Crystal structures of angiotensin converting enzyme from Anopheles gambiae in its native form and with a bound inhibitor

John S. Cashman[1], Gyles E. Cozier[1], Charlotte Harrison[1],
R. Elwyn Isaac[2] and K. Ravi Acharya[1,*]

[1]Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, U.K.

[2]School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, U.K.

[*]Correspondence: K. Ravi Acharya (bsskra@bath.ac.uk)

Abstract

The mosquitoes of the Anopheles and Aedes genus are some of the most deadly insects to humans because of their effectiveness as vectors of malaria and a range of arboviruses, including yellow fever, dengue, chikungunya, West Nile and zika. The use of insecticides from different chemical classes is a key component of the integrated strategy against An. gambiae and Ae. aegypti, but the problem of insecticide resistance means that new compounds with different modes of action are urgently needed to replace chemicals that fail to control resistant mosquito populations. We have previously shown that feeding inhibitors of peptidyl dipeptidase A to both An. gambiae and Ae. aegypti mosquito larvae leads to stunted growth and mortality. However, these compounds were designed to inhibit the mammalian form of the enzyme (angiotensin converting enzyme, ACE) and hence can have lower potency and lack selectivity as inhibitors of the insect peptidase. Thus, for the development of inhibitors of practical value in killing mosquito larvae, it is important to design new compounds that are both potent and highly selective. Here we report the first structures of AnoACE2 from An. gambiae in its native form and with a bound human ACE inhibitor fosinoprilat. Comparison of these structures with human ACE (sACE) and an insect ACE homologue from Drosophila melanogaster (AnCE) revealed that AnoACE2 structure is more similar to AnCE. In addition, important elements that differ in these structures provide information that could potentially be utilised in the design of chemical leads for selective mosquitocide development.

Introduction

Mosquitoes that flourish in tropical climates are responsible for transmitting various human viral and parasitic diseases. After a blood-meal, females lay eggs in water where the juveniles, called larvae, feed on particulate organic material. Within a few days, adults emerge from the water to continue the lifecycle. Disease transmission occurs when the adult female has multiple blood meals from different hosts. Controlling mosquito populations using insecticides is critical for reducing disease transmission but mosquitoes are showing increasing levels of resistance to existing chemicals and new insecticide classes are now needed [1-4].

In our search for a new class of insecticide to disrupt the mosquito life-cycle by targeting larval stages of the malaria and arbovirus vectors (Anopheles gambiae and Aedes aegypti respectively), we have identified an insect metallopeptidase (peptidyl dipeptidase-A), which is related to the mammalian blood pressure regulator angiotensin-1 converting enzyme (ACE), as an important molecular target [5, 6]. The insect enzyme (iACE) is a promiscuous peptidase that typically cleaves the penultimate C-terminal peptide bond of substrates and has been linked to the metabolism of peptide hormones involved in insect reproduction [6-8]. ACE inhibitors fed to adult anopheline mosquitoes (Anopheles stephensi and Anopheles gambiae) cause a drastic reduction in the numbers of eggs laid [9, 10]. In addition to their effects on adults, we have recently shown that ACE inhibitors are lethal to the developing larvae of both Anopheles gambiae and Aedes aegypti [11]. The ability of these compounds to reduce adult fecundity and to kill juvenile mosquitoes makes these enzymes attractive targets for the rational design of insect-selective inhibitors for the control of mosquito populations.

Identification of the insect ACE gene family has expanded greatly in mosquitoes (in the Anopheles gambiae genome) to nine (AnoACEs 1–9) and there is evidence that four of them (AnoACE2, AnoACE3, AnoACE7 and AnoACE9) are up-regulated after blood meals prior to egg laying and expressed during larval development [12]. AnoACE7 and AnoACE9 are predicted to have a C-terminal hydrophobic sequence that can form a membrane anchor [12]. Since the majority of the peptidyl dipeptidase activity in homogenates of larvae is soluble, it is likely that the activity results from expression of one or both of AnoACE2 and AnoACE3 [11]. The amino acid sequence of both enzymes is highly conserved in A. gambiae [13] and the larval peptidase activity is inhibited by a variety of synthetic and natural compounds [11]. Of the synthetic inhibitors, captopril is the most potent (IC_{50} 17 nM), followed by fosinoprilat (IC_{50} 35 nM). Fosinopril, the esterified prodrug form of fosinoprilat, gives an expected considerably higher IC_{50} value (326 nM) compared to the free acid. Captopril is equally potent at inhibiting mammalian and iACEs, presumably because the relatively small size of the inhibitor allows easy access to the active sites of most enzymes in this family of

proteases. In addition, studies on the toxicity of the most potent inhibitor captopril and the pro-drug fosinopril on survival of 1st, 2nd and 3rd stage larvae of Anopheles gambiae showed that within 24 h captopril had killed >90 % of the early instar. Mortality was consistently high (>80 %) within 24 h of exposure of 1st, 2nd and 3rd instars of Anopheles gambiae to fosinopril [11]. These experimental data further confirmed the potential of AnoACE inhibitors as mosquito larvicides. However, all the compounds tested were established human ACE inhibitors and were used as proof of concept. AnoACE2 specific inhibitors will need to be developed for practical use.

To determine the unique molecular features of AnoACE2, we now report the first crystal structure of AnoACE2 with the native enzyme at 2.2 Å resolution. This structure can provide the molecular basis for specific inhibition and form the basis for future inhibitor design. Comparison with the active site of the human ACE [known as somatic ACE (sACE) found in somatic tissues and with two domains (N and C) each containing an active site] (14-16) and iACE from Drosophila melanogaster (AnCE) (17), highlights unique residues and subtle differences in the orientation of the side chains at different binding pockets in the active site of AnoACE2. In addition we present AnoACE2 in complex with fosinoprilat (2.5 Å) which confirms that ACE inhibitors bind to AnoACE2 in the typical binding site seen for other ACE homologues and explains the high affinity of fosinoprilat for AnoACE2.

Materials and methods

AnoACE2 cloning

AnoACE2 (UniProt A0NFU8) is a 638 residue protein containing a 26 residue signal sequence. Residues 27-638 were cloned into pOPINE pOPINTTGneo, pOPINHBM and pOPINHA using In-Fusion Cloning technology (Takara Bio Europe) according to the manufacturer instructions. The native signal sequence (residues 1-26) was replaced with alternative signal sequences already present in pOPINTTGneo, pOPINHBM and pOPINHA vectors, to enable secretion of the expressed protein into the cell culture medium.

Protein expression and purification

Sf9 cells were maintained in suspension in Sf-900™ II serum free medium (Thermo Scientific) at 27˚C with shaking at 140 rpm. Cells were passaged every 3-4 days according to the manufacturer instructions. Baculovirus stocks were prepared by co-transfecting cells with Prof. Ian Jones bacmid and AnoACE2 expression vectors using FuGENE® HD transfection reagent (Promega). To screen for protein expression, 3 mL cultures were seeded with 1 x 10^6 viable cells/mL and infected with 3 or

30µL P1 virus. Expression was optimised using time-course experiments and large-scale expression was performed in 2-4L cultures seeded with 1 x 10^6 viable cells/mL and infected with 3mL/L P2 virus. Cultures were incubated at 27˚C with shaking at 140 rpm for 72-90 hours and AnoACE2 was purified from 500 mL batches of protein-containing cell culture medium. Cultures were harvested and filtered using Sartoclear Dynamics® Lab V devices (Sartorius). The clarified medium was loaded directly onto a 5 mL HisTrap excel column (GE Life Sciences) and non-specifically bound proteins were removed by washing with 100 mL binding buffer (50 mM HEPES; 500 mM NaCl; pH 7.4), followed by 75 mL 5 % elution buffer (50 mM HEPES; 500 mM NaCl; 500 mM imidazole; pH 7.4). Elution was effected with a single step to 100 % elution buffer for 50 mL. Peak fractions were pooled, diluted 1:4 with binding buffer (to reduce the imidazole concentration to approximately 100 mM), concentrated to a maximum volume of 750 µL and loaded onto a Superdex 10/300 column (GE Life Sciences) equilibrated with S200 buffer (50 mM HEPES; 150 mM NaCl; pH 7.4) to remove residual contaminants. Protein expression and purification were monitored by SDS PAGE and western blot with monoclonal mouse anti-his tag IgG (R&D Systems, MAB050) using a 1:5,000 dilution of the 0.5 mg/mL stock in blocking solution.

Activity and inhibition assays

Enzyme assays were performed using the internally quenched fluorogenic peptide Abz-FRK(Dnp)-P (Enzo Life Sciences Ltd, Exeter, U.K.) as substrate for 25 ng of AnoACE2 in 200 µl of 100 mM HEPES buffer pH 7.5, 50 mM NaCl and 10 µM $ZnCl_2$. The reaction was started by adding 2 µl of 5 mM Abz-FRK(Dnp)-P in dimethyl sulfoxide to the buffer in wells of a 96-well black plastic plate (Corning Life Sciences, High Wycombe, U.K.) and the release of fluorescence on hydrolysis of the substrate was monitored continuously at 20ºC using a FLUOstar Omega (BMG Labtech GmbH, Offenburg, Germany) with λ_{ex} at 340 nm and λ_{em} set at 430 nm). For studying the effect of ACE inhibitors, the enzyme in the assay buffer was pre-incubated with inhibitor for 10 min prior to the addition of substrate. IC_{50} values and 95% confidence intervals (CI) for each inhibitor were determined using the non-linear regression curve fitting software [log(inhibitor) vs response] of GraphPad Prism 7.01 GraphPad (Software, San Diego, CA, U.S.A.).

