PHD

Structure Based Inhibitor Design Studies on Angiotensin Converting Enzyme Homologues

Harrison, Charlotte

Award date:
2016

Awarding institution:
University of Bath

Structure Based Inhibitor Design Studies on Angiotensin Converting Enzyme Homologues

Charlotte Harrison

A thesis submitted for the degree of Doctor of Philosophy

University of Bath
Department of Biology and Biochemistry

September 2015

Abstract

Angiotensin converting enzyme, ACE (EC number 3. 4. 15. 1), is a zinc dependent dipeptidyl carboxypeptidase that has an essential role in mammalian blood pressure regulation as part of the renin-angiotensin aldosterone system. ACE acts to increase blood pressure through its actions on two peptides: angiotensin I and bradykinin. In light of this role ACE is a key target in the treatment of hypertension and ACE inhibitors have been widely used since the 1980s. Although these are effective drugs, a number of side effects are commonly associated with their use. These occur as a result of the inhibition of the roles of ACE in other physiological processes.

Somatic ACE consists of two homologous domains, each with a functional peptidase active site. Only one of these domains, the C-domain, is required for blood pressure homeostasis. There is a great deal of interest in developing a new generation of ACE inhibitors to selectively target the C-domain, whilst leaving the N-domain active and able to fulfil the other biological functions of ACE.

In the work presented here, AnCE, an ACE homologue from *Drosophila melanogaster*, has been used as a model to study the structural basis of ACE inhibition by a number of different inhibitory peptides. This has highlighted how interactions with active site residues could be exploited in the development of new, domain selective, inhibitors.

ACE homologues in insects are thought to have a conserved role in reproduction. The ACE-like enzymes from *Anopheles gambiae* are of particular interest due to the role of this organism in the transmission of malaria. In the work presented here the recombinant expression of an ACE homologue from *Anopheles gambiae* is reported for the first time. This is an important step towards the structural and biochemical characterisation of these enzymes.

Acknowledgements

First and foremost I would like to thank my supervisor Prof. Ravi Acharya for giving me the opportunity to work on this project and for all of his support and guidance along the way.

I also wish to thank our collaborator Prof. Elwyn Isaac (University of Leeds) for providing the initial yeast expression constructs and for all of his input on the AnCE inhibition studies.

Thank you also to Prof. Ray Owens and his group at the Oxford Protein Production Facility (Harwell Science and Innovation Campus, Didcot, Oxon, UK), particularly Dr Joanne Nettleship, for all of their help with the HEK293T and S*f*9 expression work and for making me feel very welcome on my visits. Thank you also to the staff at Diamond Light Source (Harwell Science and Innovation Campus, Didcot, Oxon, UK) for their support with x-ray diffraction data collection and to the BBSRC and the University of Bath for providing funding.

Thank you to all of the past and present members of Lab 0.34, especially Abi, Ami, Chris, Geoffrey, Jon, Shalini, Sharon, Sneha, Stef, Tram and Will, whose support, encouragement and advice has been invaluable. Particular thanks are also owed to Dr Geoffrey Masuyer for all of his help with this project and to Dr Abigail Davies and Dr Ami Miller for going the extra mile to help me when I needed it the most.

Thank you to all of my friends, especially Hannah and Bryony, who I know are only ever a phone call away.

Last but by no means least, thank you to my sister, Chloe, and to my parents. Thank you for all of the opportunities that you have given me and for always supporting me and believing in me, even when I haven't made it easy!

Declaration of Work Performed in Conjunction with Others

Much of the protein expression work detailed in Chapter 8 was performed at the Oxford Protein Production Facility, Harwell Science and Innovation Campus, Didcot, Oxon, UK with the support of the group of Prof. Ray Owens.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 8

List of Abbreviations

ACE	Angiotensin converting enzyme
Aβ42	Amyloid beta peptide 42
Ac64	*Autographa californica* nuclear polyhedrosis virus gp64 secretion signal
ACER	*Drosophila melanogaster* angiotensin converting enzyme related peptidase
AcSDKP	N-acetyl-seryl-aspartyl-lysyl-proline
AnCE	*Drosophila melanogaster* angiotensin converting enzyme
AnoACE	*Anopheles gambiae* angiotensin converting enzyme
AOX	Alcohol oxidase
B. jararaca	*Bothrops jararaca*
BMGY	Buffered glycerol-complex medium
BMMY	Buffered methanol-complex medium
BPP	Bradykinin potentiating peptide
CCP4	Collaborative computational project number 4
CHO	Chinese hamster ovary
C*oot*	Crystallographic object-oriented toolkit
CSF	Cerebrospinal fluid
D. melanogaster	*Drosophila melanogaster*
DLS	Diamond Light Source
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethly sulphoxide
DMF	Dimethylformamide
DPPIII	Dipeptidyl peptidase III
E. coli	*Escherichia coli*
EC number	Enzyme commission number
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
FBS	Foetal bovine serum
GnRH	Gonadotropin-releasing hormone
HA	Influenza A haemagglutinin secretion signal
HBM	Honeybee melittin secretion signal
HEK293T	Human embryonic kidney 293T/17 cell line
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
HIC	Hydrophobic interaction chromatography
HPLC	High performance liquid chromatography

HRP	Horseradish peroxidase
IC_{50}	Half maximal inhibitory concentration
IEX	Ion exchange chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kAF	(5*S*)-5-[(*N*-benzoyl)amino]-4-oxo-6-phenylhexanoyl-L-phenylalanine
kAW	(5S)-5-[(N-benzoyl)amino]-4-oxo-6-phenylhexanoyl-L-tryptophan
K_i	Inhibitor constant
K_m	Michaelis constant
λ	Wavelength
λ_{em}	Wavelength of emission
λ_{ex}	Wavelength of excitation
LB	Lysogeny broth
LLG	Log likelihood gain
MAD	Multi-wavelength anomalous dispersion
MDH	Minimal dextrose medium plus histidine
MMH	Minimal methanol medium plus histidine
MM	Minimal medium containing methanol
MIR	Multiple isomorphous replacement
MES	2-(*N*-morpholino)ethanesulphonic acid
MMP	Matrix metalloproteinase
MMT	Malic acid, MES monohydrate, Tris
MOPS	3-(*N*-morpholino)propanesulphonic acid
MR	Molecular replacement
Mut	Methanol utilisation phenotype
MWCO	Molecular weight cut-off
NMR	Nuclear magnetic resonance
OPPF	Oxford Protein Production Facility, Harwell Science and Innovation Campus, Didcot, Oxon, UK
PAK	Packing clashes
PBS	Phosphate buffered saline
PCA	Pyroglutamate
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulphonyl fluoride
P. pastoris	*Pichia pastoris*
RAAS	Renin-angiotensin aldosterone system

RFZ	Rotation function Z-score
RMSD	Root mean square deviation
RTPTμ	μ-phosphatase secretion signal
sACE	Somatic angiotensin converting enzyme
SAD	Single-wavelength anomalous dispersion
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SFM	Serum free media
S. frugiperda	*Spodoptera frugiperda*
*Sf*9	*Spodoptera frugiperda* cell line used for protein expression
SOC	Super optimal broth with catabolite repression
tACE	Testis angiotensin converting enzyme
TAE	Tris acetate EDTA
TE	Tris EDTA
TIMP	Tissue inhibitor of matrix metalloproteinase
TFZ	Translation function Z-score
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
YNB	Yeast nitrogen base
YPD	Yeast extract peptone dextrose

Chapter 1: Introduction to Angiotensin Converting Enzymes

1.1 Overview

Angiotensin converting enzyme (EC number 3. 4. 15. 1), or ACE as it is commonly known, is a highly conserved enzyme that is well known for its role in mammalian blood pressure regulation as an integral component of the renin-angiotensin aldosterone system (RAAS).

ACE primarily functions as a dipeptidyl carboxypeptidase, removing the two C-terminal residues from its peptide substrates. As a member of the M2 family of metalloproteases, ACE does this in a zinc dependent manner and contains the characteristic HExxH motif, the two histidines of which, plus an additional glutamate residue, co-ordinate the catalytic zinc ion (Corvol *et al.,* 1995; Williams *et al.,* 1994).

1.2 The Role of ACE in the RAAS

Skeggs *et al.,* first identified ACE in 1954 during the course of their work on the peptide angiotensin. Angiotensin is a potent vasoconstrictor and plays a key role in increasing blood pressure as part of the RAAS. Skeggs *et al*., observed that some of their samples appeared to contain two "forms" of angiotensin both of which, when injected into live animals, resulted in rapid vasoconstriction (Skeggs *et al.,* 1954).

When the different forms of the peptide were applied to isolated kidneys however, only one had this effect. This indicated that one form of angiotensin was in fact inactive and that it was undergoing rapid conversion into the second, active form, presumably catalysed by an enzyme present at high concentrations in the blood plasma. This enzyme was named angiotensin converting enzyme and the inactive and active forms of the peptide as angiotensin I and angiotensin II, respectively (Skeggs *et al.,* 1956).

The production of angiotensin II by ACE is a key process in blood pressure regulation by the RAAS (Zaman *et al.,* 2002). The RAAS, summarised in Figure 1.1, is a complex hormone system that is activated in response either to a decrease in blood pressure or blood volume, which triggers the release of the enzyme renin by the kidneys. The target for renin activity is angiotensinogen, which it cleaves to release the decapeptide angiotensin I (DRVYIHPFHL).

Angiotensin I is a substrate for ACE, which removes the C-terminal dipeptide from this inactive precursor to generate the active peptide angiotensin II (DRVYIHPF). As a potent vasoconstrictor, angiotensin II causes an increase in blood pressure and also

stimulates the release of aldosterone from the adrenal cortex. This results in increased re-absorption of water and ions by the kidneys, ultimately leading to an increase in blood volume and blood pressure.

Additionally, ACE acts on a second peptide, bradykinin (RPPGFSPFR). Bradykinin is a physiologically important peptide that has roles in a range of processes including inflammation and smooth muscle contraction. In terms of blood pressure regulation, bradykinin causes the dilation of blood vessels, and hence acts to decrease blood pressure. By removing the C-terminal dipeptide, ACE inactivates brady
kinin, preventing it from having this effect (Kuoppala *et al.,* 2000; Yang *et al.,* 1971).

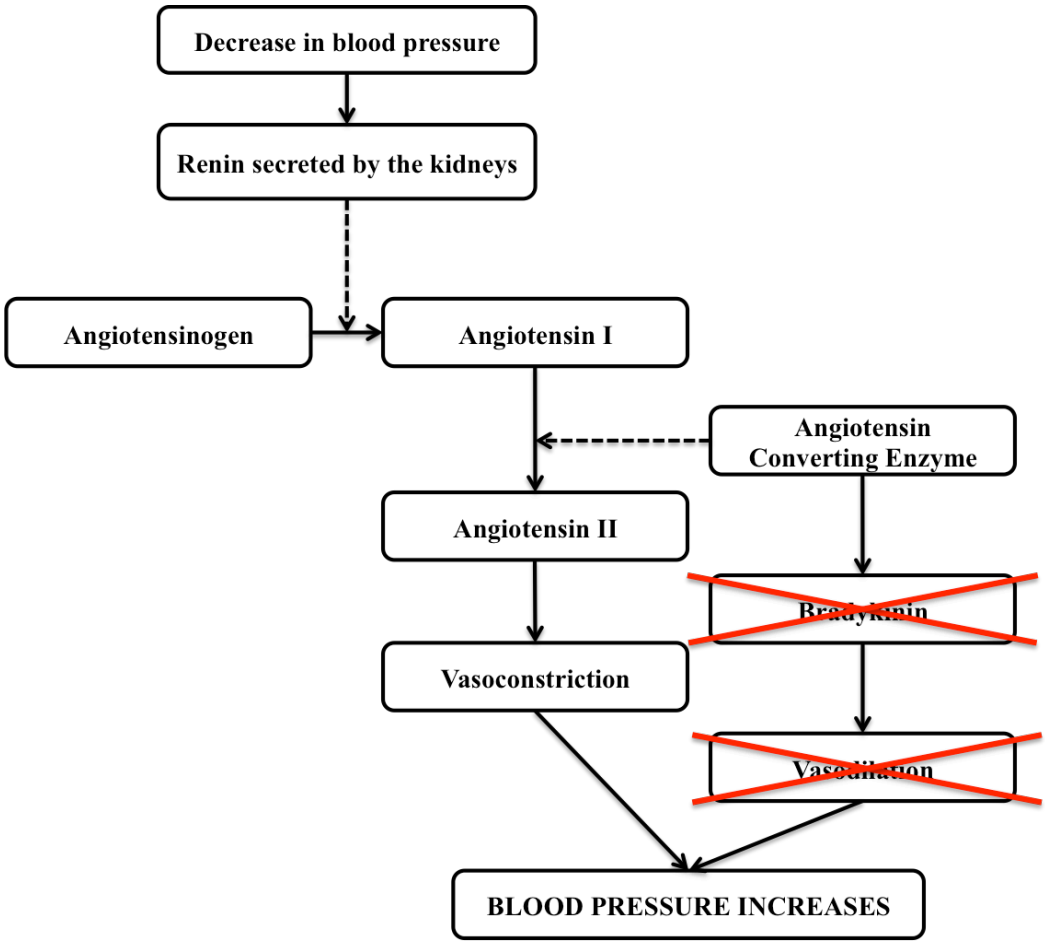


Figure 1.1: The Role of ACE in Blood Pressure Regulation.

1.3 Functions of ACE Outside of the RAAS

The role of ACE in blood pressure homeostasis has been studied for many years, but more recently, the significance of ACE in further biological processes has been realised. In addition to the systemic RAAS, the important of local, tissue-specific RAAS has become increasingly apparent in recent years. Local systems have been identified in tissues including the brain, the testes and the kidneys (Nilsson *et al.,* 2000; Paul *et al.,* 2006; Sakai and Sigmund, 2005), highlighting the roles of ACE in a wide-range of physiological processes outside of blood pressure regulation (Gonzalez-Villalobos *et al.,* 2013).

Additionally, an increasing number of peptides have now been identified as ACE substrates. This includes, but is not limited to GnRH (gonadotropin-releasing hormone), Aβ42 (amyloid beta peptide 42), angiotensin 1-9 and AcSDKP (N-acetyl-seryl-aspartyl-lysyl-proline) (Anthony *et al.,* 2012; Deddish *et al.,* 1998; Rieger *et al.,* 1993; Rousseau *et al.,* 1995). Based on these observations it is clear that ACE has a wide-range of important physiological roles that we are only beginning to understand.

1.4 Isoforms of Mammalian ACE: sACE and tACE

The concept of ACE having roles in a wide range of biological processes is supported by the fact that there are two isoforms of mammalian ACE: somatic (sACE) and testis (tACE), with different, tissue-specific expression patterns (Wei *et al.,* 1991).

sACE is an integral membrane protein mainly located on the surface of endothelial cells (Caldwell 1976; Hooper and Turner, 1987; Hooper *et al.,* 1987). It consists of two domains arranged in tandem known as the N-(terminal) and C-(terminal) domains. The two domains are connected by a short linker sequence and there is an additional transmembrane sequence located at the C-terminus of the protein. The schematic diagram in Figure 1.2 illustrates this arrangement.

The two domains of sACE are homologous, sharing 65 % overall sequence similarity, rising to 80 % in key catalytic regions. They both contain the characteristic HExxH motif and display functional peptidase activity (Soubrier *et al.,* 1988).

The second isoform, tACE, is transcribed from the same twenty-six exon gene as sACE, however tACE transcription is initiated from an intragenic promoter. Hence, the final transcript consists only of exons thirteen to twenty-six. The resulting protein has a single

catalytic domain, identical to the C-domain of sACE, including the C-terminal transmembrane sequence, Figure 1.2. In contrast to the widespread expression of sACE, tACE is confined to developing spermatids (Ehlers *et al.,* 1989; Hubert *et al.,* 1991; Sibony *et al.,* 1994) where it is thought to have a role in male fertility (Fuchs *et al.,* 2005).

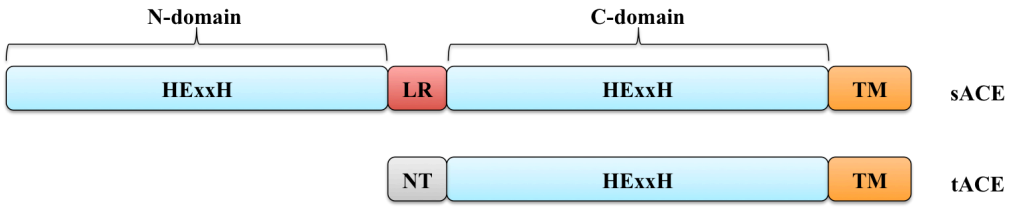


Figure 1.2: Domain Organisation in Human sACE and tACE. The conserved peptidase domain is shown in pale blue and labelled "HExxH", which is the characteristic zinc-binding motif. sACE consists of two homologous catalytic domains arranged in tandem joined by a short linker sequence (LR). tACE is identical to the sACE C-domain apart from a unique 36 residue N-terminal sequence, shown in grey and labelled "NT". sACE and tACE are both membrane bound as indicated by the C-terminal transmembrane domain, shown in orange and labelled "TM".

1.4.1 Properties of the N- and C-domains of sACE

Whilst sharing a high degree of overall sequence similarity, there are some clear distinctions between the two domains of sACE in relation to their biophysical and biochemical properties. The N-domain is significantly more thermally stable and resistant to proteolysis than the C-domain (Sturrock *et al.,* 1992; Voronov *et al.,* 2002) whilst being less sensitive to chloride activation (Wei *et al.,* 1992). This difference in sensitivity to chloride ions could be very important. It suggests that in different biological environments the two domains will display different activity profiles, which may be especially relevant when considering the tissue-specific RAAS.

Interestingly, a number of differences in substrate specificity have also been reported. Although both domains appear equally able to hydrolyse bradykinin, the C-domain processes angiotensin I to angiotensin II much more efficiently than the N-domain (Georgiadis *et al.,* 2003).

Outside of the RAAS, the N-domain hydrolyses GnRH, Aβ42, angiotensin 1-9, AcSDKP and the encephalin precursor Met^5-Enk-Arg^6-Phe^7 more efficiently than the C-

domain (Anthony *et al.,* 2012; Deddish *et al.,* 1997; Deddish *et al.,* 1998; Rieger *et al.,* 1993; Rousseau *et al.,* 1995).

Based on these observations, it seems that the C-domain alone is sufficient for blood pressure regulation *in vivo* and indeed this has been shown to be the case in mouse knock out models (Bernstein *et al.,* 2011; Fuchs *et al.,* 2004; Fuchs *et al.,* 2008). Taken together with the relationship between the C-domain of sACE and tACE, this indicates that sACE has arisen as the result of a gene duplication event. This would have enabled the N-domain to evolve and acquire new functions, exemplified by the observed differences in substrate specificity outlined above.

1.5 ACE Inhibitors

Due to its crucial role in blood pressure regulation, ACE is a key target in the treatment of hypertension and ACE inhibitors have been in widespread clinical use since the 1980s (Patchett *et al.,* 1980). Given that the first ACE structure was not reported until 2003 (Natesh *et al.,* 2003), this means that the ACE inhibitors currently in use were designed with no structural information.

ACE inhibitor design was instead based on an assumed similarity with carboxypeptidase A, which was one of the few crystal structures available at the time, and on the sequence of 'bradykinin potentiating peptides' isolated from the venom of the pit viper *Bothrops jararaca* that were observed to function as natural inhibitors of ACE (Byers and Wolfenden, 1973; Cushman *et al.,* 1977; Cushman and Ondetii, 1999; Ferreira, 1965; Ferreira *et al.,* 1970; Ondetti *et al.,* 1971; Schmid and Herriot, 1976). It was observed that all of these peptides contained a C-terminal proline residue. This was therefore assumed to be an important feature and a proline moiety was incorporated into inhibitor design.

The first widely used ACE inhibitor, captopril, is shown in Figure 1.3. It is a competitive inhibitor that co-ordinates the catalytic zinc ion through its sulphydryl group. Subsequent inhibitors have since been developed that use carboyxl rather than sulphydryl groups for zinc co-ordination. Although the carboxyl group does not form such a strong interaction with the zinc ion this is compensated for by the incorporation of additional functionalities to make further interactions with the active site. The ACE inhibitors lisinopril and enalaprilat are shown in Figure 1.3, both of which are currently in clinical use (Patchett *et al.,* 1980).

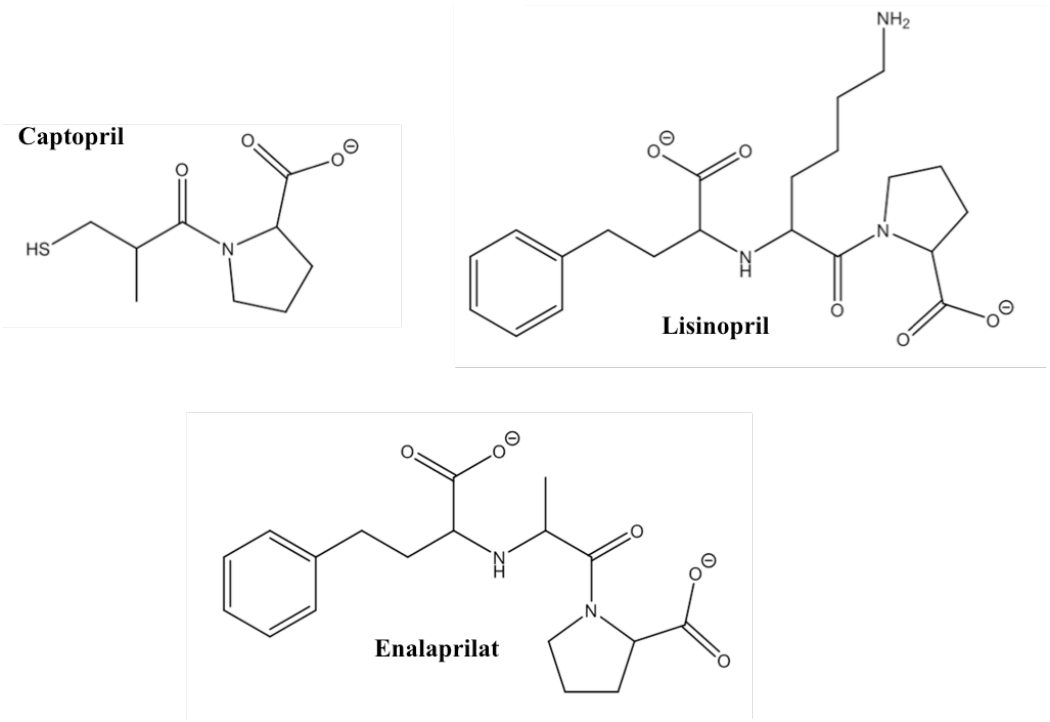


Figure 1.3: Chemical Structures of the ACE Inhibitors Captopril, Lisinopril and Enalaprilat.

1.5.1 Limitations of Current ACE Inhibitors

Although these inhibitors have been widely used for over thirty years and are effective anti-hypertensive agents, there are problems associated with their use. This includes two phenomena known as "outside in signalling" and "ACE inhibitor escape".

Outside in signalling occurs when the use of ACE inhibitors actually increases ACE expression, counter-acting the effect of the inhibitors. "ACE inhibitor escape" tends to occur following long-term treatment and describes the circumstance where levels of angiotensin II are no longer reduced. This is thought to be the result of chymase expression in the heart being upregulated as chymase is also able to convert angiotensin I into angiotensin II (Kohlstedt *et al.,* 2002; Kohlstedt *et al.,* 2005; Wei *et al.,* 2010).

By far the most widespread problem with current ACE inhibitors is their propensity to cause side effects. As many as one in every five patients treated with the current inhibitors develop a severe dry cough, whilst a smaller number suffer from angioedema, which can potentially prove fatal (Bas *et al.,* 2007; Steckelings *et al.,* 2001).

These side effects are almost certainly the result of the function of ACE in other physiological processes being inhibited. For example, bradykinin has an important role in contraction of smooth muscle in the bronchus. ACE inhibition leads to the accumulation of bradykinin, which can cause bronchoconstriction, resulting in the persistent dry cough observed in many patients (Israili and Hall, 2002; Overlack, 1996; Vegter and de Jong-van den Berg, 2010).

1.5.2 The Future of ACE Inhibitors

The N-domain of sACE appears to have evolved to acquire novel functions and the C-domain alone is apparently sufficient for *in vivo* regulation of blood pressure (Bernstein *et al.,* 2011). This indicates that it may be beneficial to specifically target the C-domain in the treatment of hypertension, thus allowing the N-domain to remain active and retain its function in physiological processes outside of the RAAS.

Crucially, given that both domains are equally able to cleave bradykinin, selective inhibition of the C-domain would prevent the accumulation of this peptide, which has been implicated in many of the most common side effects resulting from treatment with the current ACE inhibitors.

Additionally, as more is understood about the physiological function of the N-domain, it may be desirable to develop inhibitors that selectively target this domain. For example, the peptide AcSDKP, which is cleaved primarily by the N-domain (Junot *et al.,* 2001), has been shown to be involved in the regulation of proliferation of haematopoietic stem cells (HSCs) (Azizi *et al.,* 1996; Bonnet *et al.,* 1992; Lenfant *et al.,* 1989). There is evidence to suggest that inhibiting AcSDKP cleavage by ACE could protect HSCs during aggressive chemotherapy regimes (Bogden *et al.,* 1991).

Similarly, AcSDKP has a role in fibrosis and inflammation, where it functions to inhibit the proliferation of fibroblasts. Knock out studies in mice have shown that deletion of the N-domain of sACE, and hence AcSDKP hydrolysis, can reduce lung injury sustained by treatment with the chemotherapeutic agent bleomycin (Li *et al.,* 2010).

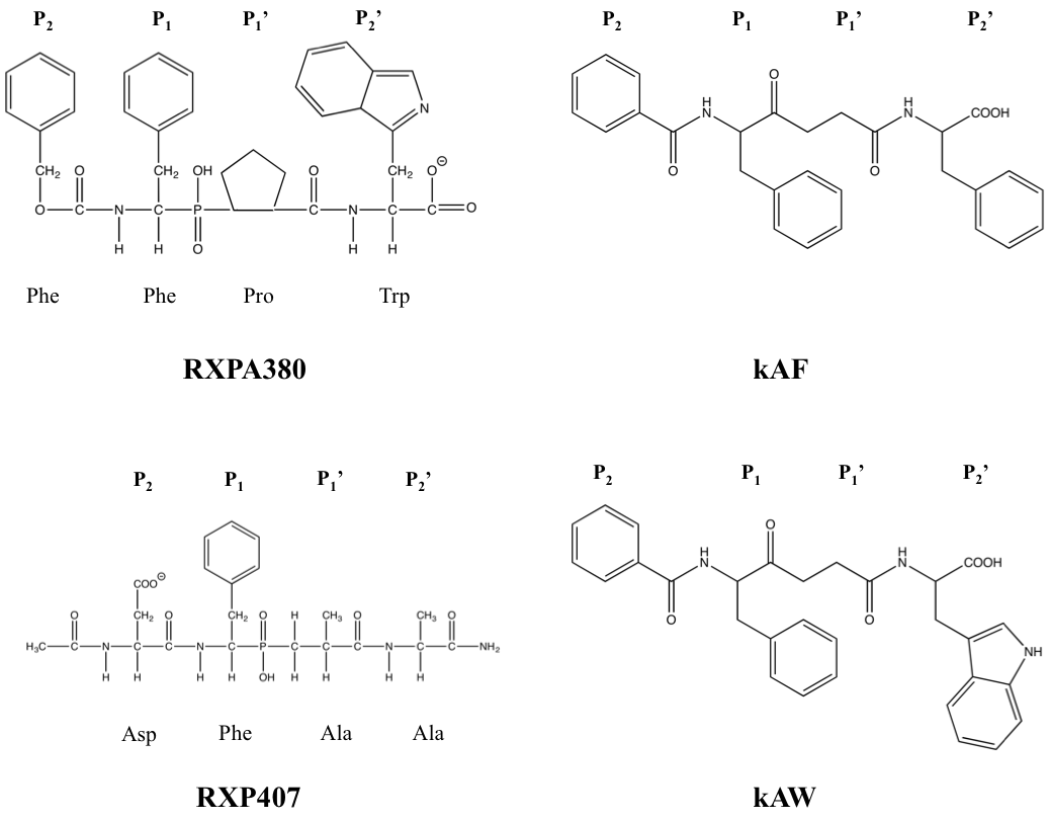


Figure 1.4: Structure of Inhibitors Selective for a Single Domain of sACE. RXPA380, kAF and kAW are all selective inhibitors of the C-domain, whereas RXP407 is a selective inhibitor of the N-domain. RXPA380 and RXP407 are phosphinic peptides and kAF and kAW are ketomethylene inhibitors.

Whilst most of the inhibitors currently in use show very little selection for either of the two domains under physiological conditions (Wei *et al.,* 1992), some domain selective inhibitors have been developed. This includes two phosphinic peptide inhibitors, RXPA380 (Georgiadis *et al.,* 2003; Georgiadis *et al.,* 2004) and RXP407 (Dive *et al.,* 1999), which are selective for the C- and N-domains respectively and two ketomethylene inhibitors kAW ((5*S*)-5-[(*N*-benzoyl)amino]-4-oxo-6-phenylhexanoyl-L-tryptophan) and kAF ((5*S*)-5-[(*N*-benzoyl)amino]-4-oxo-6-phenylhexanoyl-L-phenylalanine) that are selective inhibitors of the C-domain (Nchinda *et al.,* 2006), Figure 1.4. The development of such inhibitors provides proof of principle evidence that domain specific inhibition could be possible. Structure-based inhibitor binding studies were central to the development of these compounds, and it is clear that this method will be essential for future developments.

1.6 Structural Studies on ACE

1.6.1 Challenges in ACE Crystallisation

The first ACE structure was not reported until 2003 (Natesh *et al.,* 2003), almost fifty years after the initial discovery made by Skeggs *et al*. This is due to the fact that ACE is a very challenging enzyme to work with and a significant factor in this is the extensive glycosylation of both domains, which has been estimated at 30 % by weight and includes both N-linked and O-linked sites (Acharya and Sturrock, 2003; Ripka *et al.,* 1993).

Whilst the initial tACE structure reported in 2003 was of the native protein, truncated to remove the transmembrane domain, the crystals used to collect the diffraction data were not reproducible. This was most likely as a result of the extensive glycosylation causing heterogeneity.

In an attempt to overcome this problem, a study was performed in 2003 to investigate the importance of each of the predicted N-linked glycosylation sites in each domain of sACE; seven in the C-domain (tACE) and ten in the N-domain. By using a site-directed mutagenesis approach, it was revealed that only glycosylation at the first and third of these sites in the C-domain and at the third, eighth and ninth sites in the N-domain was required to produce stable, functional protein (Anthony *et al.,* 2010; Gordon *et al.,* 2003).

The crystal structures of both tACE and the N-domain of sACE have since been determined using constructs with the other N-linked glycosylation sites removed by mutation and with C-terminal truncations to remove the transmembrane region. As glycosylation at these sites has been shown to be essential for functional protein, the ACE constructs were unsuitable for *E. coli* expression and were instead expressed using chinese hamster ovary (CHO) cells to ensure correct post-translational processing (Corradi *et al.,* 2006; Natesh *et al.,* 2003).

1.6.2 The Structure of tACE

In 2003, the structure of human tACE (equivalent to the C-domain of human sACE) was reported (Natesh *et al.,* 2003). The overall structure is shown in Figure 1.5 and illustrates how the molecule adopts an overall ellipsoid shape, formed by twenty-seven α-helices and six short β-strands. A long substrate-binding channel extends for 30 Å across almost the entire length of the molecule, effectively dividing it into two subdomains.

It is at the centre of this channel that the active site is located. This is marked by the presence of the catalytic zinc ion, which is co-ordinated by residues His383, His387 and Glu411. Access to the active site appears to be controlled by N-terminal helices α1, α2 and α3, which form a "lid" that caps the entrance to the substrate-binding channel, preventing entry of large, folded molecules, Figure 1.5.

Two chloride ions are also visible in this structure. The first, Cl1, is located 20.7 Å from the active site zinc ion and bound by Arg186, Trp485, Arg489 and a water molecule. The second chloride ion, Cl2, is closer to the active site, located 10.4 Å from the zinc ion. It is bound to Tyr224, Arg522 and a water molecule. The co-ordination of both chloride ions is illustrated in Figure 1.5.

1.6.3 The Structure of the N-domain

The first structure of the N-domain of human sACE was reported three years later, in 2006 (Corradi *et al.,* 2006). As would be expected given the high degree of sequence identity, the overall structure is very similar to that of tACE (the C-domain), Figure 1.6.

Again the molecule adopts an overall ellipsoid shape that is formed by twenty-seven helices and six short β-strands. The substrate binding channel and N-terminal lid are both conserved, as is the position of the catalytic zinc ion, which, in this case is co-ordinated by His361, His365 and Glu389.

One of the chloride ions, Cl2 is conserved from the tACE structure and is held in place by Tyr202, Arg500 and a water molecule. No chloride ion is present at the position of Cl1 in tACE. This is due to the substitution of Arg186 in tACE by His164 in the N–domain, as illustrated in Figure 1.6. The absence of a second chloride ion in the N-domain active site goes some way towards explaining the decreased sensitivity of the N-domain to chloride concentration.

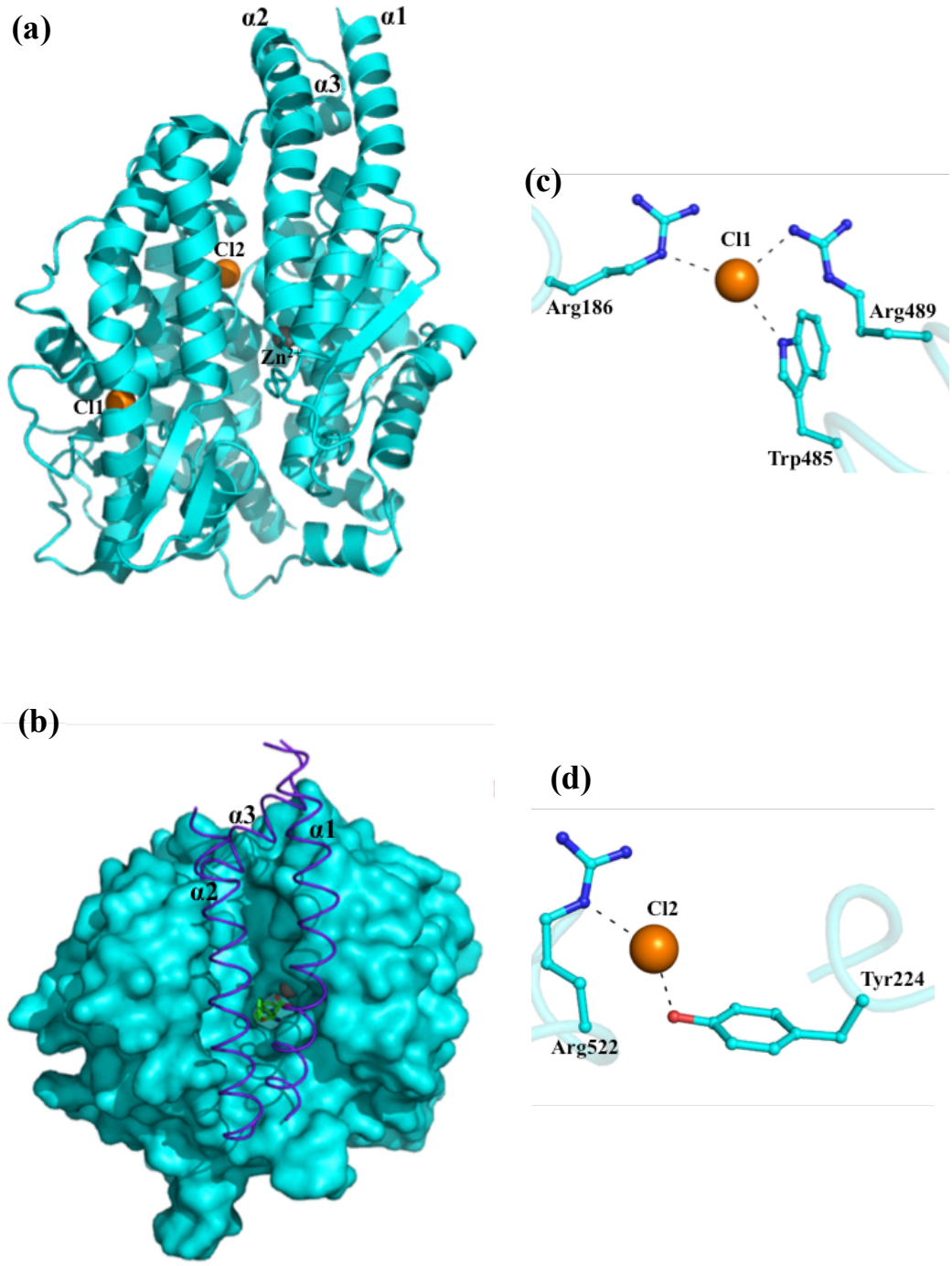


Figure 1.5: The Structure of Human tACE (PDB code 1O86). (a) Cartoon representation of the overall structure of tACE highlighting the prevalence of helical elements. The two chloride ions visible in the structure are shown as orange spheres and the catalytic zinc ion as a grey sphere. The zinc ion marks the location of the active site, at the centre of a long substrate-binding channel extending the length of the molecule. An "N-terminal lid" formed by helices α1, α2 and α3 caps the channel. (b) An alternative view of tACE to illustrate the function of the N-terminal lid. Helices α1, α2 and α3 are shown as purple helices and the rest of the molecule as a surface representation in cyan. The location of the active site at the centre of the substrate binding channel is shown by the zinc ion in grey and the inhibitor lisinopril, which is shown as green sticks. (c) and (d) Residues involved in chloride ion binding are shown as cyan sticks and the chloride ions as orange spheres.

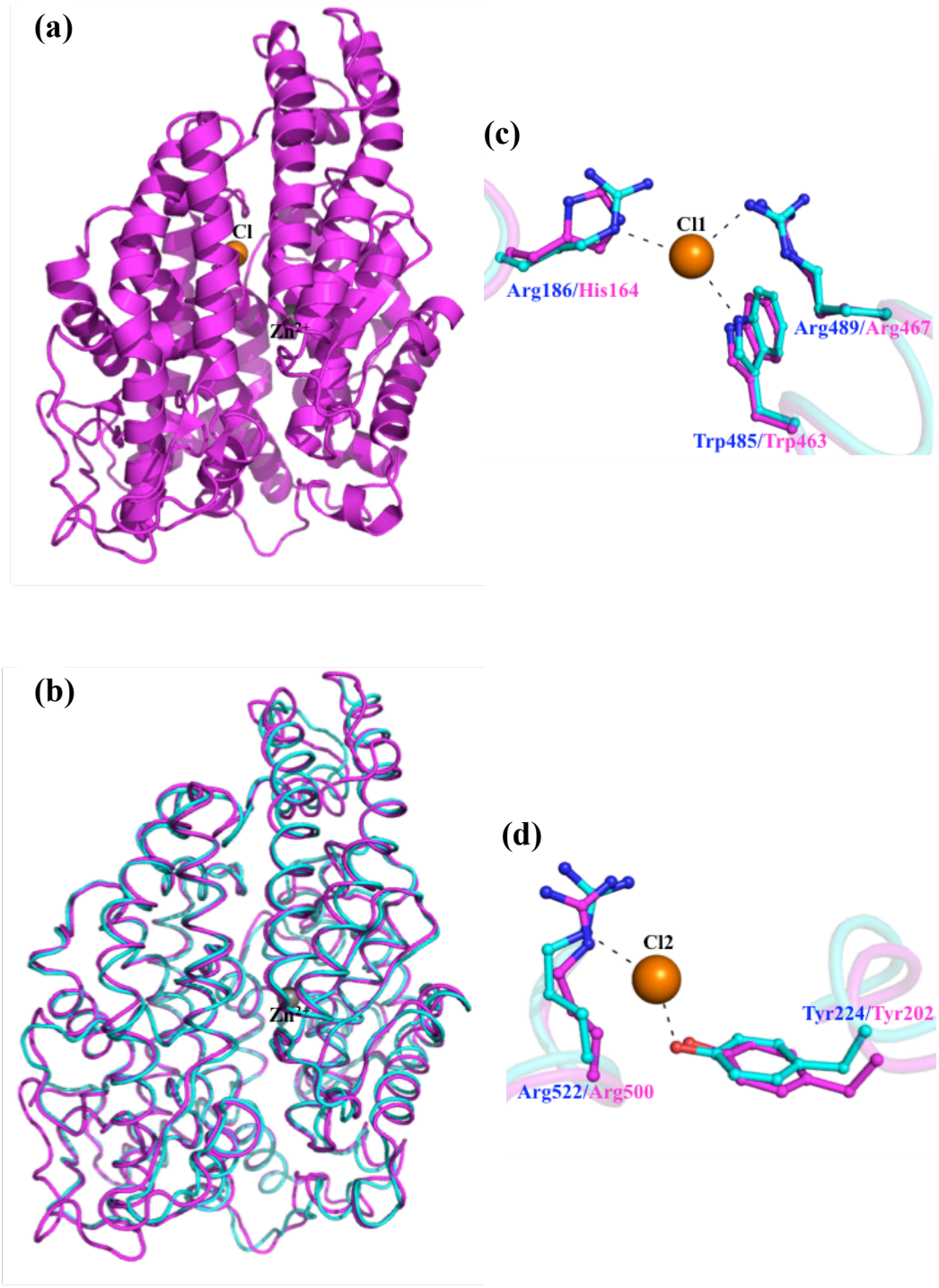


Figure 1.6: The Structure of the N-domain of Human sACE (PDB code 2C6N). (a) Cartoon representation of the overall structure of the N-domain with the active site zinc ion and the single chloride ion shown as grey and orange spheres respectively. (b) Superposition of the N-domain structure, magenta, on the structure of tACE, cyan, illustrating the similarity between the overall structures. (c) There is no chloride ion present in the N-domain structure at the site of Cl1 in tACE due to the substitution of Arg186 for His164. The other chloride binding residues are conserved. tACE residues are shown as cyan sticks and N-domain residues in magenta. Residues are labelled with the corresponding colours. (d) The second chloride ion from tACE is conserved in the N-domain. The colour scheme is as described in (c).

