-M1 (**9b**, **10o** and **10q**) are only weak inhibitors of *Tg*APN2. The differences in the S1 pocket that produce a preference for a longer, charged P1 substrate also affect inhibitor interactions, with Asp^{911} and Gln^{1362} changing the bonding capacity of the pocket. Future design could use the hydroxamic acid zinc-binding group but should focus on capturing interactions as predicted for compounds **1** and **2** tested in this study, locking the inhibitor into hydrogen bonds with multiple residues within the S1 pocket. The structure produced in this study should contribute to the design of potent inhibitors of *Tg*APN2, as well as *Tg*APN3, that would be valuable tools to assess the suitability of these enzymes as drug targets.

Conclusions

*Tg*APN2 has been proposed as a potential new drug target for future anti-toxoplasmosis agents. In this study, we have produced and analysed the recombinant ecto-domain of *Tg*APN2 and shown that it has significant

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

differences to the apicomplexan homologue *Pf*A-M1. The value of these comparative structure-activity relationships should inform the design of potent inhibitors of the *T. gondii* M1 aminopeptidases. Furthermore, the *Tg*APN2 appears to be the main intracellular M1 aminopeptidase *in T. gondii* but does have an unusual N-terminal domain that was not characterised in this study. Investigation into the biological role(s) of this enzyme would also be of great benefit to assess its capacity to be a drug target.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the National Health and Medical Research Council (Synergy Grant 1185354 to SM). The Drag laboratory is supported by National Science Centre in Poland and 'TEAM/2017-4/32' project, which is carried out within the TEAM program of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund.

Open Access

Open access for this article was enabled by the participation of Monash University in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

Author Contributions

E.M. performed experiments, analysed structure, co-wrote paper; J.A., J.C.A., M.D., C.T.W. and K.S. performed experiments; N.D. and S.M. performed experiments, analysed structure, co-wrote paper and project concept.

Acknowledgements

We thank Professor Peter J Scammells from the Monash Institute of Pharmaceutical Sciences for the provision of compounds **9b**, **10o** and **10q**. We thank Daniel E. Williams for assistance in preparing the manuscript. We thank the Australian Synchrotron (MX-2) and the beamline scientists for beamtime (CAP9648), and the Monash Technology Research Platforms (Protein Production and Crystallization) for technical assistance.

Abbreviations

ACC, 7-amino-4-carbamoylmethylcoumarin; PDB, Protein Data Bank; SAR, structure activity relationship.

References

1 Drinkwater, N., Lee, J., Yang, W., Malcolm, T.R. and McGowan, S. (2017) M1 aminopeptidases as drug targets: broad applications or therapeutic niche? *FEBS J.* **284**, 1473–1488 https://doi.org/10.1111/febs.14009

2 Florent, I., Derhy, Z., Allary, M., Monsigny, M., Mayer, R. and Schrével, J. (1998) A *Plasmodium falciparum* aminopeptidase gene belonging to the M1 family of zinc-metallopeptidases is expressed in erythrocytic stages. *Mol. Biochem. Parasitol.* **97**, 149–160 https://doi.org/10.1016/S0166-6851(98)00143-1

3 Harbut, M.B., Velmourougane, G., Dalal, S., Reiss, G., Whisstock, J.C., Onder, O. et al. (2011) Bestatin-based chemical biology strategy reveals distinct roles for malaria M1- and M17-family aminopeptidases. *Proc. Natl Acad. Sci. U.S.A.* **108**, E526–E534 https://doi.org/10.1073/pnas.1105601108

4 McGowan, S., Porter, C.J., Lowther, J., Stack, C.M., Golding, S.J., Skinner-Adams, T.S. et al. (2009) Structural basis for the inhibition of the essential *Plasmodium falciparum* M1 neutral aminopeptidase. *Proc. Natl Acad. Sci. U.S.A.* **106**, 2537–2542 https://doi.org/10.1073/pnas.0807398106