Mass spectrometry

20 µg samples of AnoACE2 were digested in solution using chymotrypsin and trypsin to increase sequence coverage. Proteins were reduced with 5 mM DTT for 30 min and alkylated with 20 mM iodoacetamide for 30 min. The samples were desalted and concentrated using chloroform-methanol concentration and digestion was performed in 1 M urea, Tris pH 7.8 using a 1:50 (enzyme:protein)

ratio. The trypsin digestions were performed at 37°C and chymotrypsin digestion was carried out at 25°C overnight. 300 nL of peptides were injected in an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Fisher Scientific) and analysed using a 1 h gradient. Raw data were processed using Byonic™ software (Protein Metrics, CA).

X-ray crystallographic analysis

An ideal concentration range for crystallisation was determined using the pre-crystallisation test, as described by Hampton Research. The sitting drop vapor diffusion method with several commercially available screens (Molecular Dimensions) was used to determine crystallisation conditions for AnoACE2. One condition which yielded crystals in the presence of 55% v/v polypropylene glycol 400, was optimised. For crystallisation with fosinopril, AnoACE2 was concentrated to between 5 and 10 mg/mL and mixed with 20 mM inhibitor at a ratio of 4:1 respectively. Complexes were allowed to form at room temperature for 1 hour before adding to screens.

Native AnoACE2 crystals were ultimately obtained using 10 mg/mL protein in 1:1 400 nL drops with 36% v/v polypropylene glycol, 10 mM ammonium sulfate, 30 mM sodium formate, 10 mM Tris pH 7.8, 0.3% w/v γ-PGA (Na^+ form, LM), 0.3% w/v PEG 20,000. Fosinopril-bound AnoACE2 crystals were grown in the same condition but containing a slightly higher polypropylene glycol concentration (43.2% v/v). Although fosinopril was used for the co-crystallisation, these prodrugs hydrolyse to their active form (in this case fosinoprilat) under the crystallisation conditions.

X-ray diffraction data were collected on station i04 at the Diamond Light Source (Didcot, UK). Crystals were kept at constant temperature (100K) under nitrogen stream during data collection. Images were collected using a PILATUS3 6M detector (Dectris, Switzerland). Raw data images were indexed and integrated with DIALS [14], and then scaled using AIMLESS [15] from the CCP4 suite [16]. Initial phases for the native structure were obtained by molecular replacement with PHASER [17] using a model AnoACE2 structure based on cACE [11], and the solved native structure was used as the search model for the fosinoprilat AnoACE2 complex. Further refinement was initially carried out using REFMAC5 [18] followed by PHENIX [19], with COOT [20] used for rounds of manual model building. Inhibitor, zinc ion and water molecules were added based on electron density in the Fo-Fc Fourier difference map. MolProbity [21] was used to help validate the structures. Crystallographic data statistics are summarized in Table 2. All figures showing the crystal structures were generated using CCP4mg [22] and schematic binding interactions are displayed using Ligplot$^+$ [23].

Results

Expression, purification of AnoACE2

AnoACE2 cDNA was cloned into various pOPIN vectors using the In-Fusion® HD Cloning Plus kit (Takara Clontech). The highest protein yields were attained from Sf9 cells using the pOPINHBM-AnoACE2 construct. The protein was estimated to be 20% glycosylated by weight, as judged by mobility on SDS polyacrylamide gels. The secreted protein following expression, was purified directly by immobilised metal affinity and size exclusion chromatography. This method reliably yields 2.1 mg protein per L of culture with >95 % purity.

AnoACE2 activity

The peptidyl dipeptidase activity of the purified mosquito protein was confirmed by using the FRET substrate Abz-FRK(Dnp)-P and by titrating various human ACE inhibitors (known ACE inhibitors have been used as proof of concept for AnoACE2 inhibition) to generate IC_{50} data (Table 1). Four inhibitors tested showed high potency as revealed by the low nM values, with lisinopril being the weakest, although still a high affinity inhibitor with an IC_{50} of 16 nM. Captopril showed a more potent IC_{50} of 5.2 nM. In contrast to the activity assay previously performed using the soluble fraction of homogenised mosquito larvae which showed that captopril was the most potent of the inhibitors tested [11], the assay using recombinant AnoACE2 showed that fosinoprilat had a slightly more potent IC_{50} (1.5 nM) than that observed for captopril. Enalaprilat had a similarly potent IC_{50} of 2.0 nM. γ-PGA, which is present in the crystallisation buffer, was found to interact with native AnoACE2 in the active site (see below). It inhibited the enzyme activity with an IC_{50} of 0.2 % w/v, which is below the 0.3 % w/v concentration used in the crystallization buffer.

AnoACE2 glycosylation

Glycosylation of human ACE proteins has been shown to be important for peptidase activity [24] and this is likely to be the case for iACEs. Sequence analysis using Prot pi to identify the number of asparagine residues present within the glycosylation consensus sequence Asn–X–Ser/Thr (where X is any residue apart from proline) showed that AnoACE2 has six potential N-glycosylation sites (residues 69, 74, 106, 187, 212 and 327) (Fig. 1). We used mass spectrometry to determine which of these sites had N-glycosylations. Digestion of AnoACE2 with chymotrypsin and trypsin allowed for 44.9 % and 45.3 % sequence coverage in the MS/MS spectra, respectively (Fig. 1A). Using both enzymes gave a total of 65.5 % sequence coverage, which included all potential N-glycosylation sites. The data revealed that five out of the six potential sites were glycosylated (residues 69, 106, 187, 212

and 327, but not residue 74) (Fig. 1B,C). This information will be useful for any future work to examine which, if any, glycosylation sites are required for enzyme activity.

Crystallisation of AnoACE2

Crystal growth was initially tested using several commercially available crystallization screens (Molecular Dimensions). One condition that yielded long, rod-like crystals (55 % v/v polypropylene glycol 400) was optimised through additive and seeding experiments, and adjustments to the protein and precipitant concentrations, which resulted in significant improvements with the quality of crystals. X-ray diffraction data were collected at Diamond Light Source beamline i04 for both native AnoACE2 and in complex with fosinoprilat.

Overall structure of AnoACE2

Both native AnoACE2 and the fosinoprilat-AnoACE2 complex structures belong to $P6_222$ space group with a monomer in the asymmetric unit (Fig. 2A,B). The native AnoACE2 was determined to 2.2 Å with a conventional R-factor (R_{cryst}) of 0.194 and a free R-factor (R_{free}) of 0.241 (Table 2), whereas fosinoprilat-AnoACE2 complex was refined to 2.5 Å with an R_{cryst} of 0.294 and R_{free} of 0.337 respectively. The R-factors for the fosinoprilat-AnoACE2 complex indicate that the X-ray diffraction data from this crystal is not of the same quality as for the native structure. Careful examination of the data with tests for twinning, alternative space groups etc were performed, but no twinning was detected, and alternative space groups such as $P3_212$ produced significantly worse R-factors and electron density maps. The lower quality data for the fosinoprilat-AnoACE2 complex is most likely due to just a lower quality crystal. The vast majority of AnoACE2 crystals tested showed low resolution (less than 4 Å) diffraction, even after extensive screening for better conditions. Many crystals of both native and ligand bound had to be tested to obtain the data presented here. This suggests that binding of fosinoprilat probably does not cause a reduction in crystal quality. To produce the best model from the fosinoprilat-AnoACE2 complex data, extensive rounds of manual rebuilding and refinement were performed, and the 'Optimize x-ray/stereochemistry weight' option in Phenix was selected for the refinement. While this increased the restraints, giving the effect of low bond length and angle root-mean-square deviation values, it did give the best model structure overall. Even so, the Ramachandran statistics are not as good as for the native structure, although they are still 99.8% within allowed regions. Analysis of the Ramachandran plot using the program MolProbity [21] showed that 98.2 and 92.6% of the residues lie in the most favourable region and 1.8 and 7.2 % are in the additionally allowed regions for the native and fosinoprilat complex respectively. Electron density

for residues 36-630 of the expressed protein was visible for both structures, although residues 97-106 of the fosinoprilat complex were too disordered to model.

The overall topology of AnoACE2 is similar to that of previously determined structures of human ACE domains [25-27] and iACE from Drosophila melanogaster (AnCE) [28, 29], arranged predominantly in a helical conformation of two subdomains (Fig. 2A,B). The conserved catalytic Zn^{2+} ion is buried deep inside the active site of AnoACE2 molecule bound to the typical His, His, Glu triad (His383, His387 and Glu411 in AnoACE2). In addition, the catalytic glutamate residue (Glu384 in AnoACE2) is conserved and adopts an equivalent position and orientation to that seen in other ACE analogues. It has previously been observed that sections of the first 100 residues form a 'lid-like' structure that has been proposed to be involved in ligand entry and exit from the binding site. This section is more flexible, and often less well defined than other regions of the structure. Analysis of the temperature factors (B-factors) indicates this increased flexibility is also present for AnoACE2 (residues 36-56 and 86-119 that comprise the 'lid-like' region have an average B-factor of 93.7 $Å^2$, compared to 53.9 $Å^2$ for the rest of the protein).