1.6.4 Structural Insights into the ACE Catalytic Mechanism

Although no definitive mechanism for the cleavage of peptide bonds by angiotensin converting enzymes has been determined, it is widely accepted that it occurs via a general base mechanism involving nucleophilic attack on the carbonyl carbon of the scissile peptide bond.

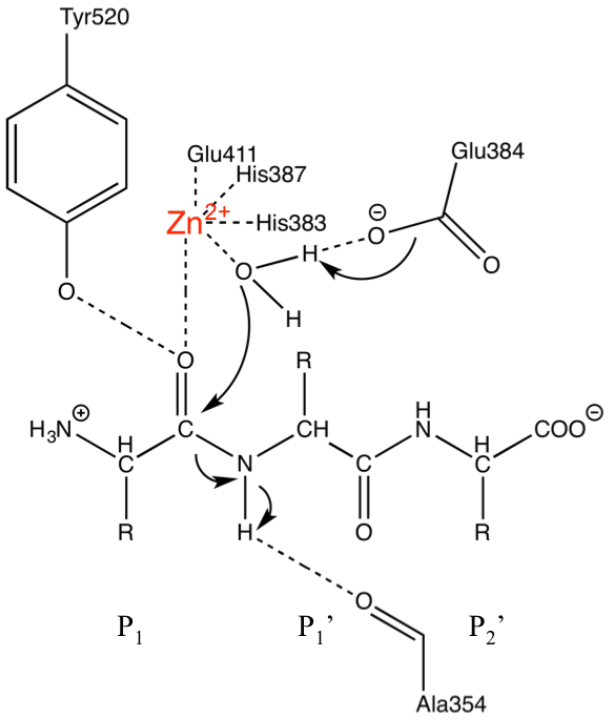


Figure 1.7: Proposed ACE catalytic mechanism. Conserved residue Glu384 abstracts a proton from a catalytic water molecule, activating it for nucleophilic attack on the carbonyl carbon of the scissile peptide bond. Nucleophilicity of the water is also enhanced by the catalytic zinc ion. The resulting transition state is stabilised by conserved residues Tyr520 and Ala354 (tACE residue numbering).

Active site residue Glu384 (tACE numbering) has been implicated in catalysis, however structural data indicate that it is not positioned suitably for direct nucleophilic attack of the scissile bond. Instead, it has been suggested that Glu384 abstracts a proton from the catalytic water molecule, thus activating it for nucleophilic attack on the carbonyl carbon of the scissile bond. This results in the formation of an oxyanion intermediate, likely stabilised by the catalytic zinc ion and conserved residues Tyr520 and Ala354 (Sturrock *et al.,* 2004; Zhang *et al.,* 2013). This is summarised in Figure 1.7.

In addition to its widely recognised role as a dipeptidyl carboxypeptidase, ACE has also been observed to release the C-terminal tripeptide from some substrates. The primary example of this being seen with the neuropeptide substance P (Skidgel *et al.,* 1985). Furthermore, ACE is also able to act as an aminopeptidase towards the substrate GnRH (Jaspard *et al.,* 1993). Neither of these activities is seen frequently, only occurring when the C-terminus of the protein is "protected". For example in GnRH the penultimate residue is a proline. The unique nature of the proline side chain means that it has a secondary rather than a primary amine group, which protects the scissile bond from cleavage. This is in agreement with the observation that all of the bradykinin potentiating peptides isolated from the venom of *B. jararaca* have two proline residues at their C-termini (Ondetti *et al.,* 1971).

The structural and mechanistic basis of the amino peptidase and C-terminal tripeptidase activity of ACE remains poorly understood and further structural and activity based studies are required to investigate this further.

1.6.5 The Structural Basis of Domain Specific Inhibition

Studying the structures of the two domains of sACE has provided information that has been invaluable in understanding how subtle differences in the active site regions could be used in the development of inhibitors that selectively target either the N- or the C-domain.

1.6.5.1 Schechter and Berger Notation for Protease Cleavage Sites

In developing domain specific inhibitors, the focus is on "subsites" that extend along the substrate-binding channel in either direction away from the catalytic zinc ion. During peptide hydrolysis, a peptide residue (P_x) occupies each of these subsites (S_x). The peptide residues are numbered from the point of cleavage such that the scissile peptide bond is between residues P_1 and P_1'. The subsites of the enzyme are labelled S_1 and S_1' accordingly so that the numbers correspond to the peptide residues by which they are occupied (Schechter and Berger, 1967). Figure 1.8 illustrates how this notation is applied to the cleavage of angiotensin I by ACE.

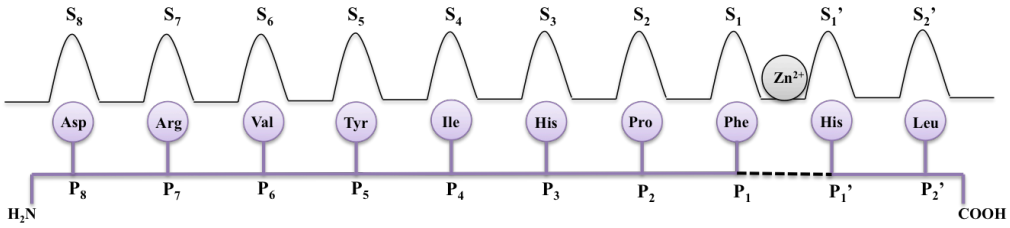


Figure 1.8: Schematic Diagram of Angiotensin I Cleavage by ACE Using the Schechter and Berger Notation. The enzyme subsites are labelled S_x and occupied by the corresponding residues of angiotensin I (P_x). The catalytic zinc ion is located between S_1 and S_1' and the peptide is positioned for cleavage of the peptide bond between the P_1 and P_1' residues, shown by the black dotted line.

1.6.5.2 RXPA380

RXPA380 is a phosphinic peptide that inhibits the C-domain of sACE with a K_i that is three orders of magnitude lower than for the N-domain. Hence it is a highly selective inhibitor of the C-domain (Georgiadis *et al.,* 2004).

The structure of RXPA380 bound to tACE (sACE C-domain) (Corradi *et al.,* 2007) indicates that it is a competitive inhibitor of ACE and co-ordinates the catalytic zinc ion through the oxygen atoms of its phosphinyl group. It binds in an elongated conformation with its phenylalanine, phenylalanine, proline and tryptophan moieties occupying the S_1, S_2, S_1' and S_2' subsites, respectively. This is illustrated in Figure 1.9.

Comparing this to the N-domain structure indicates why RXPA380 is a more potent inhibitor of the C-domain. The structure of tACE with RXPA380 illustrates that the phenylalanine residue at the P_2 position forms an aromatic interaction with Phe391 in the S_2 pocket. In the N-domain, Tyr369 replaces Phe391. Based on the structural data, the hydroxyl group of Tyr369 would be within 2 Å of the P_2 phenylalanine, thus preventing it from binding in the conformation observed with tACE. The Phe391/Tyr369 substitution is therefore likely to be a key factor contributing to the selectivity of RXPA380 for the C-domain.

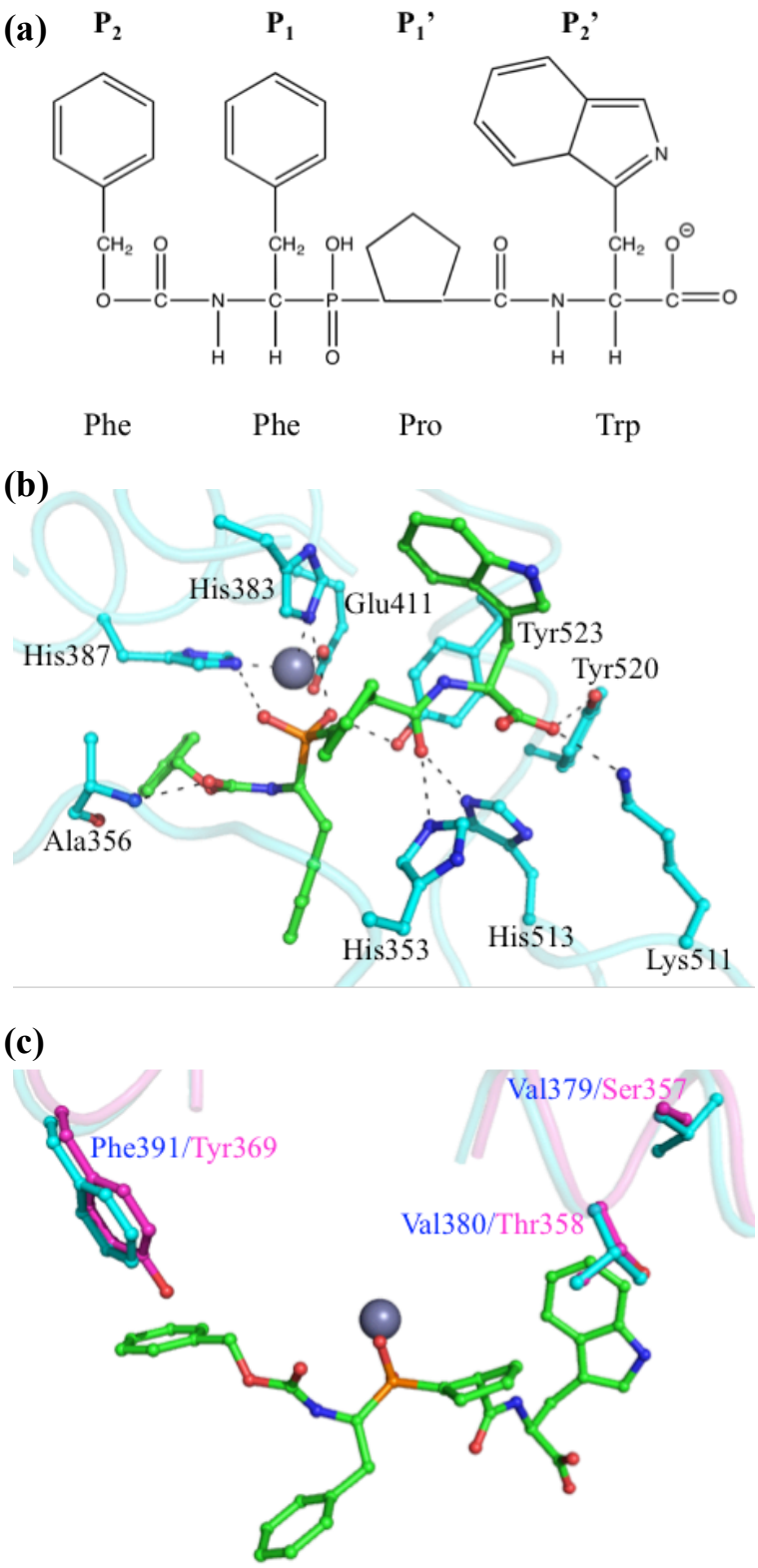


Figure 1.9: Binding of the C-domain Specific Inhibitor RXPA380 to ACE (PDB code 2OC2). (a) The structure of RXPA380 indicating the key moieties. (b) Binding of RXPA380 to the C-domain of ACE. RXPA380 is shown as green sticks and interacting C-domain residues as cyan sticks. The catalytic zinc ion is a grey sphere. (c) Key residues interacting with RXPA380 that are not conserved in the N-domain. Residues from the C-domain are shown in cyan and corresponding N-domain residues in magenta. The residues are labelled in the colours corresponding to the domain to which they belong.

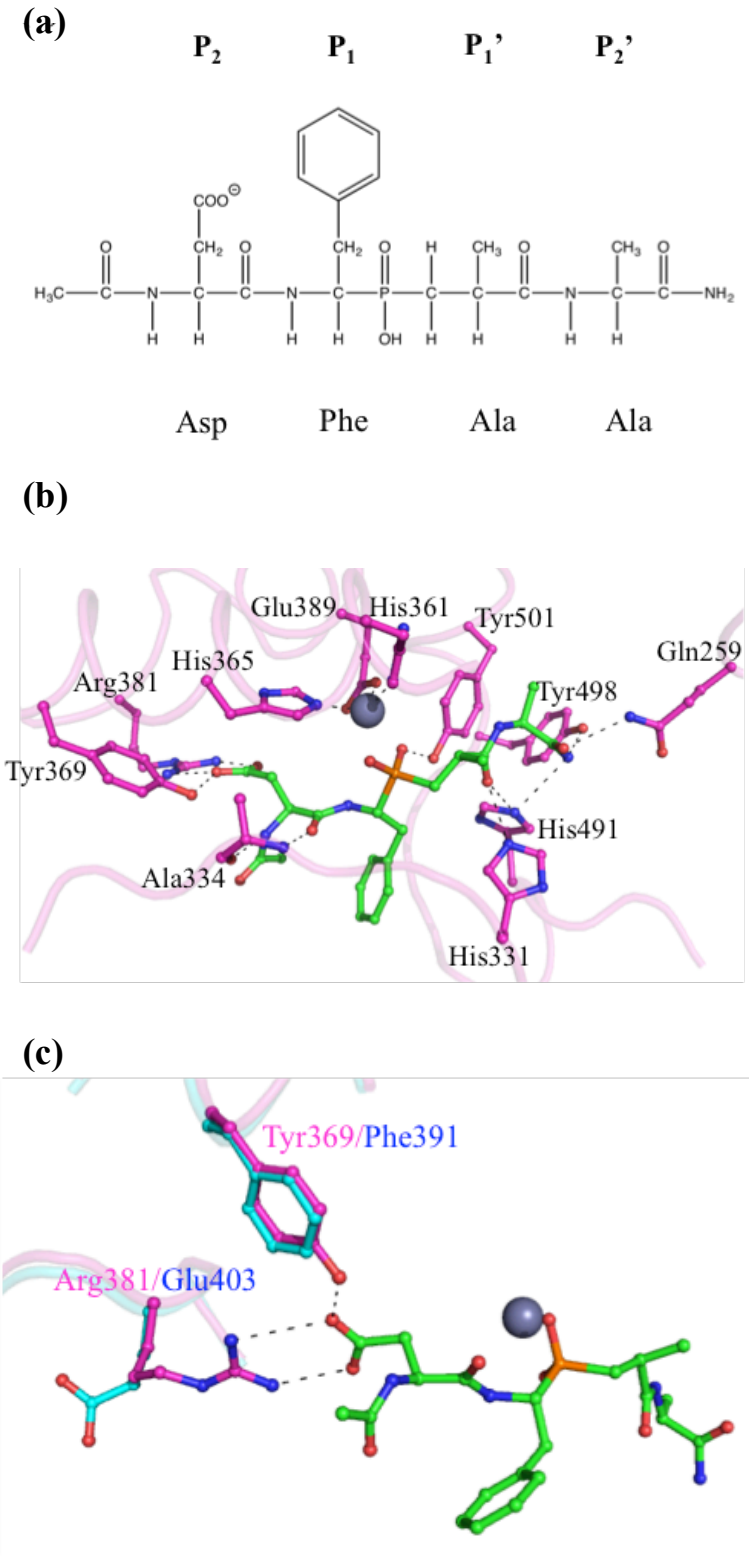


Figure 1.10: Binding of the N-domain Specific Inhibitor RXP407 to ACE (PDB code 3NXQ). (a) The structure of RXP407 indicating the key moieties. (b) Binding of RXP407 to the N-domain of ACE. RXP407 is shown as green sticks and interacting N-domain residues as magenta sticks. The catalytic zinc ion is a grey sphere. (c) Key residues interacting with RXP407 that are not conserved in the C-domain. Residues from the N-domain are shown in magenta and corresponding C-domain residues in cyan. The residues are labelled in the colours corresponding to the domain to which they belong.

At the other end of the channel lies the S_2' pocket, which, in tACE is lined by Val379 and Val380. These two valine residues contribute to a hydrophobic environment for the proline and tryptophan residues at the P_1' and P_2' positions, respectively, of the inhibitor. In the N-domain Val379 and Val380 are substituted by Ser357 and Thr358 respectively. The hydrophobic nature of this pocket is thus not conserved and suggests that the presence of hydrophobic residues at the P_1' and P_2' positions is an important factor in C-domain selectivity. Furthermore, the structure illustrates that the proline and tryptophan residues are not fully occupying this active site pocket, thus, adding further bulky, hydrophobic groups to inhibitors at this position may further increase C-domain selectivity. RXPA380 binding to the tACE active site and a comparison with the N-domain is shown in Figure 1.9.

1.6.5.3 RXP407

A second phosphinic peptide, RXP407, has been developed (Dive *et al.,* 1999) which is a selective inhibitor of the N-domain of sACE and the structure of this complex was reported in 2010 (Anthony *et al.,* 2010). Comparing this structure to the structure of tACE helps to explain the observed N-domain selectivity.

As illustrated in Figure 1.10, RXP407 forms twelve hydrogen bond interactions with N-domain active site residues, ten of which would be conserved in the C-domain. The two differences are both located in the S_2 subsite, where Glu403 and Phe391 in the C-domain replace N-domain residues Arg381 and Tyr369, respectively. The fact that these key distinguishing interactions are also located in the S_2 pocket highlights this as an important region that could be exploited in the development of domain specific inhibitors.

1.7 A New Generation of ACE Inhibitors

As discussed in 1.5.1, although current ACE inhibitors are an effective treatment for hypertension and have been in use for many years, they are not perfect. Side effects caused by the inhibition of ACE activity in other physiological processes, notably the destruction of bradykinin, are a significant problem.

It has been established that the C-domain of ACE is sufficient for *in vivo* regulation of blood pressure meaning that it would theoretically be possible to selectively target this domain and leave the N-domain active, preventing the accumulation of bradykinin. Structural studies of the binding of two phosphinic peptide inhibitors to the N- and C-

domains of human sACE have provided proof of principle data that such domain selective inhibitors could be designed.

1.7.1 Future Directions and Project Aims

Much research effort is now focused on using a structure-based approach to develop domain-specific inhibitors. A key factor in this is considering natural compounds that display some level of ACE inhibition. This includes peptides isolated from the venom of the snake *B. jararaca*, peptides derived from haemoglobin that are known as haemorphins and some flavonoids, polyphenol compounds isolated from plant extracts. This is the basis of the work discussed in Chapters 3, 4, 5 and 6. Full project aims are outlined in section 1.11.

1.8 Insect ACE homologues

The function of ACE in mammalian blood pressure regulation is well known and has been extensively studied for a number of decades. It is however, becoming increasingly apparent that ACE is involved in a wide-range of physiological processes outside of the RAAS, as illustrated by the discovery of an ever-increasing number of additional ACE substrates such as GnRH and AcSDKP.

The diverse range of functions of ACE has been brought to further prominence by the discovery of angiotensin-converting enzymes outside of mammals, notably in insects. In 1994 an ACE-like enzyme was identified in the common housefly *Musca domestica* (Lamango and Isaac, 1994) and subsequently ACE homologues have been found in every insect genome that has been sequenced (Isaac *et al.,* 2007). The conservation of angiotensin-converting enzymes throughout insects is indicative of an essential physiological role, however, insects do not have a circulatory system resembling that found in mammals. Hence, it is clear that insect ACE-like enzymes must have other essential functions. As yet it is not clear what these may be, although some proposals have been made, as discussed in section 1.9.1.1.

1.9 ACE Homologues from *Drosophila melanogaster*

The *D. melanogaster* genome codes for six potential ACE-like proteins, however based on sequence analysis it appears that only two of these would be functional, with the others lacking some essential residues. These two enzymes are known as AnCE and ACER (Cornell *et al.,* 1995; Houard *et al.,* 1998).

1.9.1 AnCE

AnCE has been studied quite extensively over the last twenty years and has provided valuable information not only about the functions of ACE homologues in insects, but also about the structural basis of ACE interactions with substrates and inhibitors.

AnCE is a single domain enzyme that shares 60 % overall sequence similarity and 40 % overall sequence identity with each of the catalytic domains of human sACE (sequence alignment shown in Figure 1.11).

Interestingly, close analysis of the sequence around the active site binding pockets indicates that AnCE shares features with both the N and C-domains of sACE and does not appear to resemble one domain more closely than the other.

sACE and tACE are both localised to the surface of endothelial cells due to the presence of a C-terminal transmembrane domain. As illustrated in Figure 1.11, this domain is not present in AnCE. Instead, there is a seventeen-residue signal sequence at the N-terminus that causes AnCE to be a secreted protein.

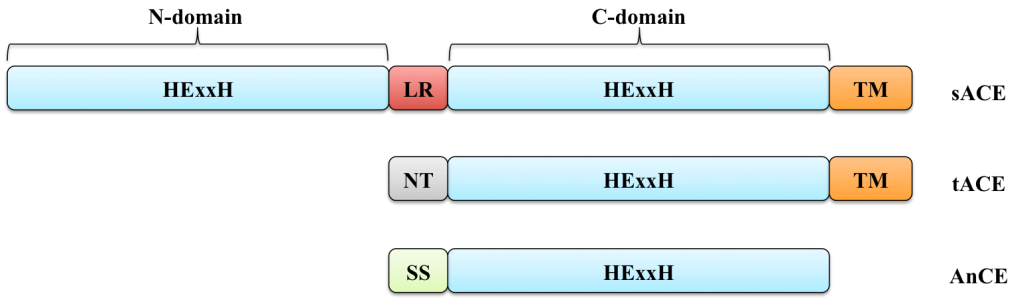


Figure 1.11: Domain Organisation in *Drosophila melanogaster* AnCE. The conserved peptidase domain is shown in pale blue and labelled "HExxH", which is the characteristic zinc-binding motif. Like tACE, AnCE consists of a single catalytic domain. However it lacks a C-terminal transmembrane domain (TM) and instead has an N-terminal secretion signal (SS), indicating that unlike sACE and tACE it is not membrane bound, but is instead secreted.

1.9.1.1 Possible Functions of AnCE

Like ACE, AnCE tends to function as a dipeptidyl carboxypeptidase, although peptidase and tripeptidase activity at both the N- and C-termini has been observed on occasion (Isaac *et al.,* 2007; Lamango *et al.,* 1996; Siviter *et al.,* 2002). AnCE has been shown *in vitro* to hydrolyse two of the main substrates of human ACE, angiotensin I and bradykinin. However given that no homologous peptides have been identified in *D. melanogaster*, it is still not clear what the true biological function of AnCE is.

AnCE must be involved in a physiological process that is essential for survival, as *Ance* mutant embryos die during early larval stages (Isaac *et al.,* 2007). Based on observed high levels of *Ance* expression in the gut epithelium and amniosera, it has been suggested that AnCE may have a role in processing peptides that are required for contraction of gut and heart muscle (Cornell *et al.,* 1995; Tatei *et al.,* 1995).

In addition to this, AnCE has also been found to be concentrated around the reproductive organs in both male and female flies, immediately suggesting a role in reproduction. A role for AnCE in spermatogenesis has been proposed, based on the fact that male *D. melanogaster* that are homozygous for a hypomorphic *Ance* allele are infertile and that this appears to be the result of sperm failing to develop properly (Isaac *et al.,* 1998; Isaac *et al.,* 1999; Schoofs *et al.,* 1998).

AnCE expression has also been shown to be concentrated in the secondary cells of the accessory glands of the testis. These glands produce peptides that mix with the sperm and seminal fluid and induce behavioural and physiological changes in the female after mating. The expression of AnCE in this region therefore suggests that it may have a role in the female after mating. To this end, Dup99B has been suggested as a potential AnCE substrate. This is because following the removal of the C-terminal dipeptide, Dup99B can induce egg laying in females (Rylett *et al.,* 2007).

Interestingly, other insect ACE homologues have also been implicated in reproduction. For example, an ACE-like enzyme in the mosquito *Anopheles stephensi* is required for egg laying where is it thought to process a peptide required for contraction of the oviduct (Ekbote *et al.,* 1999). It seems increasingly likely that insect ACE homologues have some kind of a conserved role in reproduction (Ekbote *et al.,* 2003a; Ekbote *et al.,* 2003b; Hurst *et al.,* 2003; Macours and Hens, 2004; Vercruysse *et al.,* 2004; Wijffels *et al.,* 1996).

1.9.1.2 AnCE as a Structural Model

Whilst the study of AnCE continues to provide invaluable information pertaining to the functions of angiotensin converting enzymes outside of blood pressure regulation, AnCE has also been used extensively as a structural model for human ACE, and it is in that respect that it has been used in the work presented here.

As discussed in section 1.6.1, human ACE is a very challenging enzyme to work with, namely due to its extensive glycosylation and membrane localisation. Although minimally glycosylated truncated constructs lacking the transmembrane domain have been used successfully in structural studies, the degree of glycosylation that is essential for active, stable protein still causes problems with heterogeneity during crystallisation and necessitates the use of expensive, higher eukaryotic systems for protein expression.

AnCE has been successfully expressed using the methylotropic yeast *Pichia pastoris*, and the purified protein yields consistent, diffraction quality crystals. The first crystal structure (form I) of AnCE was reported in 2003 (Kim *et al.,* 2003) and has since been followed by a series of higher resolution structures in a different crystal form (form II) (Akif *et al.,* 2010) of both the native enzyme and in complex with a wide range of inhibitors and substrates.

Given the high degree of overall sequence similarity between AnCE and each domain of sACE, it is not surprising that the overall topology of the structures is very similar. AnCE is also predominantly helical, consisting of twenty-one α-helices, nine 3_{10} helices and only two short β-strands (compared with twenty-seven α-helices and six short β-strands in tACE), which together form the ellipsoid shaped molecule that is now characteristic of angiotensin-converting enzymes, Figure 1.12.

Glycan chains are visible at three N-linked glycosylation sites in the form II structure; Asn53, Asn196 and Asn311, all of which are located on the surface of the molecule. The sequence alignment in Figure 1.13 indicates that the glycosylation site at Asn53 is conserved in the two domains of human sACE.

Like ACE, AnCE activity has also been reported to be sensitive to chloride concentration (Williams *et al.,* 1996), however unlike ACE no chloride ions are present in the AnCE structure. The Cl1 binding site that is found only in the C-domain is not conserved in AnCE, with chloride binding residues Arg186, Trp485 and Arg489 replaced by Tyr170, Phe469 and Arg473 respectively. The Cl2 chloride binding site is found in both domains of human sACE and chloride co-ordination is via Arg522/Arg500 and Tyr224/Tyr202 in the N-/C-domains, respectively. Surprisingly, as illustrated by the sequence alignment in Figure 1.12, both of these residues are conserved in AnCE as Arg506 and Tyr208, yet no chloride ion is visible in any of the AnCE structures that have been reported (Akif *et al.,* 2010).

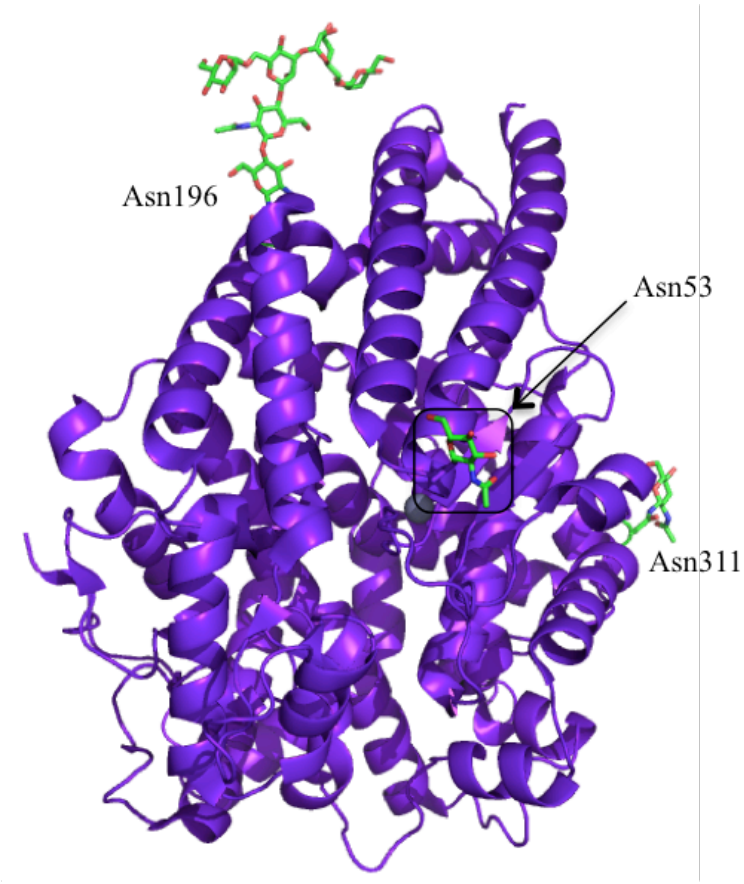


Figure 1.12: The Structure of *Drosophila melanogaster* AnCE (PDB code 2X8Y). The overall topology of the structure is well conserved from human ACE (Figures 1.5 and 1.6) and is predominantly helical. The catalytic zinc ion (grey sphere) marks the location of the active site at the centre of the substrate binding channel, capped by three helices forming an "N-terminal lid". N-linked glycosylation, shown as green sticks, is visible at Asn53, Asn196 and Asn311.

The structure of AnCE in complex with a wide range of ACE inhibitors has been reported, including lisinopril, angiotensin II and the phosphinic peptide inhibitors RXPA380 and RXP407. In all cases the general mode of binding is conserved from human ACE, indicating that AnCE is a suitable model for studying the structural basis of ACE inhibition and substrate specificity. It is in this manner that AnCE has been used in the work presented here, in order to study the binding of novel ACE inhibitors such as the natural inhibitory peptides from *B. jararaca* venom (Chapter 3) and the haemorphin derivatives spinorphin and tynorphin (Chapter 4).

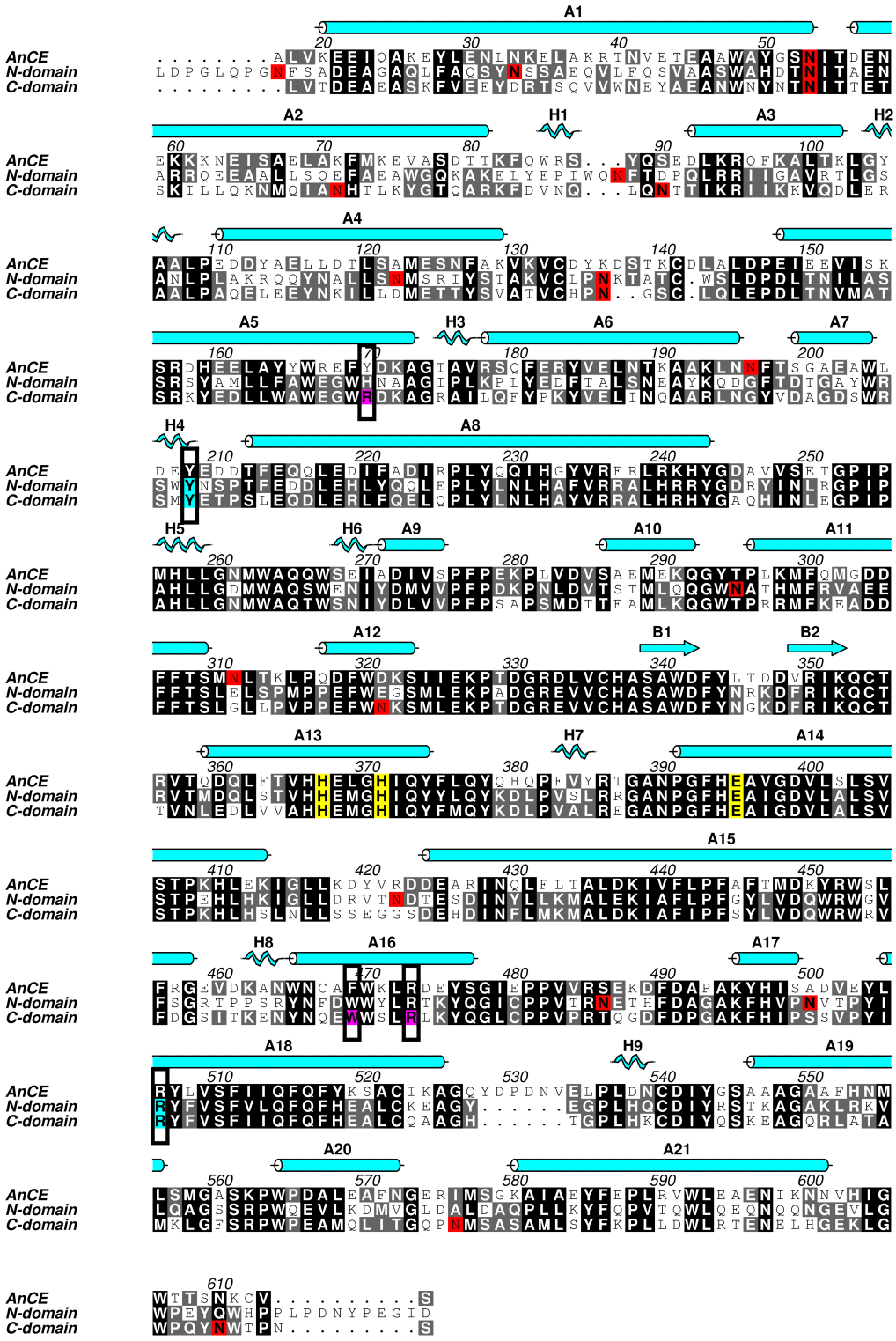


Figure 1.14: Sequence Alignment for AnCE and the N- and C-domains of Human sACE. Conserved zinc binding residues are highlighted in yellow. Putative glycosylation sites are marked in red, which highlights the conservation of AnCE site Asn53. Cl1 binding residues are shown in magenta and Cl2 in cyan, illustrating that the latter are conserved in AnCE even though no chloride is present in the structure. AnCE secondary structure elements are labelled. Residue numbering is for AnCE. Figure prepared using Aline (Bond and Schüttelkopf, 2009).

1.9.2 ACER

The second ACE homologue produced by *D. melanogaster* is distinct from AnCE in terms of both catalytic and physiological activity. The genes coding for these two enzymes show very different expression patterns, with *Acer* expression seen in neither the amniosera nor the midgut, but in heart cells, where it is known to be required for both heart development and heart function in adult flies (Houard *et al.,* 1998; Liao *et al.,* 2014; Taylor *et al.,* 1996). Further roles for ACER have also been suggested in the regulation of circadian rhythms (Carhan *et al.,* 2011; Ishimoto *et al.,* 2012), however the mechanisms underpinning these processes are still not clear.

ACER shares 66 % overall sequence similarity and 48 % sequence identity with AnCE. It also has the same domain organisation; a short signal sequence at the N-terminus followed by the conserved peptidase domain. There is no evidence of a transmembrane domain, indicating that like AnCE, and in contrast to human ACE, ACER is a secreted protein, Figure 1.14.

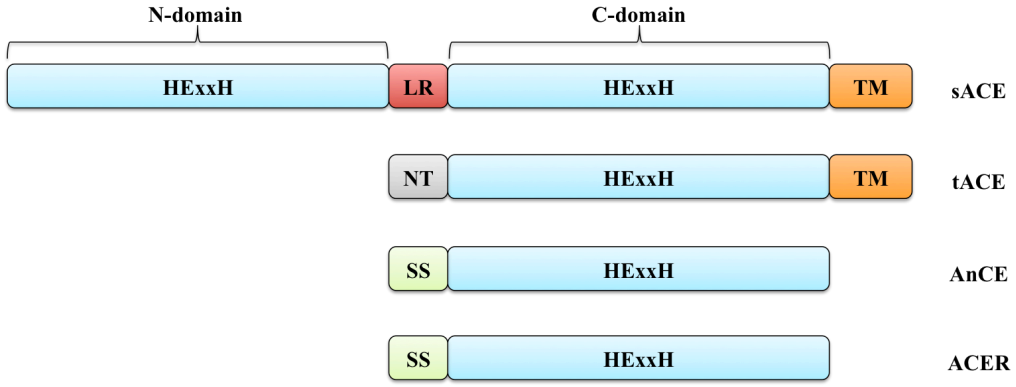


Figure 1.14: Domain Organisation in *Drosophila melanogaster* ACER. The conserved peptidase domain is shown in pale blue and labelled "HExxH", which is the characteristic zinc-binding motif. Like tACE and AnCE, ACER consists of a single catalytic domain. Like AnCE, and in contrast to sACE and tACE, it does not have a C-terminal transmembrane domain (TM) and instead has an N-terminal secretion signal (SS), indicating that ACER is a secreted rather than a membrane bound protein.

Based on the sequence alignment shown in Figure 1.15, it seems that the N-linked glycosylation site at Asn53, which is also found in the N- and C-domains of human sACE, may be conserved from AnCE, but that the other two sites, at Asn196 and Asn311, are absent.

In spite of their predicted overall similarity, AnCE and ACER do display some profound differences in substrate specificity. Crucially, whilst both are able to hydrolyse bradykinin, although AnCE does so at a significantly faster rate, only AnCE is able to cleave angiotensin I, a major substrate of human ACE. This indicates that there must be some important differences in the active site substrate binding pockets of these two enzymes (Coates *et al.,* 2000; Taylor *et al.,* 1996).

No structure has yet been reported for ACER, although a model was proposed in 2005 (Bingham *et al.,* 2005), which was used to develop hypotheses to explain the observed differences between AnCE and ACER.

The predicted conservation of glycosylation from AnCE in ACER indicates that this will be essential for protein function and stability. Hence, in order to ensure correct post-translational processing, it will be necessary to produce ACER using eukaryotic systems. Previous attempts to express ACER using *Pichia pastoris*, the system used for AnCE, have yielded mixed results. One of the aims of the work presented here is to work towards the expression of ACER using higher eukaryotic systems, with the long term aim of being able to use this protein in structural studies to try and understand the structural basis for the observed differences in substrate specificity of AnCE and ACER. This work is presented in Chapter 8.

1.10 ACE Homologues from *Anopheles gambiae*

As one of the key vectors involved in the transmission of the malaria-causing parasite *Plasmodium falciparum*, the mosquito *Anopheles gambiae* is an exceedingly important organism to study. A recent study of the *A. gambiae* genome has revealed that it codes for ten different ACE homologues, or AnoACEs. This is the highest number found in any insect genome thus far, meaning that *A. gambiae* is also an important organism to study in order to learn more about angiotensin converting enzymes, particularly in insects (Burnham *et al.,* 2005; Holt *et al.,* 2002).

The large number of ACE homologues in *A. gambiae* indicates that angiotensin converting enzymes are likely to be involved in a wide range of physiological processes. By studying these, it may be possible to learn more about the functions of ACE homologues both in insects and in mammals outside of the traditional roles in blood pressure homeostasis.

All of the ACE-like enzymes identified in insects previously are single domain, secreted enzymes like AnCE and ACER. Intriguingly, one of the ten enzymes from *A. gambiae*, AnoACE9, appears to consist of two catalytic domains arranged in tandem, as seen in human sACE (Burnham *et al.,* 2005). This discovery indicates that this arrangement has been selected for on more than one occasion, therefore must surely be in some way advantageous. By studying AnoACE9 from *A. gambiae* it may be possible to begin to understand the biological benefits of such an arrangement.

As outlined in section 1.9.1.1, there is a considerable amount of evidence implicating insect ACE homologues in reproduction. If this is indeed a conserved role adopted by one or more of the enzymes from *A. gambiae*, then it may ultimately be possible to design ACE inhibitors specific for AnoACE. These could potentially be used in insecticides to help control the mosquito population and consequently the spread of malaria.

At this stage, work on these enzymes has been limited to a single genomics based study of gene expression patterns (Burnham *et al.,* 2005). Hence, there is no indication at present as to which, if any, of the AnoACEs may have a role in reproduction.

The aim of the work on ACE homologues from *A. gambiae* that is presented in this thesis is to focus on one of the enzymes, AnoACE3, and to work towards its expression and purification. The long-term aim would then be to use this enzyme for structural and activity based studies in order to learn more about its functions and catalytic properties. Ultimately, it would be helpful to be able to do this for all of these enzymes, however AnoACE3 was chosen as a starting point because it is one of the three AnoACEs for which the expression is upregulated following a blood meal, indicating that it is likely to have an important biological role of some description (Burnham *et al.,* 2005).

1.10.1 AnoACE3

The domain organisation of AnoACE3 is shown in Figure 1.16. Like AnCE and ACER it is predicted to be a secreted protein with a single catalytic domain. This is indicated by the presence of a signal sequence towards the N-terminus and a conserved peptidase domain, including zinc binding motif, downstream of this. There is also an additional sequence at the very N-terminus of the protein that is conserved neither in AnCE nor in ACER. The function of this sequence is completely unknown and database searches have failed to identify any closely related sequences that may provide an indication of its function.

The sequence alignment in Figure 1.15 highlights the fact that the three N-linked glycosylation sites in AnCE all appear to be conserved in AnoACE3. This indicates that a eukaryotic system will be required for AnoACE3 expression so as to produce correctly folded, functional protein.

In the first instance this will be attempted using *Pichia pastoris,* as this has previously been used successfully for AnCE expression. However given the difficulties that have been encountered with ACER expression in this system, it may be necessary to move towards higher eukaryotic systems as has been the case for the expression of the human enzyme.