5 Weiss, L.M. and Dubey, J.P. (2009) Toxoplasmosis: a history of clinical observations. *Int. J. Parasitol.* **39**, 895–901 https://doi.org/10.1016/j.ijpara.2009.02.004

6 Ohrvik, V.E. and Witthoft, C.M. (2011) Human folate bioavailability. *Nutrients* **3**, 475–490 https://doi.org/10.3390/nu3040475

7 Iaccheri, B., Fiore, T., Papadaki, T., Androudi, S., Janjua, S., Bhaila, I. et al. (2008) Adverse drug reactions to treatments for ocular toxoplasmosis: a retrospective chart review. *Clin. Ther.* **30**, 2069–2074 https://doi.org/10.1016/j.clinthera.2008.10.021

8 Gras, S., Byzia, A., Gilbert, F.B., McGowan, S., Drag, M., Silvestre, A. et al. (2014) Aminopeptidase N1 (EtAPN1), an M1 metalloprotease of the apicomplexan parasite *Eimeria tenella*, participates in parasite development. *Eukaryot. Cell* **13**, 884–895 https://doi.org/10.1128/EC.00062-14

9 Lu, W., Lu, C., Zhang, Q., Cao, S., Zhang, Z., Jia, H. et al. (2020) Localization and enzyme kinetics of aminopeptidase N3 from *Toxoplasma gondii*. *Parasitol. Res.* **119**, 357–364 https://doi.org/10.1007/s00436-019-06512-6

10 Berthonneau, J., Rodier, M.H., El Moudni, B. and Jacquemin, J.L. (2000) *Toxoplasma gondii*: purification and characterization of an immunogenic metallopeptidase. *Exp. Parasitol.* **95**, 158–162 https://doi.org/10.1006/expr.2000.4524

11 Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X. et al. (2008) ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Res.* **36**, D553–D556 https://doi.org/10.1093/nar/gkm981

12 Fritz, H.M., Buchholz, K.R., Chen, X., Durbin-Johnson, B., Rocke, D.M., Conrad, P.A. et al. (2012) Transcriptomic analysis of toxoplasma development reveals many novel functions and structures specific to sporozoites and oocysts. *PLoS ONE* **7**, e29998 https://doi.org/10.1371/journal.pone.0029998