Glycosylation observed in AnoACE2 structures

The mass-spectrometry results described above indicated that the AnoACE2 molecule contains five of the six sequence predicted N-glycosylation sites. Analysis of the electron density maps of the higher resolution native AnoACE2 structure confirmed these results showing clear evidence for glycosylation on residues Asn69, Asn106, Asn187, Asn212 and Asn327, albeit the density was only clear enough to model the glycosylation on residues 69 and 327 (Fig. 3). The glycosylation on Asn69 showed the typical NAG-NAG-BMA chain. The Asn-linked NAG has a hydrogen bond with Thr71, and there is no evidence of a 1-6 linked FUC sugar. The second NAG and BMA sugars are located adjacent to symmetry related AnoACE2 molecules that help to fix the orientation of the sugars. In addition, the second NAG molecule has a hydrophobic interaction with Phe73. There are patches of electron density that indicate the glycosylation chain continues in both the 3- and 6-positions of the BMA sugar. Two NAG sugars could be modelled onto Asn327, with the Asn-linked NAG showing hydrogen bond interactions with residues Asn566 and Lys569, and some electron density for a 1-6 linked FUC sugar. The structure showed that residue Asn74 is slightly buried and is located adjacent to the glycosylated Asn69, meaning glycosylation on Asn74 would be sterically hindered. Instead the electron density maps indicating a water molecule coordinated. The lower resolution fosinoprilat complex structure showed similar results, although Asn106 is located in the section not modelled due to poor electron density.

Binding site of native AnoACE2

The large range of structures of ACE homologues have shown that their binding sites are not empty in the absence of added ligand, or solely occupied by the ligand if it is small [25, 26]. In general, the active site zinc ion will bind at least water molecules to complete its coordination sphere, and the C-terminal carboxy binding pocket always scavenges something from the expression, purification or crystallisation conditions. It was therefore not surprising that analysis of the structure of native AnoACE2 showed electron density in these regions. However, the extent of ligand observed, stretching from the S_2' to S_1 subsites, was unexpected (Fig. 4A,B). The best explanation for this region of electron density is a two residue poly-γ-glutamic acid molecule (γ-PGA) from the crystallisation condition, and this is consistent with the 0.2 % w/v IC_{50} inhibition of AnoACE2 by γ-PGA described above. Two conformations of this flexible ligand were modelled into the observed electron density (RSCC values of 0.93 for both conformations), although it appears that there may be additional conformations or even some instances of alternative ligands.

Ligplot (Fig. 5A,B) and schematic (Fig. 6A,B) representations of the two γ-PGA conformations show the interactions involved in binding to AnoACE2. The carboxy terminus of the first glutamic acid residue of both conformations of γ-PGA is bound in the typical manner to how a peptide binds to ACE homologues. There are direct hydrogen bonds to Gln281, Lys511 and Tyr520, and a second water mediated interaction with Lys511. The central section of the γ-PGA dimer has similarities to a peptide bond. In one conformation (Fig. 5A and 6A) there is a series of hydrophobic interactions with His353, His513 and Tyr523, as well as hydrogen bonds with Glu384 and the backbone of Ala354, whereas in the second conformation there is only one strong interaction, a hydrogen bond with His513 (Fig. 5B and 6B). In both conformations, what is essentially a carboxy terminus of the second glutamic acid molecule forms a bidentate interaction with the active site zinc ion, as well as a hydrogen bond with Tyr523 and interactions through a single water molecule with Ala356 and Glu411. Finally, the side chain of the second glutamic acid of γ-PGA forms interactions through a single water molecule with residues His353 and Tyr512 in one conformation (Fig. 5A and 6A), and a series of hydrophobic interactions with His353 and Tyr512 in the second conformation (Fig. 5B and 6B). γ-PGA is a new class of peptide mimic identified to bind to ACE homologues, and we have shown it inhibits AnoACE2. This backbone could be used as a starting point to design more potent and selective inhibitors.

Fosinoprilat complex with AnoACE2

AnoACE2 was also crystallized in the presence of fosinopril (the prodrug of fosinoprilat, a known inhibitor of mammalian ACE which also inhibits AnoACE2), which under crystallisation conditions

hydrolyses to fosinoprilat. Examination of the binding site revealed a large region of electron density located adjacent to the zinc ion of the complex structure (Fig. 4C,D). Due to the native structure showing that γ-PGA from the crystallization conditions can bind in the active site in the same location as expected for fosinoprilat, it is important to show that it is not γ-PGA bound, and confirm that it is indeed fosinoprilat. Firstly, the shape of the electron density observed is different from that seen for the native structure, in particular in the S_2' subsite and adjacent to the zinc ion. In addition, the electron density in the omit map next to the zinc ion is much stronger, being still clearly apparent at 7σ level. This is more consistent with the phosphinic acid group of fosinoprilat rather than a carboxylic acid group of γ-PGA. Finally, the placement of fosinoprilat fits both the mFo-DFc omit and 2mFo-DFc electron density maps well (apart from the flexible phenyl group as described below), whereas refinement with γ-PGA does not match the electron density fully and leaves large parts of it unexplained (positive density in the difference map). The final model with fosinoprilat modelled gave an RSCC value for the ligand of 0.92.

The inhibitor molecule is buried deep inside the active site pocket (Fig. 2B) and occupies the S_2' and S_1' subsites with the four carbon long chain and phenyl group stretching through the S_1 and S_2 subsites. Figure 5C and 6C show the Ligplot and schematic representations of the fosinoprilat binding interactions. The carboxy terminus mimic has typical interactions with Gln281, Lys511 and Tyr520, and a hydrophobic interaction with His513. The P_2' side chain mimic interactions extend deep into the S_2' subsite and form extensive hydrophobic interactions with Val380, His383 and Tyr523. There are additional hydrophobic interactions between His353, His383 and Tyr523 with the P_2'- P_1' peptide backbone mimic, as well as hydrogen bonds with His513 and Tyr523. The phosphinic acid group interacts directly with the zinc ion as well as forms hydrogen bonds with His383 and Tyr523. There are no close hydrophobic interactions with the four carbon chain of fosinoprilat, and while the terminal phenyl ring is sandwiched between Trp357, Ala516 and Val518, it only forms a close, single hydrophobic interaction with Trp357. Therefore, this portion of the molecule is not strongly bound and explains the relative weak density observed in the electron density maps (Fig. 4C, D).

The structural data explains the high affinity fosinopril showed for AnoACE2 in the activity assays (IC_{50} of 1.5 nM). Importantly it also showed that fosinopril bound in the ACE active site (S_2' to S_2 subsites), and therefore shows proof of concept for the development of AnoACE2 specific inhibitors targeted to the typical ACE binding site.

Discussion

Comparison of AnoACE2 Structure with Human cACE, nACE and Drosophila AnCE

A sequence alignment of the ACE domain of AnoACE2 (598 residues) shares 42.3, 43.6 and 59.7 % amino acid identity with the two human sACE domains (nACE-594 residues and cACE-594 residues) and AnCE (598 residues) respectively. This is reflected in that all these domains have the same overall topology (Fig. 5) with r.m.s. deviations of 0.77Å for AnCE (592 C^{α} atoms), 1.14Å for nACE (564 C^{α} atoms) and 1.24Å for cACE (567 C^{α} atoms). The overlay of the structures (the higher resolution native AnoACE2 structure has been used for all comparisons with other homologues described in the discussion), and the r.m.s. deviations show that AnoACE2 is more closely related to the Drosophila homologue AnCE than it is to the human ACE domains of sACE. This is highlighted by two loops (residues 148-159 and 542-553 of AnoACE2) that are of equivalent length in AnCE, but 3 and 6 residues shorter in nACE and cACE respectively (Fig. 7). The C-terminus of these homologues also show conservation between AnoACE2 and AnCE, but with nACE and cACE it adopts an orientation heading almost 180° in the opposite direction (Fig. 7). Other differences in overall structure are also located in loop regions, in particular part of the N-terminal 'lid-like' feature, and another region at the opposite end of the domain (Fig. 7).

A comparison of the glycosylation sites of AnoACE2 shows that of the five sites confirmed by mass spectrometry and crystal structure, only glycosylation of Asn69 is conserved in AnCE and both domains of sACE, Glycosylation of AnoACE2 Asn106 is only conserved in cACE, whereas Asn212 and Asn327 glycosylation is only conserved in AnCE. AnoACE2 Asn187 is only conserved in nACE, but it is not glycosylated in nACE. It has previously been shown that not all of the glycosylation sites on nACE and cACE are required for activity, and the under-glycosylated domains are routinely used for crystallography [24, 30]. It would be interesting for future work to examine the importance of individual glycosylation sites to the activity of AnoACE2.

The active site and the Zn^{2+} ion binding motif are highly conserved among the four structures. However, there are some notable differences in the substrate binding sites, with differences in the S_2, S_1, S_1' and S_2' subsites being of special interest in relation to inhibitor design as discussed below.

Structural implications of inhibitor binding

Using the collective structural data on the two sACE domains (nACE – N-domain and cACE – C-domain), AnCE and the present AnoACE2, it is possible to compare different ACE proteins to look for residue and environment differences in AnoACE2 which could be targeted to create specific AnoACE2 inhibitors. Studies involving the human nACE and cACE domains have shown that although they share a high degree of conservation in the S_2' to S_2 subsites, inhibitors can be designed that show a high specificity for individual domain. For example RXP407 is over 3 orders of magnitude more potent against nACE [31], whereas RXPA380 is over 3000 times more potent against cACE

[32]. Therefore, initially the protein residues forming the binding sub-sites (S_2', S_1', S_1 and S_2) spanning the catalytic site were compared between the different ACE proteins (Table 3). This shows that there are many residues, in addition to the zinc binding motif and the catalytic glutamate, that are conserved in all four sub-sites between different ACE homologues. These are residues that could be targeted to increase the potency of inhibitors as they are likely to be important in ligand binding, but this will not enhance specificity for AnoACE2.