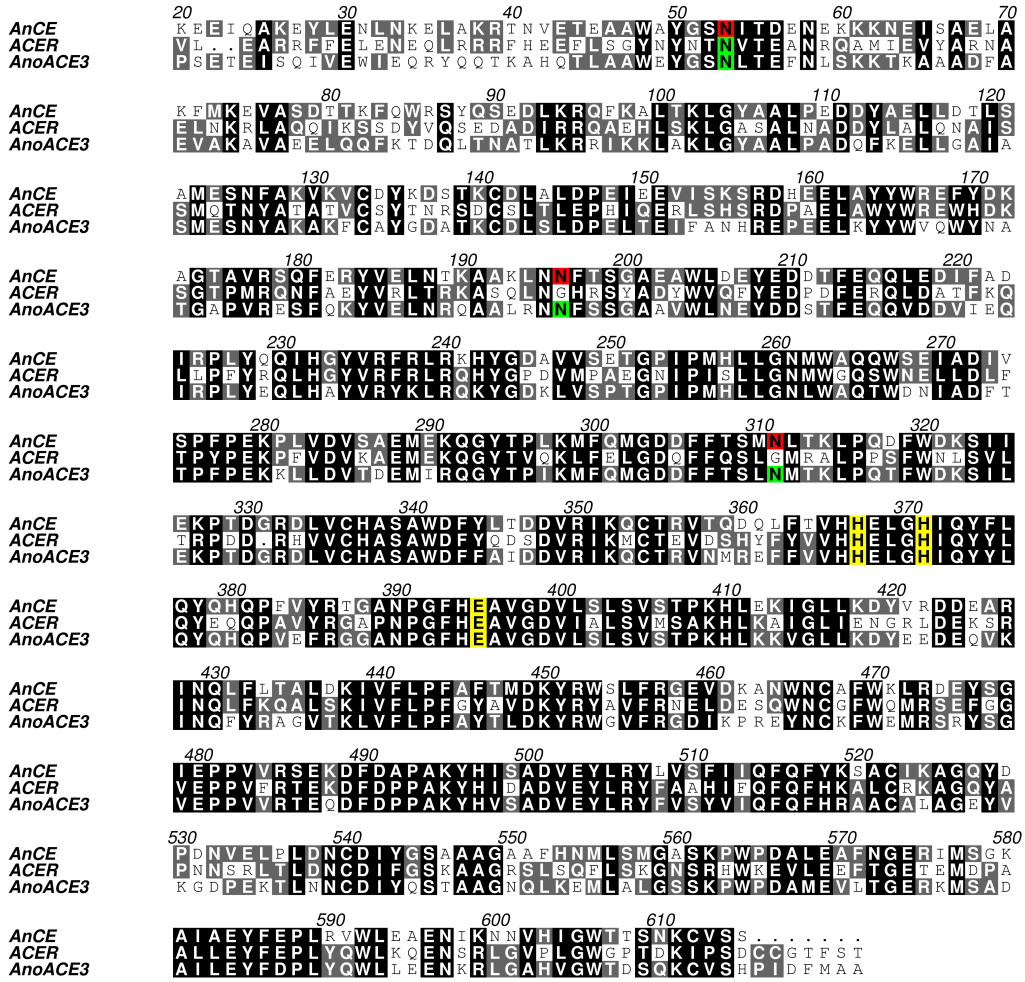


Figure 1.15: Sequence Alignment of AnCE and ACER from *Drosophila melanogaster* and AnoACE3 from *Anopheles gambiae*. Alignment is for the conserved peptidase regions only and residue numbers are based on the AnCE sequence. Conserved zinc binding residues are highlighted in yellow. The three N-linked glycosylation sites found in AnCE are shown in red and highlighted in green where they are predicted to be conserved in ACER/AnoACE3. Figure prepared using Aline (Bond and Schüttelkopf, 2009).

Figure 1.16: Domain Organisation of AnoACE3 from *Anopheles gambiae*. AnoACE3, like AnCE and ACER, is predicted to have a single catalytic domain, shown in pale blue and marked by the conserved HExxH motif. It is also predicted to be secreted, shown by the signal sequence (SS, pale green) at the N-terminus, also found in AnCE and ACER. Compared to AnCE and ACER, AnoACE3 contains an additional sixty residue sequence at its N-terminus, the function of which is unknown. This is shown in purple and labelled "X".

1.11 Summary of Project Aims

- To use the *Drosophila melanogaster* ACE homologue AnCE as a model to study the structural basis of ACE inhibition by a range of different inhibitors, particularly focusing on how interactions between inhibitors and the active site could be exploited in the development of domain specific inhibitors.

- To work towards the expression of two further insect ACE homologues: ACER from *Drosophila melanogaster* and AnoACE3 from *Anopheles gambiae*. This will involve the use both of the methylotropic yeast *Pichia pastoris* and higher eukaryotic expression systems.

Chapter 2: General Methods

2.1 Chemicals and Reagents

Unless otherwise stated, all chemicals and reagents were purchased from Sigma-Aldrich (MO, USA).

2.2 Cell Lines

2.2.1 Bacterial Strains

In the work presented here, all of the proteins were expressed using eukaryotic systems; hence bacterial strains were used only during cloning steps. For work performed at the University of Bath *E. coli* One Shot Top 10 chemically competent cells (Invitrogen™) were used for cloning and propagation. For work performed at the OPPF (Oxford Protein Production Facility, Harwell Science and Innovation Campus, Didcot, Oxon, UK) One Shot® OmniMAX™ 2 T1^{R} chemically competent *E. coli* cells (Life Technologies) were used for this purpose.

2.2.2 *P. pastoris*

All yeast expression work was performed using the *P. pastoris* GS115 strain purchased from Invitrogen.

2.2.3 Higher Eukaryotic Cell Lines

The human embryonic kidney 293T/17 (HEK293T) cell line purchased from ATCC (American Type Culture Collection) and the *Spodoptera frugiperda Sf*9 cell line purchased from Life Technologies were used for protein expression in higher eukaryotic systems.

2.3 Expression Plasmids

Vector maps for all of the plasmids used in this work are included in Appendix A.

2.3.1 *P. pastoris*

The pPIC9 plasmid (Invitrogen) was used for AnCE expression in *P. pastoris*. Key features of this plasmid include an ampicillin resistance gene for selection of positive transformants during initial cloning stages in *E. coli*, an α-factor secretion signal for secreted expression of the target protein and the *AOX1* promoter for methanol-induced target protein expression.

Trials of AnoACE3 expression in *P. pastoris* were performed using the pPICZα vector (Invitrogen). The pPICZα vector retains all of the key features of the pPIC9 vector whilst including an optional C-terminal polyhistidine tag.

2.3.2 Higher eukaryotes

Expression in HEK293T and *Sf*9 cells was performed using plasmids from the pOPIN vector suite kindly provided by Prof. Ray Owens at the OPPF. A summary of these vectors and their key features is presented in Table 2.1.

Vector	Affinity tag	Secretion signal	Antibiotic resistance	Promoter	Expression systems
pOPINEneo	C-terminal 8×His	None	Ampicillin	p10	*E. coli* Insect Mammalian
pOPINTTGneo	C-terminal 8×His	µ-phosphatase (RTPTµ)	Ampicillin	p10	*E. coli* Insect Mammalian
pOPIN-HBM	C-terminal 8×His	Honeybee Melittin (HBM)	Ampicillin	p10	*E. coli* Insect Mammalian
pOPIN-HA	C-terminal 8×His	Influenza A Haemagglutinin (HA)	Ampicillin	p10	*E. coli* Insect Mammalian
pOPIN-Ac64	C-terminal 8×His	*Autographa californica* nuclear polyhedrosis virus gp64 (Ac64)	Ampicillin	p10	*E. coli* Insect Mammalian

Table 2.1: Features of the Vectors used for Protein Expression in Higher Eukaryotes.

2.4 Cell Culture

2.4.1 *E.coli*

Three different types of media were used for *E. coli* growth: lysogeny broth (LB: 1 % w/v tryptone, 0.5 % w/v yeast extract, 0.5 % w/v NaCl), super optimal broth with catabolite repression (SOC: 2 % w/v tryptone, 0.5 % w/v yeast extract, 0.05 % w/v NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) and power broth (Athena Enzyme Systems, MD, USA). The medium type used for specific steps is clearly stated in the relevant methods sections. For growth in petri dishes the chosen media was prepared with the addition of 1 % agar. Where stated, medium was supplemented with either 100 µg ml^{-1} ampicillin or 50 µg ml^{-1} carbenicillin for selection.

2.4.2 *P. pastoris*

For protein expression *P. pastoris* was initially grown in buffered glycerol-complex medium, BMGY (1 % yeast extract, 2% tryptone, 100 mM potassium phosphate pH 6.0, 1.34 % yeast nitrogen base (YNB), 4×10^{-5} % biotin, 1 % glycerol). In order to induce expression this was replaced with buffered methanol-complex medium, BMMY (1 % yeast extract, 2% tryptone, 100 mM potassium phosphate pH 6.0, 1.34 % YNB, 4×10^{-5} % biotin, 0.5 % methanol).

When individual colonies were required, typically for inoculation of expression cultures, yeast extract peptone dextrose, YPD medium (1 % yeast extract, 2 % tryptone, 2 % dextrose) with 2 % agar was used to enable growth in petri dishes.

A number of other types of media were used for specific purposes. These are detailed as they arise in the text.

2.4.3 HEK293T Cells

HEK293T cells were cultured at 37 ^{o}C in a 5 % CO_2/95 % air environment in T75 or T25 tissue culture plates (Corning Inc.). Cells were maintained in Dulbecco's modified eagle medium (DMEM, Life Technologies) supplemented with 10 % foetal bovine serum (FBS), 1 % L-glutamine and 1 % non-essential amino acids. Following transfection, the serum content of the medium was reduced to 2 % FBS.

2.5 General Analytical Methods

2.5.1 Agarose Gel Electrophoresis of DNA Samples

1 % agarose gels were prepared by addition of agarose to Tris acetate EDTA (TAE) buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA). DNA samples were prepared for analysis by addition of 6 × blue/orange loading dye (Promega, WI, USA) in a 1:6 ratio. Electrophoresis was typically carried out for 45-60 minutes at 100 V in TAE buffer. A 1 kb or 100 bp DNA ladder standard (Promega, WI, USA) was run alongside samples to facilitate accurate estimation of their size.

2.5.2 SDS-PAGE of Protein Samples

Protein samples were prepared for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) by mixing with an equal volume of 2 × Laemmli sample buffer (0.01 % bromophenol blue, 4 % w/v sodium dodecyl sulphate (SDS), 20 % glycerol, 120 mM Tris-HCl, pH 6.8) and heating to 95 ^{o}C for 10 minutes. 15 µl of the

prepared sample were loaded into the wells of a 12 % bis-tris acrylamide gel. Gels were run at 180 V for 60 minutes in a Mini-Protean tetra cell (Bio-Rad, CA, USA) containing approximately 1 litre of running buffer (50 mM MOPS, 40 mM Tris-HCl, 1 mM EDTA, 0.1 % w/v SDS, pH 7.7). A 2-212 kDa molecular weight marker (New England BioLabs Inc.) was run concurrently to facilitate the accurate estimation of sample molecular weight.

Gels were visualised by Coomassie staining (0.25 % w/v Coomassie Brilliant Blue, 10 % v/v glacial acetic acid, 45 % v/v methanol) at room temperature for 2-3 hours with agitation by rocking. After rinsing with distilled water a destaining solution (25 % methanol, 7 % glacial acetic acid) was added prior to incubation at room temperature with agitation for a further 2-3 hours until bands on the gel were clearly visible.

2.5.3 Western Blotting

SDS-PAGE was performed as detailed above but with the addition of the PageRuler prestained protein ladder (Thermo Fisher Scientific, MA, USA) in place of the 2-212 kDa molecular weight marker. The gel was transferred to a nitrocellulose membrane (0.45 μm pore size, GE Healthcare, UK) in 25 mM Tris base, 192 mM glycine, 20 % methanol, pH 8.8 at 150 mA for 60 minutes. Following transfer membranes were blocked in a solution of 5 % milk in PBS (phosphate buffered saline) + 1 % Tween either for 1 hour at room temperature or for 16 hours at 4 ^{o}C with gentle agitation by rocking. Primary antibodies, conjugated with horseradish peroxidase (HRP), were prepared at the appropriate dilution in the blocking solution and added to the membrane. The membrane was then incubated with gentle agitation either for 1 hour at room temperature or for 16 hours at 4 ^{o}C.

After incubation with the primary antibody the membrane was washed for 3 × 10 minutes in PBS + 1 % Tween prior to application of the Pierce^{TM} ECL western blotting substrate (Thermo Fisher Scientific, MA, USA). The membrane was incubated with the substrate for 5 minutes at room temperature after which time excess substrate was removed by blotting with tissue prior to detection.

2.6 General Methods in Structural Biology

2.6.1 Introduction

Structural biology techniques enable the visualisation of biological molecules at the atomic level. Thus, they are exceedingly powerful methods that can provide a wealth of information about the molecular mechanisms underpinning biological processes. The

application of structural biology techniques to proteins has been central to many medical advances. A key example of this is the development of insulin analogues for the treatment of diabetes (Pandyarajan and Weiss, 2012).

Structural biology techniques have been especially useful in drug design and development when applied to enzymes (Hardy and Malikayil, 2003; Williams *et al.,* 2005). Key examples of drugs developed using a structure-based approach include HIV protease inhibitors (Turner *et al.,* 1998), neuraminidase inhibitors such as Tamiflu (Oseltamivir) and Relenza (Zanamivir) (Stoll *et al.,* 2003) and angiotensin converting enzyme inhibitors (Acharya *et al.,* 2003), the latter of which represents a signficiant part of the work presented in this thesis.

A number of methods exist for determining protein structures, the most widely used being electron microscopy (EM), nuclear magnetic resonance (NMR) and X-ray crystallography. Traditionally, EM has only been suitable for use on very large macromolecules and is limited in terms of the maximum resolution that can be achieved. However, recent developments have seen the number of EM entries in the Protein Data Bank (PDB) more than double in the last four years with structures now being reported at close to 3 Å resolution (Amunts *et al.,* 2014; Fischer *et al.,* 2015; Yu *et al.,* 2011). Conversely, NMR is only suitable for use with much smaller proteins, typically less than 35 kDa. This technique is also developing rapidly though and is being used increasingly for larger proteins.

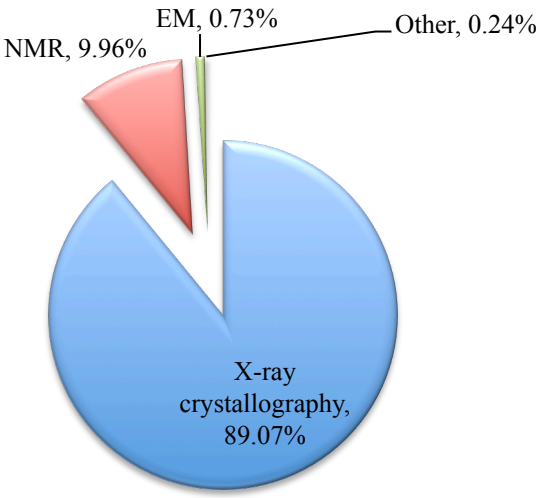


Figure 2.1: Percentage of Protein Data Bank Entries by Experimental Technique. Data correct as of August 2015.

In spite of the recent developments in NMR and EM, X-ray crystallography remains the most widely used method for structural studies of macromolecules. Of the ~111, 000 entries listed in the Protein Data Bank in August 2015, ~99, 000 were determined by X-ray crystallography, compared to ~ 11, 000 and ~ 800 for NMR and EM respectively (Figure 2.1). The main advantages of X-ray crystallography include the fact that it can be applied to samples ranging from small molecules to large macromolecular complexes, and that unlike NMR where samples must be labelled with isotopes, X-ray crystallography does not necessarily involve any such modifications, which can make sample preparation more straightforward and less expensive.

The main drawback of X-ray crystallography is that it is entirely dependent on the formation of diffraction quality crystals. Although a plethora of techniques exist for producing and optimising such crystals, it is still far from an exact science and there is no way of predicting under which conditions a given molecule will crystallise. In spite of this drawback, X-ray crystallography does remain the dominant method in structural biology and it is the method that was employed to obtain the structural data reported in this work.

The different processes involved in protein crystallography projects are described in the following sections. The flow chart in Figure 2.2 outlines these steps and the order in which they occur.

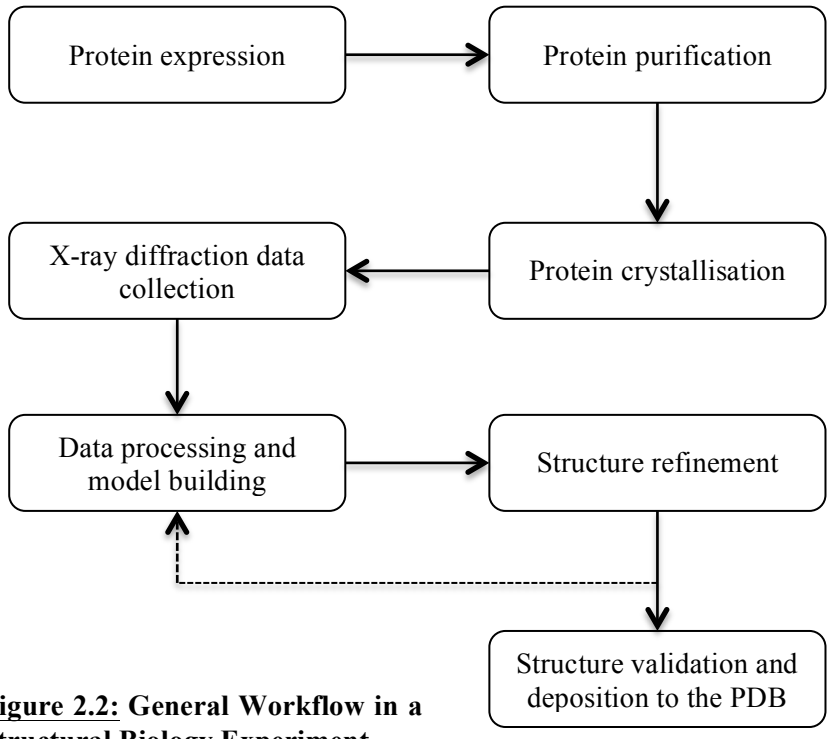


Figure 2.2: General Workflow in a Structural Biology Experiment.

2.6.2 Sample Preparation

2.6.2.1 Protein Expression

The major requirement for protein crystallography is a large quantity of pure protein. Initially, this was very often a limiting factor as target proteins had to be isolated from their original source organism. However, the advent of recombinant DNA technology has revolutionised this step with the vast majority of protein structures now being solved using proteins produced by recombinant expression.

Escherichia coli remains the most commonly used system for recombinant protein expression. As of August 2015, 77 % of all Protein Data Bank (PDB) entries cite using *E. coli* as the expression system. The major advantages of this system being that as a prokaryote genomic manipulation is relatively straight forward, it grows rapidly, is relatively inexpensive and can yield large quantities of recombinant protein. The major disadvantage of using *E. coli*, and indeed other prokaryotic systems, is that they are unable to perform many of the post-translational modifications found on eukaryotic proteins and lack chaperones that may be required for correct protein folding. It has been estimated that 65 % of human proteins do not show soluble expression in *E. coli* (Hirose *et al.,* 2011), which is particularly problematic given than human proteins are so often the targets of research for drug development.

As more protein structures are published, research is increasingly moving to focus on more challenging targets, such as large complexes, membrane proteins and those which undergo extensive post-translational modifications. This often necessitates the use of eukaryotic systems for recombinant protein expression.

One such system is the yeast *Pichia pastoris*. *P. pastoris* is in many ways an ideal expression system. It is a eukaryote, which means that it can carry out post-translational modifications, but compared to higher eukaryotes it is relatively inexpensive to grow and easy to genetically manipulate. It can be grown in a laboratory environment in shake flasks in a very similar manner to *E.coli*. *P. pastoris* is often chosen over other yeasts such as *Saccharomyces cerevisiae* due to the potential for obtaining higher protein yields.

P. pastoris is a methylotropic yeast, meaning that it can grow using methanol as its sole carbon source. It is this feature that is exploited in recombinant protein expression. The first step in methanol metabolism by *P. pastoris* is oxidation by the enzyme alcohol oxidase to form formaldehyde. In the presence of methanol and the absence of glycerol,

transcription of the alcohol oxidase gene, *AOX1* is initiated. Recombinant protein expression can be coupled to the *AOX1* promoter, meaning that target protein expression can easily be induced by addition of methanol to the growth medium (Ellis *et al.,* 1985; Kouz *et al.,* 1989; Tschopp *et al.,* 1987). Following the method outlined in section 2.7.1, *P. pastoris* was used to produce the *D. melanogaster* ACE homologue, AnCE, in this work.

Whilst *P. pastoris* is a useful system in some circumstances, there are still many instances where expression by higher eukaryotes is the only option for obtaining sufficient quantities of correctly processed, functional protein. Two of the most commonly used cell lines for this purpose are HEK293T cells (human embryonic kidney) and the *Sf*9 cell line derived from the moth *Spodoptera frugiperda,* both of which were used in this work. Although there are many advantages associated with using higher eukaryotes in terms of correct processing of the target protein, these cell lines are much more challenging and expensive to work with, requiring specialised tissue culture suites.

2.6.2.2 Protein Purification

After the target protein has been expressed using the chosen system, it must then be purified. Typically the first step in a purification process will either involve ion exchange chromatography (IEX), hydrophobic interaction chromatography (HIC) or affinity chromatography. IEX involves a protein molecule with a net surface charge interacting with an oppositely charged matrix and then being eluted by increasing ionic strength. Depending on the degree of surface charge at a given pH, different molecules will interact with the matrix to different extents, thus enabling separation. A similar principle is applied in HIC. Here a protein molecule with exposed hydrophobic groups interacts with a hydrophobic matrix and is then eluted by decreasing ionic strength. As most proteins do not have an abundance of surface hydrophobic residues this technique can be a very effective method of separation.

Affinity chromatography is currently the most widely used technique for protein purification. It exploits an interaction between the protein and a specific molecule that is immobilised on a matrix. This is a ligand that the protein is able to bind to either naturally or due to the addition of an affinity tag. A common example of such a tag is the polyhistidine tag, which binds to nickel and other divalent cations. Affinity chromatography is a very powerful tool that can yield very pure protein samples even from crude cell lysates.

After an initial purification step it is quite common for a final "polishing" step to be performed to ensure that the sample is as homogenous as possible. Typically this is in the form of size exclusion chromatography (SEC) whereby molecules are separated based on their size.

After purification it is important to confirm the identity of the protein before beginning crystallisation trials. Techniques for this include but are not limited to: sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), which separates proteins based on size, western blotting, which uses antibodies to identify target proteins, and mass spectrometry, which, when combined with trypsin or pepsin digestion, can be used to determine protein sequence in a technique known as peptide mass fingerprinting.

2.6.2.3 Protein Crystallisation

Crystals consist of a regular arrangement of identical repeating units in three dimensions forming a lattice. In a perfect crystal, each repeating unit is in an identical environment forming identical non-covalent interactions with neighbouring units. The building block of the crystal lattice is known as the unit cell, and crystals are essentially repeating arrangements of identical unit cells. Unit cells are related to each other by symmetry operations, which define the space group of the crystal. There are 230 possible crystal space groups, however the chirality of protein molecules means that protein crystals can adopt only 65 of these.

Whilst crystals can be thought of as being built-up from repeating arrangements of unit cells, the unit cell is not necessarily the empirical building block of the lattice. The smallest unit from which the entire lattice can be built by applying only symmetry operations is known as the asymmetric unit, and unit cells often contain more than one asymmetric unit. It is also important to recognise that in the case of protein crystals it does not necessarily follow that the asymmetric unit comprises one protein molecule.

Obtaining diffraction quality crystals is very often the rate-limiting step in structure determination. Protein crystals will only grow under very specific conditions. These vary widely between different proteins and are impossible to predict.

The process of crystallisation can be understood by studying the phase diagram in Figure 2.3. This is an indicator of the solubility of a protein at a given concentration for a specific adjustable parameter. This is often temperature, pH or precipitant concentration. The phase diagram is divided into two main regions separated by the

solubility curve. Below the solubility curve the solution is undersaturated; protein molecules are fully solvated hence it is energetically favourable for them to remain in solution. Above the curve the solution is supersaturated and the protein will be driven out of solution in a bid to reduce the free energy of the system.

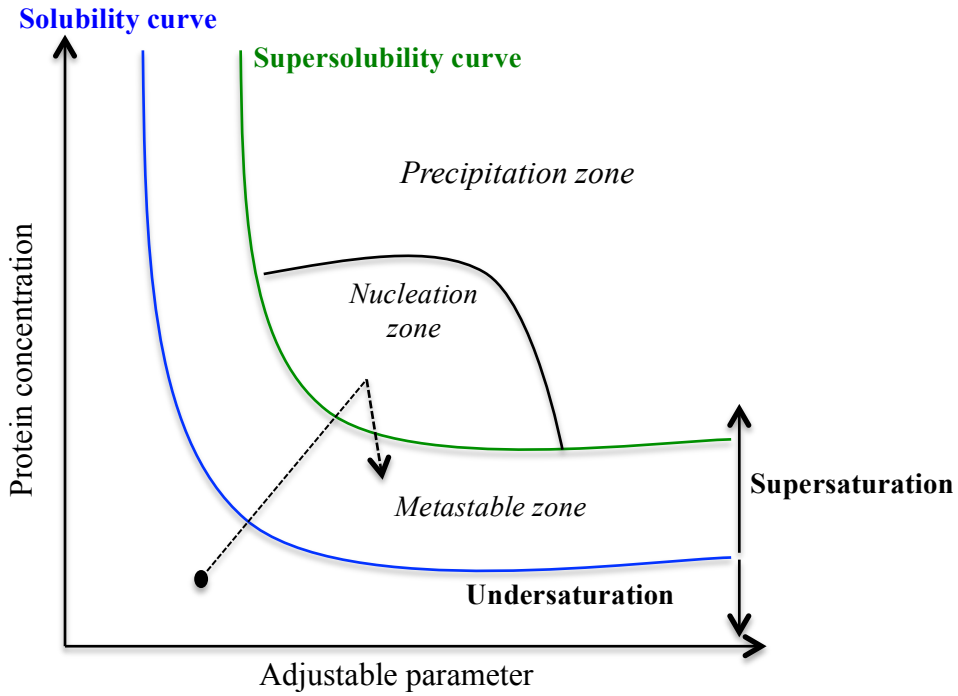


Figure 2.3: Phase Diagram for Protein Crystallisation. A diagram illustrating the effect of protein concentration and adjustable parameters on protein solubility. The plot is separated by the solubility curve (blue) into two regions. Below the curve the solution is undersaturated and the protein will remain in solution. Above the curve the solution is supersaturated. The supersaturated region can be further divided by the supersolubility curve (green). Below the supersolubility curve is the metastable zone where crystal growth can occur. Directly above this is the nucleation zone where crystal nuclei can begin to form. Above this is the precipitation zone where the solution is so saturated that the protein will precipitate out of solution. The aim of crystallisation experiments is to gradually increase the effective protein concentration to bring the solution into the nucleation zone so that crystal nuclei form and for this to then reduce the saturation bringing the solution into the metastable zone for crystal growth. This is illustrated by the dotted lines on the phase diagram.

The aim of crystallisation experiments is to gradually alter the conditions of a protein solution so that the solubility limit of the protein is exceeded and crystals can form. The critical event in crystal growth is the formation of stable nuclei that are the precursors of crystals. These will only form in a very specific region of the phase diagram known as the nucleation zone. Above this zone the solution is too saturated and the protein molecules will aggregate and form precipitate. Below this region the solution is undersaturated and crystallisation will never occur. Ideally, the occurrence of nucleation will then cause the solution to be less saturated and drop into the metastable zone. It is here that crystal growth can occur as protein molecules add to the nuclei in an ordered manner.

A number of different methods exist for achieving protein crystallisation. Currently the most commonly used is the vapour diffusion technique. Here the protein is mixed with a solution containing a precipitant to form a drop. This drop is then sealed in an enclosed chamber with a reservoir containing a higher concentration of said precipitant. This establishes a concentration gradient between the drop and reservoir. As the system moves towards equilibrium, water will gradually diffuse from the protein drop to the reservoir, increasing the effective protein concentration and gradually moving the protein solution towards the supersaturated region of the phase diagram.

A number of different variants of the vapour diffusion technique exist. The most common are hanging and sitting drop vapour diffusion, illustrated in Figure 2.4. In hanging drop vapour diffusion the protein/precipitant drop is suspended over a well containing the reservoir solution, whilst in the sitting drop set-up the drop is placed on a pedestal next to the reservoir.

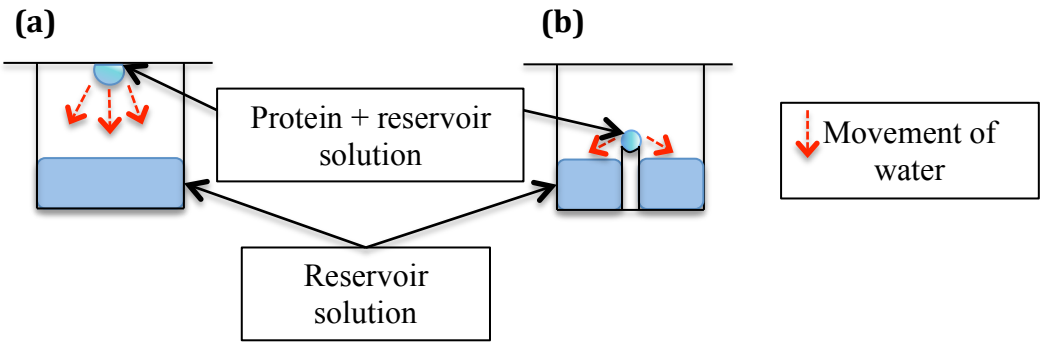


Figure 2.4: Hanging (a) and Sitting (b) Drop Vapour Diffusion Experiment Set-ups.

The conditions that will give rise to crystals vary widely between different proteins and are impossible to predict. For this reason it is now commonplace to begin by screening hundreds, if not thousands of conditions using high throughput robots and commercial screens. Experiments are checked regularly for any sign of crystal formation, at which point the conditions can be optimised to try and improve crystal growth. Typically, this involves iteratively varying parameters such as protein concentration, precipitant concentration, pH, temperature, salt and buffer.

As well as being the critical step in crystal formation, nucleation is often also the most difficult state to achieve. Seeding is a very useful technique that effectively bypasses this process. By artificially introducing nuclei into the drop, the level of supersaturation that is required is decreased, reducing the likelihood of the protein forming aggregates. There are three types of crystal seeding: macroseeding, where a single large crystal is introduced into the drop, microseeding, where small crystal fragments are introduced and streak seeding where an implement, typically a whisker, is used to transfer crystal nuclei into the drop.

2.6.3 X-ray Diffraction Data Collection and Processing

2.6.3.1 X-rays and Diffraction

X-rays are a form of electromagnetic radiation with wavelengths (λ) in the range 10^{-11} to 10^{-8} m, or 0.1 to 100 Å. Depending on their wavelength, X-rays are classified as either hard or soft, with soft X-rays having wavelengths greater than ~ 2 Å, and hard X-rays below this value. Macromolecular crystallography typically uses hard X-rays for which λ = 0.5-1.6 Å. This is the same order of magnitude as most of the relevant interatomic bonds (e.g. C-C ~ 1.5 Å) allowing visualisation of structures at the atomic level.

The wave nature of X-rays means that they can be diffracted and it is this property that is exploited in X-ray crystallography. During data collection, the majority of X-rays will pass straight through the crystal and be captured by the beam stop, providing no useful information. The X-rays that are important are those that are diffracted by atoms in the crystal. Inside the crystal many diffraction events occur, resulting in a large number of diffracted X-rays that interfere with each other. If diffracted rays are not in phase, then the result will be destructive interference and the waves will cancel each other out so that no signal is detected.

In contrast, if two waves are in phase with each other, constructive interference will occur and the signal will be amplified. This is where the repeating nature of the crystal lattice is so important. As illustrated in Figure 2.5, for there to be constructive interference between two waves their path lengths must differ by $n\lambda$, where n is a whole number integer. That is to say, that the difference between the path lengths must be a multiple of the wavelength. Applying this to a crystal lattice, this means that X-rays diffracted from planes throughout the crystal that are separated by a specific distance, d, will all be in phase. This will result in constructive interference and a greatly amplified signal that can be detected. These planes are known as Bragg planes and Bragg's law, Figure 2.5, describes this concept (Bragg and Bragg, 1913).

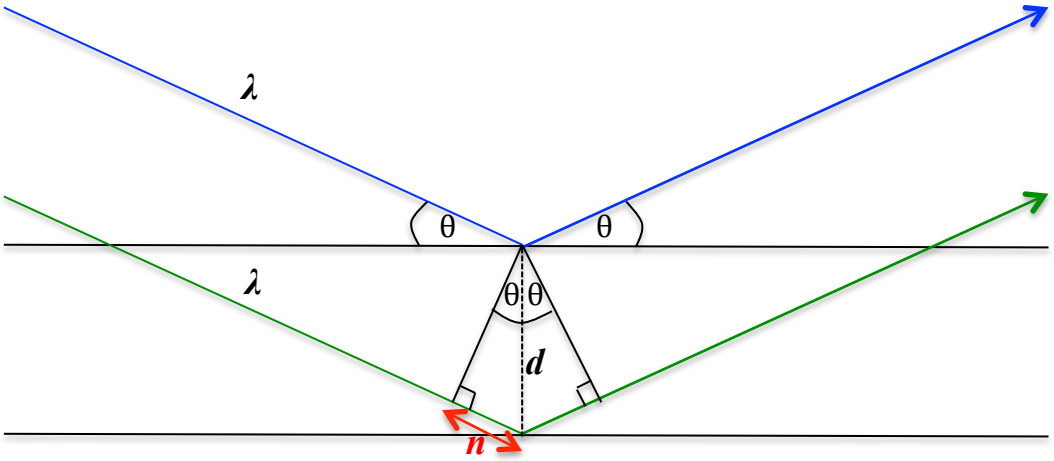


Figure 2.5: Bragg's Law. For two waves with wavelength λ to remain in phase after diffraction from two planes in a crystal separated separated by a distance d, the additional distance travelled by the second wave (green) must be an integer multiple of the wavelength. This is expressed by Bragg's law: $n\lambda = 2d \sin \theta$

The wavelength of the X-rays is unchanged following diffraction but their phase and amplitude will differ from that of the original incident wave. This is governed by the distribution of scattering matter, electron density, in the crystal. The diffracted waves can be defined by structure factors, which are complex numbers that describe their amplitude and phase.

By performing a mathematical operation known as a Fourier transform, it is possible to calculate the distribution of electron density within the crystal, based on the amplitude and phase of the scattered waves.

2.6.3.2 X-ray Diffraction Data Collection

2.6.3.2.1 X-ray Sources

Diffraction data are collected from protein crystals using X-rays that are generated either from an "in-house" source or by a synchrotron.

In-house X-ray sources are very often rotating copper anodes. These work by accelerating a beam of electrons towards a copper anode in a vacuum. The energy transfer that occurs when the electrons reach the anode ejects an electron from the inner shell of the atom. Consequently, an electron from a higher energy level is then able to drop down and occupy the now vacant, lower energy level position. This is accompanied by the release of energy in the form of an X-ray photon.

Each metal produces X-rays of a characteristic wavelength, which is governed by the energy levels of its electrons. For copper this wavelength is 1.54 Å.

The intensity of the X-rays produced by such sources is limited by the rate at which heat can be conducted away from the site of bombardment. For this reason, the copper anode is usually a rotating, hollow cylinder that is cooled from inside by jets of water.

Although in-house sources can be used to collect good quality X-ray diffraction data, increased access to synchrotrons in the last few decades has lead to synchrotron radiation becoming the dominant source of X-rays for protein crystallography experiments. Synchrotron radiation is produced by using a magnetic field to accelerate electrons, which are travelling at close to the speed of light, in a circular path. The result is an intense X-ray beam that is emitted at a tangent to the path of the electrons.

There are a number of advantages that come from using synchrotrons over in-house X-ray sources. The radiation that is emitted is much more intense than that produced by laboratory sources, consequently, complete datasets can be detected in a matter of minutes. Furthermore, whereas the wavelength of the X-rays produced by rotating anode sources is fixed and determined by the choice of metal, the wavelength of X-rays produced by synchrotrons can be adjusted using undulators and wigglers. This is crucial

for methods such as multi-wavelength anomalous dispersion (MAD), a powerful tool for solving the phase problem, which is described in section 2.6.3.3.1.3.

All of the diffraction data presented in this thesis were collected at the United Kingdom's national synchrotron facility: Diamond Light Source (DLS), Harwell Science and Innovation Campus, Didcot, Oxon, UK.

2.6.3.2.2 Radiation Damage

Radiation damage to the crystal presents a significant problem in diffraction experiments and data are often collected at 100 K in order to minimise this effect. Prior to data collection it is commonplace to rapidly cool the crystals. It is important that this is a very fast process so as to avoid temperatures at which ice formation can occur, as this can disrupt the crystal lattice. The addition of cryoprotectant solutions, often containing glycerol, glucose or polyethylene glycols (Berejnov *et al.,* 2006; Garman and Mitchell, 1996) can help to prevent ice formation and protect crystals during freezing.

2.6.3.2.3 Measures of Data Quality

The final structure can only ever be as good as the data that are collected hence it is exceedingly important to have an indication of the quality of the raw data. A number of different statistical values are generally used for this purpose.

2.6.3.2.3.1 *R* factors

Depending on the symmetry of the crystal, there are certain reflections that should have the same intensity. This is described by Friedel's law and these reflections are known as Friedel pairs.

R_{sym} is a statistical measure of the differences between the intensity of symmetry related measurements. A low R_{sym} is indicative of good agreement between these measurements and hence reliable data. A second value, R_{merge} is often reported and is closely related to R_{sym}. R_{merge} takes into consideration the fact that normally during data collection the same reflection will be measured many times and so also includes these in the calculation.

A significant problem with R_{merge} exists though in that it is dependent on the redundancy of the data. If the redundancy is low than the value for R_{merge} will automatically be lower, but at the same time the data are less accurate. Similarly, excluding reflections will improve R_{merge} whilst having a detrimental effect on the quality of the final structure. A

slightly improved value, R_{pim} (precision indicated merging R factor) can be used which also takes into account the precision of the averaged measurement (Diederichs and Karplus, 1997; Weiss and Hilgenfield, 1997; Weiss, 2000).

2.6.3.2.3.2 Completeness and Redundancy

As outlined above, *R* factors are intrinsically linked to the completeness and the redundancy of the data hence it is essential that these values be considered.

The completeness of the data is a measure of the total number of unique reflections that are recorded as a fraction of the total number of unique reflections that are required to generate a complete set of reflections. This is of course dependent on crystal symmetry. Generally speaking, overall completeness should be greater than or equal to 95 %.

Redundancy is a slightly different concept, which presents the total number of reflections that are measured as a fraction of the total number of unique reflections that are measured. This is to take into consideration the fact that measuring a number of different, symmetry related, reflections will give more reliable data than measuring the same reflection many times.

2.6.3.3 Data Processing and Structure Solution

Having collected the diffraction data, the first step in data processing is to index and then integrate this data. During indexing the unit cell dimensions are determined, which are then used during integration, when the intensity of the diffraction spots are measured. In the work presented in this thesis MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) was used to perform indexing and integration.

Following integration the data are then scaled and merged. This involves merging reflections that are equivalent, scaling the data to account for errors during collection and combining all of the integrated values to produce the structure factors. The program Aimless (Evans, 2006; Evans, 2007), part of the CCP4 suite (Winn *et al*., 2011) was used to scale and merge all of the data collected in this thesis.

2.6.3.3.1 The Phase Problem

As detailed in section 2.6.3.1, the distribution of electron density in the crystal can be calculated by performing an inverse Fourier transform on the structure factors of the diffracted X-rays. This requires knowledge of the amplitude and phase of these waves.

The amplitude is proportional to the square of the intensity, hence is calculated based on the intensity recorded by the detector. The detector cannot measure the phase of the waves, thus this information is lost. This is known as the phase problem and is one of the biggest issues faced in X-ray crystallography.

2.6.3.3.1.2 Multiple Isomorphous Replacement

Multiple isomorphous replacement (MIR) is a method for solving the phase problem that was pioneered by Max Perutz (Green *et al.,* 1954; Perutz *et al.,* 1960). In MIR heavy atoms, typically mercury (binds to thiol groups on cysteine residues), uranium (binds carboxyl groups on aspartic and glutamic acid residues) or platinum (binds to histidine residues), are incorporated into the crystals. This can either be at the point of setting up the crystallisation experiment or by soaking the crystals once grown in solutions containing the heavy metal atoms. It is imperative for the success of this method that binding of the heavy metals does not alter the crystal lattice; the native and derivative crystals must be isomorphous. Data are collected from both the native crystal and the heavy atom derivative. Comparing the data enables the position of the heavy atoms to be established, meaning that the amplitude and phase can be calculated. This information is then used to calculate the phases for the native data.

2.6.3.3.1.3 Multi-wavelength Anomalous Dispersion

A related method for solving the phase problem, also using heavy metals, is multi-wavelength anomalous dispersion (MAD). The major advantage of MAD is that it only requires data from a single crystal, which can negate the problem of non-isomorphism following heavy atom soaks.

At wavelengths close to the absorption edge of the heavy metal of choice, strong absorption of the incident X-ray means that the phase of X-rays diffracted from the heavy metals is shifted relative to those diffracted from the other atoms in the crystal. This is known as anomalous scattering. In MAD, multiple (typically three) datasets are collected from a single crystal at different wavelengths, one of which must be close to the absorption edge of the heavy metal (Hendrickson *et al.,* 1988; Hendrickson, 1991).

MAD has become much more accessible with the advent of synchrotron beamlines with tuneable wavelengths. Selenium is very often the atom of choice for MAD phasing. This is because it can be easily and reliably incorporated into proteins by recombinant expression in selenomethionine containing media by methionine auxotrophs. Single-

wavelength anomalous dispersion (SAD) is a related technique that only requires data collection at a single wavelength (Hendrickson *et al.,* 1990).

2.6.3.3.1.4 Molecular Replacement

The method used for solving the phase problem in the work presented here is molecular replacement (MR). This is a very different technique to those outlined above and is entirely dependent on an existing structure of a related protein, typically a homologue, which is used as a model. The basic premise of MR is to determine the rotation and translation functions that must be applied to the search model in order to correctly orient it in the unit cell of the target protein. Once this has been achieved phases can be calculated and used to produce an electron density map for the target protein (Rossmann and Blow, 1962).