13 Sidik, S.M., Huet, D., Ganesan, S.M., Huynh, M.H., Wang, T., Nasamu, A.S. et al. (2016) A genome-wide CRISPR screen in toxoplasma identifies essential apicomplexan genes. *Cell* **166**, 1423–1435 e1412 https://doi.org/10.1016/j.cell.2016.08.019
14 Li, Q., Jia, H., Cao, S., Zhang, Z., Zheng, J. and Zhang, Y. (2017) Biochemical characterization of aminopeptidase N2 from *Toxoplasma gondii*. *J. Vet. Med. Sci.* **79**, 1404–1411 https://doi.org/10.1292/jvms.17-0119
15 Poreba, M., McGowan, S., Skinner-Adams, T.S., Trenholme, K.R., Gardiner, D.L., Whisstock, J.C. et al. (2012) Fingerprinting the substrate specificity of M1 and M17 aminopeptidases of human malaria, *Plasmodium falciparum*. *PLoS ONE* **7**, e31938 https://doi.org/10.1371/journal.pone.0031938
16 Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 https://doi.org/10.1016/j.pep.2005.01.016
17 Vinh, N.B., Drinkwater, N., Malcolm, T.R., Kassiou, M., Lucantoni, L., Grin, P.M. et al. (2019) Hydroxamic acid inhibitors provide cross-species inhibition of *Plasmodium* M1 and M17 aminopeptidases. *J. Med. Chem.* **62**, 622–640 https://doi.org/10.1021/acs.jmedchem.8b01310
18 Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 https://doi.org/10.1107/S0907444909047337
19 Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292 https://doi.org/10.1107/S090744491003982X
20 Collaborative Computational Project Network. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 https://doi.org/10.1107/S0907444994003112
21 McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. and Read, R.J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 458–464 https://doi.org/10.1107/S0907444905001617
22 Bompard-Gilles, C., Villeret, V., Davies, G.J., Fanuel, L., Joris, B., Frere, J.M. et al. (2000) A new variant of the Ntn hydrolase fold revealed by the crystal structure of L-aminopeptidase D-Ala-esterase/amidase from *Ochrobactrum anthropi*. *Structure* **8**, 153–162 https://doi.org/10.1016/S0969-2126(00)00091-5
23 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 https://doi.org/10.1107/S0907444909052925
24 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 https://doi.org/10.1107/S0907444904019158
25 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 https://doi.org/10.1016/j.jmb.2007.05.022
26 Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2256–2268 https://doi.org/10.1107/S0907444904026460
27 Heinig, M. and Frishman, D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–W502 https://doi.org/10.1093/nar/gkh429
28 McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N. et al. (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 https://doi.org/10.1093/nar/gkt376
29 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 https://doi.org/10.1038/msb.2011.75
30 Oates, M.E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M.J., Xue, B. et al. (2013) D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508–D516 https://doi.org/10.1093/nar/gks1226
31 Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. and Godzik, A. (2005) FFAS03: a server for profile–profile sequence alignments. *Nucleic Acids Res.* **33**, W284–W288 https://doi.org/10.1093/nar/gki418
32 Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 https://doi.org/10.1093/nar/25.17.3389
33 Ito, K., Nakajima, Y., Onohara, Y., Takeo, M., Nakashima, K., Matsubara, F. et al. (2006) Crystal structure of aminopeptidase N (proteobacteria alanyl aminopeptidase) from *Escherichia coli* and conformational change of methionine 260 involved in substrate recognition. *J.Biol. Chem.* **281**, 33664–33676 https://doi.org/10.1074/jbc.M605203200
34 Santiago, C., Mudgal, G., Reguera, J., Recacha, R., Albrecht, S., Enjuanes, L. et al. (2017) Allosteric inhibition of aminopeptidase N functions related to tumor growth and virus infection. *Sci. Rep.* **7**, 46045 https://doi.org/10.1038/srep46045
35 Kochan, G., Krojer, T., Harvey, D., Fischer, R., Chen, L., Vollmar, M. et al. (2011) Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc. Natl Acad. Sci. U.S.A.* **108**, 7745–7750 https://doi.org/10.1073/pnas.1101262108
36 Addlagatta, A., Gay, L. and Matthews, B.W. (2006) Structure of aminopeptidase N from *Escherichia coli* suggests a compartmentalized, gated active site. *Proc. Natl Acad. Sci. U.S.A.* **103**, 13339–13344 https://doi.org/10.1073/pnas.0606167103
37 Nocek, B., Mulligan, R., Bargassa, M., Collart, F. and Joachimiak, A. (2008) Crystal structure of aminopeptidase N from human pathogen *Neisseria meningitidis*. *Proteins* **70**, 273–279 https://doi.org/10.1002/prot.21276
38 Kannan Sivaraman, K., Paiardini, A., Sieńczyk, M., Ruggeri, C., Oellig, C.A., Dalton, J.P. et al. (2013) Synthesis and structure-activity relationships of phosphonic arginine mimetics as inhibitors of the M1 and M17 aminopeptidases from *Plasmodium falciparum*. *J. Med. Chem.* **56**, 5213–5217 https://doi.org/10.1021/jm4005972
39 Mistry, S.N., Drinkwater, N., Ruggeri, C., Sivaraman, K.K., Loganathan, S., Fletcher, S. et al. (2014) Two-pronged attack: dual inhibition of *Plasmodium falciparum* M1 and M17 metalloaminopeptidases by a novel series of hydroxamic acid-based inhibitors. *J. Med. Chem.* **57**, 9168–9183 https://doi.org/10.1021/jm501323a
40 Drinkwater, N., Vinh, N.B., Mistry, S.N., Bamert, R.S., Ruggeri, C., Holleran, J.P. et al. (2016) Potent dual inhibitors of *Plasmodium falciparum* M1 and M17 aminopeptidases through optimization of S1 pocket interactions. *Eur. J. Med. Chem.* **110**, 43–64 https://doi.org/10.1016/j.ejmech.2016.01.015
41 Velmourougane, G., Harbut, M.B., Dalal, S., McGowan, S., Oellig, C.A., Meinhardt, N. et al. (2011) Synthesis of new (-)-bestatin-based inhibitor libraries reveals a novel binding mode in the S1 pocket of the essential malaria M1 metalloaminopeptidase. *J. Med. Chem.* **54**, 1655–1666 https://doi.org/10.1021/jm101227t