There are a series of AnoACE2 residues (Glu140, Val380, Tyr391, Tyr512, Ala516 and Val518) in the S_1', S_1 and S_2 sub-sites that are only conserved with one or two of the other homologues. Targeting a few, or even all, of these residues may help in improving the specificity of potential inhibitors for AnoACE2.

Of particular interest to this study, some residues are unique to AnoACE2, and targeting these differences are the most likely route for the rational development of specific inhibitors. These variations of the sub-sites have already been shown to give rise to different substrate and inhibitor specificities between the different ACE homologues. For example, differences in the S_2 sub-site between nACE and cACE have been identified as the cause of certain substrate and inhibitor specificity differences. In particular the substitution of Phe391 and Glu403 in cACE, to Tyr369 and Arg381 in nACE, have been attributed to C-domain selectivity of RXPA380 (Phe391 implicated) [33] and specificity for nACE of 33RE (importance of Tyr369) [34] and RXP407 (involvement of both Tyr369 and Arg381) [24]. As mentioned above, AnoACE2 has residue Tyr391, which is identical to nACE Tyr369, and therefore targeting this residue would only give specificity against cACE (Phe391) and AnCE (Phe375). However, Glu403 of cACE and Arg381 of nACE are replaced by Gly403 in AnoACE2 and Thr387 in AnCE. While targeting interactions with Gly403 of AnoACE2 could only be with conserved backbone atoms, the lack of charge, and extra space available due to a glycine residue could be exploited.

In addition to these differences in the S_2 sub-site, the S_2' pocket also shows unique differences in AnoACE2. Firstly Tyr527 of AnoACE2 is replaced by a phenylalanine residue in the other homologues (Phe505 in nACE, Phe527 in cACE and Phe511 in AnCE). In the structures of nACE, cACE and AnCE an acidic residue is conserved (Asp354, Glu376 and Asp360 respectively), although they are too far for strong interactions with a typical P_2' substituent. In contrast, this acidic residue is replaced by the longer, basic Arg376 residue in AnoACE2, although in both AnoACE2 structures presented here, this arginine residue points away from the S_2' sub-site. However, it is conceivable that an acidic P_2' moiety would cause a different arginine conformation allowing a salt bridge interaction, making Arg376 of AnoACE2 a strong candidate to target for specificity.

While typical ACE inhibition has concentrated on the S_2', S_1', S_1 and S_2 sub-sites, primarily because this would give the usual maximum sized drug-like compound, varying this approach could possibly be beneficial by utilizing other regions of the binding pocket that are less well conserved than these central sub-sites. For example, captopril is a potent inhibitor of ACE homologues, yet its small size only occupies the S_2' and S_1' sub-sites along with interaction to the active site zinc ion. Instead of the typical mode of extending inhibitors into the non-prime binding lobe to try and gain specificity, lengthening the P_2' moiety would allow interactions with a whole series of poorly conserved residues beyond the usual S_2' sub-site (Table 3, denoted as extended ′ residues). Mutagenesis studies have shown that many of these residues are implicated in inhibition specificity between nACE and cACE, although this is not through direct interaction with inhibitors, but instead having an effect on the hinge region of the ACE domains [35, 36]. Therefore, targeting interactions in this region of AnoACE2 residues offer an attractive prospect for potent, selective inhibition.

Another approach would be to anchor one end of an inhibitor by interaction with the active site zinc ion, and extend further into the non-prime binding lobe than the typical end at the S_2 sub-site. Beyond the S_3 pocket, residues are much less conserved, making this region an ideal target for specificity enhancement. Table 3 highlights four of the more attractive AnoACE2 residues to target, with Lys56, Gln59 and Phe360 being unique and Lys78 only being conserved in AnCE.

Conclusion

The first AnoACE2 structures presented here highlight the differences between the insect and mammalian enzymes in order to begin the design of a range of 'iACE-*enhanced'* inhibitors. They additionally provide information to allow specificity between Anopheles and Drosophila insect species. This level of detailed knowledge makes AnoACE2 an attractive target for applying structure-based drug design to developing potent and, importantly, highly selective inhibitors which could be used as insecticides. These could be presented to the larvae Anopheles gambiae encapsulated in a particulate form to maximize oral uptake in these filter feeders.

Database Depositions

The atomic co-ordinates and structure factors have been deposited in the RCSB Protein Data Bank with accession numbers 6S1Y and 6S1Z. The atomic co-ordinates and experimental data will be released upon article publication.

Abbreviations

ACE - Angiotensin-1 converting enzyme; sACE - somatic ACE; iACE - insect ACE; tACE - testis ACE; nACE - sACE N-domain; cACE - sACE C-domain; NAG - N-acetylglucosamine; BMA, β-D-mannose; FUC - D-fucose; γ-PGA - poly-γ-glutamic acid

Author contribution

J.S.C. performed all the protein expression, purification and preliminary crystallographic study and edited the manuscript. G.E.C. performed all the crystallography data analysis, wrote the first draft and edited the manuscript. C.H. performed the initial protein expression and purification studies. R.E.I. performed the enzyme kinetics study, analysed the data and edited the manuscript. K.R.A. conceptualized and supervised the study, analysed the data, wrote and edited the manuscript. All authors reviewed the manuscript.

Acknowledgements

The pOPIN plasmids were gifts from Professor Ray Owens (OPPF-UK), Mass spectrometry experiments were completed by Drs. Roman Fischer, Marie-Laetitia Thezenas and George Berridge at the Target Discovery Institute, University of Oxford (UK). We thank the scientists at station i04 (Proposal Number mx17212) of Diamond Light Source, Didcot, Oxfordshire (UK), for their support during X-ray diffraction data collection.

Competing Interests

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

References

1 Enayati, A. and Hemingway, J. (2010) Malaria management: past, present, and future. Annu Rev Entomol. **55**, 569-991

2 Hemingway, J. (2014) The role of vector control in stopping the transmission of malaria: threats and opportunities. Philos Trans R Soc Lond B Biol Sci. **369**, 20130431

3 Ranson, H. and Lissenden, N. (2016) Insecticide resistance in African Anopheles mosquitoes: A worsening situation that needs urgent action to maintain malaria control. Trends Parasitol. **32**, 187-196

4 Thomsen, E. K., Strode, C., Hemmings, K., Hughes, A. J., Chanda, E., Musapa, M., Kamuliwo, M., Phiri, F. N., Muzia, L., Chanda, J., Kandyata, A., Chirwa, B., Poer, K., Hemingway, J., Wondji, C. S., Ranson, H. and Coleman, M. (2014) Underpinning sustainable vector control through informed insecticide resistance management. PLoS One. **9**, e99822

5 Cornell, M. J., Williams, T. A., Lamango, N. S., Coates, D., Corvol, P., Soubrier, F., Hoheisel, J., Lehrach, H. and Isaac, R. E. (1995) Cloning and expression of an evolutionary conserved single-domain angiotensin converting enzyme from Drosophila melanogaster. J Biol Chem. **270**, 13613-13619

6 Isaac, R. E., Lamango, N. S., Nachman, R. J., Strey, A. and Hayes, T. K. (1997) Angiotensin-converting enzyme and the metabolism of regulatory peptides in insects. Ann Ny Acad Sci. **814**, 339-341

7 Ekbote, U., Coates, D. and Isaac, R. E. (1999) A mosquito (Anopheles stephensi) angiotensin I-converting enzyme (ACE) is induced by a blood meal and accumulates in the developing ovary. FEBS Lett. **455**, 219-222

8 Isaac, R. E. and Shirras, A. D. (2013) Peptidyl-Dipeptidase A (Invertebrate). Handbook of Proteolytic Enzymes, Vols 1 and 2, 3rd Edition, 494-498

9 Ekbote, U. V., Weaver, R. J. and Isaac, R. E. (2003) Angiotensin I-converting enzyme (ACE) activity of the tomato moth, Lacanobia oleracea: changes in levels of activity during development and after copulation suggest roles during metamorphosis and reproduction. Insect Biochem Mol Biol. **33**, 989-998

10 Isaac, R. E., Ekbote, U., Coates, D. and Shirras, A. D. (1999) Insect angiotensin-converting enzyme. A processing enzyme with broad substrate specificity and a role in reproduction. Ann N Y Acad Sci. **897**, 342-347

11 Abu Hasan, Z. I., Williams, H., Ismail, N. M., Othman, H., Cozier, G. E., Acharya, K. R. and Isaac, R. E. (2017) The toxicity of angiotensin converting enzyme inhibitors to larvae of the disease vectors Aedes aegypti and Anopheles gambiae. Sci Rep. **7**, 45409

12 Burnham, S., Smith, J. A., Lee, A. J., Isaac, R. E. and Shirras, A. D. (2005) The angiotensin-converting enzyme (ACE) gene family of Anopheles gambiae. BMC Genomics. **6**, 172

13 Isaac, R. E., Lamango, N. S., Ekbote, U., Taylor, C. A., Hurst, D., Weaver, R. J., Carhan, A., Burnham, S. and Shirras, A. D. (2007) Angiotensin-converting enzyme as a target for the development of novel insect growth regulators. Peptides. **28**, 153-162

14 Waterman, D. G., Winter, G., Gildea, R. J., Parkhurst, J. M., Brewster, A. S., Sauter, N. K. and Evans, G. (2016) Diffraction-geometry refinement in the DIALS framework. Acta Crystallogr D Struct Biol. **72**, 558-575