2.6.3.4 Structure Refinement and Validation

2.6.3.4.1 Refinement

After the phase problem has been solved, a model of the structure is produced based on interpretation of the electron density map. This model must then be refined; adjustments are made to the structure to ensure that it is the best possible representation of the experimental data and to minimise errors in the final structure. Refinement is usually carried out using computational software, with REFMAC (Murshudov *et al.,* 2011), part of the CCP4 suite (Winn *et al.,* 2011), being a widely used example.

The first step in the refinement process is typically rigid body refinement. Here the model is moved as a single, rigid entity to find the position in which it best fits the electron density. At this stage the positions of every atom relative to each other are fixed.

Following rigid body refinement it is common to perform iterative rounds of model building, where manual adjustments are made to improve the "fit" of the structure to the density, and restrained refinement. Whereas in rigid body refinement the positions of atoms relative to each other are fixed, in restrained refinement a certain degree of movement is permitted. This is governed by a set of restraints that are imposed, primarily to ensure that the model remains physically reasonable. Typically this includes, but is not limited to, restraints on bond length, bond angle, planarity and distances between atoms (to ensure that they do not clash).

In addition to visual inspection of the model, the progress of refinement is followed using statistical analysis. The *R* (reliability) factor is an exceedingly important value that measures the agreement between the model and the experimental data, and R_{cryst} is frequently reported for protein structures.

Broadly speaking, a lower *R* factor is indicative of better agreement between the model and the experimental data, with a theoretically "perfect" model having an *R* factor of zero. Thus, R_{cryst} is expected to decrease during the refinement process.

Unfortunately, R_{cryst} is by no means a perfect measure of model quality. A significant problem exists in that R_{cryst} is calculated using information that is also used in refinement. This means that it is susceptible to large systematic errors and over-fitting. For example, a study performed in 1995 demonstrated that tracing a protein backbone completely incorrectly, in this example backwards, could, if data were not of very high resolution, still result in seemingly acceptable *R* factors (Kleywegt *et al.,* 1995).

A second measure, R_{free}, was introduced in 1992 in order to overcome this problem (Brünger, 1992; Brünger, 1993). The R_{free} value is calculated using a subset, normally 5-10 %, of reflections that are set aside at the outset and hence never used in refinement. Considering the difference in the values reported for R_{free} and R_{cryst} can therefore provide information as to whether the model is actually being improved through refinement. The R_{free} value is always expected to be greater than the R_{cryst} value, but a large disparity indicates a problem with the refinement process. There is no firm rule as to what is an acceptable difference between R_{free} and R_{cryst}, which is significantly influenced by resolution (Tickle *et al.,* 1998). Generally speaking the R_{free} value should not exceed ~ 35 % and should be approximately 1.2 times greater than R_{cryst} (Kleywegt and Brünger, 1996).

2.6.3.4.2 Structure Validation

After refinement, it is commonplace to perform some form of geometric validation of the final model. This routinely includes calculating the root mean square deviation (RMSD) of bond lengths and angles from ideal values and checking for adherence to the Ramachandran plot (Ramachandran *et al.,* 1963). More thorough geometric validation is often performed using software such as Molprobity, which performs functions that include analysis of contacts between atoms to check for clashes, checking adherence to the Ramachandran plot, detecting cis-peptides and identifying unusual rotamers (Davis *et al.,* 2007; Chen 2010).

2.7 General Protocols for AnCE

AnCE is one of two angiotensin converting enzyme homologues from *Drosophila melanogaster* (section 1.9.1). It is used for the structural studies presented in Chapters 3, 4, 5 and 6. General protocols for AnCE expression, purification and crystallisation that were used for this work are outlined in the following sections. Variations to this protocol for individual experiments are detailed as they arise in the appropriate chapters.

2.7.1 Recombinant Expression by *P. pastoris*

P. pastoris GS115 cells transformed with the *Ance* coding sequence in the pPIC9 vector were kindly provided by Prof. Elwyn Isaac, University of Leeds. Recombinant expression of AnCE by *P. pastoris* was performed according to the previously established protocol (Houard *et al.,* 1998; Akif *et al.,* 2010).

Briefly, a single colony of the transformed GS115 cells was used to inoculate 25 ml BMGY media in a 100 ml baffled flask. This culture was incubated for 24 hours at 30 ^{o}C with orbital shaking at 250 rpm. Two 500 ml subcultures in 2 litre baffled flasks were prepared by inoculation with 1 % of the overnight culture and incubated at 30 ^{o}C, 250 rpm for a further 24 hours until the optical density at 600 nm was between 2 and 6. This culture was harvested by centrifugation at 3000 g, 18 ^{o}C for 5 minutes using the Beckman Coulter^{TM} Avanti J-25 centrifuge and JLA 16.250 rotor. The supernatant was discarded and the cell pellets resuspended in a total volume of 200 ml BMMY media to induce expression. The 200 ml culture was returned to the incubator at 30 ^{o}C, 250 rpm for 48 hours. In order to maintain the concentration of methanol required for induction, every 24 hours methanol was added directly to the culture to a final concentration of 0.5 %, compensating for losses through evaporation and consumption.

Following 48 hours of induction the culture was harvested by centrifugation at 3000 g, 4 ^{o}C using the Beckman Coulter^{TM} Avanti J-25 centrifuge and JLA 16.250 rotor. The cell pellets were discarded and the supernatant retained for immediate use in purification steps.

2.7.2 Purification by HIC

The AnCE expression supernatant was treated with an equal volume of 3 M ammonium sulphate, 20 mM Tris pH 8.0 (to give a final ammonium sulphate concentration of 1.5 M) and incubated for 10 minutes at room temperate before centrifugation at 4000 g, 4 ^{o}C for 10 minutes using the Beckman Coulter^{TM} Avanti J-25 centrifuge and the 16.250 rotor.

10 ml phenyl sepharose 6 fast flow high sub resin (GE Healthcare, UK) were packed in an XK 16/60 column and equilibrated with buffer A (1.5 M ammonium sulphate, 20 mM Tris pH 8.0). The supernatant from the ammonium sulphate treatment was applied to the column and washed with buffer A until the absorbance at 260 nm reached a stable baseline. Elution was performed with a step gradient of decreasing salt: 0.45 M and then 0 M ammonium sulphate. The absorbance at 280 nm was recorded and used to monitor elution. 3 ml fractions were collected and analysed by SDS-PAGE. As illustrated in Figure 2.6 AnCE elution occurs at 0.45 M ammonium sulphate.

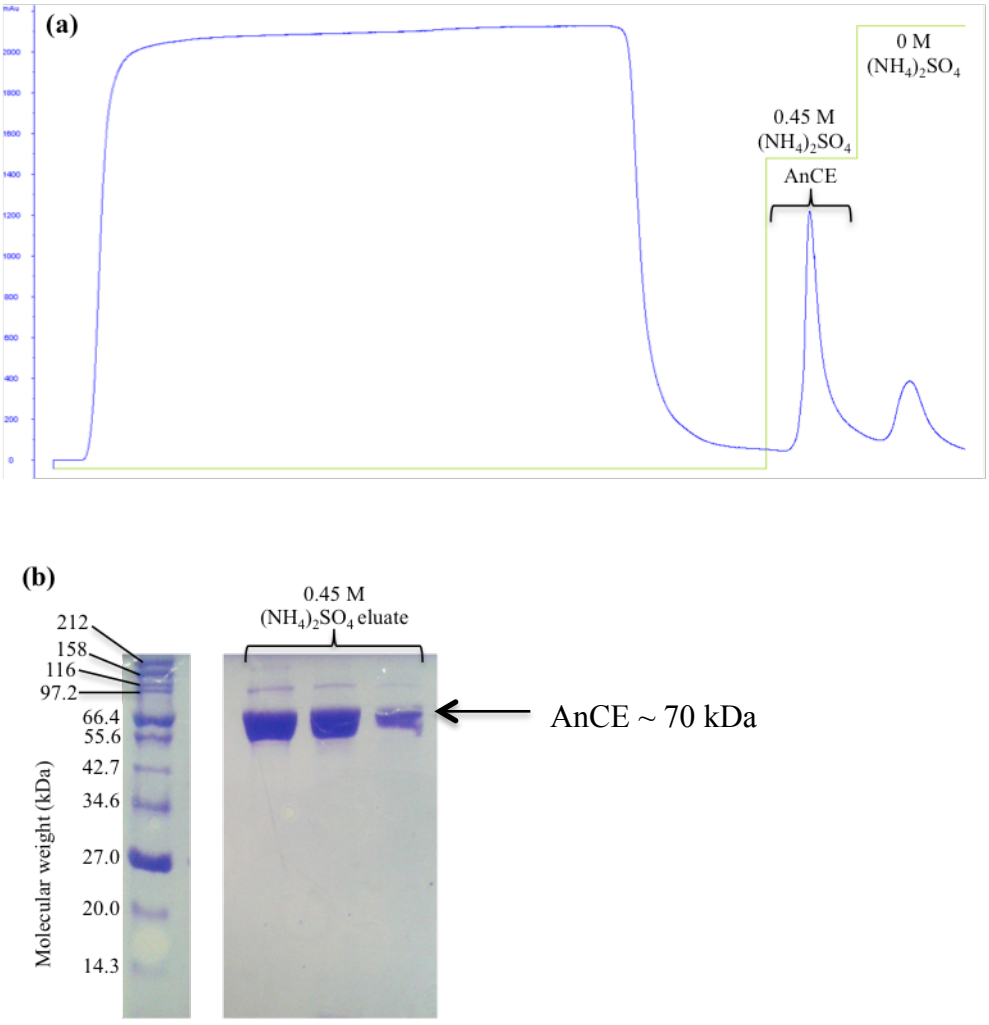


Figure 2.6: AnCE Purification by HIC. (a) Chromatogram for purification of AnCE by HIC. The blue trace shows the absorbance at 280 nm, used to monitor protein elution. The green trace shows the concentration of elution buffer. The peak eluted at 0.45 M ammonium sulphate was analysed by SDS-PAGE. This gel is shown in (b) and confirms that the elution peak contains AnCE, illustrated by the bands at the expected molecular weight of ~70 kDa.

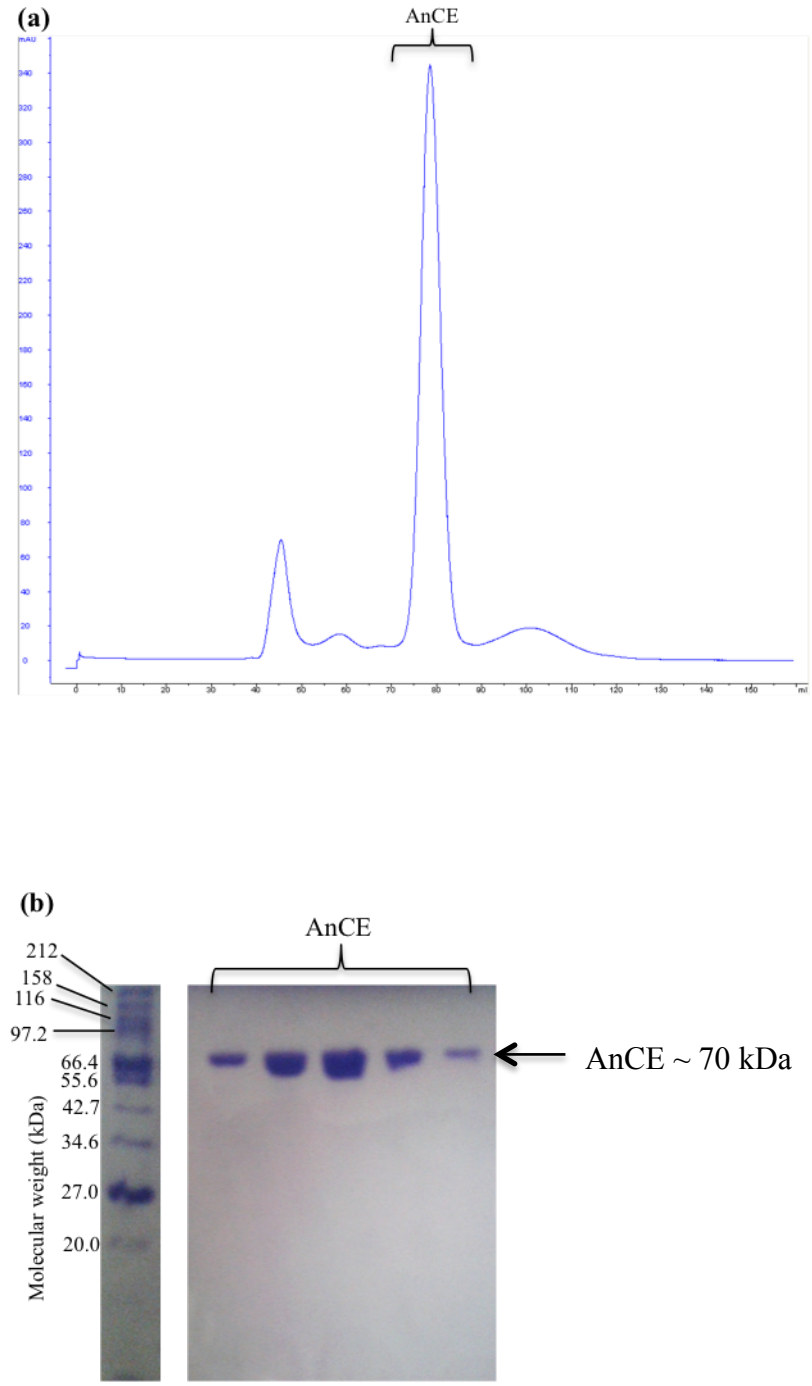


Figure 2.7: AnCE Purification by SEC. (a) Chromatogram for the application of pooled fractions from HIC to the SEC column. The blue trace shows the absorbance at 280 nm, used to monitor protein elution. Fractions from the peak labelled "AnCE" were analysed by SDS-PAGE. This gel is shown in (b) and confirms that this peak contains pure AnCE, shown by the bands at ~70 kDa, the expected molecular weight of AnCE.

2.7.3 Purification by SEC

Fractions from HIC confirmed by SDS-PAGE to contain AnCE were pooled together and dialysed for 18 hours at 4 ^{o}C against an excess of size exclusion buffer (150 mM NaCl, 20 mM Tris, pH 8) using 12-14 kDa molecular weight cut-off dialysis membrane (Medicell International Ltd). The dialysed sample was concentrated to a final volume not exceeding 1.5 ml by centrifugation at 3000 g, 4 ^{o}C, in a Heraeus centrifuge (Thermo Fisher Scientific, MA, USA) using 30 kDa molecular weight cut-off Amicon Ultra-15 centrifugal filter units (Merck Millipore). The concentrated sample was then applied to a Superdex-200 16/60 column (GE Healthcare, UK) pre-equilibrated with gel filtration buffer. Monitoring the absorbance at 280 nm was used to follow the elution. 1.5 ml fractions were collected and the presence of pure AnCE confirmed by SDS PAGE analysis, with an example gel shown in Figure 2.7. Using the technique described above, pure AnCE was concentrated to 17.5 mg ml^{-1} before being split into 50 µl aliquots and rapid frozen in liquid nitrogen prior to storage at -80 ^{o}C.

2.7.4 Crystallisation

AnCE crystallisation was performed using the hanging drop vapour diffusion technique that is described in section 2.6.2.3. Previous experiments by Akif *et al.,* used high throughput crystallisation screens to identify a condition in which AnCE crystallises: 1.2-1.3 M sodium citrate, 0.1 M HEPES, pH 7.5; this is the condition that was used to grow the AnCE crystals used in Chapters 3, 4 and 5 (Akif *et al.,* 2010).

Crystallisation experiments were performed in 24 well plates (Molecular Dimensions Ltd., Newmarket, UK). 500 µl of the reservoir solution (1.2-1.3 M sodium citrate, 0.1 M HEPES pH 7.5) were used in each case. A drop was prepared containing 1 µl reservoir solution and 1 µl 17.5 mg ml^{-1} AnCE. In the case of inhibitor binding studies, the inhibitor was pre-incubated with AnCE before the drop was prepared. The drops were suspended over the reservoirs on 20 mm siliconised cover slips and sealed using high vacuum grease. Experiments were incubated at 21 ^{o}C.

Typically diffraction quality crystals such as those shown in Figure 2.8 grew within 7 days. These crystals belong to the space group *h3* and contain one AnCE molecule per asymmetric unit. Due to the high concentration of citrate in the condition, complete datasets can be collected at 100 K without the use of a cryoprotectant. Diffraction from these crystals is routinely recorded to at least 2.0 Å, as illustrated by the example diffraction image in Figure 2.9.

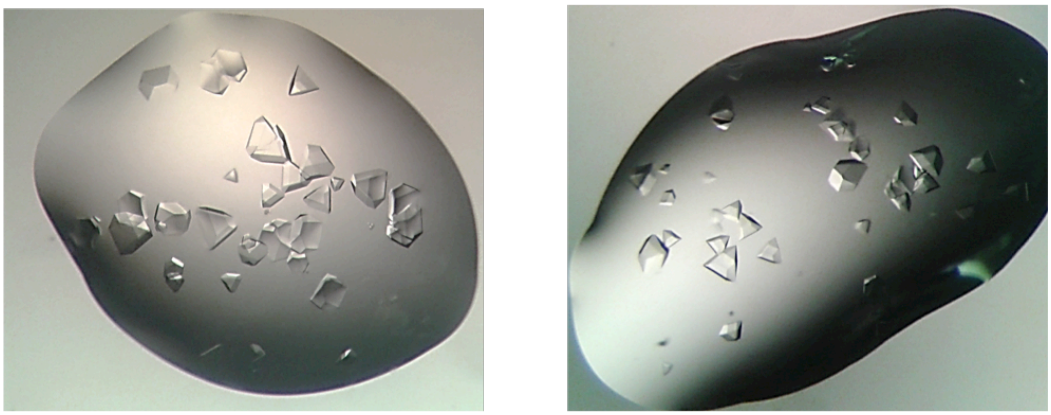


Figure 2.8: AnCE Crystals Grown using the Hanging Drop Vapour Diffusion Technique. The crystals shown are in 2 µl drops containing 1 µl 17.5 mg ml^{-1} AnCE and 1 µl reservoir solution. The reservoir solution contains 1.26 M sodium citrate and 0.1 M HEPES pH 7.5. Crystallisation experiments were performed at 21 ^{o}C and crystal typically appeared within 7 days.

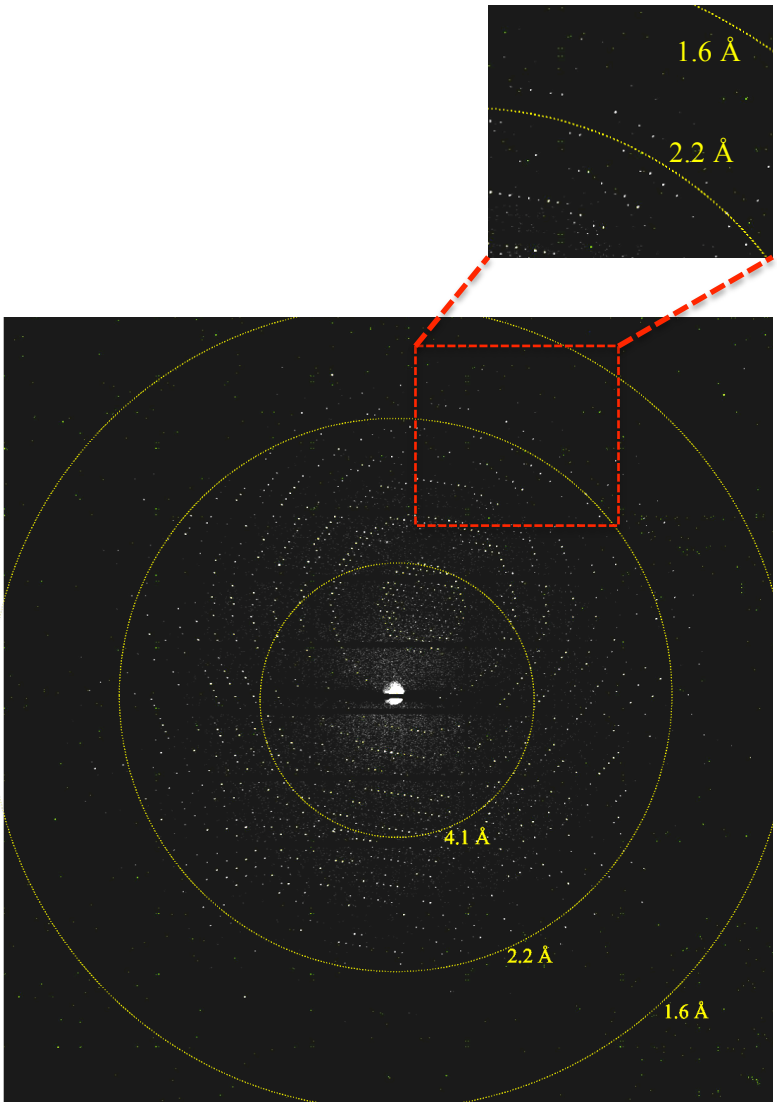


Figure 2.9: Example of X-ray Diffraction Data collected from an AnCE crystal. Diffraction data were collected to ~1.7 Å, as highlighted by the inset image.

Chapter 3: Crystal Structures of AnCE in Complex with Natural Inhibitory Peptides from the Venom of *Bothrops jararaca*

3.1 Introduction

3.1.1 Requirement for a New Generation of ACE Inhibitors

Angiotensin converting enzyme is a well-established target in the treatment of hypertension. This is primarily achieved through the use of ACE inhibitors; small molecules that bind to the enzyme's active site and competitively inhibit its activity (Cushman and Ondetti, 1999; Zaman *et al*., 2002). ACE inhibitors such as captopril have been widely used since the 1980s, however, as outlined in section 1.5.1, these drugs are frequently associated with side effects caused by the inhibition of the roles of ACE in other physiological processes (Israili and Hall, 1992; Nussberger *et al.,* 1998; Steckelings *et al*., 2001).

It is increasingly being recognised that the C- and N-domains of sACE have evolved separately to acquire divergent functions. It is the C-domain that is primarily responsible for blood pressure regulation *in vivo*, converting angiotensin I into the hypertensive peptide angiotensin II much more efficiently than the N-domain (Georgiadis *et al.,* 2003).

Both domains of ACE have been shown to hydrolyse the peptide bradykinin with equal efficiency. This is important, because the accumulation of bradykinin associated with ACE inhibition has been implicated in many of the side effects associated with the use of ACE inhibitors, not least the persistent dry cough that afflicts so many patients (Bas *et al*., 2007).

Additionally, it is apparent that the N-domain is involved in many other physiological processes outside of blood pressure regulation, cleaving a number of other peptide substrates more effectively than the C-domain. This includes peptides such as Aβ42, which has been implicated in Alzheimer's disease, and AcSDKP (Rousseau *et al*., 1995), which is involved in the regulation of the proliferation of HSCs (haematopoietic stem cells) (Lenfant *et al*., 1989).

With this in mind, a significant amount of research is currently focused on the design and development of a new generation of ACE inhibitors that selectively target only one of the catalytic domains. Inhibitors specific for the C-domain could be used to treat hypertension whilst potentially avoiding many of the side effects currently associated with ACE inhibitors. Similarly, in some cases, it may be beneficial to selectively target the N-domain, leaving the C-domain active in blood pressure regulation. One example of this is in aggressive chemotherapy treatments, where there is evidence to suggest that

inhibiting AcSDKP hydrolysis, a function of the N-domain, can protect HSCs (Bogden *et* al., 1991).

The development of two phosphinic peptide inhibitors, RXP407 and RXPA380, which selectively inhibit the N- and C-domains of sACE, respectively, has provided proof of principle data that such domain specific inhibition could be possible (Dive *et* al., 1999; Georgiadis *et al*., 2004). A key process in developing selective inhibitors is identifying and exploiting different features of the active site pockets of the two domains. As detailed in section 1.6.5, information obtained from structural studies was crucial in understanding the basis for domain-specific inhibition by RXPA380 and RXP407 and will no doubt be invaluable in future inhibitor development.

3.1.2 Bradykinin Potentiating Peptides

The bradykinin potentiating peptides (BPPs) from the venom of the pit viper *Bothrops jararaca* are natural inhibitors of angiotensin converting enzymes. They have been studied since the 1970s and analysis of the sequences of these peptides played a pivotal role in the design of the original ACE inhibitors. All of the BPPs have a conserved IPP motif at their C-terminus that is critical for ACE inhibition (Ferreira, 1965; Ferreira, 1970; Ondetti *et al*., 1971).

In 2002 a study was performed by Cotton *et al.* to investigate the inhibition of the two domains of sACE by three BPPs. The Ki values obtained for each peptide with each domain are illustrated in Table 3.1 and the sequences of the peptides are included in Table 3.2 (Cotton *et al*., 2002).

The data obtained indicate that whilst all three peptides are effective inhibitors of the C-domain, only BPPc inhibits the N-domain with equal potency. Both BPP9a and BPPb are selective inhibitors of the C-domain and BPPb is an exceedingly poor inhibitor of the N-domain.

Based on sequence analysis, Cotton *et al.* proposed that the residue at the P_2 position of the BPPs was crucial for domain selectivity. BPPc, which inhibits both domains equally, has a small proline side chain at this position, whereas the glutamine and lysine residues at P_2 in BPPb and BPP9a have much larger side chains that may in some way impede binding to the N-domain whilst enhancing C-domain binding and hence inhibition.

Peptide	Ki for N-domain (nM)	Ki for C-domain (nM)
BPP9a	100	1
BPPb	10, 000	30
BPPc	80	80

Table 3.1: K_i Values for the Inhibition of Each Domain of sACE by Three Different BPPs, as Determined by Cotton *et al*., 2002.

Peptide	P_9	P_8	P_7	P_6	P_5	P_4	P_3	P_2	P_1	P_1'	P_2'
BPP9a			PCA	Trp	Pro	Arg	Pro	Gln	Ile	Pro	Pro
BPPb	PCA	Gly	Leu	Pro	Pro	Arg	Pro	Lys	Ile	Pro	Pro
BPPc	PCA	Gly	Leu	Pro	Pro	Gly	Pro	Pro	Ile	Pro	Pro
BPP12b	PCA	Gly	Pro	Pro	Pro	Arg	Pro	Gln	Ile	Pro	Pro
BPPn	PCA	Gly	Leu	Trp	Pro	Arg	Arg	Lys	Ile	Pro	Pro

Table 3.2: Sequences of BPPs Indicating the Residues at Each Position According to the Notation Developed by Schechter and Berger Described in Section 1.6.5.1 (Schechter and Berger, 1967). PCA denotes N-terminal pyroglutamate residues.

3.1.3 Aims

The aim of this work is to use X-ray crystallography to study the binding of a number of BPPs to the *D. melanogaster* ACE homologue AnCE and in doing so to learn more about how these peptides interact with the active site residues and how this may explain the domain selectivity that has been reported. AnCE is an ideal model for this as the active site pockets share some common features with both the N- and C-domains of human sACE, thus it will be possible to make informed comparisons.

To date, the only structures of angiotensin converting enzymes in complex with BPPs are of AnCE (Akif *et al.,* 2012) and human tACE (Masuyer *et al*., 2012), both bound to BPPb. Comparisons with these structures will form an important part of the analysis of the structures of AnCE with other BPPs presented here.

The sequences of the peptides used in this study are detailed in Table 3.2. In addition to the naturally occurring peptides BPP9a, BPPc and BPP12b, an additional peptide, BPPn has been designed. As shown by the sequences in Table 3.2, all of the naturally occurring BPPs used here have a proline residue at the P_3 position. In BPPn this has been substituted by an arginine. The aim of doing this is to investigate the significance of this position on the inhibitory properties of the peptide. The adjacent P_2 position has already been implicated in domain selectivity, hence, extending this to the next position

is a logical next step towards understanding the key features of these peptides required for ACE inhibition.

3.2 Methods

3.2.1 Protein Expression, Purification and Crystallisation

AnCE expression and purification was achieved following the general protocol that is described in section 2.7. *P. pastoris* was used to produce recombinant AnCE, which was then purified in a two-step process consisting of hydrophobic interaction chromatography followed by size exclusion chromatography.

Purified AnCE was concentrated to 17.5 mg ml^{-1} and co-crystallised with the peptides BPP9a, BPPc, BPP12b and BPPn (Table 3.2) using the hanging drop vapour diffusion method, as described in section 2.7.4. All of the BPPs were purchased from Sigma-Aldrich with the exception of BPPn, which was purchased from GL Biochem (Shanghai) Ltd.

Immediately prior to crystallisation, 20 mM of each peptide was added to 17.5 mg ml^{-1} AnCE in a 1:4 ratio of peptide:protein and incubated at 4 ^{o}C for approximately 30 minutes. Hanging drops were prepared with 1 µl of the protein and peptide mixture and 1 µl of the reservoir solution (1.2-1.3 M sodium citrate, 0.1 M HEPES, pH7.5). These drops were suspended on glass coverslips over wells containing 500 µl of the reservoir solution and sealed using high vacuum grease. Crystallisation experiments were incubated at 21 ^{o}C and observed regularly.

Where crystals did not appear within the normal time frame for AnCE under these conditions (7 to 10 days) the crystallisation experiments were repeated with slight variations to the experimental parameters. This included decreasing the concentration of the protein (5-17.5 mg ml^{-1}) and the peptide (2-20 mM), altering the ratio of peptide: protein (1:4 to 1:19) and reducing the incubation temperature to 19 ^{o}C.

Where these variations still did not yield crystals, seeding was used to try and induce nucleation. Two methods were tried: streak seeding and micro-seeding. In both cases the "seeds" were derived from native AnCE crystals grown under the conditions outlined in section 2.7.4.

For the streak seeding experiments, drops were prepared as outlined above, placed over the reservoir solution and allowed to equilibrate at 21 ^{o}C for 4 to 6 hours. Following this,

a whisker was used to transfer "seeds" from the native AnCE crystals to the new drops by passing through the length of the drop. The drops were then once again sealed over the reservoirs and incubated at 21 ^{o}C.

For the micro-seeding experiments, a seed stock was prepared by thoroughly resuspending a 2 µl drop containing native AnCE crystals in 10 µl of reservoir solution. The crystals in the drop were then crushed by vigorous agitation using a "seed-bead" (Hampton Research, CA, USA). Dilutions of the resulting suspension were prepared in the range 1×10^{-1} to 1×10^{-3} using the reservoir solution. 1 µl of these solutions was then added to drops prepared as detailed above that had been allowed to equilibrate at 21 ^{o}C for 4-6 hours. The drops containing the seeding solution were incubated at 21 ^{o}C and observed regularly.

3.2.2 X-ray Diffraction Data Collection

The crystallisation condition used to grow the crystals of AnCE with the BPPs has been used on many previous occasions (Akif *et al.,* 2010). It has previously been established that the relatively high concentration of sodium citrate present in the condition (1.2-1.3 M) provides adequate cryo-protection, thus no further cryo-protectant solutions were added to the crystals prior to data collection.

In place of a cryo-protectant, 1 µl of the reservoir solution was added to the drop prior to crystal mounting. This was to prevent the drop from drying out whilst also maximising the citrate concentration. Crystals were mounted directly into litholoops (Molecular Dimensions Ltd., Newmarket, UK) before rapid freezing in liquid nitrogen for transportation to the synchrotron.

All X-ray diffraction data were collected using beamline i03 at DLS. Crystals were mounted for data collection using the sample changer robot and diffraction data were recorded using a Pilatus3 6M detector. All data were collected at 100 K.

3.2.3 Data Processing and Structure Solution

Diffraction data were integrated using MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) prior to scaling with Aimless (Evans, 2006; Evans, 2007), part of the CCP4 suite (Winn *et al*., 2011). Phases were calculated by molecular replacement using Phaser (McCoy *et al*., 2007) with the native AnCE structure, PDB code 2X8Y (Akif *et al.,* 2010), as the search model.

Following production of initial structural models by molecular replacement, iterative rounds of model building and refinement were performed using C*oot* (Emsley *et al.,* 2011) and REFMAC (Murshudov *et al.,* 2011), respectively. 5 % of residues were separated for use as the R_{free} set in cross-validation.

A key part of the model building process was placing peptide residues into the difference density in the substrate-binding channel. Water molecules were added where Fo-Fc electron density peaks exceeded 3σ and potential hydrogen bonds could be made.

Molprobity (Chen *et al.,* 2010) was used to validate the quality of the final structures and LigPlot^{+} (Laskowski and Swindells, 2011) was used to study interactions between the peptides and the active site residues. RMSD calculations were performed using the molecular graphics program Pymol (The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC), which was also used to create all of the figures.

3.3 Results

3.3.1 Co-crystallisation of AnCE with the BPPs

Crystals of AnCE with both BPP9a and BPPn appeared in all drops (1.2-1.3 M sodium citrate, 0.1 M HEPES, pH 7.5) within 10 days of incubation at 21 ^{o}C. Figure 3.1 shows an example of a typical crystal mounted in a litholoop immediately prior to data collection.

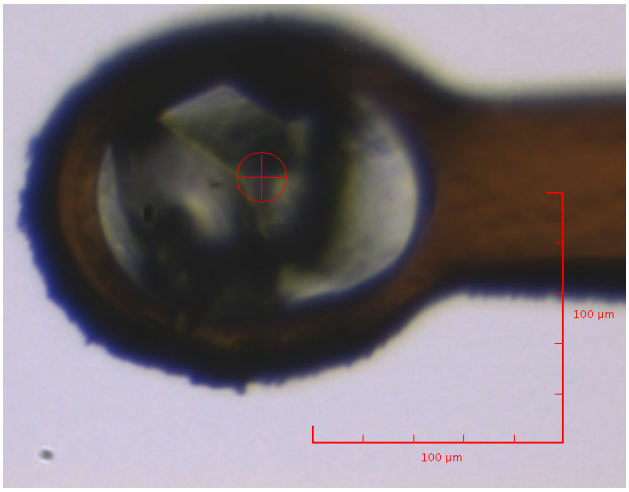


Figure 3.1: A Crystal of AnCE with BPPn Mounted in a Litholoop Immediately Prior to Data Collection. The 100 µm scale bars provide an indication as to the size of the crystal.

Crystallisation with BPP12b and BPPc proved to be much more challenging. Initial experiments using the standard conditions failed to yield any diffraction quality crystals and only precipitate was visible in the drops. Attempts to overcome this by decreasing

protein and peptide concentration, reducing incubation temperature and varying the ratio of peptide: protein, as outlined in section 3.2.1 all failed to yield crystals of any description.

3.3.2 X-ray Diffraction Data Collection

Diffraction data for AnCE with BPP9a and BPPn were collected at DLS on 21st November 2014 using beamline i03, as summarised in Table 3.3. A total of six datasets were collected for AnCE with BPP9a (AnCE_BPP9a_1-6) and three for AnCE with BPPn (AnCE_BPPn_1-3), an example diffraction image is shown in Figure 3.2. All datasets were initially processed in the space group *h3* using MOSFLM and Aimless. The statistics obtained are summarised in Table 3.4.

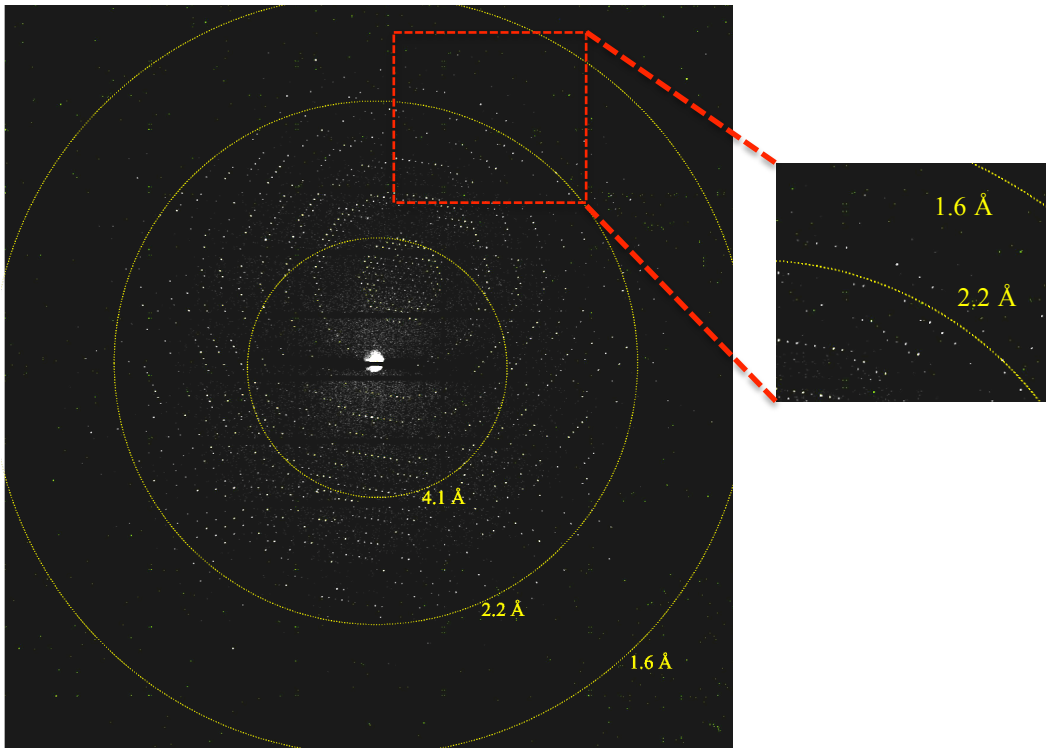


Figure 3.2: X-ray Diffraction Pattern Produced by a Crystal of AnCE with BPP9a. Resolution rings are labelled. Inset image shows diffraction to ~ 1.7 Å.

Dataset name	Collection date	Beamline	Detector	Number of images	Phi range (°)	Resolution range (Å)	Overall completeness (%)
AnCE_BPP9a_1	24th November 2014	i03	Pilatus3 6M	600	87-207	86.72-1.70	93.0
AnCE_BPP9a_2	24th November 2014	i03	Pilatus3 6M	600	0-120	49.98-2.21	99.9
AnCE_BPP9a_3	24th November 2014	i03	Pilatus3 6M	900	0-180	28.98-1.96	99.6
AnCE_BPP9a_4	24th November 2014	i03	Pilatus3 6M	600	160-280	39.15-2.67	99.6
AnCE_BPP9a_5	24th November 2014	i03	Pilatus3 6M	600	85-205	38.17-2.04	99.4
AnCE_BPP9a_6	24th November 2014	i03	Pilatus3 6M	600	0-120	84.75-2.00	99.7
AnCE_BPPn_1	24th November 2014	i03	Pilatus3 6M	600	0-120	84.64-1.96	99.6
AnCE_BPPn_2	24th November 2014	i03	Pilatus3 6M	600	0-120	86.82-1.72	98.2
AnCE_BPPn_3	24th November 2014	i03	Pilatus3 6M	1000	150-300	38.15-1.83	99.9

Table 3.3: Summary of Data Collected for AnCE with the Peptides BPP9a and BPPn and Statistics following initial processing with MOSFLM and Aimless. The two "best" datasets that were used for further downstream processing are highlighted in yellow.

3.3.3 Data Processing and Structure Solution

Data for AnCE_BPP9a_1 and AnCE_BPPn_2 were initially processed in the space group *h3* using MOSFLM and Aimless. Phases were then calculated by molecular replacement performed using Phaser with the native AnCE structure, PDB code 2X8Y, as the search model. In both cases there was one molecule per asymmetric unit of the crystal. Full data processing statistics are included in Table 3.4.

	AnCE_BPP9a	**AnCE_BPPn**
Beamline at DLS	i03	i03
Detector	Pilatus3 6M	Pilatus3 6M
Wavelength (Å)	0.976	0.976
Space group	*h3*	*h3*
Unit cell parameters		
a, b, c (Å)	173.40, 173.40, 102.60	173.01, 173.01, 101.85
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Molecules per asymmetric unit	1	1
Resolution range (Å)	86.72-1.70	86.82-1.72
$^{\#}R_{merge}$ (%)	6.0 (61.1)	23.4 (70.00)
$^{+}R_{pim}$ (%)	5.1 (51.7)	23.4 (70.00)
Mean *I/σ(I)*	7.8 (1.3)	2.4 (0.6)
Completeness (%)	93.0 (96.5)	98.2 (94.3)
Number of reflections:		
Total	333, 555	278, 942
Unique	117, 755	118, 354
Multiplicity	2.8 (2.7)	2.4 (2.1)
Wilson *B* factor (Å^2)	29.28	27.69
Average *B* factor (Å^2)		
Protein	28.16	26.88
Peptide	49.71	25.99
Zinc	31.74	22.11
Glycosylated sugars	50.44	48.62
Solvent	37.89	36.52
$^{\S}R_{cryst}$/$^{\#\#}R_{free}$ (%)	17.54/19.68	17.26/19.73
R.M.S deviation from ideal values		
Bond lengths (Å)	0.008	0.007
Bond angles (°)	1.289	1.116
Ramachandran plot statistics (%)		
Favoured	98.67	98.34
Disallowed	0.00	0.00

$^{\#}$ $R_{merge} = \Sigma_{hkl}\Sigma_i|\boldsymbol{I_i}\,(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I_i}(hkl)$, where $\boldsymbol{I}_i$(hkl) is the intensity of *i*th measurement and <***I****(hkl)*> is the average of symmetry-related observations of a unique reflection.

$^{+}R_{pim} = \Sigma_{hkl}\sqrt{\frac{1}{n-1}}\Sigma_{i=1}^{n}|\boldsymbol{I}_i(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I}_i(hkl)$

$^{\S}R_{cryst} = \Sigma_h|F_o - F_c|/\Sigma_h F_o$, where F_o and F_c are observed and calculated structure factor amplitudes of reflection *h*, respectively.

$^{\#\#}R_{free}$ is equal to R_{cryst} for a randomly selected 5 % subset of reflections.

Table 3.4: Data Processing Statistics for AnCE with BPP9a and BPPn. Outer shell values are shown in parentheses.

The structures of the resulting models of AnCE bound to each of the two peptides are identical to that of the native AnCE structure, as illustrated by the superposition in Figure 3.3. The structures are predominantly helical and form an ellipsoid shape. Electron density is visible for N-linked glycan chains at Asn53, Asn196 and Asn311, the same three locations as in the original AnCE structure.