Biochemical Journal (2020) **477** 3819–3832
https://doi.org/10.1042/BCJ20200569

42 Alday, P.H. and Doggett, J.S. (2017) Drugs in development for toxoplasmosis: advances, challenges, and current status. *Drug Design Dev. Ther.* **11**, 273–293 https://doi.org/10.2147/DDDT.S60973
43 Drag, M., Bogyo, M., Ellman, J.A. and Salvesen, G.S. (2010) Aminopeptidase fingerprints, an integrated approach for identification of good substrates and optimal inhibitors. *J. Biol. Chem.* **285**, 3310–3318 https://doi.org/10.1074/jbc.M109.060418
44 Zervoudi, E., Papakyriakou, A., Georgiadou, D., Evnouchidou, I., Gajda, A., Poreba, M. et al. (2011) Probing the S1 specificity pocket of the aminopeptidases that generate antigenic peptides. *Biochem. J.* **435**, 411–420 https://doi.org/10.1042/BJ20102049
45 Evnouchidou, I., Momburg, F., Papakyriakou, A., Chroni, A., Leondiadis, L., Chang, S.C. et al. (2008) The internal sequence of the peptide-substrate determines its N-terminus trimming by ERAP1. *PLoS ONE* **3**, e3658 https://doi.org/10.1371/journal.pone.0003658
46 Hearn, A., York, I.A. and Rock, K.L. (2009) The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum. *J. Immunol.* **183**, 5526–5536 https://doi.org/10.4049/jimmunol.0803663
47 Chang, S.C., Momburg, F., Bhutani, N. and Goldberg, A.L. (2005) The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. *Proc. Natl Acad. Sci. U.S.A.* **102**, 17107–17112 https://doi.org/10.1073/pnas.0500721102
48 Nguyen, T.T., Chang, S.C., Evnouchidou, I., York, I.A., Zikos, C., Rock, K.L. et al. (2011) Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat. Struct. Mol. Biol.* **18**, 604–613 https://doi.org/10.1038/nsmb.2021
49 Gandhi, A., Lakshminarasimhan, D., Sun, Y. and Guo, H.C. (2011) Structural insights into the molecular ruler mechanism of the endoplasmic reticulum aminopeptidase ERAP1. *Sci. Rep.* **1**, 186 https://doi.org/10.1038/srep00186
50 Weglarz-Tomczak, E., Poreba, M., Byzia, A., Berlicki, L., Nocek, B., Mulligan, R. et al. (2013) An integrated approach to the ligand binding specificity of *Neisseria meningitidis* M1 alanine aminopeptidase by fluorogenic substrate profiling, inhibitory studies and molecular modeling. *Biochimie* **95**, 419–428 https://doi.org/10.1016/j.biochi.2012.10.018
51 Byzia, A., Haeggström, J.Z., Salvesen, G.S. and Drag, M. (2014) A remarkable activity of human leukotriene A4 hydrolase (LTA4H) toward unnatural amino acids. *Amino Acids* **46**, 1313–1320 https://doi.org/10.1007/s00726-014-1694-2