15 Evans, P. R. and Murshudov, G. N. (2013) How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. **69**, 1204-1214

16 Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. **50**, 760-763

17 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J Appl Crystallogr. **40**, 658-674

18 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. **53**, 240-255

19 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. **66**, 213-221

20 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. **60**, 2126-2132

21 Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. **66**, 12-21

22 McNicholas, S., Potterton, E., Wilson, K. S. and Noble, M. E. (2011) Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr. **67**, 386-394

23 Laskowski, R. A. and Swindells, M. B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model. **51**, 2778-2786

24 Anthony, C. S., Corradi, H. R., Schwager, S. L., Redelinghuys, P., Georgiadis, D., Dive, V., Acharya, K. R. and Sturrock, E. D. (2010) The N domain of human angiotensin-I-converting enzyme: the role of N-glycosylation and the crystal structure in complex with an N domain-specific phosphinic inhibitor, RXP407. J Biol Chem. **285**, 35685-35693

25 Corradi, H. R., Schwager, S. L. U., Nchinda, A. T., Sturrock, E. D. and Acharya, K. R. (2006) Crystal structure of the N domain of human somatic angiotensin I-converting enzyme provides a structural basis for domain-specific inhibitor design. J Mol Biol. **357**, 964-974

26 Natesh, R., Schwager, S. L., Sturrock, E. D. and Acharya, K. R. (2003) Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. Nature. **421**, 551-554

27 Natesh, R., Schwager, S. L. U., Evans, H. R., Sturrock, E. D. and Acharya, K. R. (2004) Structural details on the binding of antihypertensive drugs captopril and enalaprilat to human testicular angiotensin I-converting enzyme. Biochemistry. **43**, 8718-8724

28 Akif, M., Georgiadis, D., Mahajan, A., Dive, V., Sturrock, E. D., Isaac, R. E. and Acharya, K. R. (2010) High-Resolution Crystal structures of Drosophila melanogaster Angiotensin-converting enzyme in complex with novel inhibitors and antihypertensive drugs. J Mol Biol. **400**, 502-517

29 Kim, H. M., Shin, D. R., Yoo, O. J., Lee, H. and Lee, J. O. (2003) Crystal structure of Drosophila angiotensin I-converting enzyme bound to captopril and lisinopril. FEBS Letters. **538**, 65-70

30 Watermeyer, J. M., Sewell, B. T., Schwager, S. L., Natesh, R., Corradi, H. R., Acharya, K. R. and Sturrock, E. D. (2006) Structure of testis ACE glycosylation mutants and evidence for conserved domain movement. Biochemistry. **45**, 12654-12663

31 Dive, V., Cotton, J., Yiotakis, A., Michaud, A., Vassiliou, S., Jiracek, J., Vazeux, G., Chauvet, M. T., Cuniasse, P. and Corvol, P. (1999) RXP 407, a phosphinic peptide, is a potent inhibitor of angiotensin I converting enzyme able to differentiate between its two active sites. Proc Natl Acad Sci U S A. **96**, 4330-4335

32 Georgiadis, D., Beau, F., Czarny, B., Cotton, J., Yiotakis, A. and Dive, V. (2003) Roles of the two active sites of somatic angiotensin-converting enzyme in the cleavage of angiotensin I and bradykinin - Insights from selective inhibitors. Circ Res. **93**, 148-154

33 Corradi, H. R., Chitapi, I., Sewell, B. T., Georgiadis, D., Dive, V., Sturrock, E. D. and Acharya, K. R. (2007) The structure of testis angiotensin-converting enzyme in complex with the C-domain specific inhibitor RXPA380. Biochemistry. **46**, 5473-5478

34 Douglas, R. G., Sharma, R. K., Masuyer, G., Lubbe, L., Zamora, I., Acharya, K. R., Chibale, K. and Sturrock, E. D. (2014) Fragment-based design for the development of N-domain-selective angiotensin-1-converting enzyme inhibitors. Clin Sci (Lond). **126**, 305-313

35 Lubbe, L., Sewell, B. T. and Sturrock, E. D. (2016) The influence of angiotensin converting enzyme mutations on the kinetics and dynamics of N-domain selective inhibition. FEBS J. **283**, 3941-3961

36 Sturrock, E. D., Lubbe, L., Cozier, G. E., Schwager, S. L. U., Arowolo, A. T., Arendse, L. B., Belcher, E. and Acharya, K. R. (2019) Structural basis for the C-domain-selective angiotensin-converting enzyme inhibition by bradykinin-potentiating peptide b (BPPb). Biochem J. **476**, 1553-1570

Table 1. IC_{50} and confidence interval (CI) values for the inhibition of AnoACE2 activity.

NB. γ-PGA is written as % w/v due to it being a mix of different length polymers.

Inhibitor	**IC_{50} (nM or % w/v)**	**95% CI (nM or % w/v)**
Fosinoprilat	1.5 nM	0.7-3.2 nM
Lisinopril	16 nM	13-19 nM
Captopril	5.2 nM	3.9-6.8 nM
Enalaprilat	2.0 nM	0.92-4.3 nM
γ-PGA	0.2 %	0.067-0.62 %

Table 2. X-ray data collection and refinement statistics. Inner shell, overall and outer shell statistics are given in square brackets, un-bracketed and round brackets respectively.

	AnoACE2 Native	**AnoACE2 Fosinoprilat**
Resolution (Å)	[139.41-9.07] (2.27-2.20)	[117.46-9.01] (2.60-2.50)
Space group	$P6_22$ 2	$P6_22$ 2
Cell dimensions (a,b,c) angles (α,β,γ)	161.0, 161.0, 121.9 Å 90.0, 90.0, 120.0°	166.7, 166.7, 117.5 Å 90.0, 90.0, 120.0°
Molecules/asymmetric unit	1	1
Total / Unique reflections	1,821,323/47,663	2,623,791/33,809
Completeness (%)	[99.9] 100.0 (100.0)	[100.0] 100.0 (100.0)
R_{merge}	[0.075] 0.316 (6.955)	[0.038] 0.149 (1.220)
R_{pim}	[0.014] 0.052 (1.121)	[0.005] 0.017 (0.137)
<I/σ(I)>	[39.5] 11.0 (0.9)	[114.0] 24.8 (4.7)
$CC_{1/2}$	[0.998] 0.998 (0.551)	[1.000] 1.000 (0.622)
Multiplicity	[30.8] 38.2 (39.2)	[63.7] 77.3 (79.8)
Refinement statistics		
R_{work}/R_{free}	0.194/0.241	0.294/0.337
Rmsd in bond lengths (Å)	0.008	0.002
Rmsd in bond angles (°)	0.960	0.496
Ramachandran statistics (%)		
Favoured	98.2	92.6
Allowed	1.8	7.2
Outliers	0.0	0.2
Average B- factors ($Å^2$)		
Protein	57.7	95.1
Ligand	74.5	101.7
Water	47.7	51.2
Number of atoms		
Protein	9605	9397
Ligand	233	136
Water	199	25
PDB code	6S1Y	6S1Z

Table 3. Comparison of binding site residues. Residues involved in forming the sub-sites for AnoACE2, cACE and nACE of sACE, and AnCE. Typical ACE inhibitors are designed to target the S_2', S_1', S_1 and S_2 sub-sites, although there are many structural differences beyond a typical S_2' sidechain length and the S_3 sub-site. A selection of these residues are included below. Ex′ stands for Extended′.

	AnoACE2	AnCE	cACE	nACE
S_2', S_1', S_1, S_2 sub-sites				
Fully conserved target to increase affinity	N277 (S_2') Q281 (S_2') D415 (S_2') K454 (S_2') F457 (S_2') W279 (S_1') A354 (S_1') E384 (S_1') V351 (S_1) S355 (S_1) K368 (S_1) A356 (S_2) H410 (S_2) R522 (S_2)	N261 Q265 D399 K438 F441 W263 A338 E368 V335 S339 K352 A340 H394 R506	N277 Q281 D415 K454 F457 W279 A354 E384 V351 S355 K368 A356 H410 R522	N255 Q259 D393 K432 F435 W257 A332 E362 V329 S333 K346 A334 H388 R500
Partially conserved target to give a degree ficity	V380 (S_1') E140 (S_1/S_2) Y512 (S_1) A516 (S_1) V518 (S_1) Y391 (S_2)	T364 E124 Y496 A500 V502 F375	V380 E143 F512 S516 V518 F391	T358 S119 F490 N494 T496 Y369
Non-conserved giving space	G403 (S_2)	T387	E403	R381
Non-conserved targets.	R376 (S_2') Y527 (S_2')	D360 F511	E376 F527	D354 F505
Other more distant binding cavity residues				
Non-conserved targets	D162 (Ex′) T166 (Ex′) T282 (Ex′) D284 (Ex′) E377 (Ex′) F379 (Ex′) S418 (Ex′) T453 (Ex′) K78 (S_4) F360 (S_5) K56 (S_5/S_6) Q59 (S_6)	D146 E150 Q266 S268 Q361 F363 S402 D437 K62 Y344 T40 E43	E162 T166 T282 S284 D377 V379 A418 D453 L81 Y360 W59 Y62	D140 T144 S260 E262 Q355 S357 A396 E431 Q54 Y338 L32 S35