The active site, marked by the location of the zinc ion, is located at the centre of the catalytic channel, which extends almost the entire length of the molecule and is capped by a "lid" formed by helices α1, α2 and α3. Following rigid body refinement clear electron density was visible at the centre of the active site channel in both models, and this was deemed to be due to the binding of BPP9a and BPPn, as appropriate.

The high degree of overall similarity between the native AnCE structure and the structures with BPP9a and BPPn (RMSD values of 0.138 and 0.142, respectively) indicates that there is no significant movement of secondary structure elements upon peptide binding. Careful comparison of the models with the native structure also shows that there is no movement of active site residues induced by peptide binding. The details of the interactions made by the peptides with the enzyme are outlined in the following sections.

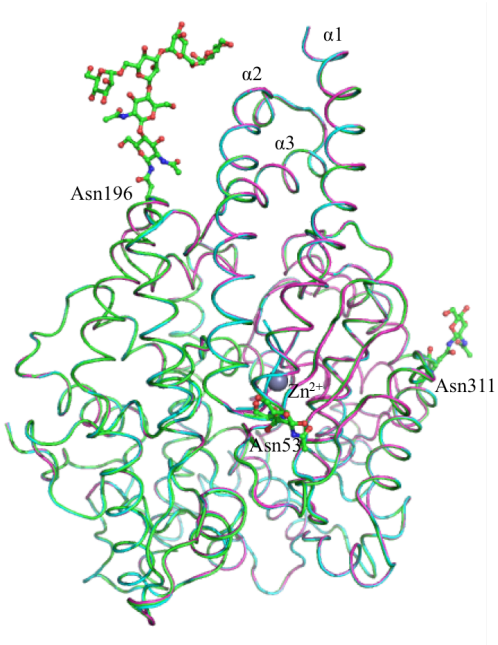


Figure 3.3: Superposition of the Structure of AnCE with BPP9a (cyan) and BPPn (magenta) and the Native AnCE Structure (PDB code 2X8Y) (green). All key structural features are conserved, including N-linked glycosylation at Asn53, Asn196 and Asn311, which is shown as green sticks. The location of the active site at the centre of the substrate- binding channel is indicated by the zinc ion, shown as a grey sphere. The channel is capped by helices α1, α2 and α3, which form the "N-terminal lid", limiting access to the active site.

3.3.4 Description of BPP9a Binding to the AnCE Active Site

Following molecular replacement and initial rounds of refinement, a significant amount of difference density was visible in the AnCE active site region. As illustrated in Figure 3.4, this density extends in both directions away from the catalytic zinc ion and appears to be long enough to accommodate a six-residue peptide backbone. Based on this, the six C-terminal residues of BPP9a were modelled such that the arginine, proline, glutamine, isoleucine, proline and proline residues occupied the S_4, S_3, S_2, S_1, S_1' and S_2' pockets respectively.

Figure 3.4 illustrates that whilst the side chains of the five C-terminal residues fit the difference density very well, the side chain of the final residue, the P_4 arginine is not accommodated by the difference density. Removing this residue completely resulted in a significant amount of difference density where the main chain of the arginine is located. This is a clear indication that this residue is indeed present in the structure and so should be included. It is likely that the arginine is present but that the side chain is not visible because it is highly mobile. No further density was visible beyond the S_4 region of the active site hence it was not possible to add any further residues to the N-terminus of BPP9a.

As outlined in Table 3.5, BPP9a bound at this position forms a number of interactions with the AnCE active site. Beginning at the C-terminus of the peptide, the oxygen atoms belonging to the carboxyl group of the P_2' proline residue form hydrogen bond interactions with the side chains of Gln265, Lys495 and Tyr504. These interactions are a common feature of peptide binding to the active site of angiotensin converting enzymes and act as an anchor to hold the C-terminus of the peptide in place.

The P_1' proline residue also participates in conserved peptide binding interactions with the active site; the proline carbonyl oxygen atom interacts with the side chains of His337 and His497.

Moving slightly further along the peptide chain, the carbonyl oxygen atom of the P_1 isoleucine forms a hydrogen bond interaction with the hydroxyl group of Tyr507 and also completes the co-ordination of the catalytic zinc ion.

The N-terminal portion of the peptide makes fewer interactions with the active site residues. The main point of contact in this region is between the peptide backbone at the

P_2 glutamine and the protein backbone at Ala340. This is another interaction that is commonly found in peptide binding to ACE homologues (Akif *et al.,* 2012).

Beyond this, the side chain of the last peptide residue for which electron density is visible, the arginine at P_4, appears to interact with the side chain of Asp501 and the protein backbone at Ala500. However, due to the poor difference density for the arginine side chain it was very difficult to model accurately, hence it is not clear how significant these interactions may actually be. A complete list of interactions occurring between BPP9a and the AnCE catalytic site is presented in Table 3.5 and summarised in Figure 3.5.

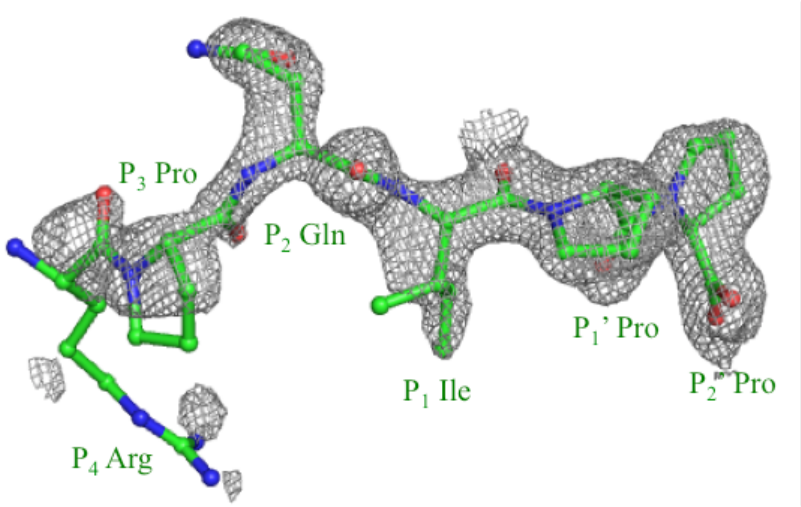


Figure 3.4: Fo-Fc Map for BPP9a Binding to AnCE. Fo-Fc map is shown as a grey mesh contoured at 3σ and the peptide BPP9a as green sticks. Density for the peptide backbone and side chains of P_2' to P_3 is unambiguous. There is little density visible for the P_4 Arg side chain, most likely due to the highly flexible nature of the side chain.

BPP9a atom	Interacting atom	Distance (Å)
P_2' Pro OXT	Water	2.84
P_2' Pro O	Lys495 NZ	2.72
P_2' Pro O	Gln265 NE2	3.07
P_2' Pro O	Tyr504 OH	2.60
P_1' Pro O	His337 NE2	2.94
P_1' Pro O	His497 NE2	2.95
P_1 Ile O	Tyr507 OH	2.68
P_1 Ile O	Zn^{2+}	2.53
P_2 Gln O	Ala340 N	2.79
P_2 Gln OE1	Water	2.85
P_2 Gln N	Ala340 O	2.95
P_4 Arg NH1	Water	2.69
P_4 Arg NH1	Asp501 O	2.80
P_4 Arg NH1	Ala500 O	3.06
P_4 Arg NH2	Ala500 O	3.01

Table 3.5: Interactions Formed by the BPP9a Peptide with AnCE Active Site Residues.

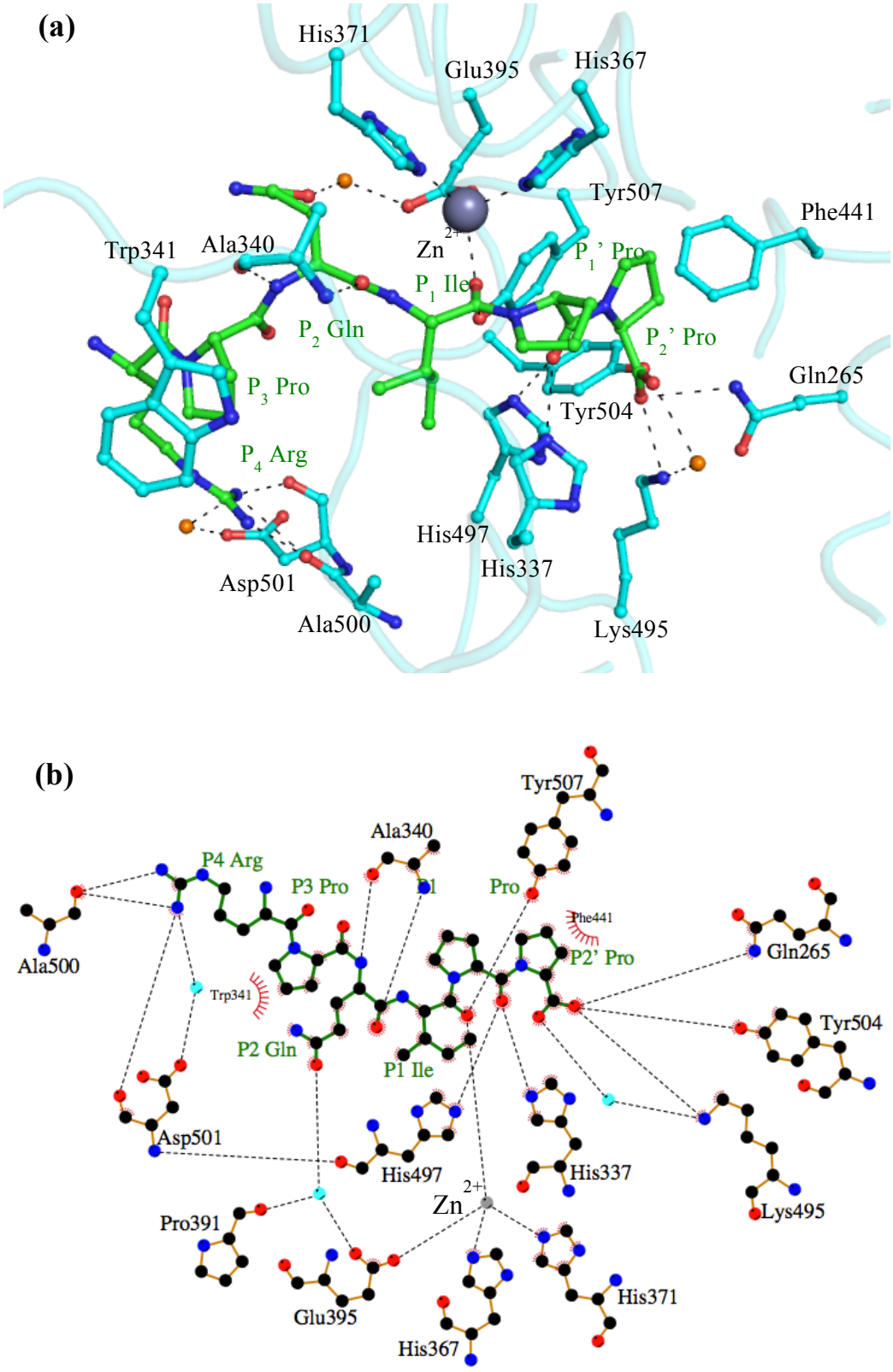


Figure 3.5: BPP9a Binding to AnCE. (a) Visualisation of the interactions between BPP9a and the AnCE active site residues. Peptide residues are shown as green sticks and labelled in green, AnCE residues are shown as cyan sticks with black labels. The active site zinc ion is shown as a grey sphere and key water molecules as orange spheres. (b) Schematic representation of the interactions shown in (a), here water molecules are shown as turquoise rather than orange spheres and protein residues in brown.

3.3.5 Description of BPPn Binding to the AnCE Active Site

The difference density present at the active site of the structure of AnCE with BPPn is far less extensive than that seen for AnCE with BPP9a. The density is located close to the catalytic zinc ion and whilst it appears to occupy a region around the S_1' and S_2' pockets it does not extend very far along the catalytic channel at all.

Based on these observations, the C-terminal tripeptide of BPPn was modelled into the active site such that the isoleucine, proline and proline residues adopted the P_1, P_1' and P_2' positions, respectively. As illustrated in Figure 3.6 the tripeptide is well accommodated by the density, indicating that it has been correctly positioned.

Following placement of the tripeptide there was some residual density remaining around the S_1 and S_2 subsites, Figure 3.6. Attempts were made to place an additional lysine residue at the N-terminus of the peptide, however there was insufficient density to justify this. Removing the flexible side chain of the lysine residue and reducing the occupancy of the peptide backbone to 0.5 both failed to provide a solution. Ultimately, water molecules were placed in the density.

The C-terminal tripeptide of BPPn adopts a very similar position in the AnCE active site to the equivalent residues in the BPP9a peptide. The two C-terminal proline residues form identical interactions with the S_2' and S_1' pockets of the enzyme. This includes interaction of the C-terminal carboxyl group with Lys495 and Tyr504 and of the P_1' proline carbonyl oxygen with the side chains of His337 and His497. BPP9a and BPPn both interact with the catalytic zinc ion via the carbonyl oxygen of their P_1 isoleucine residue.

The interactions formed by the P_1 residues in the two structures vary somewhat and this is a reflection of the slightly different positions adopted by the peptides at this point. The P_1 isoleucine in BPPn is considerably closer to the zinc ion and is able to make an additional interaction via its amide nitrogen. This change in position also means that it makes additional contacts with His371 and Glu368 that are not seen in the BPP9a structure. A complete list of interactions formed by BPPn with the AnCE active site is included in Table 3.6 and illustrated in Figure 3.7.

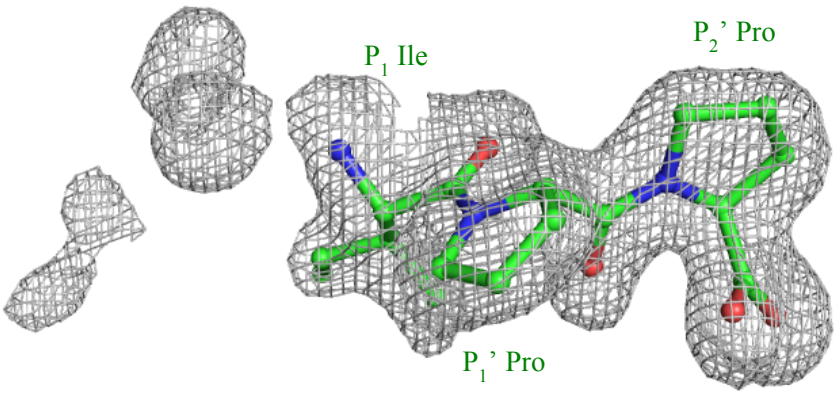


Figure 3.6: Fo-Fc Map for BPPn Binding to AnCE. Fo-Fc map is shown as a grey mesh contoured at 3σ. The three C-terminal residues of BPPn; isoleucine, proline and proline are shown as green sticks and are accommodated well by the density. Some residual density remains at the N-terminus of the peptide, however this was not sufficient to justify fitting more residues at the N-terminus of the peptide.

BPPn atom	Interacting atom	Distance (Å)
P_2' Pro OXT	Water	2.77
P_2' Pro O	Tyr504 OH	2.65
P_2' Pro O	Lys495 NZ	2.69
P_1' Pro O	His 497 NE2	3.12
P_1' Pro O	His337 NE2	2.82
P_1 Ile O	Tyr507 OH	2.59
P_1 Ile O	His367 NE2	3.20
P_1 Ile O	Zn^{2+}	2.46
P_1 Ile N	Zn^{2+}	2.30
P_1 Ile N	His371 NE2	3.23
P_1 Ile N	Glu368 OE2	2.82
P_1 Ile N	Water	2.60

Table 3.6: Interactions Formed by the BPP9n Peptide with AnCE Active Site Residues.

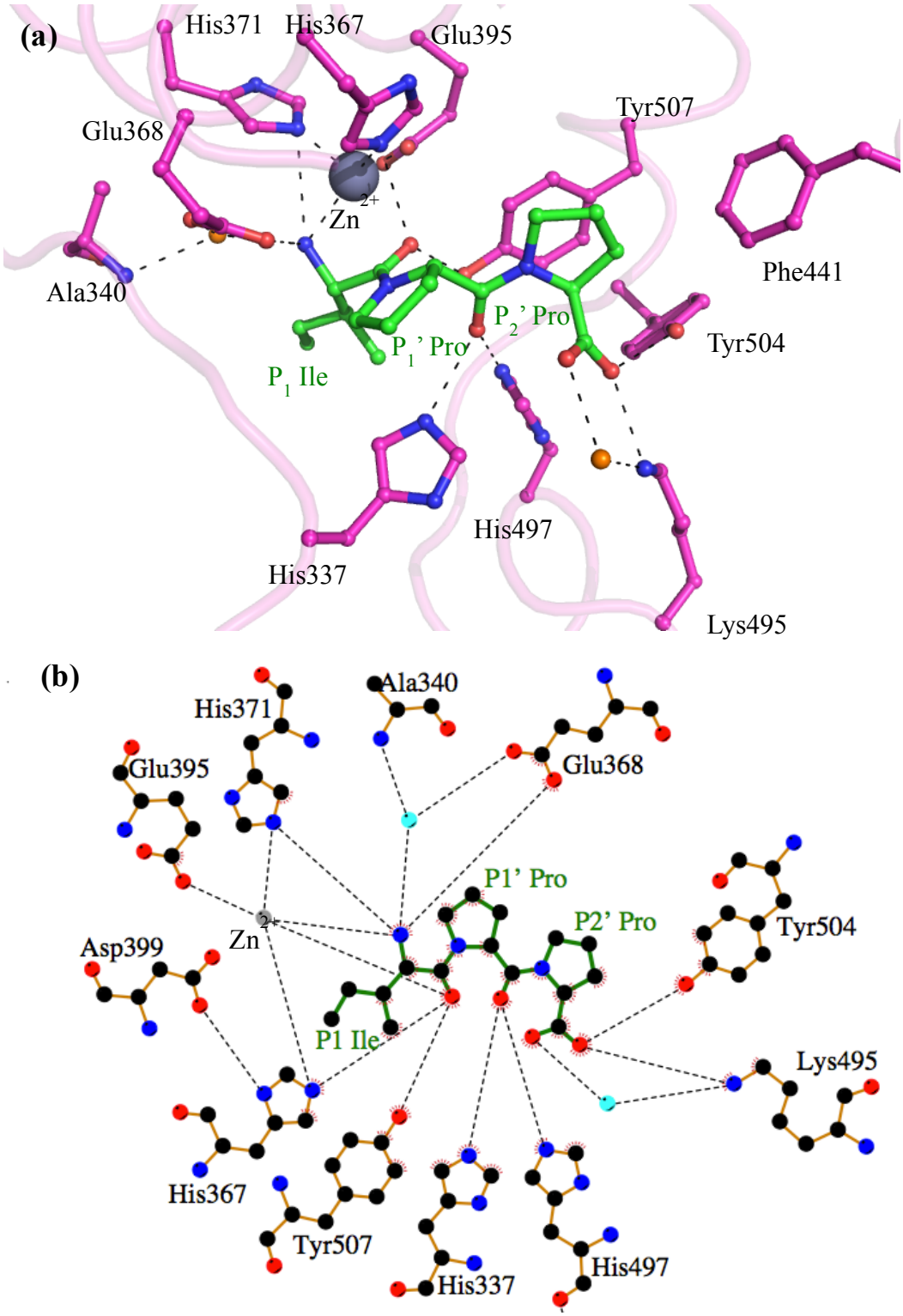


Figure 3.7: Binding of BPPn to the AnCE Active Site. (a) and (b) display the key interactions occurring between peptide and protein residues. BPPn residues are shown as green sticks and labelled accordingly. AnCE residues are shown as magenta sticks in (a) and brown sticks in (b) and labelled in black. The catalytic zinc ion is displayed as a grey sphere and water molecules involved in binding are shown as orange spheres in (a) and turquoise spheres in (b).

3.4 Discussion

3.4.1 The Structure of AnCE with BPP9a

It is clear by studying the difference density in Figure 3.4 that a significant portion of the peptide BPP9a is present in the AnCE active site in this structure. The five C-terminal residues (PQIPP) are accommodated comfortably by the density and are observed to form a number of interactions with active site residues. This includes the conserved interactions of the C-terminus of the peptide with Gln265, Lys495 and Tyr504 that acts as an anchor to hold the peptide in place.

The difference density for the sixth residue, the arginine at P_4 is not quite so unambiguous. Whilst there is clear density to accommodate the peptide backbone, there is very little at all for the side chain. Removing this residue completely resulted in a significant amount of difference density remaining following refinement. This indicates that this residue is definitely present and that the peptide is not being cleaved at this position.

The absence of observable density for the side chain is almost certainly due to the flexible nature of the long arginine side chain. In the absence of strong interactions holding it in a specific conformation, it is free to sample all of the space in the rather large S_4 cavity. This means that it will be present in multiple conformations throughout the crystal and hence not visible in the final structure.

The N-terminal portion of the BPP9a peptide may also not be visible for a similar reason. As illustrated in Figure 3.5, the majority of interactions between BPP9a and the active site involve the C-terminal portion of the peptide and are concentrated in the S_1' and S_2' pockets. In fact, beyond the isoleucine at the P_1 position, interactions are limited to two hydrogen bonds between the P_2 glutamine and Ala340 and a possible stacking interaction between P_3 proline and Trp341.

After this point, the substrate-binding channel opens out quite significantly. It is therefore likely that the remaining peptide residues, not constrained by strong interactions with the active site residues, are free to exist in multiple different conformations in this space. Hence, like the side chain of the P_4 arginine, they are not visible in the crystal structure, even though they are still present.

A second possible explanation for the absence of the N-terminal residues also exists, whereby AnCE has acted as an aminopeptidase and proteolytically removed these

residues. Although angiotensin converting enzymes normally function as dipeptidyl carboxypeptidases to remove the C-terminal dipeptide from their substrates, as discussed in section 1.6.4, ACE has been reported to act as an aminopeptidase towards certain substrates, such as GnRH (Jaspard *et al.,* 1993; Skidgel *et al.,* 1985).

This aminopeptidase activity normally occurs when the C-terminus of the protein is "protected" from hydrolysis. The presence of proline residues at the C-terminus of BPP9a means that that is the case here, as the unique nature of the proline side chain means that there is a secondary rather than a primary amine group, protecting the scissile bond from cleavage.

Although it is not impossible that the results observed here are as a consequence of ACE acting as an aminopeptidase towards BPP9a, it is quite unlikely. This is because the final residue that is visible at the N-terminus of the peptide chain, arginine, is flanked on either side by proline residues at the P_3 and P_5 positions. In the same way that the C-terminal proline residues make the BPPs resistant to the carboxypeptidase actions of ACE, surely a proline residue at the P_5 position would protect the peptide bond between positions P_4 and P_5 residues from any aminopeptidase activity.

3.4.1.1 Comparison of the Structure of AnCE with the Peptides BPP9a and BPPb

The structure of AnCE in complex with the related peptide BPPb has previously been determined (Akif *et al*., 2012). The sequence of BPPb is very similar to that of BPP9a, with small variations at the P_2 and P_6 positions where glutamine and tryptophan in BPP9a are replaced by lysine and proline in BPPb, respectively (Table 3.2).

Interestingly, both peptides bind to the AnCE active site in the same conformation and form identical interactions with the protein residues. This includes at the P_2 position, where the main chain atoms of lysine and glutamine from the two peptides both interact with the protein backbone at Ala340, Figure 3.8.

In both structures, only the six C-terminal residues of the peptide are visible (RPKIPP in BPPb and RPQIPP in BPP9a). The fact that this is also the case in the BPPb structure supports the hypothesis that the residues are not visible because they are disordered, rather than as a result of peptide bond cleavage by AnCE. This is because the additional proline residue at the P_6 position of BPPb, replacing tryptophan in BPP9a, would surely provide further protection to the N-terminus against aminopeptidase activity.

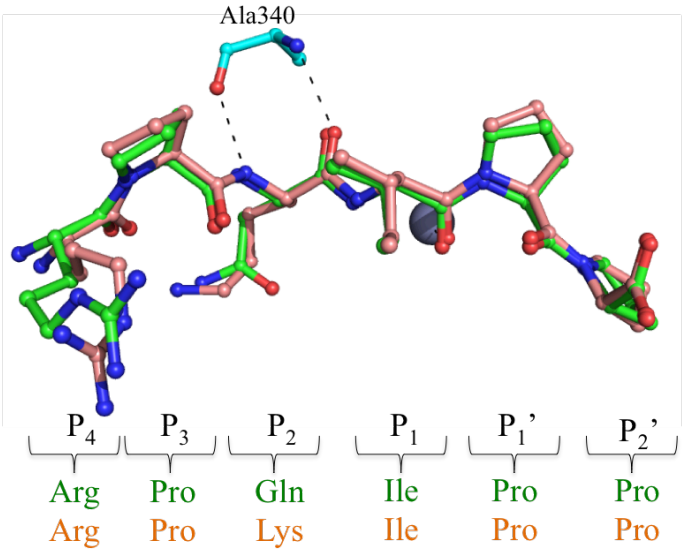


Figure 3.8: Binding of BPP9a and BPPb to AnCE. The peptides BPP9a and BPPb are shown as green and coral sticks, respectively and the residues at each position are annotated in the corresponding colours. Both peptides adopt the same conformation on binding to AnCE and hence form identical interactions with active site residues. This includes at the P_2 position, where glutamine in BPP9a is substituted for lysine in BPPb. Here the only contact between the peptides and the protein is between the peptide backbone and the main chain of Ala340. The Ala340 residue is included as cyan sticks to illustrate this.

3.4.1.2 Structural Insights into Domain Specific Inhibition by BPPs

The structure of human tACE (equivalent to the C-domain of sACE) with BPPb has previously been reported (Masuyer *et al.,* 2012) and is the only structure of an ACE homologue and a bradykinin potentiating peptide in which the entire peptide is visible. Peptide binding in this instance is accompanied by some significant structural changes: the zinc ion is absent and helices α1 and α2 of the N-terminal lid and the proximal loop region located at the entrance to the substrate binding channel have moved to accommodate the N-terminal residues of the peptide.

By comparing this structure to the structures of AnCE with BPP9a and BPPb (Akif *et al.,* 2012) it is possible to understand more about the importance of specific residues that are involved in active site interactions. Consequently, this provides some explanation as to why the full-length peptide is visible in the tACE structure but not in the AnCE structure. Furthermore, by comparing the interactions formed by these peptides in the AnCE and tACE active sites and considering the conservation of these residues in the N-domain, it is possible to suggest which residues may be involved in the domain selectivity of these peptides.

3.4.1.2.1 The Importance of the P_2 Peptide Residue in Domain Selective Inhibition

Both BPPb and BPP9a have been identified as C-domain selective inhibitors and the P_2 position has been implicated in this (Cotton *et al.,* 2002). The structures of AnCE with BPPb and BPP9a indicate that the interaction with the active site formed by the lysine and glutamine residues at this position, respectively, is limited to two hydrogen bonds between the peptide main chain and Ala340. Thus, this interaction is largely sequence independent and should be conserved in both domains of sACE.

The lysine at P_2 of BPPb forms an additional salt bridge with tACE active site residue Glu403. Glu403 is not conserved in AnCE and is instead replaced by Thr387, which would be unable to replicate this interaction. Comparison with the structure of the human N-domain indicates that Arg381 replaces Glu403, which would also not be able to replicate this interaction, Figure 3.9. This explains why both BPP9a and BPPb have a higher affinity for the C-domain of sACE compared to the N-domain and confirms the importance of the P_2 residue.

In AnCE, the side chain of residue Thr387 is relatively small, and whilst unable to form a salt bridge with the P_2 lysine of BPPb like that seen with Glu403 in tACE, it does not appear to obstruct peptide binding in any way (Figure 3.9).

Conversely, the large, positively charge side chain of Arg381 in the N-domain is likely to clash with the P_2 lysine of BPPb, decreasing the affinity of the peptide for the N-domain active site. The presence of the smaller glutamine side chain at P_2 in BPP9a is unlikely to pose such a problem, which would explain why BPP9a is a slightly better inhibitor of the N-domain than BPPb. These interactions are illustrated in Figure 3.9.

Interestingly, bradykinin, a peptide substrate that is hydrolysed with equal efficiency by both the N- and C-domains of human sACE (Georgiadis *et al.,* 2003) has a serine residue at the P_2 position. The small serine side chain would not clash with Arg381 in the N-domain whilst also not forming a salt bridge with Glu403 in the C-domain. This may help to explain the observed absence of domain selectivity. Angiotensin I, which is hydrolysed more efficiently by the C-domain than the N-domain, has a proline residue at the P_2 position. The P_2 proline would not be expected to interact with Glu403, but may clash with the long, polar side chain of Arg381, and this may contribute to the domain selectivity of angiotensin I.

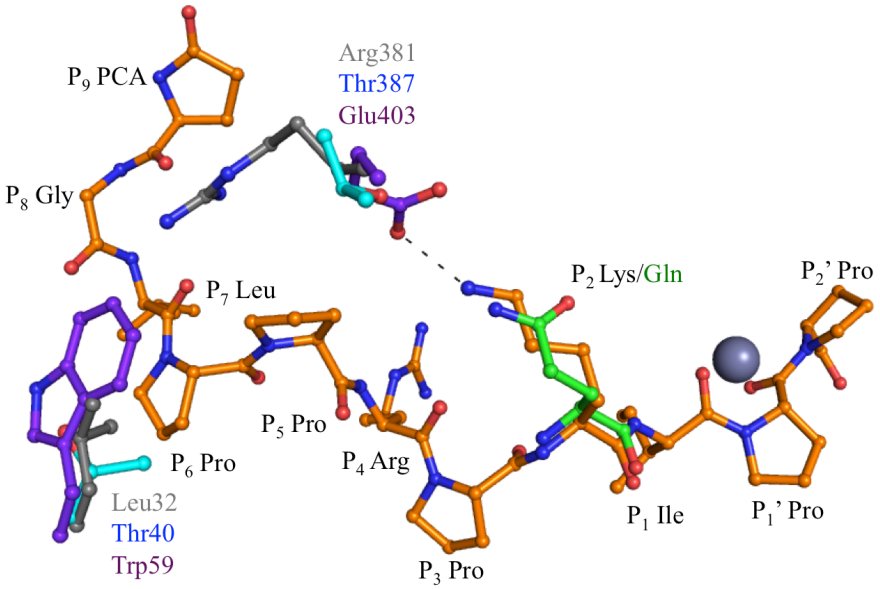


Figure 3.9: Key Features of BPP Binding to the N- and C-domains of sACE and to AnCE. The BPPb peptide from the structure with tACE is shown as orange sticks with the residues labelled in black. The P_2 lysine forms a salt bridge with tACE Glu403 (purple sticks) that is not replicated in AnCE where Glu403 is replaced by Thr387 (cyan sticks). Arg381 at this position in the N-domain (grey sticks) is likely to clash with the positive lysine side chain, explaining why BPPb is a poor inhibitor of the N-domain. In BPP9a the P_2 lysine is replaced by glutamine, shown as green sticks. The smaller glutamine side chain should reduce the clash with Arg381 in the N-domain, hence why BPP9a is a slightly better inhibitor of the N-domain than BPPb. The P_6 proline residue in BPPb appears to form a stacking interaction with Trp59 (purple sticks) in the C-domain, helping to stabilise the N-terminal portion of the peptide. This interaction is not conserved in AnCE and the N-domain, where Trp59 is replaced by Thr40 (blue sticks) and Leu32 (grey sticks), respectively. The absence of this stabilising interaction may explain why the N-terminal portion of BPP9a is not visible in the structure with AnCE (Figure 3.4). The zinc ion is shown as a grey sphere as a point of reference.

3.4.1.2.2 The Importance of the P_6 Peptide Residue in Domain Selective Inhibition

The only structure of an ACE homologue with a BPP in which the N-terminal portion of the peptide is visible is that of tACE with BPPb. Studying the N-terminal portion of BPPb, there are surprisingly few interactions with tACE active site residues. Those that are present are all water-mediated and only involve the protein backbone, thus should be conserved in AnCE, and indeed in the N-domain.

The only exception to this is an interaction between the hydroxyl group of the Tyr360 side chain and the carbonyl oxygen of the proline at the P_5 position. This tyrosine residue is conserved in both AnCE (Tyr344) and the N-domain (Tyr338) thus this interaction should be conserved and so does not explain the observed differences in peptide binding, nor the selective inhibition of the C-domain.

The structure presented by Masuyer *et al.,* indicates that a stacking interaction exists between tACE Trp59 and the proline residue at P_6 of BPPb. This tryptophan residue, and consequently this interaction, is conserved neither in AnCE nor in the N-domain and is replaced by Thr40 and Leu32, respectively, Figure 3.9.

All of the BPPs have either a tryptophan (BPP9a and BPPn) or a proline (BPPb, BPP12b and BPPc) residue at the P_6 position, indicating that this may be an important stabilising interaction, helping to keep the peptide bound in a single conformation. The absence of an equivalent interaction between AnCE and BPP9a and BPPb may explain why the N-termini of these peptides is not visible in the crystal structures.

It is possible that the absence of this interaction between BPPs and the N-domain may contribute to the C-domain selectivity observed for BPP9a and BPPb. However a proline residue at this position is conserved in BPPc, which is an equally effective inhibitor of both domains.

Perhaps more relevant is the observation that BPP9a has a tryptophan residue at P_6. BPP9a is an exceptionally potent inhibitor of the C-domain (Table 3.1) and the increased strength of the interaction between the two tryptophan residues compared to a tryptophan and a proline may well be a contributory factor.

3.4.2 The Structure of AnCE with BPPn

Analysis of the difference density shown in Figure 3.6 indicates that the C-terminal tripeptide of BPPn can confidently be placed in the active site such that the S_1, S_1' and S_2' pockets are occupied by the isoleucine, proline and proline residues respectively. This is in keeping with the common mode of peptide binding seen in previous ACE structures (Masuyer *et al.,* 2012).

In contrast to the structure of AnCE with BPP9a, it was not possible to place any further residues at the N-terminus of the peptide. There are two possible reasons for this; the peptide is being hydrolysed or the N-terminal portion is too flexible to see in the crystal structure.

As discussed in relation to the structure of AnCE with BPP9a (section 3.4.1), it is possible that the presence of only the C-terminal tripeptide is due to hydrolysis of BPPn by AnCE. In this case this would most likely be due to AnCE displaying C-terminal

tripeptidase activity. This is theoretically possible, as ACE has been reported to cleave the C-terminal tripeptide from substrates when the C-terminal dipeptide is protected. The primary example of this is seen with substance P (Skidgel *et al.,* 1985).

In the case of BPPn, the C-terminal proline residues should prevent AnCE acting as a dipeptidyl carboxypeptidase, hence, it is feasible that it has instead cleaved the peptide bond between the P_1 and P_2 residues. This would have produced two peptide fragments; PCA-LWPRRK and IPP, the latter of which has remained bound to the active site.

Analysis of the difference density in the active site suggests that this is unlikely to have occurred. There is some remaining density following the N-terminal isoleucine residue (Figure 3.6). Although, as detailed in section 3.3.5, this density is not sufficient to accommodate an additional peptide residue, it does indicate that there is "something" there. Therefore, the most likely explanation for the results observed is that the remainder of the peptide is present, but due to a lack of interactions with the active site channel, it is highly mobile and hence not visible in the structure.

This is the same effect that is observed in the structures of AnCE with BPPb and BPP9a, albeit occurring with much more of the peptide. This indicates that BPPn is unable to replicate some of the interactions formed by BPP9a and BPPb with the active site that are important in stabilising the peptide.

3.4.2.1 The Importance of the P_3 Proline Residue in ACE Inhibition by BPPs

As indicated in Table 3.2, the sequence of BPPn differs from that of BPPb and BPP9a at the P_3 position. The proline residue that is conserved in BPPs was replaced by an arginine in BPPn. The aim of doing this was to investigate the role of this position in inhibitor binding. The fact that the P_3 proline is conserved in all of the BPPs indicates that it may be an important inhibitory feature; so mutating it to arginine, which has a large positively charged side chain, should help to clarify whether this is the case.

The fact that only the C-terminal tripeptide of BPPn is visible in the AnCE active site indicates that the proline residue at the P_3 position is indeed an important conserved feature of BPPs. Considering the structure of AnCE with BPPb and BPP9a, this residue potentially forms a stacking interaction with Trp341 that would be conserved in both the N- and C-domains with Trp335 and Trp357, respectively (Figure 3.10). The large polar side chain of arginine would be unable to replicate this, and the loss of this stabilising interaction appears to result in the N-terminal portion of the peptide being disordered.

Interestingly, an aromatic residue is conserved at the P_3 position in two of the major ACE substrates: histidine in angiotensin I and phenylalanine in bradykinin.

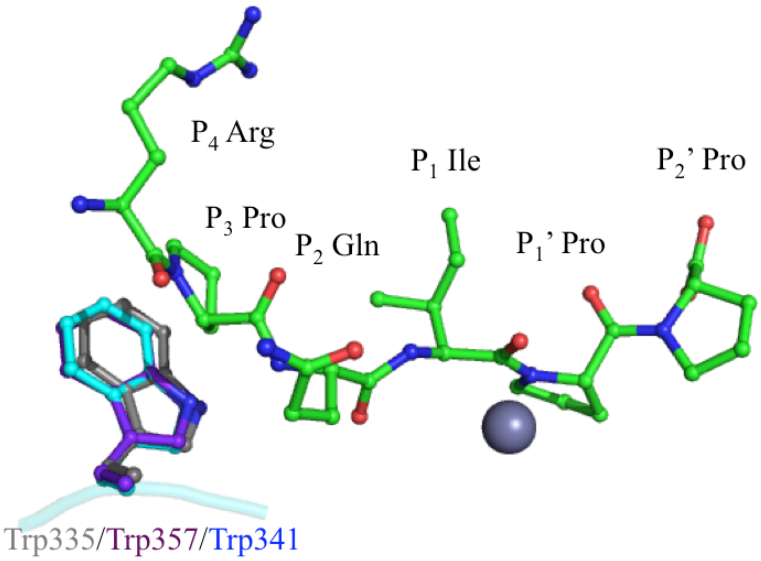


Figure 3.10: **The Importance of the P_3 BPP Proline Residue.** The six C-terminal residues of BPP9a from the structure with AnCE are shown as green sticks and the active site zinc ion as a grey sphere. Trp341 in AnCE, cyan sticks, forms a stacking interaction with the proline at position P_3. Structural and sequence analysis indicates that this would be conserved in the N-domain with Trp335, grey sticks, and in the C-domain with Trp341, purple sticks.

This is only a hypothesis, based on the structural data that are available. The next step in this work would be to obtain kinetic data to investigate the effect that this mutation has on the potency of BPPn as an ACE inhibitor. Based on the structural information available, it is likely that it would be a very poor inhibitor of the N-domain due to the P_3 substitution being compound by the P_2 lysine clashing with Arg381, and a moderate to weak inhibitor of the C-domain and AnCE.

An interesting future experiment would also be to replace the P_3 proline residue with a larger aromatic side chain, for example tryptophan to see if this increases the potency of the peptide as an inhibitor of ACE.

3.4.3 Failure to Crystallise AnCE with BPPc and BPP12b

Unfortunately, no crystals were obtained for AnCE in complex with BPPb and BPPc. This is very unusual; the crystallisation condition used here typically yields diffraction

quality crystals of AnCE within only a few days and has previously been used to crystallise AnCE with a wide range of inhibitors (Akif *et al.,* 2010).

The fact that crystals could not be obtained using this condition for AnCE with BPP12b and BPPc indicates that binding of these peptides may be causing some significant structural rearrangements. This in turn may affect crystal contacts between molecules, altering the conditions under which the protein will crystallise.

This hypothesis is supported by the crystal structure of tACE with BPPb reported in 2012 (Masuyer *et al.,* 2012). Here, as illustrated in Figure 3.11, binding of the peptide has caused significant movement of secondary structure elements in order that the full peptide may be accommodated in the substrate-binding channel.

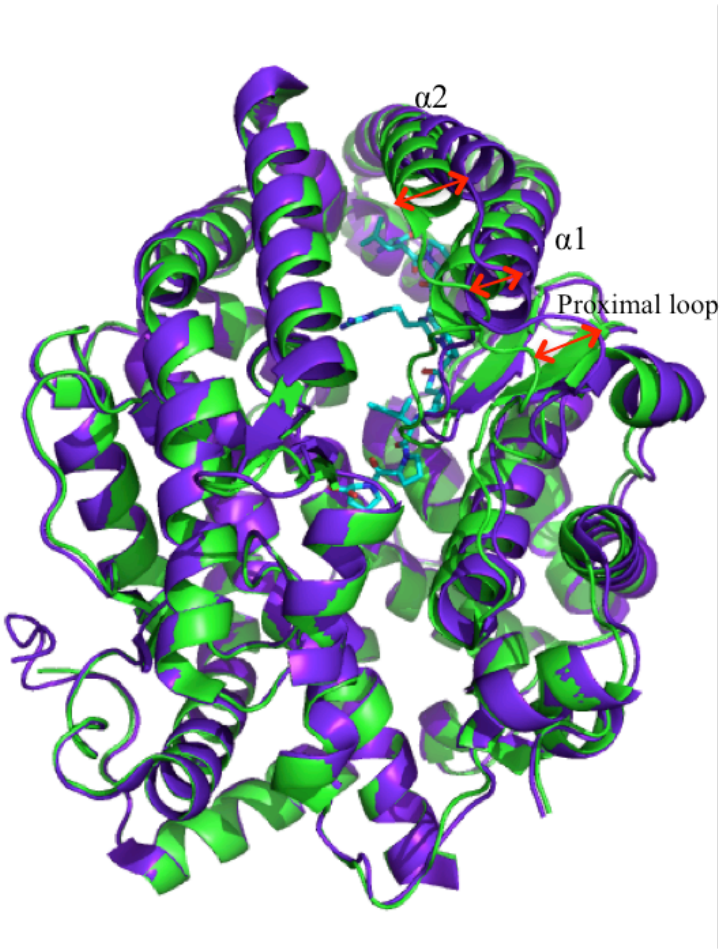


Figure 3.11: Structural Changes Observed on BPPb Binding to tACE, as Reported by Masuyer *et al.*, 2012 (PDB code 4AA2). Cartoon representation of the structure of tACE bound to BPPb (purple) is superposed on the native tACE structure (green). BPPb residues occupying the substrate-binding channel are shown as cyan sticks. Clear movement of helices α1 and α2, components of the N-terminal lid, and the proximal loop region occur on peptide binding, as indicated by the red arrows.