a

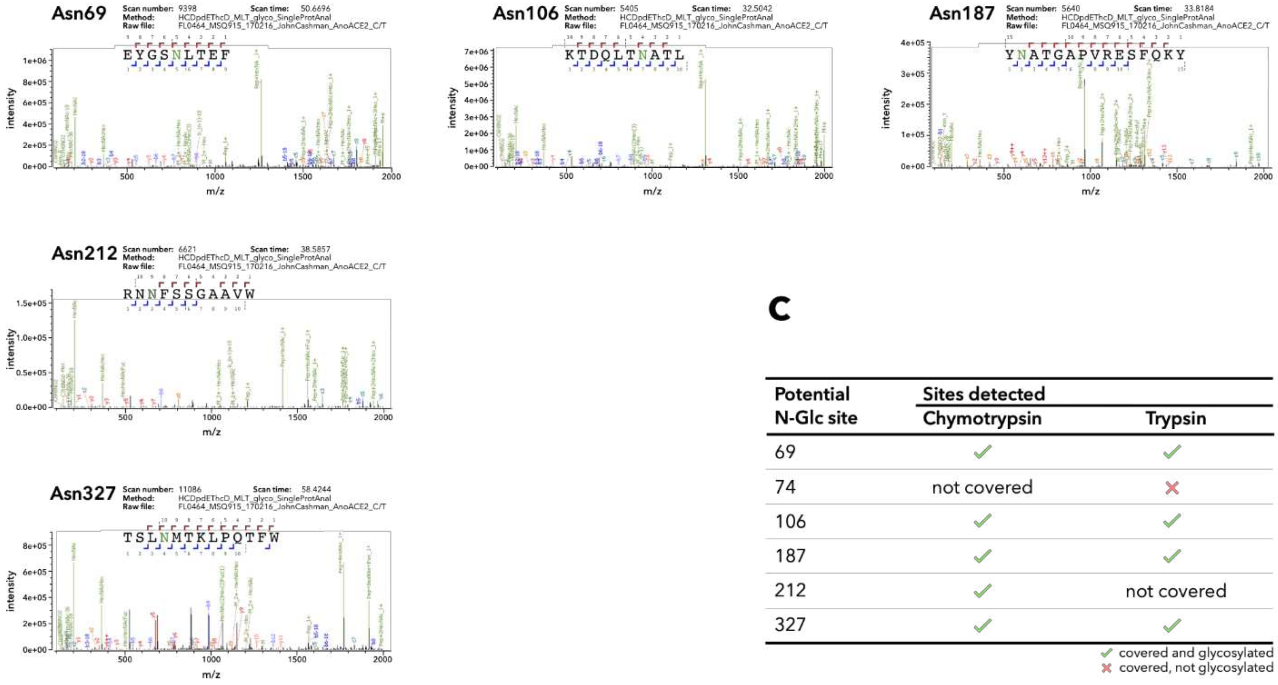
Chymotryptic peptides (mature protein)

```
RRSTESEKAP SETEISQIVE WIEQRYQQTK AHQTLAAWEY GSNLTEFNLS   76
KKTKAAADFA EVAKAVAEEL QQFKTDQLTN ATLKRRIKKL AKLGYAALPA  126
DQFKELLGAI ASMESNYAKA KFCAYGDATK CDLSLDPELT EIFANHREPE  176
ELKYYWVQWY NATGAPVRES FQKYVELNRQ AALRNNFSSG AAVWLNEYDD  226
STFEQQVDDV IEQIRPLYEQ LHAYVRYKLR QKYGDKLVSP TGPIPMHLLG  276
NLWAQTWDNI ADFTTPFPEK KLLDVTDEMI RQGYTPIKMF QMGDDFFTSL  326
NMTKLPQTFW DKSILEKPTD GRDLVCHASA WDFFAIDDVR IKQCTRVNMR  376
EFFVVHHELG HIQYYLQYQH QPVEFRGGAN PGFHEAVGDV LSLSVSTPKH  426
LKKVGLLKDY EEDEQVKINQ FYRAGVTKLV FLPFAYTLDK YRWGVFRGDI  476
KPREYNCKFW EMRSRYSGVE PPVVRTEQDF DPPAKYHVSA DVEYLRYFVS  526
YVIQFQFHRA ACALAGEYVK GDPEKTLNNC DIYQSTAAGN QLKEMLALGS  576
SKPWPDAMEV LTGERKMSAD AILEYFDPLY QWLLEENKRL GAHVGWTDSQ  626
KCVSHPIDFM AAKHHHHHHH H                                 647
```

Tryptic peptides (mature protein)

```
RRSTESEKAP SETEISQIVE WIEQRYQQTK AHQTLAAWEY GSNLTEFNLS   76
KKTKAAADFA EVAKAVAEEL QQFKTDQLTN ATLKRRIKKL AKLGYAALPA  126
DQFKELLGAI ASMESNYAKA KFCAYGDATK CDLSLDPELT EIFANHREPE  176
ELKYYWVQWY NATGAPVRES FQKYVELNRQ AALRNNFSSG AAVWLNEYDD  226
STFEQQVDDV IEQIRPLYEQ LHAYVRYKLR QKYGDKLVSP TGPIPMHLLG  276
NLWAQTWDNI ADFTTPFPEK KLLDVTDEMI RQGYTPIKMF QMGDDFFTSL  326
NMTKLPQTFW DKSILEKPTD GRDLVCHASA WDFFAIDDVR IKQCTRVNMR  376
EFFVVHHELG HIQYYLQYQH QPVEFRGGAN PGFHEAVGDV LSLSVSTPKH  426
LKKVGLLKDY EEDEQVKINQ FYRAGVTKLV FLPFAYTLDK YRWGVFRGDI  476
KPREYNCKFW EMRSRYSGVE PPVVRTEQDF DPPAKYHVSA DVEYLRYFVS  526
YVIQFQFHRA ACALAGEYVK GDPEKTLNNC DIYQSTAAGN QLKEMLALGS  576
SKPWPDAMEV LTGERKMSAD AILEYFDPLY QWLLEENKRL GAHVGWTDSQ  626
KCVSHPIDFM AAKHHHHHHH H                                 647
```

b



c

Potential N-Glc site	Sites detected	
	Chymotrypsin	Trypsin
69	✓	✓
74	not covered	✗
106	✓	✓
187	✓	✓
212	✓	not covered
327	✓	✓

✓ covered and glycosylated
✗ covered, not glycosylated

Figure 1. AnoACE2 glycosylation sites were identified by mass spectrometry. (A) AnoACE2 was digested with chymotrypsin (left) or trypsin (right). Chymotrypsin and trypsin peptides are shown bold with potential N-linked glycosylation sites highlighted (green: covered, glycosylated; red: covered, not glycosylated; yellow: not covered). (B) Mass spectra for each potential glycosylation site. (C) Summary of glycosylation sites in AnoACE2.

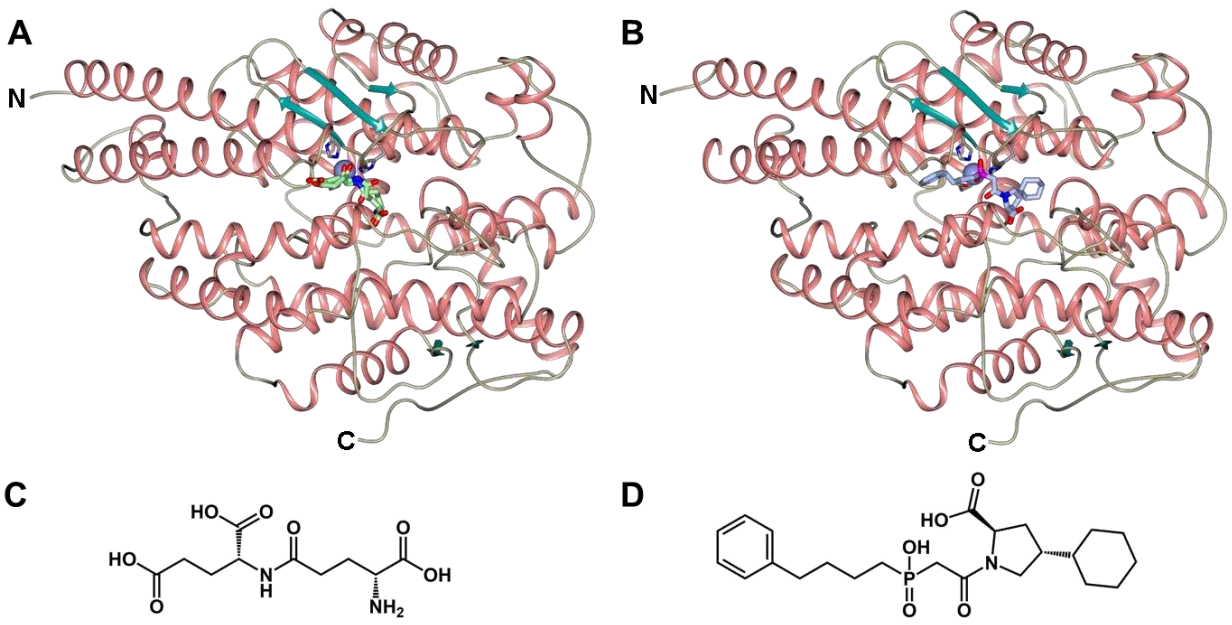


Figure 2. AnoACE2 crystal structures. Schematic representation of the overall fold of (A) native and (B) fosinoprilat complex AnoACE2 structures. γ-PGA dimer and fosinoprilat are depicted with light green and light blue sticks respectively. Zinc ions are shown as lilac spheres, with α-helices and β-strands shown in rose and dark cyan respectively. Loop regions have been smoothed for clarity. Chemical structures of (C) γ-PGA dimer and (D) fosinoprilat.

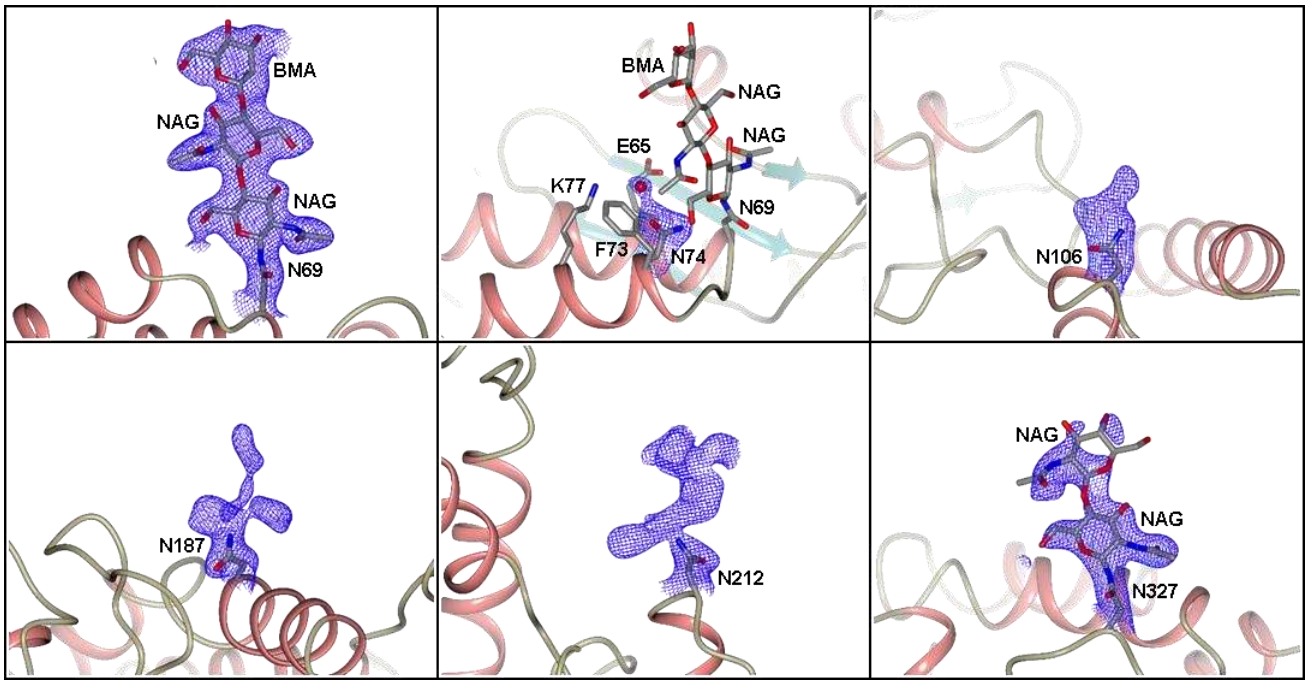


Figure 3. Predicted AnoACE2 N-linked glycosylation sites. The native AnoACE2 crystal structures shows evidence for glycosylation on five of the six predicted N-linked glycosylation sites (on asparagines 69, 106, 187, 212 and 327), with Asn74 unglycosylated being partially buried and coordinated to a water molecule. Sufficient density is observed to model sugars for Asn69 and Asn327 glycosylation. α-helices and β-strands shown in rose and dark cyan respectively, with the 2mFo-DFc electron density map shown in blue (contoured at 1σ level). NAG and BMA are N-acetylglucosamine and β-D-mannose residues respectively.

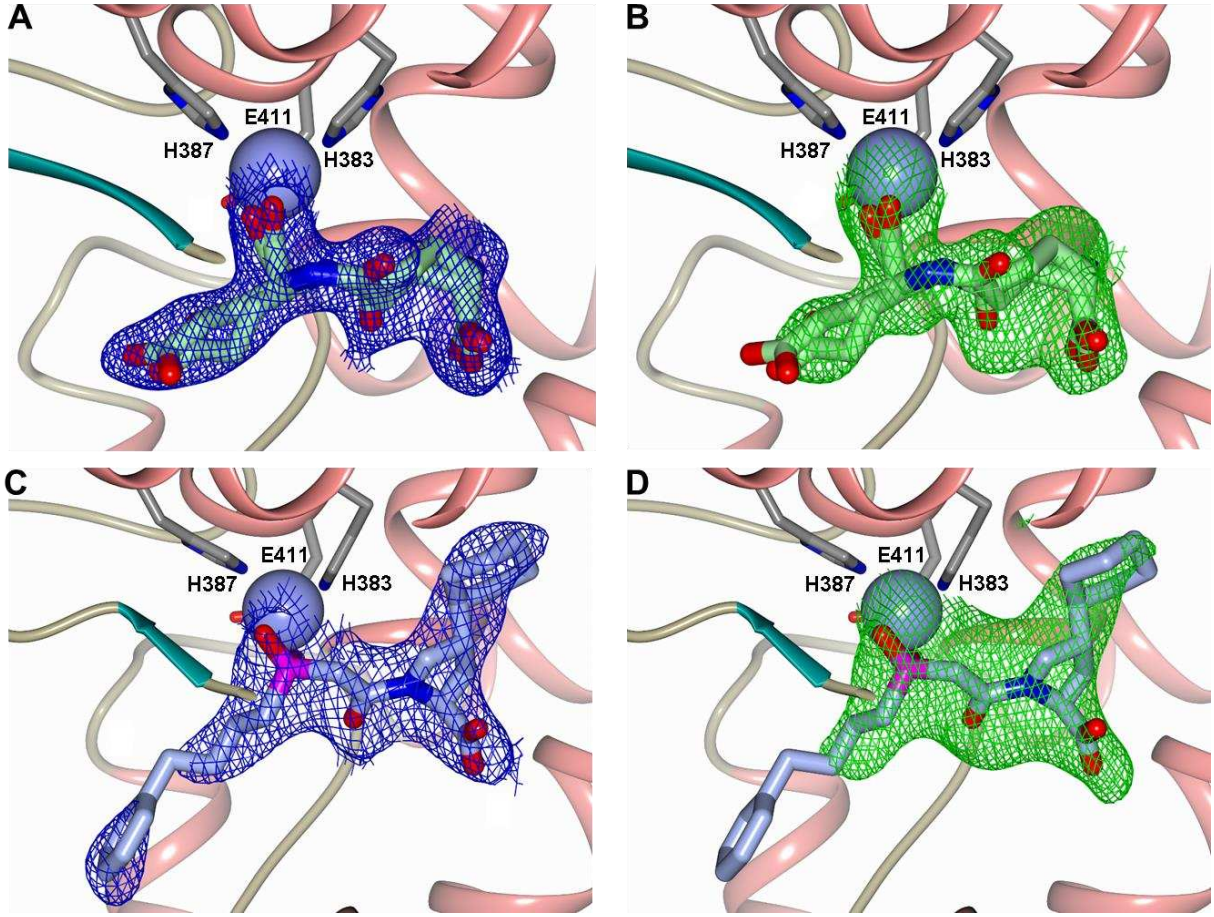


Figure 4. AnoACE2 ligand binding sites. Close up views of (A+B) γ-PGA dimer (two alternate conformations) and (C+D) fosinoprilat binding sites overlayed with the final 2mFo-DFc electron density map (blue, contoured at 1σ level) and the mFo-DFc electron density omit map (green, contoured at 3 and 2.5σ level for γ-PGA and fosinoprilat respectively). γ-PGA dimer and fosinoprilat are depicted with light green and light blue sticks respectively. Zinc ions are shown as lilac spheres, with α-helices and β-strands shown in rose and dark cyan respectively.