Examination of the sequences of BPP12b and BPPc compared to the other BPPs (Table 3.2) indicates that both of these peptides are very proline rich. In addition to the conserved proline residues at positions P_2', P_1', P_3 and P_5, BPPc has two further proline residues at P_2 and P_6 and BPP12b at P_6 and P_7.

The unique cyclic nature of the proline side chain means that free rotation around the peptide bond is greatly restricted. The abundance of proline residues in BPPc and BPP12b is therefore likely to make these peptides very rigid in nature and it is quite possible that substantial conformational changes may be required to accommodate these peptides in the catalytic channel, explaining why complexes of AnCE with BPPc and BPP12b have not been observed to crystallise under "standard" conditions.

Further attempts have been made to crystallise AnCE with BPPc and BPP12b and recently, a second condition in which AnCE forms diffraction quality crystals has been identified. This work is outlined in Chapter 6.

3.5 Conclusions and Future Directions

The properties of bradykinin potentiating peptides isolated from the venom of *B. jararaca* as potent ACE inhibitors are well known. Indeed, these peptides formed the basis of the development of the very first ACE inhibitors (Cushman and Ondetti, 1999).

More recently, two of these peptides, BPP9a and BPPb, have been identified as selective inhibitors of the C-domain of sACE, and the residue at the P_2 position has been implicated in this (Cotton *et al.,* 2002).

Here, the structure of the *D. melanogaster* ACE homologue, AnCE has been reported with BPP9a. Comparing this structure to the structures of tACE and AnCE with BPPb has provided further evidence of the importance of this position in C-domain selectivity.

The N-terminal residues of BPP9a were not visible in the structure with AnCE and this was thought to be due to high flexibility resulting from a lack of interactions with the active site. Comparison with the structure of tACE with BPPb highlighted an interaction between proline at P_6 and Trp59, which is conserved neither in AnCE nor in the N-domain. This suggests that the P_6 peptide position may also have an important role in ACE inhibition and domain selectivity, the first time that this position has been implicated in this way.

A new peptide, BPPn, was designed in order to investigate the role of the conserved proline at position P_3 in the BPPs. The structure of AnCE with BPPn only showed clear density for the three C-terminal residues, thus indicating that interactions of the P_3 proline with the ACE active site might be important for peptide binding and inhibition. A conserved tryptophan residue has been identified that is likely to be responsible for this.

Having established that peptide residues P_2, P_3 and P_6 might be involved in peptide binding and inhibition, future work should include inhibition assays to study the effects of mutating these on inhibition of AnCE and each domain of sACE. Similarly, it would be interesting to extend this work to look at the effects of mutating other peptide residues to investigate their roles in inhibition.

This could include studying the peptides BPP12b and BPPc, which are both particularly proline rich. Initial attempts to crystallise AnCE with BPP12b and BPPc have proved unsuccessful. This could be as the result of structural rearrangements occurring on peptide binding, similar to those observed for tACE with BPPb (Masuyer *et al.,* 2012). In an attempt to overcome this, experiments have been initiated to try and find a new condition in which AnCE crystals can be obtained. This is detailed in Chapter 6.

In conclusion, the work presented here provides more information about the structural basis of ACE inhibition by bradykinin potentiating peptides from *B. jararaca* that could be used in work towards the development of new inhibitors. It is also an illustration of how the *D. melanogaster* homologue, AnCE, can be used as a model in structural studies.

Chapter 4: Crystal Structures of AnCE in Complex with the ACE Inhibitors Spinorphin and Tynorphin

4.1 Introduction

The requirement for a new generation of ACE inhibitors that selectively target only one of the two domains of the enzyme has been discussed in some detail in sections 1.5.1 and 3.1.1.

The design of the original ACE inhibitors was greatly influenced by studying the natural inhibitory peptides from the venom of the pit viper *B. jararaca*. In the same way, analysis of the structural basis of domain inhibition by these peptides has now been able to provide important information about ways in which subtle differences in the active site pockets of the N- and C-domains could be exploited to develop domain specific inhibitors (Chapter 3).

Haemorphins are a group of naturally occurring peptides that are produced by the enzymatic cleavage of the β-chain of haemoglobin (Brantl *et al.,* 1986). Originally isolated from bovine blood, they are also present in humans and some have been shown to function as ACE inhibitors (Lantz *et al.,* 1991). The fact that they are derived from haemoglobin suggests that they may have a role in local regulation of ACE activity *in vivo* as part of tissue specific RAAS (Fruitier-Arnaudin *et al.,* 2002).

4.1.1 Spinorphin

Spinorphin was the first haemorphin shown to have ACE inhibitory properties. It is a heptapeptide with the sequence LVVYPWT, corresponding to residues 32-38 of the β-chain of human haemoglobin. Spinorphin was originally isolated in 1993 from bovine spinal cord as an inhibitor of enkephalin degrading enzymes, but it is also found in human cerebrospinal fluid (CSF). During the course of this initial work spinorphin was identified as an inhibitor of ACE, with an IC_{50} of 2.7 μM reported (Nishimura and Hazato, 1993).

In spite of the lack of sequence identity between spinorphin and the BPPs from *B. jararaca*, the reported IC_{50} values are of the same order of magnitude. It would therefore be interesting to study ACE inhibition by spinorphin from a structural perspective, to try and learn more about the mechanism of inhibition and whether this could be applied to the design of new, more effective ACE inhibitors.

4.1.2 Tynorphin

A second peptide, tynorphin, was developed in 2000. Tynorphin is a truncated derivative of spinorphin consisting of the five central residues: VVYPW. Tynorphin was designed

in an attempt to develop a more potent inhibitor of one of the other enzymes inhibited by spinorphin, DPPIII (dipeptidyl peptidase III). Whilst removal of the N-terminal leucine and C-terminal threonine residues did make tynorphin a more potent inhibitor of DPPIII, tynorphin has actually been shown to be a much weaker inhibitor of ACE than spinorphin, thus indicating that these terminal residues are in some way important for ACE inhibition (Yamamoto *et al.,* 2000). Structural studies on ACE inhibition by tynorphin could be used to help to understand this, thereby providing more information about essential features of ACE inhibitors.

4.1.3 Aims

The aim of this work is to use the *D. melanogaster* ACE homologue AnCE as a model to study the structural basis of ACE inhibition by spinorphin and tynorphin. In particular to understand the importance of the terminal residues in spinorphin that appear to have such a profound effect on its inhibitory properties. The long-term aim would be to use this information to aid the design of a new generation of ACE inhibitors.

4.2 Methods

4.2.1 Protein Expression, Purification and Crystallisation

Recombinant AnCE was expressed by *P. pastoris* and purified using a two step protocol involving hydrophobic interaction chromatography followed by size exclusion chromatography as detailed in section 2.7.

Purified AnCE was co-crystallised with spinorphin or tynorphin (Tocris Bioscience, Bristol, UK) using the hanging drop vapour diffusion technique following the general protocol and conditions for AnCE, outlined in section 2.7.4.

Immediately prior to crystallisation 17.5 mg ml^{-1} AnCE was incubated with 10 mM of spinorphin or tynorphin in a 1:4 peptide:protein ratio and incubated on ice for approximately 30 minutes. Hanging drops were then prepared containing 1 µl of the protein and peptide mixture and 1 µl of the reservoir solution. Drops were suspended over 500 µl of reservoir solution and incubated at 21 ^{o}C.

4.2.2 X-ray Diffraction Data Collection

Previous diffraction experiments with AnCE have shown that the relatively high concentration of sodium citrate (1.2-1.3 M) used in the crystallisation solution is sufficient for cryo-protection (Akif *et al.,* 2010). For this reason no additional cryo-protectant solutions were added to the drop prior to crystal harvesting. Instead 1 µl of the

reservoir solution was added to maximise sodium citrate concentration and to prevent the drop from drying out.

Crystals were mounted directly into litholoops (Molecular Dimensions Ltd., Newmarket, UK) and flash frozen in liquid nitrogen for transportation to the synchrotron. All data were collected at DLS.

Four separate data collection experiments were performed using crystals grown under identical conditions. In all cases samples were mounted using the sample changer robots on the beamlines and diffraction data were collected at 100 K.

The first data were collected on 30th June 2012 using beamline i04 and an ADSC CCD Q315 detector. A single dataset for AnCE with tynorphin was collected. The crystallisation and data collection for this sample were both performed by previous group member Dr Geoffrey Masuyer.

The remaining data were collected on 24th May 2015 on beamline i03, 2nd July 2015 on beamline i04 and 25th July 2015 on beamline i02. These datasets were recorded using Pilatus3 6M, Pilatus 6M-F and Pilatus 6M detectors respectively.

4.2.3 Data Processing and Structure Solution

All datasets collected were initially processed using the xia2 pipeline at DLS (Winter, 2010). The "best" datasets from each visit were then selected based primarily on resolution and completeness. For these datasets integration was performed using MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) and scaling using Aimless (Evans, 2006; Evans, 2007). Initial phases were calculated by molecular replacement with Phaser (McCoy *et al.,* 2007) using the native AnCE structure, PDB code 2X8Y (Akif *et al.,* 2010), as a search model.

Iterative rounds of model building and refinement were performed using C*oot* (Emsley *et al.,* 2011) and REFMAC (Murshudov *et al.,* 2011), respectively with 5 % of residues separated for use as the R_{free} set in cross-validation. This process included manually building spinorphin and tynorphin residues into difference density in the active site. Water molecules were added where Fo-Fc electron density peaks exceeded 3σ and potential hydrogen bonds could be made. The quality of the final structure was validated using MolProbity (Chen *et al.,* 2010). LigPlot^{+} (Laskowski and Swindells, 2011) was used to study interactions between the inhibitors and the active site. All figures were

prepared using Pymol (The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC), which was also used to perform RMSD calculations.

4.3 Results

4.3.1 Co-crystallisation of AnCE with Spinorphin and Tynorphin

Crystals appeared in all drops (1.2-1.3 M sodium citrate, 0.1 M HEPES, pH 7.5) within 7 days of incubation at 21 ^{o}C. An image of one of these crystals mounted in a litholoop immediately prior to data collection is shown in Figure 4.1.

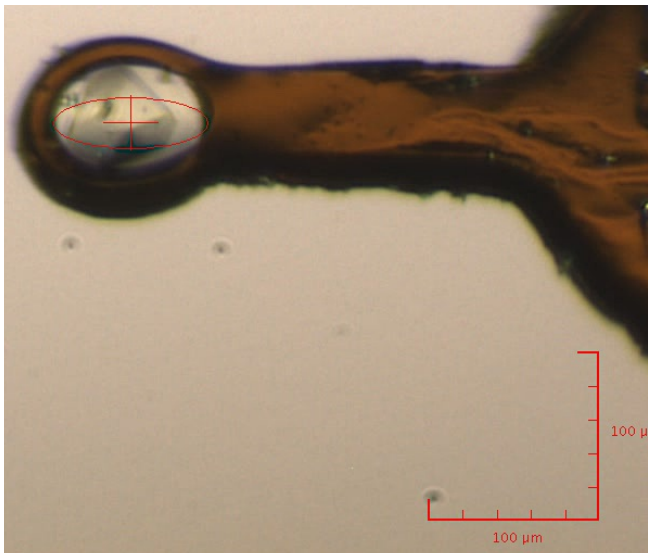


Figure 4.1: A Crystal of AnCE with Spinorphin Mounted in a Litholoop Immediately Prior to Data Collection. The size of the crystal is indicated by the 100 µm scales bars in the bottom right of the image.

4.3.2 X-ray Diffraction Data Collection

X-ray diffraction data for the two complexes were collected on four visits to DLS. Four datasets were collected for AnCE in complex with tynorphin: AnCE_Tyn_1-4, and seven for AnCE with spinorphin: AnCE_Spin_1-7. All datasets were successfully processed in the space group *h3* using the xia2 pipeline (Winter, 2010) at DLS. A summary of the data collected and the statistics from xia2 processing are presented in Table 4.1 and an example of a diffraction image is shown in Figure 4.2.

Based on the statistics obtained from the processing with xia2, one dataset for each complex from each synchrotron visit was selected for further processing. This selection was based on the maximum resolution that could be achieved whilst retaining acceptable values for completeness and $R_{merge.}$ The datasets chosen were AnCE_Tyn_1, AnCE_Tyn2_2, AnCE_Tyn_3, AnCE_Spin_4, AnCE_Spin_5 and AnCE_Spin_6.

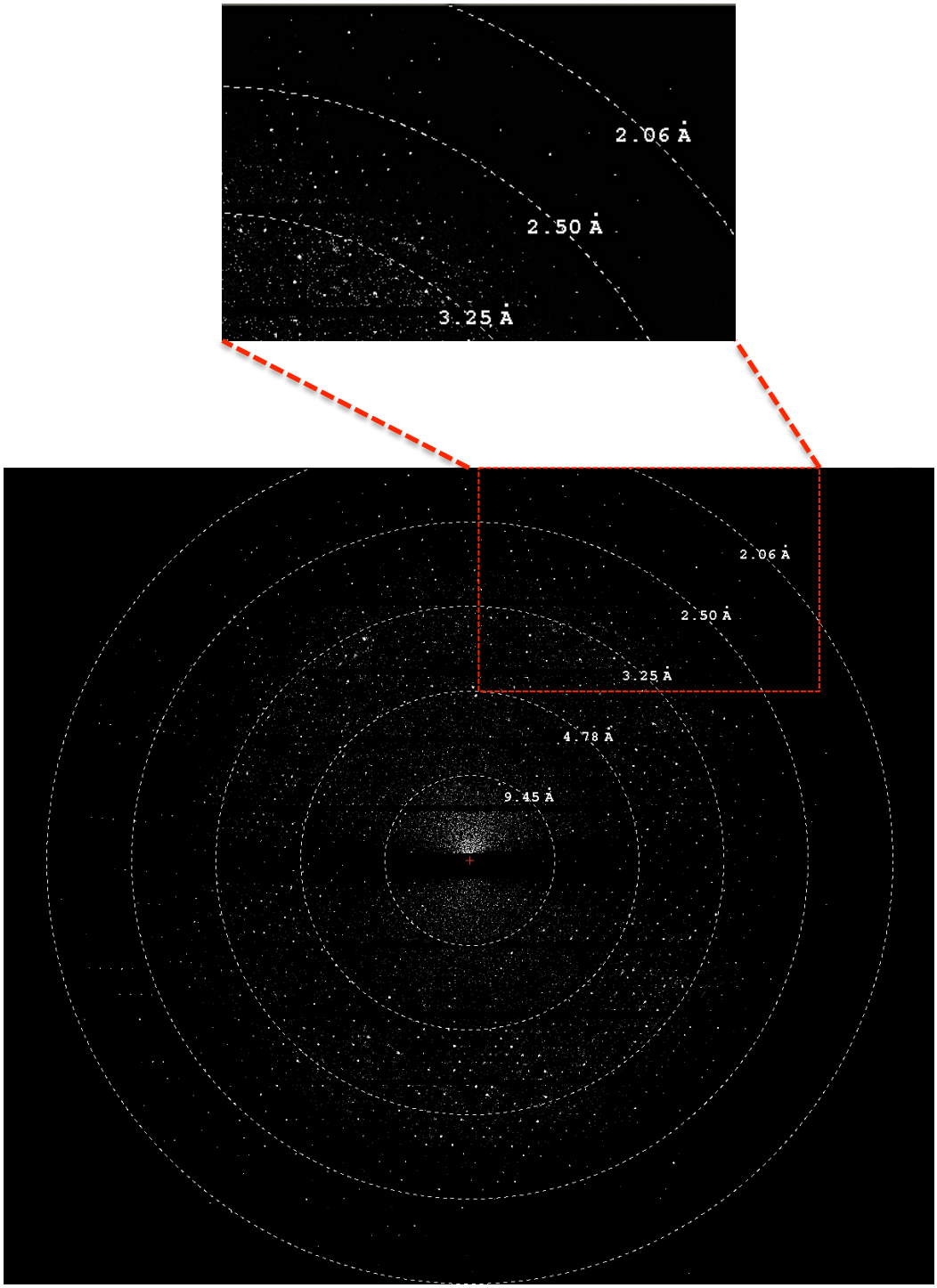


Figure 4.2: An Example of a Diffraction Image Collected for AnCE with Tynorphin Showing Diffraction to Beyond 2.06 Å.

Dataset name	Collection date	Beamline	Detector	Number of images	Phi range (°)	Resolution range (Å)	Overall completeness (%)
AnCE_Tyn_1	30th June 2012	i04	ADSC CCD Q315	120	0-120	85.84-2.16	99.7
AnCE_Tyn_2	2nd July 2015	i04	PILATUS 6M-F	720	150-330	86.68-2.01	99.7
AnCE_Tyn_3	25th July 2015	i02	PILATUS 6M	774	111-227.1	49.66-1.71	99.5
AnCE_Tyn_4	25th July 2015	i02	PILATUS 6M	774	99-215.1	49.64-1.99	98.7
AnCE_Spin_1	24th May 2015	i03	PILATUS3 6M	1800	0-180	60.85-3.06	99.7
AnCE_Spin_2	24th May 2015	i03	PILATUS3 6M	1800	0-180	49.97-2.30	99.9
AnCE_Spin_3	24th May 2015	i03	PILATUS3 6M	1800	180-360	28.51-4.23	97.1
AnCE_Spin_4	24th May 2015	i03	PILATUS3 6M	1800	0-180	86.15-2.00	99.8
AnCE_Spin_5	2nd July 2015	i04	PILATUS6M-F	900	20-200	88.76-2.07	99.9
AnCE_Spin_6	25th July 2015	i02	PILATUS 6M	920	160-252	86.68-1.90	95.8
AnCE_Spin_7	25th July 2015	i02	PILATUS 6M	1500	57-282	60.75-2.10	97.7

Table 4.1: Summary of Data Collection and Processing Using Xia2 for AnCE in Complex with Tynorphin and Spinorphin. The datasets highlighted in yellow are those that were selected for further processing based on resolution and completeness.

4.3.3 Data Processing and Structure Solution

The raw data images from the six selected datasets highlighted in Table 4.1 were processed using MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) and Aimless (Evans, 2006; Evans, 2007) in the space group *h3*. Phases were calculated by molecular replacement performed using Phaser (McCoy *et al.,* 2007) with the native AnCE structure (Akif *et al.,* 2010) as the search model. In each case there was one molecule per asymmetric unit in the crystal.

The models resulting from molecular replacement were all subjected to an initial round of rigid body refinement using REFMAC (Murshudov *et al.,* 2011). Following this, clear density for glycosylated sugars was visible at residues Asn53, Asn196 and Asn311 and for the catalytic zinc ion co-ordinated by residues His367, His371 and Glu395. These were added to the model accordingly. Spinorphin and tynorphin residues were also placed in the active site based on interpretation of the electron density. This process is discussed in detail for each dataset in the following sections. Further rounds of refinement were then performed prior to structure validation using Molprobity (Chen *et al.,* 2010). Data processing statistics are summarised in Tables 4.2 and 4.3.

The overall topology of each structure is identical to that of 2X8Y (RMSD values 0.115-0.228 Å) with no evidence of any large-scale movement of secondary structure elements. Each molecule therefore has a predominantly helical structure. Key features include the characteristic N-terminal lid over the active site that is formed by helices α1, α2 and α3 and a long substrate-binding channel extending almost the entire length of the molecule. The catalytic zinc ion is found at the centre of this channel, marking the location of the active site.

4.3.4 Analysis of Active Site Difference Density and Inhibitor Binding

All structures contained some difference density in the active site into which spinorphin and tynorphin residues were placed. The extent of this density varied significantly between datasets and the individual cases are discussed in detail in the following sections.

	AnCE_Tyn_1	AnCE_Tyn_2	AnCE_Tyn_3
Beamline at DLS	i04	i04	i02
Detector	ADSC CCD Q315	PILATUS 6M-F	PILATUS 6M
Wavelength (Å)	0.980	0.980	0.979
Space group	*h3*	*h3*	*h3*
Unit cell parameters **a, b, c (Å)** **α, β, γ (°)**	 171.68, 171.68, 102.46 90.00, 90.00, 120.00	 173.40, 173.40, 102.80 90.00, 90.00, 120.00	 173.49, 173.49, 102.35 90.00, 90.00, 120.00
Molecules per asymmetric unit	1	1	1
Resolution range (Å)	85.84-2.16	86.68-2.01	86.74-1.80
$^{\#}R_{merge}$ (%)	18.3 (65.0)	6.2 (42.7)	6.7 (57.2)
$^{+}R_{pim}$ (%)	17.5 (61.8)	4.8 (35.5)	5.8 (49.8)
Mean *I/σ(I)*	2.9 (1.4)	11.8 (2.3)	7.3 (1.4)
Completeness (%)	99.7 (97.9)	99.7 (95.3)	99.1 (99.3)
Number of reflections: **Total** **Unique**	 190, 357 60, 489	 344, 197 76, 464	 322, 060 105, 479
Multiplicity	3.1 (3.0)	4.5 (3.0)	3.1 (3.1)
Wilson *B* factor (Å^2)	23.65	35.19	29.28
Average *B* factor (Å^2) **Protein** **Peptide** **Zinc** **Glycosylated sugars** **Solvent**	 23.17 23.64 20.75 43.77 26.72	 34.49 43.75 29.69 58.56 40.60	 28.44 39.01 24.84 50.39 36.97
$^{\S}R_{cryst}$/$^{\#\#}R_{free}$ (%)	18.75/22.27	17.47/20.62	18.36/21.13
R.M.S deviation from ideal values **Bond lengths (Å)** **Bond angles (°)**	 0.009 1.301	 0.009 1.245	 0.009 1.248
Ramachandran plot statistics (%) **Favoured** **Disallowed**	 97.84 0.17	 98.16 0.00	 98.49 0.00

$^{\#}R_{merge} = \Sigma_{hkl}\Sigma_i|\boldsymbol{I_i}\,(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I_i}(hkl)$, where $\boldsymbol{I}_i$(hkl) is the intensity of *i*th measurement and <***I(hkl)***> is the average of symmetry-related observations of a unique reflection.

$^{+}R_{pim} = \Sigma_{hkl}\sqrt{\frac{1}{n-1}}\Sigma_{i=1}^{n}|\boldsymbol{I}_i(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I}_i(hkl)$

$^{\S}R_{cryst} = \Sigma_h|F_o - F_c|/\Sigma_h F_o$, where F_o and F_c are observed and calculated structure factor amplitudes of reflection *h*, respectively.

$^{\#\#}R_{free}$ is equal to R_{cryst} for a randomly selected 5 % subset of reflections.

Table 4.2: X-ray Data Collection, Processing and Refinement Statistics for AnCE with Tynorphin. Outer shell values are shown in parentheses.

	AnCE_Spin_4	AnCE_Spin_5	AnCE_Spin_6
Beamline at DLS	i03	i04	i02
Detector	PILATUS3 6M	PILATUS 6M-F	PILATUS 6M
Wavelength (Å)	0.980	0.980	0.979
Space group	*h3*	*h3*	*h3*
Unit cell parameters **a, b, c (Å)** **α, β, γ (°)**	 172.30, 172.30, 101.73 90.00, 90.00, 120.00	 173.50, 173.50, 103.10 90.00, 90.00, 120.00	 173.36, 173.36, 103.12 90.00, 90.00, 120.00
Molecules per asymmetric unit	1	1	1
Resolution range (Å)	86.15-2.00	88.76-2.07	86.68-1.90
$^{\#}R_{merge}$ (%)	9.6 (68.4)	5.4 (31.3)	11.8 (66.3)
$^{+}R_{pim}$ (%)	8.6 (60.4)	4.2 (26.4)	11.8 (66.1)
Mean *I/σ(I)*	6.3 (1.5)	13.9 (3.6)	3.8 (1.1)
Completeness (%)	99.5 (99.6)	99.9 (99.2)	95.8 (95.0)
Number of reflections: **Total** **Unique**	 249, 149 75, 693	 329, 325 69, 977	 164, 478 87, 243
Multiplicity	3.3 (3.3)	4.7 (3.9)	1.9 (1.8)
Wilson *B* factor (Å^2)	33.52	35.70	26.26
Average *B* factor (Å^2) **Protein** **Peptide** **Zinc** **Glycosylated sugars** **Solvent**	 33.44 39.79 30.07 61.25 37.35	 34.85 68.04 30.90 58.56 41.09	 25.42 35.26 22.59 46.83 35.14
$^{\S}R_{cryst}$/$^{\#\#}R_{free}$ (%)	18.17/21.10	16.47/19.25	17.27/20.14
R.M.S deviation from ideal values **Bond lengths (Å)** **Bond angles (°)**	 0.010 1.362	 0.008 1.184	 0.008 1.179
Ramachandran plot statistics (%) **Favoured** **Disallowed**	 98.66 0.00	 97.83 0.00	 98.66 0.00

$^{\#}$ $R_{merge} = \Sigma_{hkl}\Sigma_i|\boldsymbol{I_i}\,(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I_i}(hkl)$, where $\boldsymbol{I}_i$(hkl) is the intensity of *i*th measurement and <***I(hkl)***> is the average of symmetry-related observations of a unique reflection.

$^{+}R_{pim} = \Sigma_{hkl}\sqrt{\frac{1}{n-1}}\Sigma_{i=1}^{n}|\boldsymbol{I}_i(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I}_i(hkl)$

$^{\S}R_{cryst} = \Sigma_h|F_o - F_c|/\Sigma_h F_o$, where F_o and F_c are observed and calculated structure factor amplitudes of reflection *h*, respectively.

$^{\#\#}R_{free}$ is equal to R_{cryst} for a randomly selected 5 % subset of reflections.

Table 4.3: X-ray Data Collection, Processing and Refinement Statistics for AnCE with Spinorphin. Outer shell values are shown in parentheses.

4.3.4.1 AnCE_Tyn_1

4.3.4.1.1 Position 1

Data for AnCE_Tyn_1 were processed to 2.16 Å resolution. Although a total of 120 diffraction images were collected, only the first 100 of these were used in data processing. Including the final 20 images caused the R_{merge} to increase to an unacceptably high value and by excluding these images it was possible to reduce this whilst retaining the 2.16 Å resolution. Removing these images did not have an adverse effect on the completeness of the data, which remained at 99.7 % overall. A complete summary of the statistics from data processing is included in Table 4.2.

Initial analysis of the electron density at the active site, after placement of the zinc ion, indicated that the full tynorphin peptide might be bound. This was based on the observation that the density appeared to extend for the length of a pentapeptide. To this end an attempt was made to fit tynorphin into the active site electron density, with the valine, valine, tyrosine, proline and tryptophan residues occupying the S_3, S_2, S_1, S_1' and S_2' subsites respectively. This arrangement positioned the peptide bond between the tyrosine and proline residues directly above the catalytic zinc ion with the proline carbonyl oxygen partaking in zinc ion co-ordination.

Following a round of restrained refinement it was clear that there was insufficient electron density for the peptide bond between the tyrosine and proline residues. To try and overcome this, the occupancy of the peptide was reduced from 1.0 to 0.5 however this still did not solve the problem. The conclusion was therefore drawn that although it is possible that some intact peptide is bound, in the majority of molecules the enzyme has cleaved the tyrosine-proline bond. The pentapeptide was therefore fitted in the density as two fragments: VVY and PW (Figure 4.4). The placement of the dipeptide and tripeptide fragments outlined below is referred to as "position 1".

In position 1, the PW dipeptide sits with the proline and tryptophan residues occupying the S_1' and S_2' subsites respectively. The terminal oxygen of the tryptophan residue forms a water-mediated interaction with the side chain of Lys495. Further interactions between the second tryptophan oxygen and the side chains of Lys495, Gln265 and Tyr504 anchor the C-terminus of the dipeptide in this position in the active site. Additionally, the carbonyl oxygen of the proline forms hydrogen bond interactions with the side chains of His497 and His337. The interactions described here are routinely found in peptide binding to ACE homologues.

The S_1' and S_2' region of the active site is very hydrophobic. The side chains of Phe441 and Phe511 form a hydrophobic pocket that accommodates the hydrophobic PW dipeptide well. Furthermore, the proline side chain is able to form stacking interactions with His337 and His367, which will further contribute to dipeptide binding. The interactions between the PW dipeptide and the active site are summarised in Table 4.4 and Figure 4.3.

The VVY tripeptide sits directly along from the PW dipeptide, occupying the S_1, S_2 and S_3 subsites, respectively. The terminal oxygen of the tyrosine, introduced by cleavage of the Tyr-Pro peptide bond, appears to co-ordinate the catalytic zinc ion and also interacts with the side chains of His367 and His371. The second tyrosine oxygen atom forms water-mediated interactions with the side chains of Glu395 and Tyr507, as well as a direct interaction with the hydroxyl group of Tyr507. Interactions between the enzyme and the valine residues are mainly water mediated, however these residues do occupy a hydrophobic pocket formed by Ser339, Trp341, Phe375, Tyr496 and Val502. These interactions are summarised in Table 4.4 and Figure 4.3.

VVY atom	Interacting atom	Distance (Å)	PW atom	Interacting atom	Distance (Å)
P_1 Tyr OXT	Zn^{2+}	2.03	P_2' Trp NE1	Gln266 OE1	2.75
P_1 Tyr OXT	His367 NE2	3.28	P_2' Trp OXT	Water	2.96
P_1 Tyr OXT	His371 NE2	3.16	P_2' Trp O	Water	2.96
P_1 Tyr O	Water	3.20	P_2' Trp O	Lys495 NZ	2.64
P_1 Tyr O	Tyr507 OH	2.36	P_2' Trp O	Gln265 NE2	3.11
P_1 Tyr N	Water	2.84	P_2' Trp O	Tyr504 OH	3.00
P_2 Val O	Ala340 N	3.06	P_1' Pro O	His497 NE2	3.16
P_2 Val N	Ala340 O	3.22	P_1' Pro O	His337 NE2	2.90
P_3 Val O	Water	2.74			
P_3 Val O	Water	2.90			
P_3 Val N	Water	2.74			
P_3 Val N	Water	2.75			

Table 4.4: List of Interactions Formed by the VVY Tripeptide and PW Dipeptide at Position 1 with AnCE Active Site Residues (AnCE_Tyn_1 data).

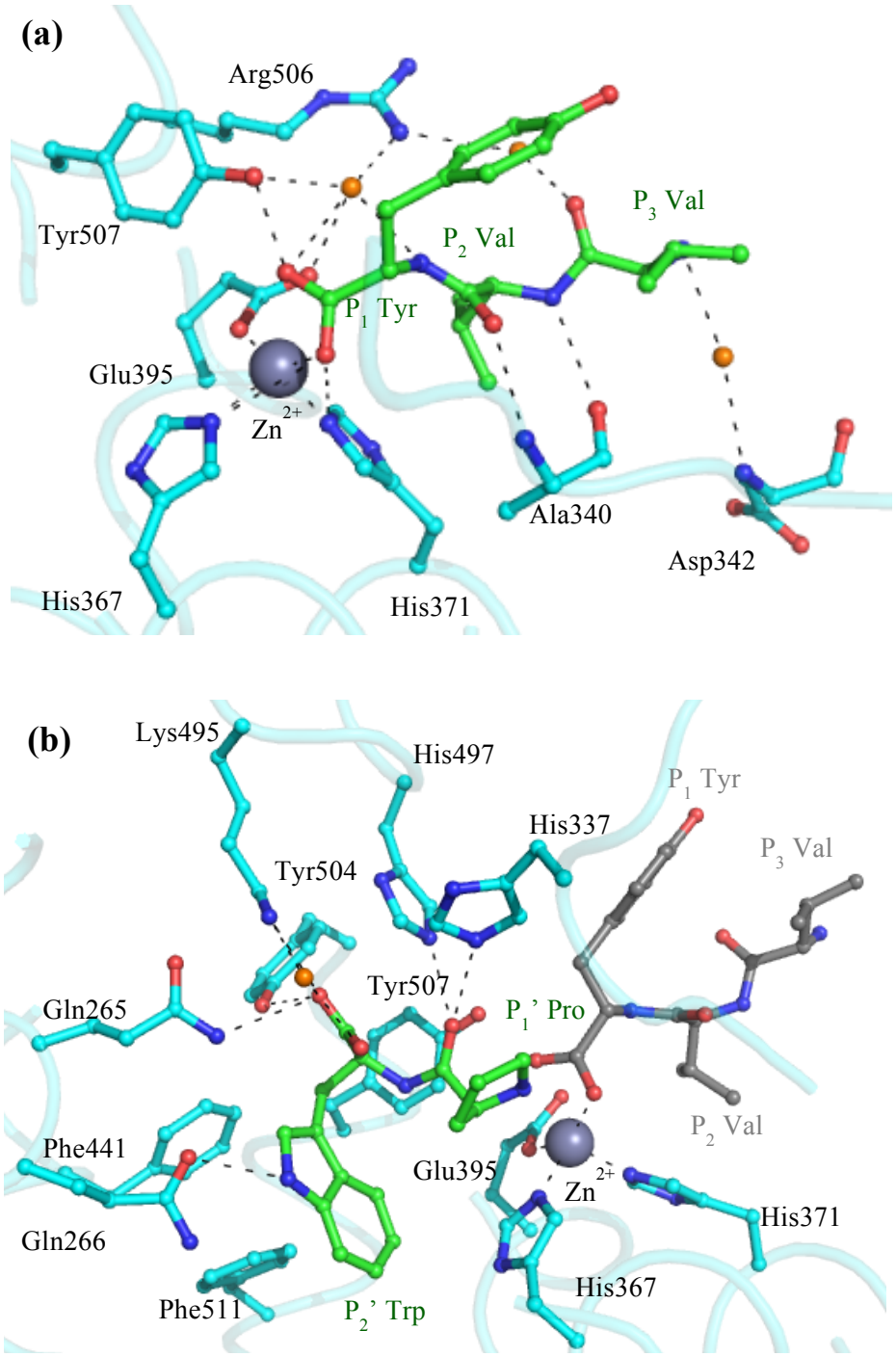


Figure 4.3: Interactions Formed by the VVY Tripeptide and PW Dipeptide Fragments of Tynorphin at "Position 1" in the AnCE Active Site (AnCE_Tyn_1 data). (a) Interactions formed by the VVY fragment of tynorphin at the S_1, S_2 and S_3 subsites. Key water molecules are shown in orange and the active site zinc ion is in grey. The tripeptide and protein residues are shown as green and cyan sticks, respectively. (b) Interactions between the PW fragment of tynorphin and the S_1' and S_2' subsites. The colour scheme is as described for (a). For reference the VVY tripeptide from (a) is shown in (b) as grey sticks.

4.3.4.1.2 Position 2

Fitting the two fragments in position 1 meant that the residues were well accommodated by the difference density however unoccupied electron density remained around some of the side chains, Figure 4.4. This was most noticeable at the P_2 and P_1' positions, currently occupied by valine and proline respectively, and indicated that fragments of the peptide were likely to be bound in more than one position.

The difference density map in Figure 4.4 shows where this extra side chain density was observed. It is clear from this that some form of aromatic side chain must also occupy the S_2 subsite and the terminal tynorphin tryptophan residue fitted well here. A number of different attempts were made to place the tynorphin peptide, and different fragments thereof, in this residual density. Ultimately, the peptide was broken down into the same two fragments as for position 1: PW and YVV. The PW dipeptide was placed at the P_3 and P_2 positions and VVY at the P_1, P_1' and P_2' positions, respectively. All four tynorphin fragments were assigned an occupancy of 0.5 and following refinement the model was observed to fit the electron density well, Figure 4.4. The peptide fragments bound in this way are referred to as being in "position 2".

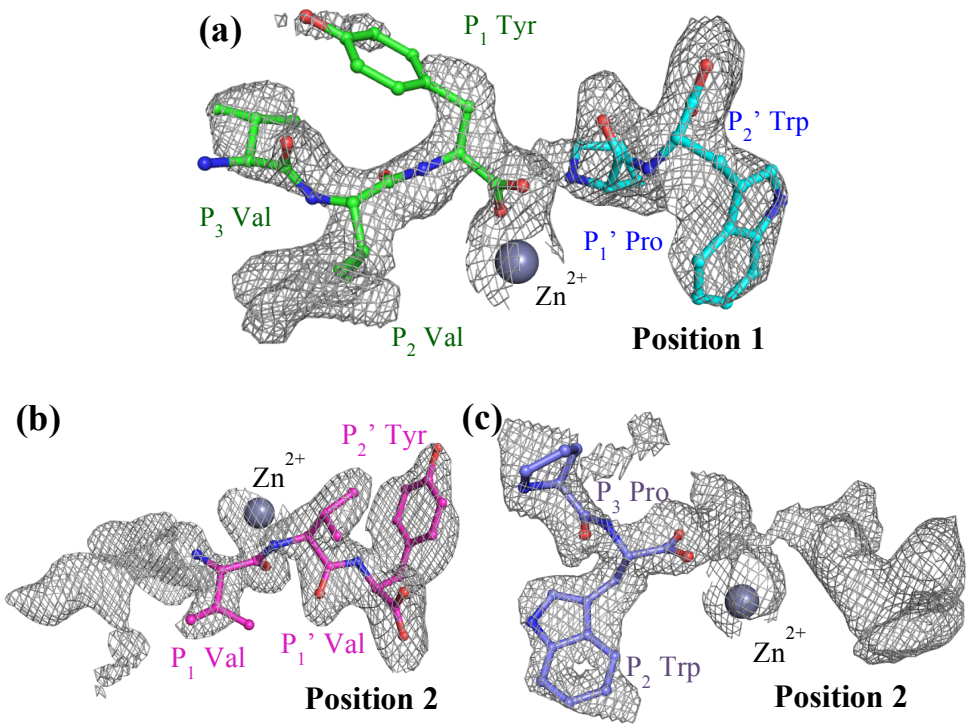


Figure 4.4: Placement of Tynorphin Residues in the Electron Density at the AnCE_Tyn_1 Active Site. (a) The original placement of two tynorphin fragments VVY (green) and PW (cyan) at "position 1" left residual difference density, most evident at the P_2 position. The two fragments were placed in a second orientation, "position 2", shown in (b) and (c) to rectify this. The Fo-Fc density maps are all contoured at 3σ.

In the position 2 arrangement the VVY tripeptide is located at the opposite end of the active site compared to in position 1, such that the residues occupy the S_1, S_1' and S_2' subsites respectively. The C-terminal tyrosine forms similar, although not as extensive, interactions with the active site residues Gln265, Lys495 and Tyr504 as the tryptophan from the PW dipeptide in position 1. Additionally the hydroxyl oxygen of the tyrosine side chain forms water mediated interactions with the side chains of Gln265 and Gln266. The hydrogen bonds formed by the carbonyl oxygen of proline in the dipeptide with the side chains of His367 and His371 are replicated by the P_1' valine here.

The additional valine residue at P_1, which completes the tripeptide, co-ordinates the catalytic zinc ion through its carbonyl oxygen, thus appearing to position the tripeptide for cleavage of the peptide bond between the two valine residues.

In the position 2 arrangement the PW dipeptide now occupies the S_3 and S_2 subsites where, like the two valine residues in the tripeptide, it predominantly interacts with the active site through hydrophobic interactions. In particular the tryptophan forms stacking interactions with the side chain of His394, as illustrated in Figure 4.5. All of these interactions are summarised in Table 4.5 and Figure 4.5.

VVY atom	Interacting atom	Distance (Å)	PW atom	Interacting atom	Distance (Å)
P_2' Tyr OH	Water	3.16	P_2 Trp OXT	Water	2.90
P_2' Tyr O	Gln265 NE2	3.06	P_2 Trp O	Ala340 N	3.16
P_2' Tyr O	Lys495 NZ	2.73	P_2 Trp NE1	Water	2.83
P_2' Tyr OXT	Water	2.41	P_2 Trp N	Ala340 O	2.66
P_1' Val O	His337 NE2	2.64	P_3 Pro O	Water	3.26
P_1' Val O	His497 NE2	2.88			
P_1' Val N	Ala338 O	2.93			
P_1 Val O	Zn^{2+}	2.77			
P_1 Val O	Tyr507 OH	2.40			

Table 4.5: List of Interactions Formed by the VVY Tripeptide and PW Dipeptide at Position 2 with AnCE Active Site Residues (AnCE_Tyn_1 Data).