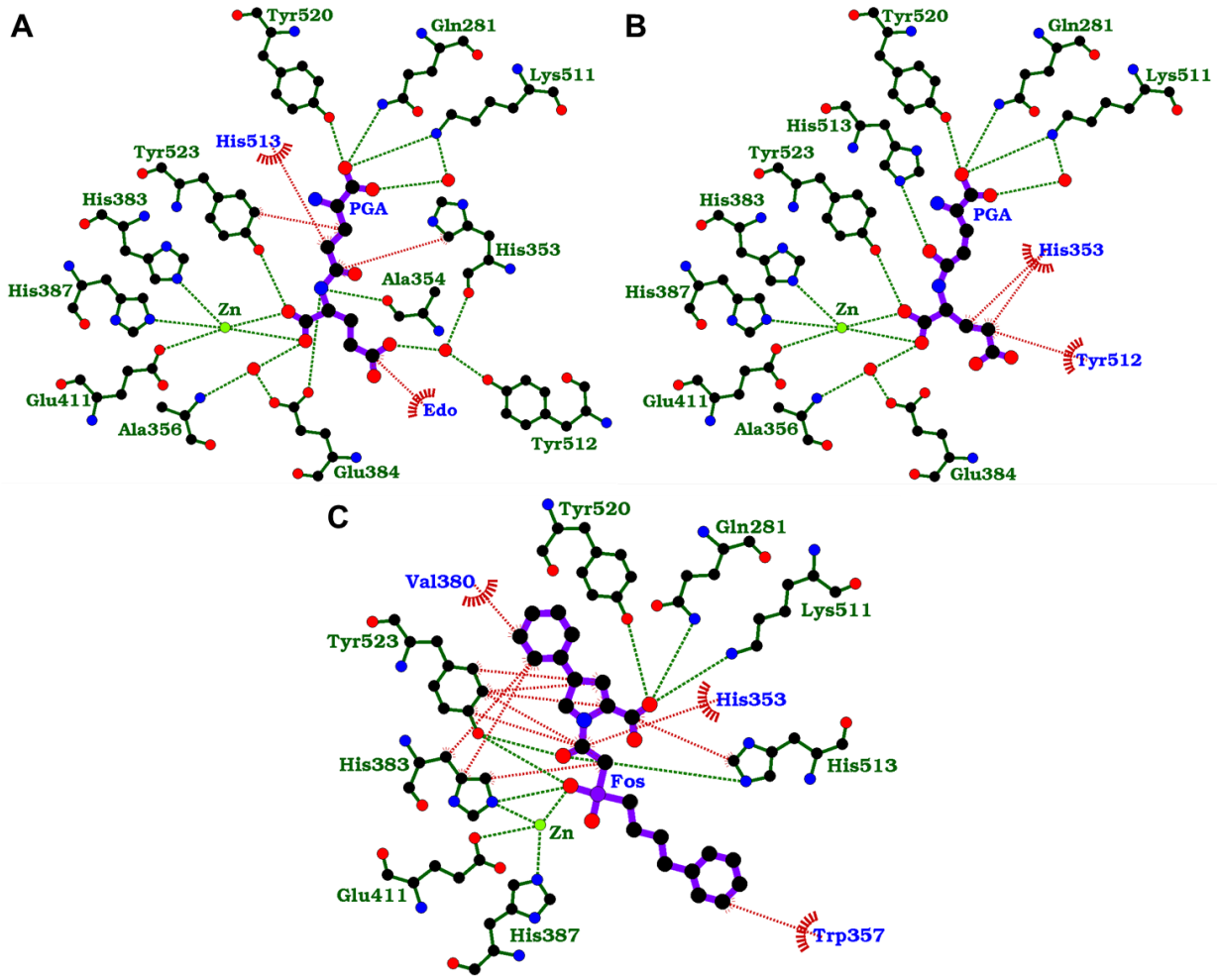


Figure 5. Ligplot representation for γ-PGA dimer and fosinoprilat AnoACE2 complexes. Binding site interactions of (A) γ-PGA conformation 1, (B) γ-PGA conformation 2 and (C) fosinoprilat. H-bond/electrostatic and hydrophobic interactions are shown in green and red dotted lines respectively, water molecules as red spheres, and residues solely involved in hydrophobic interactions are depicted by red, semi-circular symbols.

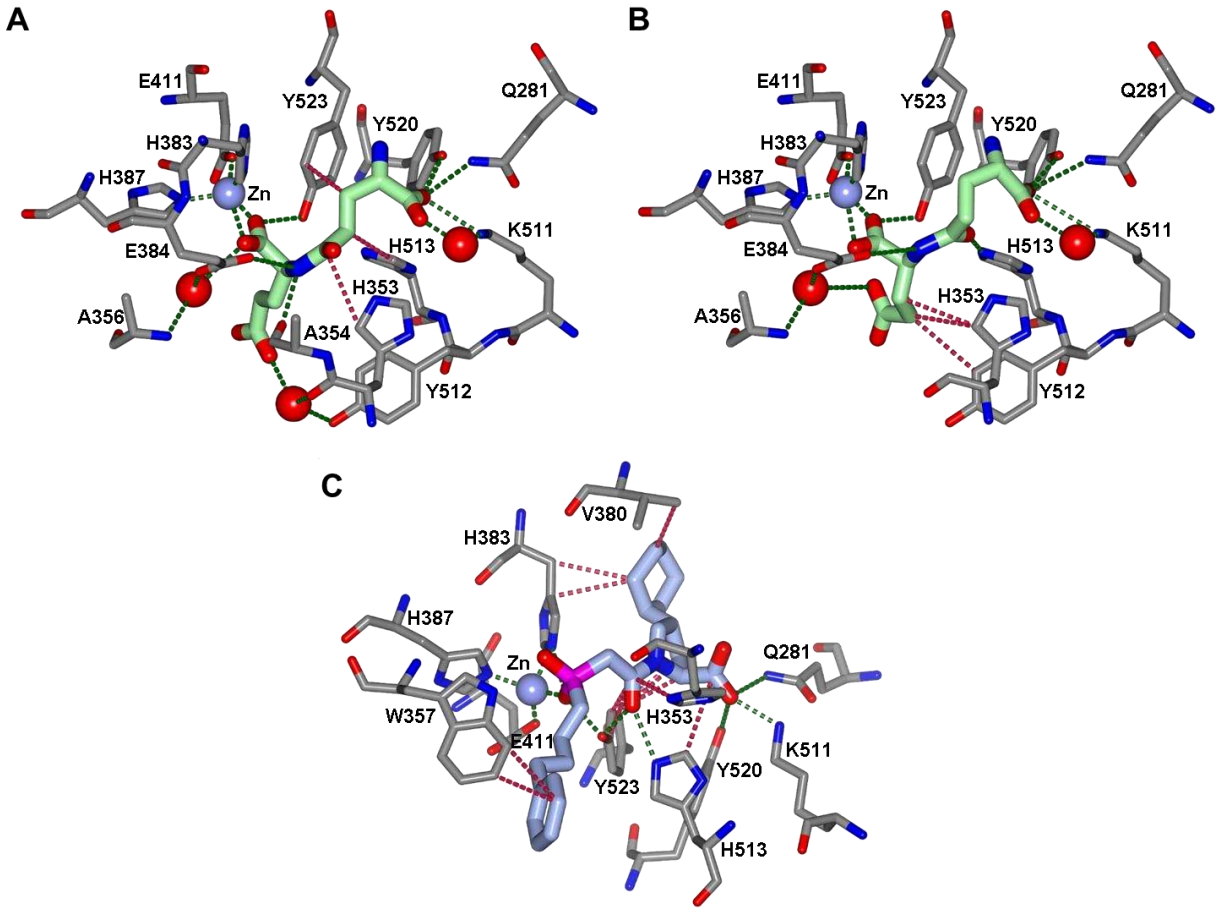


Figure 6. Binding site interactions for γ-PGA dimer and fosinoprilat. AnoACE2 complexes. Schematic representation of (A) γ-PGA conformation 1, (B) γ-PGA conformation 2 and (C) fosinoprilat binding sites showing the interactions involved with the ligands. H-bond/electrostatic and hydrophobic interactions are shown in green and red dotted lines respectively, and water molecules as red spheres.

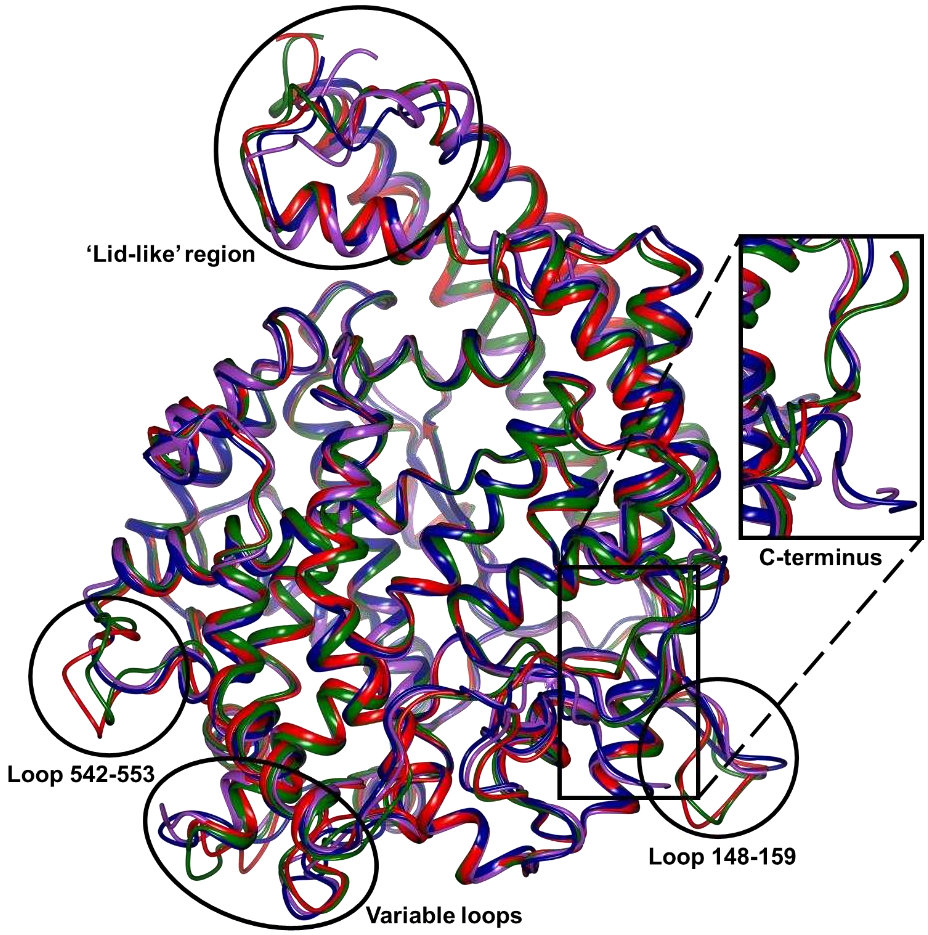


Figure 7. AnoACE2 overall structure comparison with ACE homologues. AnoACE2 green (Native structure), AnCE (2X8Y) red, nACE (6H5X) purple, tACE (6H5W) blue. Inset is the C-terminus region that has been enlarged and rotated for clarity.