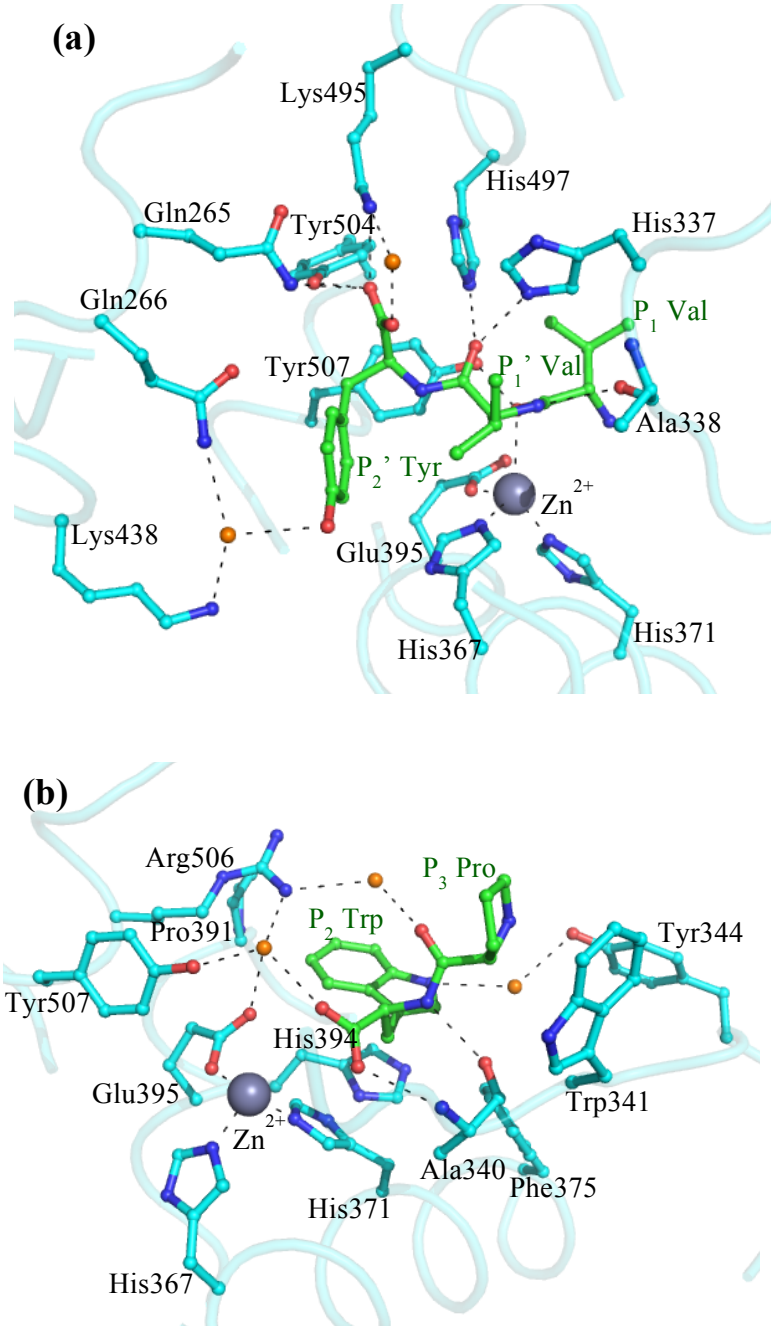


Figure 4.5: Interactions Formed by the VVY Tripeptide and PW Dipeptide Fragments of Tynorphin at "Position 2" in the AnCE Active Site (AnCE_Tyn_1 Data). (a) Representation of the hydrogen bond interactions formed by the VVY fragment of tynorphin at the S_1, S_1' and S_2' subsites. Key water molecules are shown in orange and the active site zinc ion is in grey. The tripeptide and protein residues are shown as green and cyan sticks, respectively. (b) Representation of the interactions between the PW fragment of tynorphin and the S_2 and S_3 subsites. The colour scheme is as described for (a). Key hydrophobic residues are shown as cyan sticks.

4.3.4.2 AnCE_Tyn_2

The second dataset collected for AnCE with tynorphin, AnCE_Tyn_2, consisted of 720 images. All 720 images were used to process the data to 2.01 Å resolution. A complete summary of data processing statistics is presented in Table 4.2.

As was the case for AnCE_Tyn_1, after placing the catalytic zinc ion in the active site some difference density remained. It was immediately apparent however that this density was far less extensive than that seen in AnCE_Tyn_1 and could not possibly accommodate all five tynorphin residues. Furthermore, the electron density was not continuous, with a clear break close to the catalytic zinc ion, Figure 4.6.

The combination of reasonably high-resolution data (2.01 Å) and the relatively limited number of potential residues meant that a VY dipeptide was quickly and confidently placed in one half of the density. As illustrated in Figure 4.6, the bulky tyrosine side chain fits the density here very well.

The position of the VY dipeptide is almost identical to that of the C-terminal valine and tyrosine residues of the VVY tripeptide bound in position 2 of the AnCE_Tyn_1 structure. Valine occupies the S_1' subsite and tyrosine the S_2' subsite. This is shown in Figure 4.7 and a full list of interactions with active site residues is detailed in Table 4.6, for the most part these are identical to those in AnCE_Tyn_1.

The main difference is that whereas in AnCE_Tyn_1 the catalytic zinc ion is co-ordinated directly by the carbonyl oxygen of the additional valine residue at the P_1 position, the absence of a third peptide residue here means that the zinc co-ordination is completed by a water molecule. It is through this water molecule that the amide nitrogen of the P_1' valine interacts with the catalytic zinc ion, Figure 4.7.

The second piece of active site electron density is quite clearly separated from the first; there was no possibility of placing a continuous peptide backbone here to include both regions. Instead a PW dipeptide was placed to occupy the S_3 and S_2 subsites in an almost identical manner to that seen in AnCE_Tyn_1 position 2 (Figures 4.6 and 4.7).

A full list of hydrogen bonding interactions for this dipeptide with active site residues is included in Table 4.6, however direct contacts are minimal and those that are present are conserved in the AnCE_Tyn_1 position 2 structure. There are a number of additional water mediated interactions not present in AnCE_Tyn_1, likely due to increased active

site water content in the absence of residues at a second position. As in AnCE_Tyn_1, hydrophobic interactions play a key role in holding this dipeptide in place.

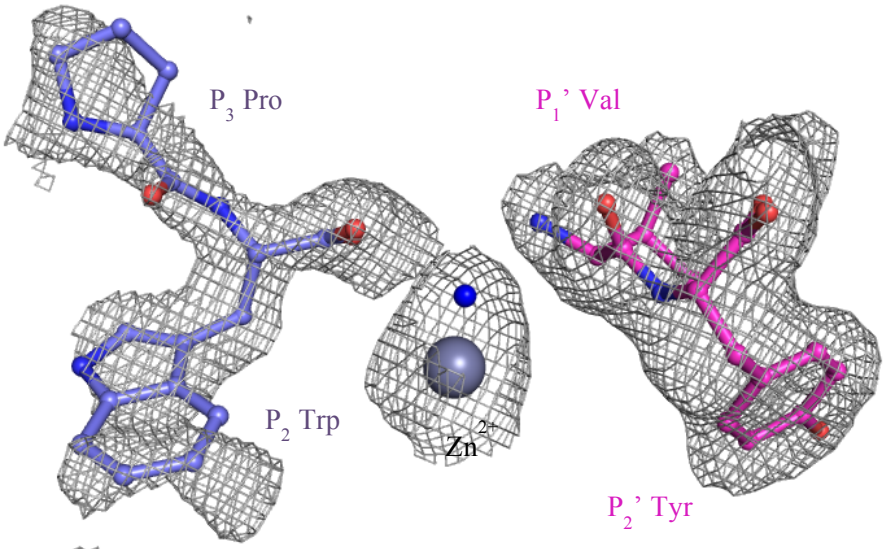


Figure 4.6: Placement of Tynorphin Fragments in the AnCE_Tyn_2 Active Site. The Fo-Fc map is shown as a grey mesh, contoured at 3σ. The difference density observed here is not as extensive as for AnCE_Tyn_1. Two dipeptides, VY (magenta) and PW (purple), were fitted in the density in very similar locations to AnCE_Tyn_1 position 2. A water molecule (blue sphere) is seen to co-ordinate the catalytic zinc ion (grey sphere).

VY atom	Interacting atom	Distance (Å)	PW atom	Interacting atom	Distance (Å)
P_2' Tyr OH	Water	3.01	P_2 Trp OXT	Water	2.99
P_2' Tyr O	Tyr504 OH	2.64	P_2 Trp OXT	Water	2.79
P_2' Tyr O	Gln265 NE2	3.19	P_2 Trp OXT	Water	2.40
P_2' Tyr O	Lys495 NZ	2.68	P_2 Trp O	Ala340 N	3.12
P_2' Tyr OXT	Water	2.66	P_2 Trp NE1	Water	2.96
P_1' Val O	His337 NE2	2.69	P_2 Trp N	Ala340 O	2.73
P_1' Val O	His497 NE2	3.13			
P_1' Val N	Glu368 OE2	2.69			
P_1' Val N	Ala338 O	3.12			
P_1' Val N	Water	2.95			
P_1' Val N	Water	2.60			

Table 4.6: List of Interactions Formed by the VY and PW Dipeptides with AnCE Active Site Residues (AnCE_Tyn_2 data).

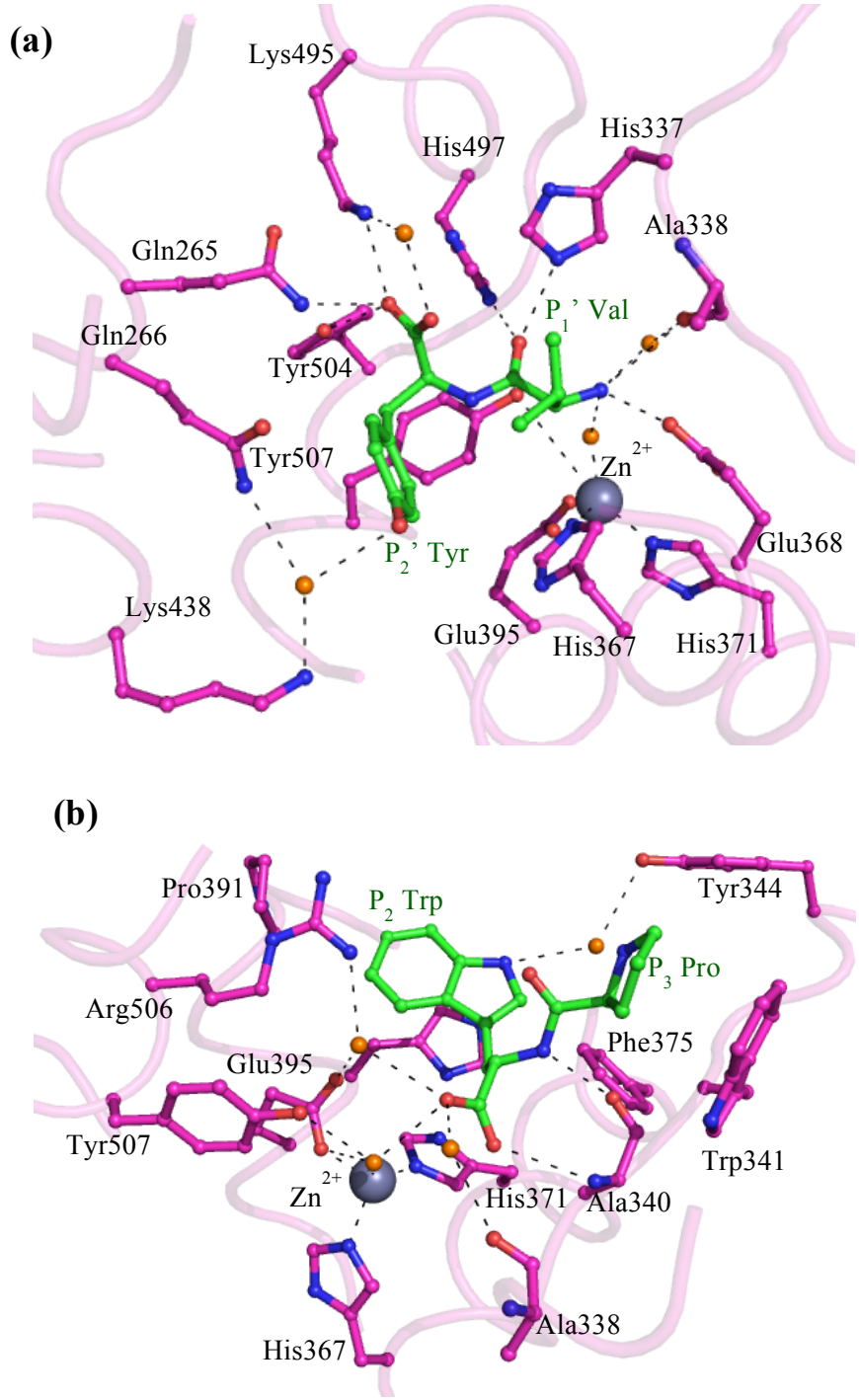


Figure 4.7: Interactions Formed by the VY and PW Dipeptide Fragments of Tynorphin in the AnCE Active Site, Based on Interpretation of the Electron Density Map for AnCE_Tyn_2. (a) The hydrogen bonding interactions formed by the VY fragment of tynorphin at the S_1' and S_2' subsites. Key water molecules are shown in orange and the active site zinc ion in grey. The dipeptide and protein residues are shown as green and magenta sticks, respectively. (b) The interactions between the PW fragment of tynorphin and the S_2 and S_3 subsites. The colour scheme is as described for (a).

4.3.4.3 AnCE_Tyn_3

A total of 774 diffraction images were collected for AnCE_Tyn_3 covering 116.1°. Data were collected to 1.71 Å resolution and all 774 images were used for data processing. Data processing statistics are summarised in Table 4.2.

Following molecular replacement using the native AnCE structure as a model there was once again additional electron density visible in the active site. However this density was far less extensive than that previously seen in AnCE_Tyn_1 and AnCE_Tyn_2, and it was only possible to fit a single dipeptide, occupying the S_2' and S_1' subsites. Based on interpretation of the visible density two valine residues were placed here, as illustrated in Figure 4.8. As was the case in the AnCE_Tyn_2 structure, a water molecule was placed in the difference density to co-ordinate the catalytic zinc ion.

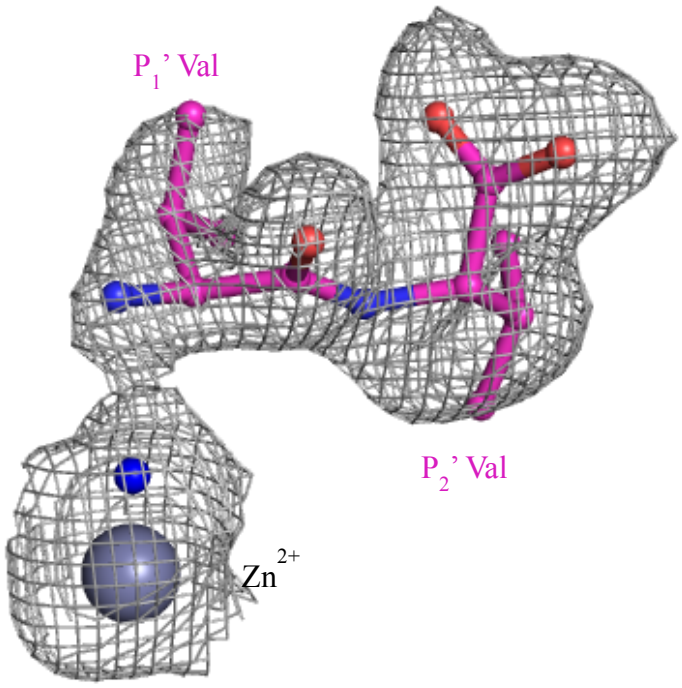


Figure 4.8: Placement of Tynorphin Fragments in the AnCE_Tyn_3 Active Site. Only a small amount of difference density is observed in the AnCE_Tyn_3 active site, illustrated by the Fo-Fc map contoured at 3 σ, shown as a grey mesh. A single VV dipeptide was placed in the same location as the VY dipeptide in AnCE_Tyn_2. As in AnCE_Tyn_2 a water molecule (blue sphere) is seen to co-ordinate the active site zinc ion (grey sphere).

A complete list of interactions between the VV dipeptide and active site residues is presented in Table 4.7 and summarised in Figure 4.9. The backbone of this dipeptide is positioned in an almost identical manner to the backbone of the VY dipeptide in the AnCE_Tyn_2 structure. Consequently, many of the interactions with active site residues are conserved. This includes the key interactions of the C-terminal carboxylate group with the side chains of Gln265 and Lys495, which anchor the peptide in the active site, and the water mediated interaction with the catalytic zinc ion formed by the amide nitrogen of the P_1' valine.

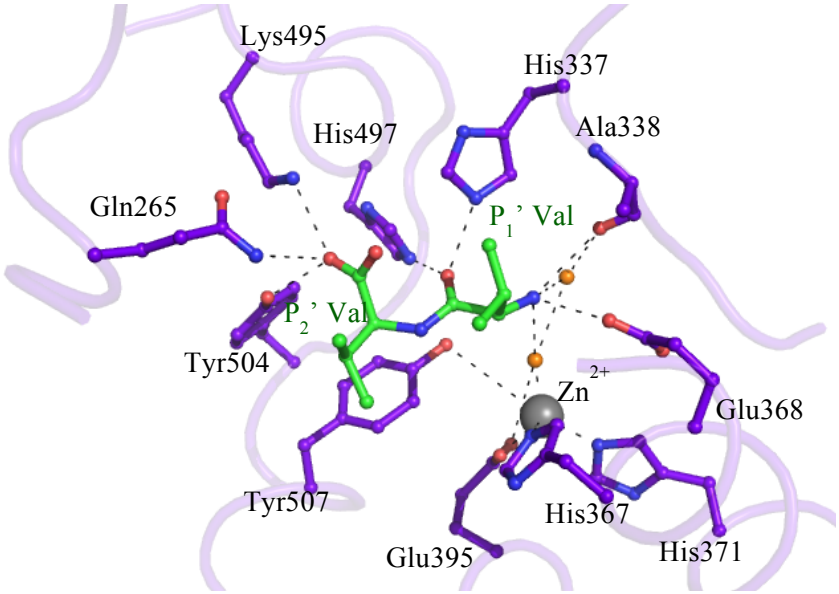


Figure 4.9: Interactions Formed by the VV Dipeptide Fragment of Tynorphin in the AnCE Active Site, Based on Interpretation of the Electron Density Map for AnCE_Tyn_3. Interactions formed by the VV fragment of tynorphin at the S_1' and S_2' subsites. Key water molecules are shown in orange and the active site zinc ion in grey. Residues of the dipeptide and AnCE are shown as green and purple sticks, respectively.

VV atom	Interacting atom	Distance (Å)
P_2' Val O	Tyr504 OH	2.62
P_2' Val O	Gln265 NE2	3.25
P_2' Val O	Lys495 NZ	2.77
P_1' Val O	His337 NE2	2.59
P_1' Val O	His497 NE2	3.01
P_1' Val N	Ala338 O	2.98
P_1' Val N	Glu368 OE2	2.92
P_1' Val N	Water	2.91
P_1' Val N	Water	2.50

Table 4.7: List of Interactions Formed by the VV Dipeptide Fragment of Tynorphin with AnCE Active Site Residues (AnCE_Tyn_3 data).

4.3.4.4 AnCE_Spin_4

A total of 1800 diffraction images at 2.00 Å resolution were collected from the AnCE_Spin_4 crystal. Following initial processing using MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) and Aimless (Evans, 2006; Evans, 2007) with all 1800 images, the R_{merge} value was unacceptably high. Rather than imposing a limit on the resolution, the number of images used in the processing was reduced such that only images 1-1250 were included. This greatly improved the R_{merge} value (9.6 % overall and 68.4 % outer shell) without adversely affecting the completeness of the data, which remained at 99.5 % overall. Complete data processing statistics are shown in Table 4.3.

Following molecular replacement using the native AnCE structure (Akif *et al.,* 2010*)* as a search model, a significant amount of extra electron density was visible in the active site. Efforts were made to place the spinorphin peptide into this density, however it was immediately apparent that the density did not extend far enough along the catalytic channel to be able to accommodate the full-length peptide.

Ultimately, the LVV tripeptide was modelled into the density. This equates to the first three residues of the spinorphin peptide. The residues were placed such that the S_1, S_1' and S_2' subsites of the enzyme were occupied by leucine, valine and valine respectively. After placing these residues further rounds of refinement were performed using REFMAC (Murshudov *et al.,* 2011). This confirmed that the residues fitted the density well, and also that there was definitely no further density in which it may have been possible to model other peptide residues. Figure 4.10 illustrates the fit of the tripeptide in the active site difference density.

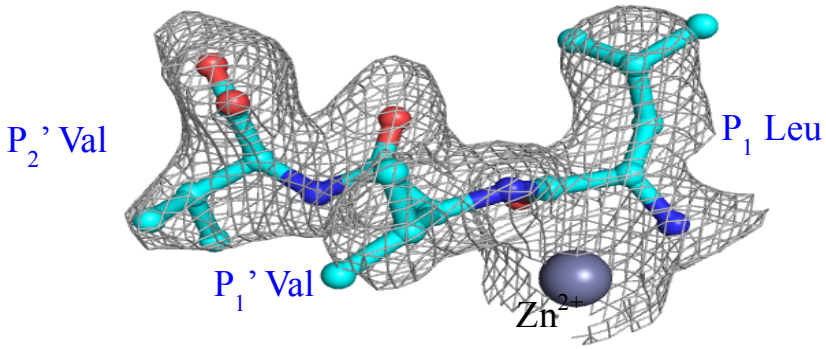


Figure 4.10: Placement of Spinorphin Fragments in Electron Density at the AnCE_Spin_4 Active Site. A tripeptide consisting of leucine, valine, and valine occupying the S_1, S_1' and S_2' subsites, respectively is shown as cyan sticks in the AnCE_Spin_4 Fo-Fc map, contoured at 3σ. The catalytic zinc ion is shown as a grey sphere.

Most of the interactions between the tripeptide and the ACE active site involve the peptide backbone. This is a common feature of peptide binding to ACE as it allows the enzyme to act on a wide range of peptide substrates.

As is normally the case in structures of ACE homologues bound to peptides, a key point of contact with the enzyme is via the C-terminal carboxyl group of the tripeptide. Here the P_2' valine interacts through its carboxyl group directly with the side chains of Lys495 and Tyr504. There is also a water-mediated interaction with the side chain of Gln265.

Further hydrogen bond interactions exist between the carbonyl oxygen of the P_1' valine and the side chains of His337 and His497. The amide nitrogen of this valine also interacts with the side chain of Glu369 and contacts the protein backbone at Ala338.

The N-terminal residue of the tripeptide, the P_1' leucine, contacts the side chains of Tyr504 and Glu368. Interestingly, it also appears to co-ordinate the catalytic zinc ion via both its carbonyl oxygen and amide nitrogen, as illustrated in Figure 4.11. A complete list of interactions between the peptide and active site residues in presented in Table 4.8 and summarised in Figure 4.11.

LVV Atom	Interacting atom	Distance (Å)
P_2' Val OXT	Water	2.75
P_2' Val O	Lys495 NZ	2.64
P_2' Val O	Tyr504 OH	2.62
P_1' Val O	His337 NE2	2.70
P_1' Val O	His497 NE2	3.11
P_1' Val N	Ala338 O	3.34
P_1' Val N	Glu369 OE2	2.92
P_1 Leu O	Zinc	2.28
P_1 Leu O	Tyr504 OH	2.64
P_1 Leu N	Zinc	2.55
P_1 Leu N	Glu368 OE1	2.88
P_1 Leu N	Water	2.75

Table 4.8: List of Interactions Formed by the LVV Tripeptide Fragment of Spinorphin with AnCE Active Site Residues in the AnCE_Spin_4 Structure.

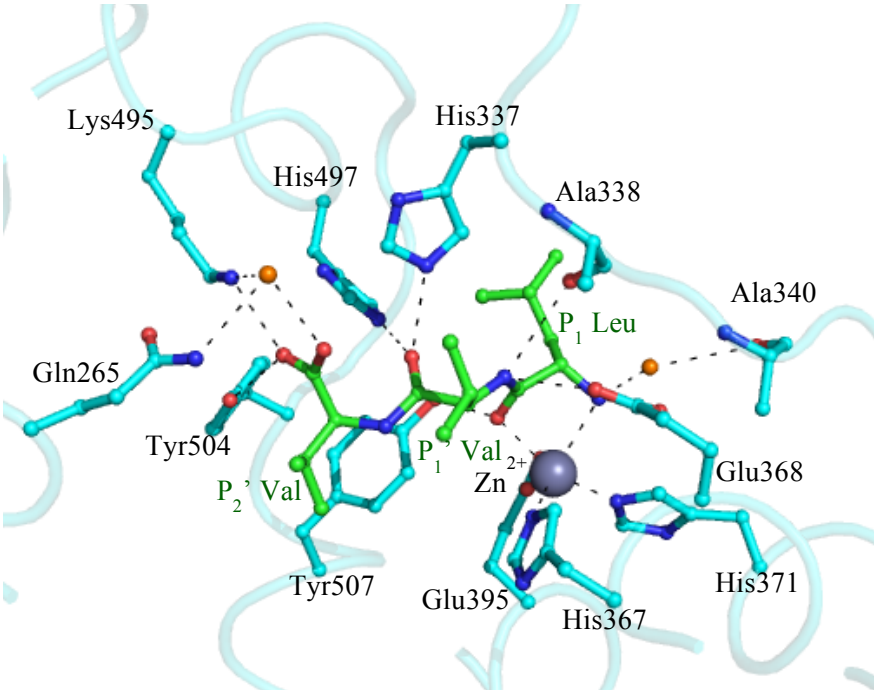


Figure 4.11: Interactions Formed by the LVV Tripeptide Fragment of Spinorphin in the AnCE Active Site Based on Interpretation of the Electron Density Map for AnCE_Spin_4. Interactions formed by the LVV fragment of spinorphin at the S_1, S_1' and S_2' subsites. Key water molecules are shown in orange and the active site zinc ion in grey. The tripeptide and protein residues are shown as green and magenta sticks, respectively.

4.3.4.5 AnCE_Spin_5

The AnCE_Spin_5 dataset comprised 900 diffraction images collected to 2.07 Å resolution. All 900 images were used for data processing, with statistics summarised in Table 4.3.

Following molecular replacement and initial structure refinement it was clear that there was a significant portion of difference density in the active site of the structure, much more than was seen in AnCE_Spin_4. Four alanine residues were initially fitted into the density to build a tetrapeptide backbone. Following a round of restrained refinement it was then possible to build the appropriate side chains into the remaining density. This resulted in the peptide YPWT with the tyrosine, proline, tryptophan and threonine side chains adopting the P_2, P_1, P_1' and P_2' positions, respectively. As illustrated in Figure 4.12, the tetrapeptide filled the difference density well and it was quite clear that no substantial further density was present, thus the peptide was not extended beyond these four residues.

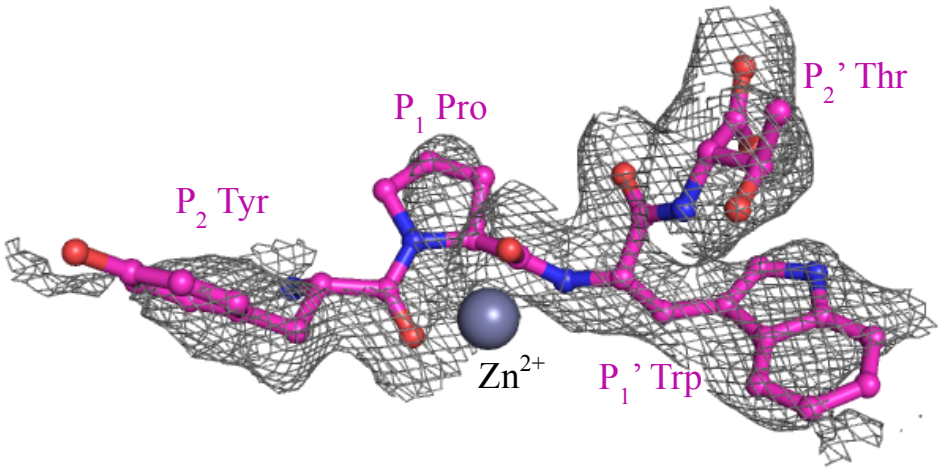


Figure 4.12: Placement of Spinorphin Fragments in Electron Density at the AnCE Active Site in the AnCE_Spin_5 Structure. A peptide consisting of four spinorphin residues: tyrosine, proline, tryptophan and threonine, is shown as magenta sticks in the AnCE_Spin_5 Fo-Fc density map, contoured at 3σ, occupying the S_2, S_1, S_1' and S_2' subsites. The peptide co-ordinates the catalytic zinc ion, grey sphere, via the proline and tyrosine carbonyl oxygen atoms.

The position of the YPWT fragment of spinorphin in this way places the peptide bond between the tryptophan and proline above the catalytic zinc ion with the carbonyl oxygen atoms from both residues appearing to contact the zinc ion, Figure 4.13.

The C-terminus of the YPWT peptide is held in place through interactions between the P_2' threonine carboxyl group and active site residues Gln265, Lys495, Tyr504. The tryptophan at the P_1' position interacts with the side chains of His337 and His497 through it carbonyl oxygen atom, and with Glu368 and Ala338 via its amide nitrogen. There are fewer interactions between the N-terminus of the peptide and the active site; the amide nitrogen of the P_2 tyrosine contacts the protein backbone at Ala340, whilst the hydroxyl group of its side chain interacts with the side chain of Thr387.

As has previously been described for other spinorphin and tynorphin fragments, the hydrophobicity of this region of the AnCE active site also appears to play some role in YPWT binding. Figure 4.13 and Table 4.9 illustrate the key interactions between YPWT and the AnCE active site.

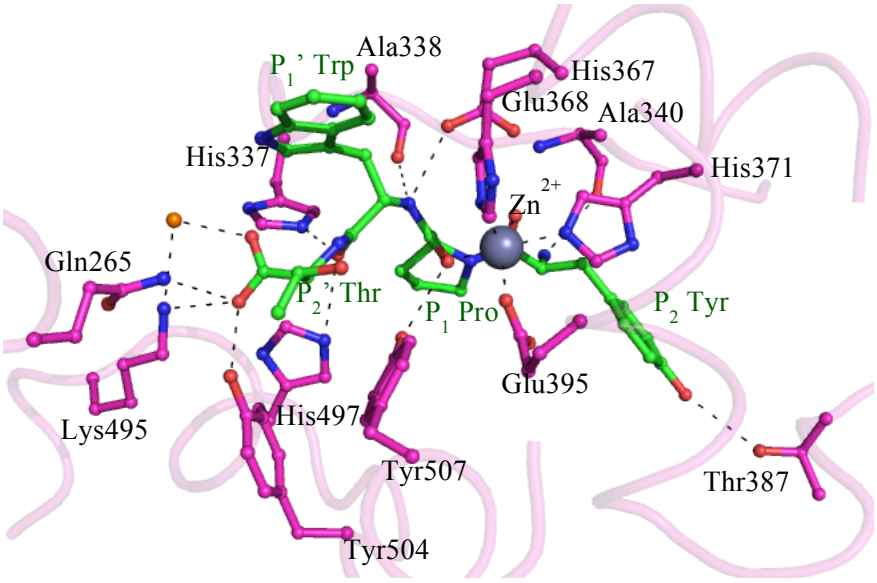


Figure 4.13: Interactions Formed by the YPWT Tetrapeptide Fragment of Spinorphin in the AnCE Active Site, Based on Interpretation of the Electron Density Map for AnCE_Spin_5. Interactions formed by the YPWT fragment of spinorphin at the S_2, S_1, S_1' and S_2' subsites. Key water molecules are shown in orange and the active site zinc ion in grey. The peptide and protein residues are shown as green and magenta sticks, respectively.

YPWT atom	Interacting atom	Distance (Å)
P_2' Thr OXT	Water	2.65
P_2' Thr O	Gln265 NE2	3.15
P_2' Thr O	Lys495 NZ	2.81
P_2' Thr O	Tyr504 OH	2.59
P_1' Trp O	His497 NE2	3.12
P_1' Trp O	His337 NE2	2.64
P_1' Trp N	Ala338 O	3.17
P_1' Trp N	Glu368 OE2	2.81
P_1 Pro O	Tyr507 OH	2.86
P_1 Pro O	Zinc	2.08
P_2 Tyr O	Zinc	3.29
P_2 Tyr N	Ala340 O	2.78
P_2 Tyr OH	Thr387 OG1	3.09

Table 4.9: List of Interactions Formed by the YPWT Tetrapeptide Fragment of Spinorphin with AnCE Active Site Residues in the AnCE_Spin_5 Structure.

4.3.4.6 AnCE_Spin_6

The AnCE_Spin_6 dataset consisted of 920 images collected to 1.6 Å resolution. Attempts to process the data at such high resolution using all of the images resulted in unacceptably high values for R_{merge} (>100 %). In an initial attempt to overcome this problem only the first 700 images were used. Although this reduced the R_{merge} value significantly it was still too high. Whilst removing further images may have further improved the R_{merge} this would have had an adverse effect on the completeness of the data.

The data were therefore processed again, still using images 1-700, but imposing a high-resolution limit of 1.90 Å. This resulted in an overall R_{merge} of 11.8 % (66.3 % for outer shell data) and overall completeness of 95.8 % (95.0 % for the outer shell). Full data processing statistics are included in Table 4.3.

Following molecular replacement and initial rounds of refinement, a relatively small amount of difference density was visible in the active site. The extent of this density was considerably less than seen for either AnCE_Spin_4 or AnCE_Spin_5 and was only able to accommodate a single dipeptide, Figure 4.14.

Although the density was limited in terms of length, due to the high resolution of the data it was very clear which residues formed the dipeptide and proline and tyrosine were place at the P_2' and P_1' positions respectively, Figure 4.14.

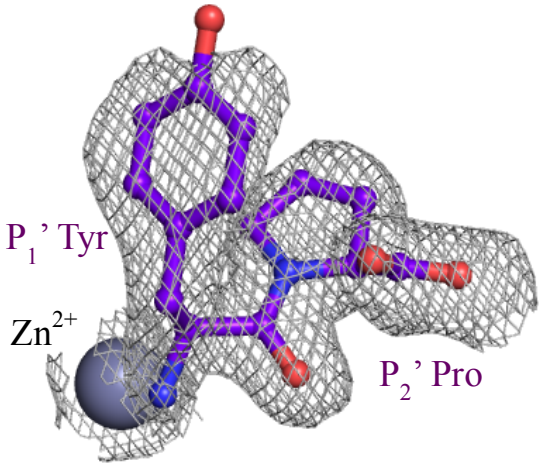


Figure 4.14: Placement of Spinorphin Fragments in Electron Density at the AnCE Active Site in the AnCE_Spin_6 Structure. A dipeptide fragment of spinorphin consisting of tyrosine and proline is shown as purple sticks in the Fo-Fc difference density map, contoured at 3σ, derived from the AnCE_Spin_6 data. The tyrosine and proline residues occupy the S_1' and S_2' subsites of the AnCE active site, respectively. The catalytic zinc ion does not interact directly with the dipeptide but is shown as a grey sphere for reference.

The majority of interactions between the YP dipeptide and the AnCE active site involve the carboxyl group of the P_2' proline, which interacts with the side chains of Gln265, Lys495 and Tyr504. Additionally, the carbonyl oxygen of the tyrosine residue forms hydrogen bond interactions with His337 and His497.

The dipeptide does not appear to interact directly with the active site zinc ion. Instead, the dipeptide interacts with the zinc via a water molecule and the tyrosine carbonyl oxygen. This oxygen also interacts directly with active site residues Glu368 and Ala338. These interactions are illustrated in Figure 4.15 and detailed in Table 4.10.

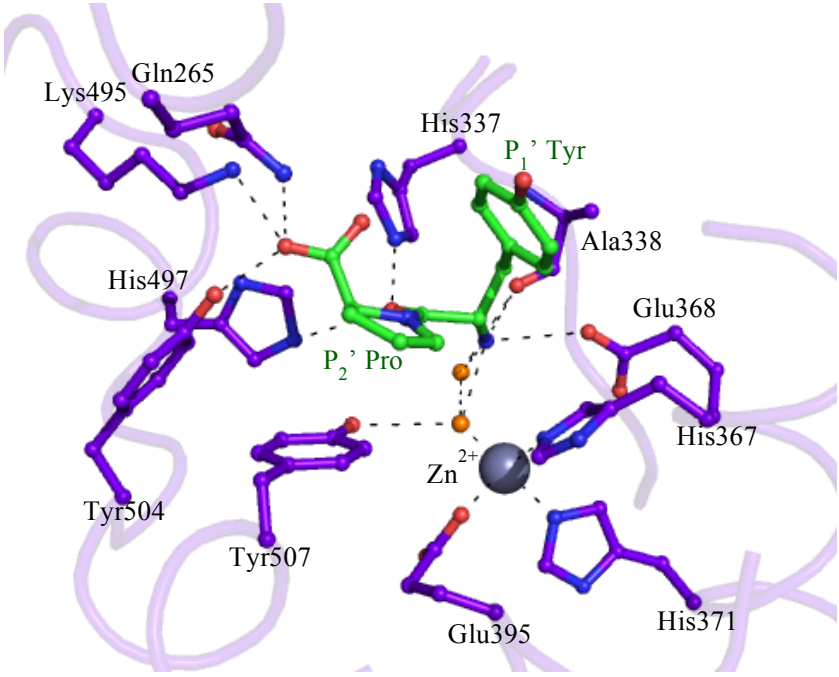


Figure 4.15: Interactions Formed by the YP Dipeptide Fragment of Spinorphin in the AnCE Active Site, Based on Interpretation of the Electron Density Map for AnCE_Spin_6. Interactions formed by the YP fragment of spinorphin at the S_2' and S_1' subsites. Key water molecules are shown in turquoise and the active site zinc ion in grey. The peptide and protein residues are shown as green and purple sticks, respectively.

PY atom	Interacting atom	Distance (Å)
P_2' Pro O	Gln265 NE2	3.18
P_2' Pro O	Lys495 NZ	2.66
P_2' Pro O	Tyr504 OH	2.53
P_1' Tyr O	His337 NE2	2.73
P_1' Tyr O	His497 NE2	3.03
P_1' Tyr N	Water	2.38
P_1' Tyr O	Water	3.04
P_1' Tyr O	Ala338 O	3.27
P_1' Tyr O	Glu368 OE2	2.83

Table 4.10: List of Interactions Formed by the YP Dipeptide Fragment of Spinorphin with AnCE Active Site Residues in the AnCE_Spin_6 Structure.

4.4 Discussion

Following incubation with spinorphin and tynorphin, AnCE crystals were readily obtained from the "standard" crystallisation condition (1.2-1.3 M Na citrate, 0.1 M HEPES, pH 7.5). When taken to DLS for data collection, these crystals diffracted well and a number of complete datasets were collected to good, ~2 Å, resolution (Table 4.1). Although no cryo-protecting solutions were used there were no problems with ice, illustrated by the absence of ice rings on the diffraction image in Figure 4.2. The crystals appeared to be relatively stable during data collection.

All data were processed in the space group *h3*. This is typical of AnCE crystals grown in this condition and indicates that there is unlikely to have been any significant conformational change occurring on spinorphin and tynorphin binding. This was confirmed both by visual inspection of the final structures and by calculating the RMSD compared to the native AnCE structure, PDB code 2X8Y, (Akif *et al.,* 2010). RMSD values for all structures were in the range 0.115-0.228 Å.

4.4.1 Structures of AnCE with Tynorphin

All of the data collected for AnCE with tynorphin contained some active site difference density into which different fragments of the tynorphin peptide were modelled. However the extent of this density varied greatly across the three datasets. Furthermore, it was not possible to model the full-length tynorphin peptide into the active site electron density of any of the structures.

4.4.1.1 AnCE_Tyn_1

The greatest amount of difference density was visible in the AnCE_Tyn_1 structure and initially it seemed that the full-length tynorphin pentapeptide was bound here. However, following refinement, it was clear that there was insufficient difference density in the region around the peptide bond between the tyrosine and proline residues at the S_1 and S_1' positions for this to be the case. Even after reducing the occupancy of the peptide quite significantly this issue remained; therefore tynorphin was placed in the active site as two fragments; VVY and PW.

Further analysis of the density after placement of the VVY and PW peptides in what is referred to as "position 1" revealed residual difference density at a number of positions. This was indicative of tynorphin residues binding to the active site at multiple locations. Following careful analysis of the density and attempts at positioning residues in many different combinations, ultimately the same two fragments were placed in the active site

in a second orientation: position 2. Here the VVY tripeptide residues are now at the P_1, P_1' and P_2' positions and the PW dipeptide residues at P_3 and P_2, respectively.

These results suggest that the enzyme is cleaving the peptide bond between the tyrosine and proline residues. This at first does not seem surprising, given that angiotensin-converting enzymes typically cleave the C-terminal dipeptide from their substrates. In this case however the penultimate peptide residue is a proline.

As discussed in section 3.1.2, proline residues at the P_1' position have been identified as key features of the BPPs, which are natural ACE inhibitory peptides. The unique nature of the proline side chain is thought to protect the scissile bond from cleavage by restricting access to the amide nitrogen. Thus, it is very surprising to see the tynorphin peptide cleaved into these fragments. However in all three structures of AnCE with tynorphin the sequences of the fragments of peptide present in the active site indicate that this has occurred. This strongly suggests that this is a valid observation, rather than based on incorrect interpretation of the electron density maps.

One possible explanation for this is that AnCE is acting as an aminopeptidase. ACE has previously been shown to remove residues from the N-terminus of peptides if the C-terminus is "protected", a notable example being towards GnRH, which also has a proline residue at the penultimate, P_1' position (Jaspard *et al.,* 1993; Skidgel *et al.,* 1985). The basis of aminopeptidase activity by ACE is not yet fully understood, however such activity could potentially explain the results seem here.

An alternative, and perhaps more likely, explanation might be that AnCE is able to remove the C-terminal dipeptide from substrates with a P_1' proline, but that this occurs much more slowly than "normal" hydrolysis. On a biological scale, this activity may therefore not be relevant. In the crystal environment however, the peptides are exposed to AnCE at very high concentrations for extended periods of time and are quite possibly constrained to the active site for days or weeks.

Although it is not possible to draw a definitive conclusion based on a single diffraction experiment using a single crystal, the positions and orientations of the different tynorphin fragments do indicate a probable sequence of events occurring in the AnCE active site.

It is likely that full-length tynorphin initially binds in the orientation that the peptide fragments adopt in position 1. That is to say that the C-terminal tryptophan residue occupies the S_2' subsite. This orientation is likely to be favourable, because in this position the C-terminal carboxyl group can form a number of key interactions with active site residues (namely Lys495, Tyr504 and Gln265) that anchor the peptide in place. This is a common feature of peptide interactions with angiotensin converting enzymes. Furthermore, this orientation places the aromatic side chains of tryptophan and proline in a hydrophobic pocket, with the proline residue effectively "sandwiched" between two histidine residues.

Tynorphin binding in this way would position the peptide bond between the tyrosine and proline residues for cleavage, resulting in the two fragments: VVY and PW. If the PW dipeptide were then released, the VVY dipeptide could "slide" along the active site and now occupy the S_1, S_1' and S_2' subsites, as seen in position 2. The newly generated carboxyl group would then form the same interactions with the S_2' subsite as the tryptophan residue had previously. This concept of "sliding" in the ACE active site has previously been observed with angiotensin (Masuyer *et al.,* 2012).

Following this movement the released PW dipeptide would then be free to occupy the S_3 and S_2 subsites, as seen in position 2. It is likely that this is a weaker interaction, restricted mainly to hydrophobic and water mediated interactions. It should also be noted that the VVY tripeptide is now positioned ideally for cleavage of the peptide bond between the two valine residues. This sequence of events is illustrated in Figure 4.16 (a) and (b).

4.4.1.2 AnCE_Tyn_2

The difference density in the AnCE_Tyn_2 active site was considerably less ambiguous than that seen in AnCE_Tyn_1. As such, two dipeptides, PW and VY were quickly placed in the active site. The PW dipeptide adopts an almost identical position in the S_3 and S_2 subsites as in position 2 of the AnCE_Tyn_1 structure. Similarly, the valine and tyrosine residues of the dipeptide are located at the P_1' and P_2' positions, respectively, as they were in the VVY tripeptide in position 2 of the AnCE_Tyn_1 structure.

The most likely explanation for the arrangement seen here is that further cleavage of the tripeptide has occurred. The VVY tripeptide in AnCE_Tyn_1 position 2 was positioned ideally for cleavage of the peptide bond between the two valine residues. The alternative explanation is that the P_1 valine residue is present but not visible in the structure due to

being mobile. The electron density around the valine residue of the dipeptide was examined carefully, however there was no evidence at all of it continuing past the P_1' position (Figure 4.6), hence it is highly likely that the VVY tripeptide has undergone further cleavage. An illustration of this process is shown in Figure 4.16 (b) and (c).

The AnCE_Tyn_1 and AnCE_Tyn_2 datasets were collected three years apart from different crystals grown using different batches of AnCE. The consistency of the data is therefore an indication that this is an accurate representation of what is happening to tynorphin at the AnCE active site. If time had permitted, it would have been interesting to collect data from the crystals used for the AnCE_Tyn_1 data collection a few days or weeks after the initial collection. If the VVY tripeptide at position 2 were still intact, it would have been an indication that the additional cleavage seen in AnCE_Tyn_2 may have occurred prior to crystal growth. If however the tripeptide were observed to have undergone further cleavage it would illustrate that this may be occurring within the crystal itself.

4.4.1.3 AnCE_Tyn_3

The AnCE_Tyn_3 structure contained far less difference density in the active site than either of the previous two structures. Whereas in AnCE_Tyn_1 and AnCE_Tyn_2 there were two regions of density, in AnCE_Tyn_3 only a single dipeptide, VV, could be placed in the active site, occupying the S_2' and S_1' subsites.

It is quite apparent from studying the electron density in Figure 4.8 that no further residues could have been added to the dipeptide, and that the two valine residues fit the density very well. However there is a small possibility that the VV dipeptide could actually be VY, as seen in AnCE_Tyn_2, but that the tyrosine side chain density is not entirely visible.

Although the presence of a tyrosine residue cannot be ruled out entirely, the most probable explanation for the results seen in the AnCE_Tyn_3 structure is that AnCE has cleaved tynorphin into multiple fragments and individual residues. It is possible that the PW dipeptide present in AnCE_Tyn_1 and AnCE_Tyn_2 has been reduced to individual residues, which have either been released from the enzyme or are not bound in any defined position in the active site and hence are not visible in the structure.

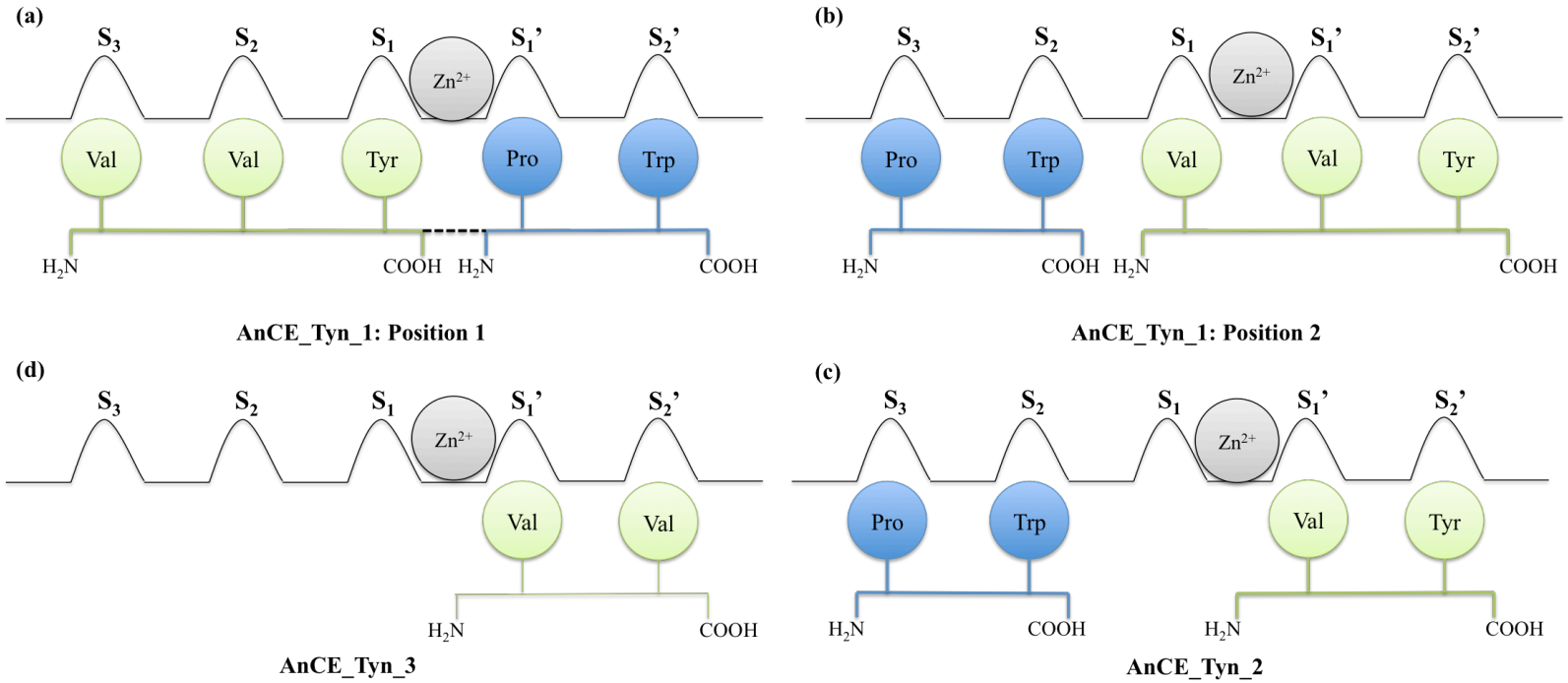


Figure 4.16: Schematic Representation of Tynorphin Binding at the AnCE Active Site. In all figures the locations of the subsites of the enzymes are labelled as S_3, S_2, S_1, S_1' and S_2'. The location of the catalytic zinc ion between the S_1 and S_1' subsites is also indicated. The tynorphin residues occupying each subsite are labelled. (a) In position 1 of the AnCE_Tyn_1 structure two peptide fragments, VVY (green) and PW (blue), are bound. It is likely that this is due to the enzyme cleaving the peptide bond between the tyrosine and proline residues of the full peptide, indicated by a black dotted line. (b) In position 2 of the AnCE_Tyn_1 structure the PW dipeptide has been released and re-bound at the S_3 and S_2 positions. The VVY fragment has moved along the active site and is now positioned for cleavage of the peptide bond between the two valine residues. (c) In the AnCE_Tyn_2 structure this cleavage has occurred with the release of a valine residue. (d) in the AnCE_Tyn_3 structure only a VV dipeptide is bound at the S_1' and S_2' positions, likely due to cleavage of the peptide bond between the valine and tyrosine residues in VVY or full length tynorphin.

The AnCE_Tyn_3 data is interesting because it was obtained from crystals grown at the same time as those used to collect the AnCE_Tyn_2 data. These two data sets were collected three weeks apart, meaning that the tynorphin in the AnCE_Tyn_3 data was incubated with the AnCE crystal for considerably longer. The results would appear to suggest that during this time further cleavage of the peptide has occurred.

4.4.2 Stability of Tynorphin in the Presence of Angiotensin Converting Enzymes

What is clear from the data presented here is that AnCE is cleaving the tynorphin peptide, apparently into multiple fragments. Tynorphin is a synthetic peptide, which was designed based on the haemorphin spinorphin with the aim of creating a better inhibitor of the enzyme DPPIII (Yamamoto *et al.,* 2000). Like ACE, DPPIII is a zinc metalloprotease, but whereas ACE removes residues from the C-terminus of peptide substrates, DPPIII targets residues at the N-terminus (Bezerra *et al.,* 2012).

Whilst tynorphin is a more potent inhibitor of DPPIII than spinorphin, it is actually a weaker inhibitor of ACE. The percentage ACE inhibition by tynorphin is approximately 50 % that seen at the same concentration of spinorphin (Yamamaoto *et al*., 2000).

When the initial properties of tynorphin as an inhibitor were first reported by Yamamoto *et al*, the stability of this peptide in human serum was also measured. Tynorphin was incubated at 37 ^{o}C with human serum and after 2 and 4 hours the survival of the peptide was analysed using HPLC. After 2 hours very little full-length peptide remained, appearing to have been cleaved into VYP, YPW and VYPW. After a further 2 hours almost all of the peptide had been cleaved to a background level and only small amounts of VYP and individual amino acids were detected.

Studying the sequences of the fragments detected after incubation with human serum suggests that a component in the serum is removing residues sequentially from either end of the peptide. This is somewhat different to the results of the AnCE crystallisation experiments, where it is the C-terminal dipeptide that appears to be removed.

When leuhistin was added to the human serum and tynorphin mixture, tynorphin remained stable for at least 24 hours. Leuhistin is an inhibitor of aminopeptidases (Yoshida *et al.,* 1991), indicating that it is aminopeptidases present in the human serum that are responsible for the observed rapid degradation of tynorphin.

There is no indication as to what the levels of ACE were in the serum, however it is not likely to have been present in very large quantities because ACE is predominantly located on endothelial cell membranes (Soubrier *et al.,* 1988). There is therefore no strong evidence to contradict the observations made from the crystal structures of AnCE with tynorphin; that AnCE, and hence ACE, cleave tynorphin at multiple positions, and that this appears to begin with cleavage of the tyrosine-proline peptide bond.

What is not clear is the time-scale on which these events occur. The proline residue at the P_1' position should protect the tyrosine-proline peptide bond from hydrolysis by AnCE in its role as a dipeptidyl carboxypeptidase, but the data presented here indicated that this may not be the case in the crystal environment. Further research is required to fully understand the stability of tynorphin towards angiotensin converting enzymes. A good starting point for this would be HPLC assays to understand what fragments AnCE cleaves tynorphin into and what the time frame is for these cleavage events, and hence whether they are biologically relevant.

4.4.3 Structures of AnCE with Spinorphin

All three datasets processed for AnCE with spinorphin displayed some difference density in the active site due to the presence of spinorphin fragments. As was the case with tynorphin, it was not possible to fit the full spinorphin heptapeptide (LVVYPWT) into any of the active sites; furthermore, different peptide fragments were visible in each structure.

4.4.3.1 AnCE_Spin_4

As illustrated in Figure 4.10, it was quite clear by studying the difference density in the AnCE_Spin_5 active site that the LVV tripeptide should be placed such that the residues occupied the S_1, S_1' and S_2' subsites of the enzyme respectively. Placed in this way the catalytic zinc ion is co-ordinated by the carbonyl oxygen of the leucine and the amide nitrogen of the P_1' valine. This would appear to ideally position the tripeptide for cleavage of the leucine-valine peptide bond. Interestingly, this arrangement is very similar to the inhibition of MMPs (matrix metalloproteinases) by TIMPs (tissue inhibitors of metalloproteinases). Here there is bidentate chelation of the MMP zinc ion by the amino and carbonyl groups of the N-terminal TIMP residue (Brew *et al.,* 2000).

The electron density stops abruptly after the P_2' valine residue (Figure 4.10) and there was certainly no possibility of adding any additional spinorphin residues beyond this point. This strongly suggests that AnCE has cleaved spinorphin between the valine and

tyrosine residues, resulting in the LVV tripeptide. Further evidence for this is provided by the fact that a C-terminal carboxyl group that would have been introduced to the P_2' valine following peptide bond cleavage fits the difference density very well and forms conserved interactions with AnCE active site residues.

Based on a single crystal structure, it is not possible to determine the sequence of events that led to the LVV tripeptide binding to the active site in this way. ACE homologues tend to remove the C-terminal dipeptide from their substrate, therefore it seems probable that the full spinorphin peptide may have bound to the active site with the C-terminal threonine residues at the P_2' position. The enzyme could then have removed the four C-terminal residues by sequential cleavage of the WT and YP dipeptides. This would leave the LVV tripeptide able to slide along the active site to reach the position that it is in in the AnCE_Spin_4 structure. This is illustrated in Figure 4.17.

4.4.3.2 AnCE_Spin_5

The second structure of AnCE with spinorphin contains more extensive density in the active site than is seen in the AnCE_Spin_4 structure. As illustrated in Figure 4.12, the YPWT tetrapeptide was modelled in this density with the tyrosine, proline, tryptophan and threonine residues at the P_2, P_1, P_1' and P_2' positions, respectively. There was little ambiguity in the placement of these residues as the bulky side chains of tyrosine and tryptophan, the unique cyclic nature of proline and the terminal carboxyl group of threonine matched the density well, as shown in Figure 4.12. There was also no possibility of extending the N-terminus of the peptide beyond the tyrosine. The difference density stopped abruptly at this point, indicating that the peptide had been cleaved between the valine and tyrosine residues.

This complements the AnCE_Spin_4 structure well, providing further evidence of the cleavage of the valine-tyrosine peptide bond by AnCE. It does however contradict the proposed sequence of events. Given that the YPWT peptide has remained intact here, the enzyme appears to be cleaving the valine-tyrosine bond in the full-length peptide, rather than following cleavage of the proline-tryptophan bond. Although angiotensin-converting enzymes most commonly act as dipeptidyl carboxypeptidases, N-terminal tripeptidase activity has previously been reported not only for human ACE, but also for AnCE (Isaac *et al.,* 2007; Lamango *et al.,* 1996; Siviter *et al.,* 2002) thus, this would not be unprecedented.

4.4.3.3 AnCE_Spin_6

The AnCE_Spin_6 data shows much less difference density in the active site than in either of the other two spinorphin structures. In spite of this, it is very clear from the density that it is due to a dipeptide, consisting of tyrosine at the P_1' position and proline at the P_2' position. This is illustrated in Figure 4.14, where the YP dipeptide is shown in the electron density.

The most likely explanation for the YP dipeptide at this position is that the YPWT peptide seen in AnCE_Spin_5 has had its C-terminal dipeptide (WT) removed by AnCE. The remaining two residues, tyrosine and proline, have then been able to slide along the active site and occupy the S_1' and S_2' positions. Here the newly introduced proline carboxyl group interacts with active site residues Gln265, Lys495 and Tyr504, stabilising the dipeptide in this position. This sequence of events is depicted in Figure 4.17.

Interestingly, the crystals used for AnCE_Spin_6 data collection were grown at the same time as those used for the AnCE_Spin_5 data. Hence, it seems reasonable to suggest, that during the intervening three weeks further peptide cleavage occurred within the crystal.

This does not correspond quite so well with the interpretation of the AnCE_Spin_4 and AnCE_Spin_5 structures, where it was proposed that AnCE was cleaving the N-terminal tripeptide (LVV) from spinorphin. It is not possible to form a conclusion based only on the three structures presented here, however it is likely that AnCE is cleaving both the N-terminal tripeptide and the C-terminal dipeptide. It is not possible to ascertain from the structures alone the sequence or relative rates of these processes, although it is probable that the N-terminal tripeptide is removed first. It is also quite possible, based on the position of the LVV tripeptide in AnCE_Spin_4, that this fragment undergoes further cleavage to leucine and a dipeptide consisting of VV.

Irrespective of the precise sequence of events, it is clear that AnCE is cleaving spinorphin at multiple sites and that the resulting peptide fragments are able to move and occupy different positions within the catalytic site. This is very similar to what was observed with tynorphin. Peptide hydrolysis at multiple sites is likely a consequence of the broad range of biological functions of ACE homologues. The majority of interactions formed between angiotensin converting enzymes and peptide substrates only involve the peptide backbone, resulting in relatively broad substrate specificity.

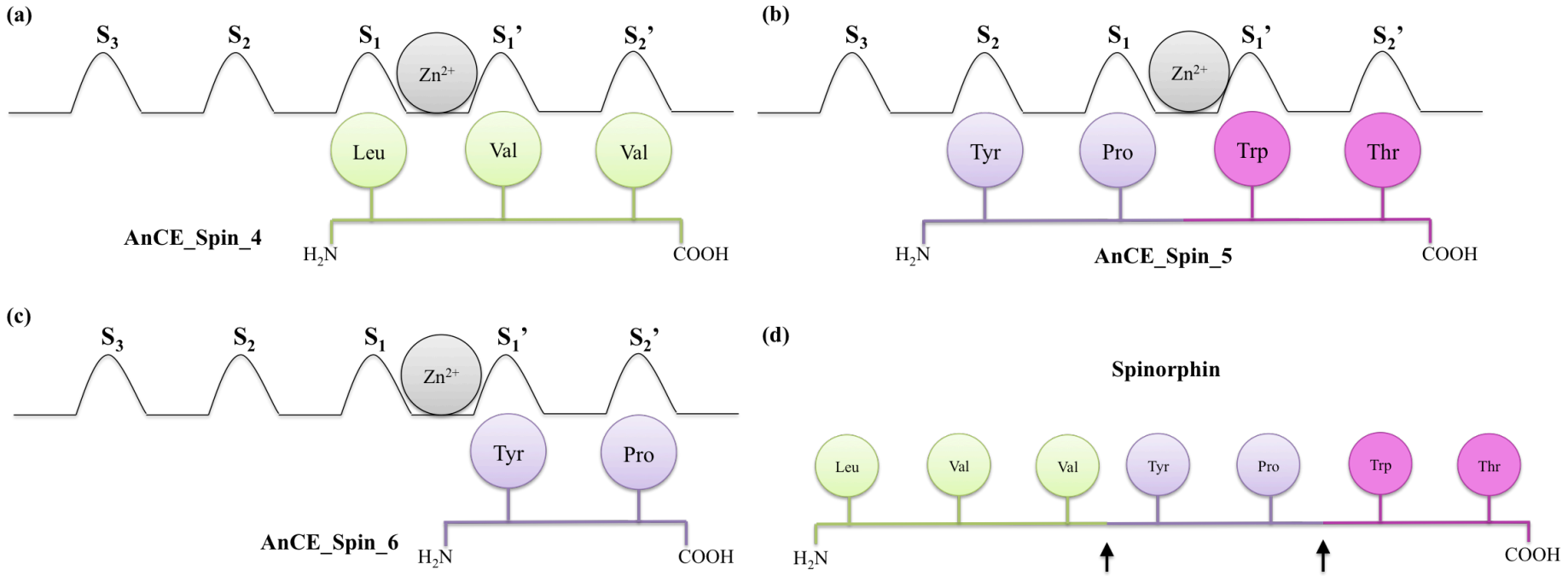


Figure 4.17: Schematic Representation of Spinorphin Binding at the AnCE Active Site. In all figures the locations of the subsites of the enzymes are labelled as S_3, S_2, S_1, S_1' and S_2'. The location of the catalytic zinc ion between the S_1 and S_1' subsites is also indicated. The spinorphin residues occupying each subsite are labelled. (a) In the AnCE_Spin_4 structure the tripeptide LVV, shown in green, occupies the S_1, S_1' and S_2' subsites. (b) In the AnCE_Spin_5 structure a four residue fragment of spinorphin, YPWT, occupies the S_2, S_1, S_1' and S_2' subsites. The peptide is positioned for cleavage of the peptide bond between the proline and tryptophan residues. (c) In the AnCE_Spin_6 structure a single dipeptide, YP (shown in purple), is found in the S_1' and S_2' subsites. This is likely to be the result of cleavage of the YPWT tetrapeptide shown in (b) and release of WT from the active site, enabling the YP dipeptide to move and occupy this position. (d) Schematic showing the full-length spinorphin peptide with residues colour coded as in (a), (b) and (c) to illustrate the different cleavage products. Arrows highlight sites where peptide cleavage has been observed to occur.

4.4.4 Stability of Spinorphin in the Presence of Angiotensin Converting Enzymes

Two studies have previously investigated the stability of spinorphin in the presence of ACE. The first, performed by Lantz *et al*. in 1991, used a related haemorphin with the sequence LVVYPWTQR. This is identical to spinorphin with the addition of glutamine and arginine at the C-terminus. Following incubation with lung extract, a source of ACE, the C-terminal QR dipeptide was cleaved leaving LVVYPWT, i.e. spinorphin. Following further incubation the LVV tripeptide was released leaving YPWT. YPWT did not undergo further cleavage and remained stable for the 7 hours of the experiment.

These results support the removal of the LVV N-terminal tripeptide seen here. However, the experiment was not performed with pure ACE. The lung extract would have contained many different membrane localised enzymes hence it is impossible to determine which enzymes were responsible for the cleavage. The fact that the YPWT peptide remained intact does appear to contradict the results from the crystal structures.

The second time that this was investigated was by Zhao *et al*., in 1994. Here the same related haemorphin, LVVYPWTQR, was used and this time incubated directly with ACE. Here once again the peptide bond between threonine and glutamine was cleaved, resulting in the release of the QR dipeptide. Further incubation for 7 hours did not lead to further peptide cleavage, indicating that under these conditions ACE is neither acting as an N-terminal tripeptidase nor is it removing further residues from the C-terminus of the peptide.

The results from the structural studies with AnCE are a direct contradiction of this, however there are a number of important points to consider. The first is that the nature of the two systems is exceedingly different. The stability experiments performed by Zhao *et al*., took place over 7 hours at 37 ^{o}C in a dilute solution. In contrast the crystallisation experiments were performed at 21 ^{o}C, the set-up is by its very nature highly concentrated, the components of the solution (buffer, salt and precipitant) were different and the peptide was incubated with the enzyme for 30 minutes prior to crystallisation and then in the crystals for up to 3 months. It is possible that ACE does display further activity towards spinorphin, but the rate of this reaction is so slow that it would not be detected under the conditions of the experiment performed by Zhao *et al*., it is also therefore unlikely to be physiologically relevant.

The second consideration is that this work was performed using the *D. melanogaster* ACE homologue, AnCE, rather than the human enzyme. Although AnCE has been used

very successfully as a model for human ACE it is important not to overlook the limitations of this system. It is possible that whilst AnCE may cleave spinorphin, and indeed tynorphin, into multiple fragements, ACE may display much more limited peptidase activity towards these peptides. The only way to investigate this further would be to perform further experiments with both ACE and AnCE.

4.4.5 Spinorphin and Tynorphin as ACE Inhibitors

Spinorphin is a naturally occurring peptide formed by the cleavage of the β-chain of haemoglobin that was identified as an inhibitor of enkephalin degrading enzymes, including ACE (Nishimura *et al.,* 1993). Tynorphin is a related peptide designed to be a better inhibitor of DPPIII and is a weaker inhibitor of ACE than spinorphin (Yamamoto *et al.,* 2000).

The sequences of spinorphin and tynorphin are closely related; spinorphin has an additional leucine residue at the N-terminus and an additional threonine at its C-terminus. It is the additional leucine that is mainly thought to be responsible for the increased potency of spinorphin as an ACE inhibitor; a study performed using the closely related peptide VVYPWTQF showed that addition of an N-terminal leucine residue, as seen in spinorphin, decreased the IC_{50} for ACE inhibition from 10.2 μM to 1.7 μM (Zhao *et al.,* 1994; Zhao and Piot*,* 1997).

Based on the structural data presented here it is not immediately apparent why the N-terminal extension has such a profound effect on the inhibitory properties of these peptides. The most straightforward explanation would be that the additional residue allows the peptide to form more extensive interactions with the active site. However it is unlikely that the explanation is as simple as this. Primarily, this is because ACE tends to interact mainly with the peptide backbone of its substrates and the additional leucine residue has quite a small, non-polar side chain that is unlikely to form particularly extensive interactions with the active site.

Beyond this, the mechanism by which spinorphin and tynorphin function as ACE inhibitors also remains unclear. Based on the data presented here they appear to be competitive inhibitors that bind to the active site through interactions that are well conserved in peptide binding to ACE homologues.

4.4.5.1 Comparison of Spinorphin and Tynorphin with the BPPs

The BPPs are a group of natural ACE inhibitory peptides originally isolated from the venom of the pit viper *B. jararaca*. As detailed in Chapter 3, the properties of these peptides as ACE inhibitors have been studied quite extensively. The BPPs are all proline rich peptides and there is very little relationship between the sequences of these peptides and spinorphin and tynorphin. This raises the question as to how spinorphin and tynorphin function as ACE inhibitors.

The P_1' proline is recognised as an important feature of the BPPs as it protects the scissile peptide bond from cleavage, allowing them to act as competitive inhibitors of ACE. Interestingly, a proline residue is conserved at this position in tynorphin, but not in spinorphin, which is a more potent ACE inhibitor. Furthermore the data obtained here indicate that whilst spinorphin is cleaved by AnCE into many fragments, this does not include removal of the C-terminal threonine residue, which would leave the proline at the P_1' positon.

It is clear that much more work is required to understand the mechanisms by which spinorphin and tynorphin function as ACE inhibitors. This may involve further crystallisation experiments with ACE rather than AnCE and if possible with inactive AnCE or ACE mutants. The latter would allow the study of the binding mechanism of the full-length peptides, which would provide more information about any additional interactions formed by the N-terminus of spinorphin compared to tynorphin.

It would also be very beneficial to obtain kinetic data for ACE and AnCE inhibition by both peptides, and to perform further analysis of their stability in the presence of ACE. This would likely involve the use of HPLC and it would be particularly important to look at stability over different timescales, in different solution conditions and at different temperatures.

Based on the crystallographic data presented here, it is possible that small, two or three residue, fragments of spinorphin and tynorphin are acting as competitive ACE inhibitors. This is an interesting concept and suggests that it may be possible to create potent ACE inhibitors by focusing on maximising the interactions with a small area of the active site, likely the S_1' and S_2' pockets, rather than along the length of the substrate binding channel. This idea is pursued further in Chapter 5.

4.5 Conclusions and Further work

The peptide spinorphin, which is produced by enzymatic cleavage of the β-chain of haemoglobin, and the closely related synthetic peptide tynorphin, were characterised as ACE inhibitors a number of years ago. Prior to this study there was very little data to indicate how these peptides acted as ACE inhibitors or to explain the observed differences in their ability to inhibit ACE.

Here three structures are presented for each peptide in complex with the *D. melanogaster* ACE homologue, AnCE. The structures indicate that both peptides are being cleaved by AnCE into multiple fragments, some of which are being retained in the active site and presumably acting as competitive inhibitors of enzyme activity.

The possible sequence of cleavage events is discussed but further experimental evidence is required to form a firm conclusion. From the data presented here it is also difficult to explain why spinorphin is a more potent ACE inhibitor than tynorphin and the further work that would be required to investigate this is outlined.

The crystal structures indicate that small peptide fragments derived from enzymatic cleavage of longer peptides may be able to function as good ACE inhibitors. This is investigated further in Chapter 5.

Chapter 5: Crystal Structures of AnCE in Complex with Dipeptide Inhibitors

5.1 Introduction

In addition to the well-studied actions of certain naturally occurring peptides as ACE inhibitors, there is an increasing amount of evidence to suggest that much smaller dipeptide fragments can also have this effect (Lunow *et al.*, 2013; Marczak *et al.,* 2003; Martin *et al.,* 2008). This is an important development, as small dipeptides are likely to be biologically available, for example as products of digestion, and could have an important role in the *in vivo* regulation of ACE activity.

The work presented in Chapter 4 on the structural basis of ACE inhibition by the haemorphins spinorphin and tynorphin indicated that small peptide fragments, arising from hydrolysis by ACE, might be at least partly responsible for the inhibitory activity of these two peptides.

In recent years the action of dipeptides as ACE inhibitors has begun to be characterised. In 2013 a study conducted by Lunow *et al.* reported that the dipeptides IW and AW were able to inhibit sACE activity with IC_{50} values of 0.7 μM and 20 μM, respectively (Lunow *et al.,* 2013). In a further study, performed in 2015, the same group reported that some dipeptides displayed selective inhibition of the two domains of sACE (Lunow *et al.,* 2015).

Taken together with the potential bioavailability of such compounds, these developments indicate that research into the inhibitory actions of dipeptides could be an important area to pursue both in terms of understanding more about the *in vivo* activities of this enzyme, but also how it may be possible to use the way that these molecules interact with the two domains of sACE in the development of a new generation of domain specific inhibitors.

In the work presented by Lunow *et al.* in 2015, a C-terminal tryptophan residue was implicated in selective inhibition of the C-domain by dipeptides. It has been suggested that this is as a result of the increased hydrophobicity of the S_2' pocket in the C-domain compared to the N-domain, a feature that has previously been exploited in the development of the C-domain selective inhibitors RXPA380 (Georgiadis *et al.,* 2004) and kAW ((5S)-5-[(N-benzoyl)amino]-4-oxo-6-phenylhexanoyl-L-tryptophan) (Almquist *et al.,* 1980; Deddish *et al.,* 1998; Nchinda *et al.,* 2006; Redelinghuys *et al.,* 2006; Watermeyer *et al.,* 2008). Similarly, an aliphatic residue at the P_1' position improved C-domain inhibition, with the dipeptides IW and VW being better inhibitors

than YW. Conversely, the data collected indicated that a C-terminal arginine residue might be associated with N-domain selectivity (Lunow *et al.,* 2015).

5.1.1 Aims

The aim of the work presented in this chapter is to use the *D. melanogaster* ACE homologue AnCE as a model to study the structural basis of ACE inhibition by four dipeptides: WR, VW, IW and YW in order to learn more about how these molecules function as inhibitors and how this could be built upon in the design of future, domain specific inhibitors.

AnCE is a good system to use as a model for this. The hydrophobicity of the S_2' pocket in the C-domain has been suggested as the cause of the observed domain selectivity seen with the tryptophan containing dipeptides. In the C-domain two valine residues, Val379 and Val380 are primarily responsible for this. In the N-domain, Ser357 and Thr358 replace these, respectively, reducing the hydrophobic nature of this site. Interestingly, in AnCE, Phe363 and Thr364 replace Val379 and Val380, respectively. This indicates that the S_2' pocket in AnCE shares some features with that of the N-domain, whilst also retaining some of the hydrophobicity present in the C-domain. It should therefore be possible to use AnCE to provide an insight into how these dipeptides interact with both of the domains of sACE.

5.2 Methods

5.2.1 AnCE Expression, Purification and Crystallisation

Expression and purification of AnCE was performed as outlined in 2.7. Recombinant AnCE was expressed using the methylotropic yeast *P. pastoris* and purified by hydrophobic interaction chromatography followed by size exclusion chromatography. The identity and purity of the protein were confirmed by SDS-PAGE analysis.

Co-crystallisation experiments with the dipeptides WR, VW, IW and YW were performed using the hanging drop vapour diffusion technique following the general protocol for AnCE crystallisation that is detailed in section 2.7.4. The four dipeptides were all purchased from GL Biochem (Shanghai) Ltd.

Immediately prior to crystallisation, AnCE at a concentration of 17.5 mg ml^{-1} was incubated with 20 mM of each dipeptide in a 1:4 ratio of dipeptide:protein at 4 ^{o}C for approximately 30 minutes. 1 µl of the protein and dipeptide mixture was then mixed

with 1 µl of reservoir solution to form a 2 µl drop. Each drop was suspended on a coverslip over 500 µl of the appropriate reservoir solution (1.2-1.3 M sodium citrate, 0.1 M HEPES pH 7.5) and the wells sealed with high vacuum grease. The crystallisation experiments were incubated at 21 ^{o}C and observed regularly for any signs of crystal formation.

5.2.2 X-ray Diffraction Data Collection

Previous experiments with AnCE have shown that the high sodium citrate concentration (1.2-1.3 M) used during crystallisation provides sufficient cryo-protection, negating the need for additional cryo-protectants (Akif *et al.,* 2010). Based on this, prior to crystal mounting an additional 1 µl of the reservoir solution was added to the drop so as to maximise the sodium citrate concentration and to help to stop the drop from becoming too dry.

Crystals were mounted directly into litholoops (Molecular Dimensions Ltd., Newmarket, UK) and rapidly frozen in liquid nitrogen for transportation to the synchrotron.

All diffraction data were collected at DLS using beamline i04. Data were captured using a Pilatus3 6M detector and all diffraction experiments were performed at 100 K. A summary of key data collection parameters is presented in Table 5.1.

5.2.3 Data Processing and Structure Solution

All datasets collected were initially processed using the xia2 pipeline (Winter, 2010) at DLS. Following this the "best" dataset collected for each dipeptide was identified based on the resolution and completeness of the data. These four datasets were then used for further processing and structure solution.

Integration for the four best datasets was performed using MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) followed by scaling using Aimless (Evans, 2006; Evans, 2007). Initial phases were calculated by molecular replacement using Phaser (McCoy *et al.,* 2007) with the native AnCE structure, PDB code 2X8Y (Akif *et al.,* 2010), as the search model.

Following production of the initial models, iterative rounds of model building in C*oot* (Emsley *et al.,* 2011) and refinement using REFMAC (Murshudov *et al.,* 1997; Murshudov *et al.,* 2011) were performed. A key part of this was modelling dipeptide

residues into active site difference density. Water molecules were added where Fo-Fc peaks in electron density exceeded 3σ and where potential hydrogen bonds could be made.

Molprobity (Chen *et al.,* 2010) was used to validate the quality of the final structures and LigPlot^{+} (Laskowski and Swindells, 2011) to study interactions between the dipeptides and the active site. The molecular graphics program Pymol (The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC) was used to perform RMSD calculations and to prepare the figures.

5.3 Results

5.3.1 Co-crystallisation of AnCE with IW, VW, YW and WR

Following co-crystallisation with the four dipeptides, crystals appeared in every drop (1.2-1.3 M sodium citrate, 0.1 M HEPES, pH 7.5) within 10 days of incubation at 21 ^{o}C. These crystals had the same appearance as those that are normally obtained for AnCE in this condition and some examples are shown in Figure 5.1.

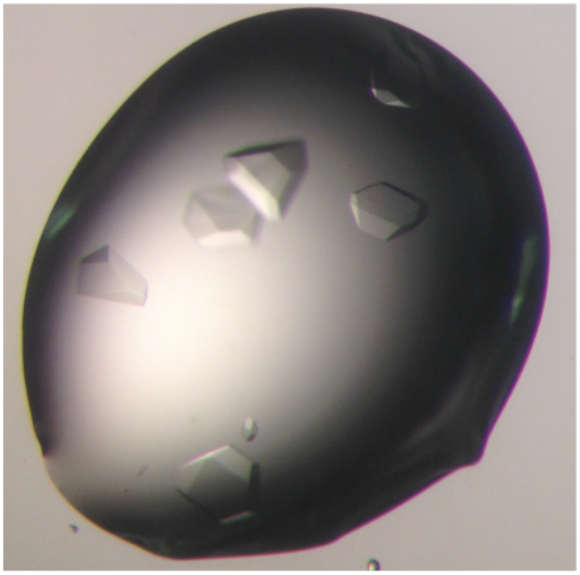


Figure 5.1: Examples of AnCE Crystals Obtained Following Co-crystallisation with the Dipeptides IW, VW, YW and WR. Crystals are shown in a 2 µl drop and were grown at 21 ^{o}C in a condition containing 1.26 M sodium citrate, 0.1 M HEPES, pH 7.5.

5.3.2 X-ray Diffraction Data Collection

X-ray diffraction data for AnCE in complex with each of the dipeptides were collected at DLS using beamline i04 on 16th April 2015. Following this, a further two datasets were collected for AnCE with IW and WR, also using beamline i04 at DLS, on 2nd July 2015. All data were collected using a Pilatus3 6M detector. An example of the diffraction images obtained is shown in Figure 5.2.

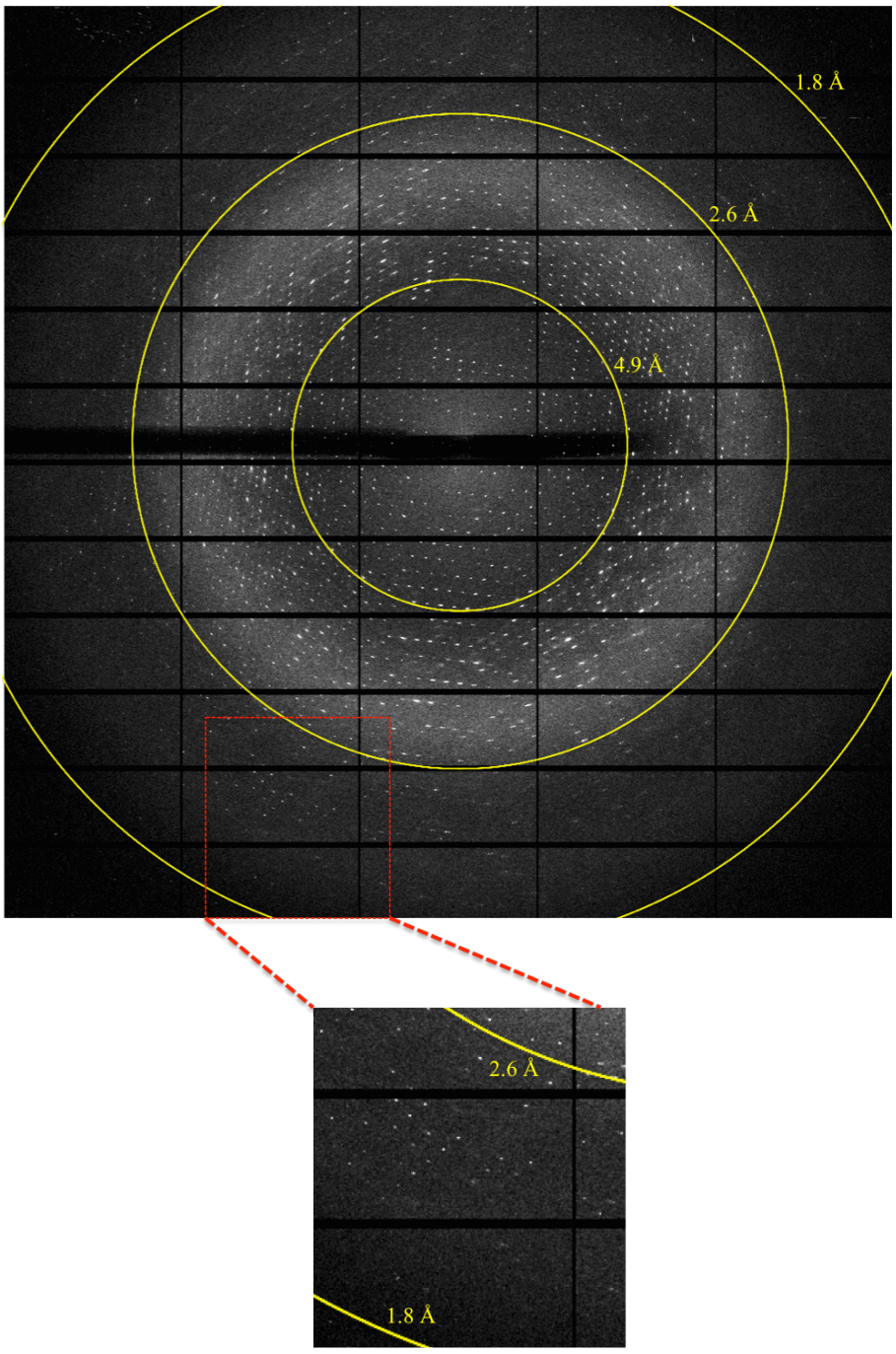


Figure 5.2: Example of a Diffraction Image Obtained from a Crystal of AnCE with IW. Resolution rings are shown in yellow and labelled. Diffraction beyond 2.6 Å is clearly visible on the inset image.

Dataset name	Collection date	Beamlline	Detector	Number of images	Phi range (°)	Resolution range (Å)	Overall completeness (%)
AnCE_IW_1	16th April 2015	i04	Pilatus3 6M	1800	0-180	28.80-1.95	81.8
AnCE_VW_1	16th April 2015	i04	Pilatus3 6M	900	0-180	86.54-1.90	99.6
AnCE_YW_1	16th April 2015	i04	Pilatus3 6M	900	0-180	86.33-1.72	98.4
AnCE_WR_1	16th April 2015	i04	Pilatus3 6M	900	0-180	86.25-1.72	98.9
AnCE_IW_2	2nd July 2015	i04	Pilatus3 6M	1200	63-243	86.69-2.15	99.5
AnCE_WR_2	2nd July 2015	i04	Pilatus3 6M	900	123-303	86.66-1.71	99.9

Table 5.1: Summary of Data Collection and Initial Processing Statistics for AnCE with the Dipeptides IW, VW, YW and WR. Where more than one dataset was collected for the same dipeptide the data used in downstream processing is highlighted in yellow.

The six datasets collected were all successfully processed in the first instance using the xia2 pipeline at DLS (Winter, 2010) in the space group *h3*. The statistics obtained from this are included in Table 5.1.

Where two datasets were collected for the same dipeptide, namely IR and WR, the "best" dataset was identified and used in further processing steps. For WR this selection was straightforward and the second dataset collected (AnCE_WR_2) was chosen because it was collected to a higher resolution than the first.

In the case of IW the resolution of the first dataset, AnCE_IW_1, was slightly higher than that of AnCE_IW_2, 1.90 Å and 2.15 Å, respectively. However during data processing the maximum completeness that could be achieved for the AnCE_IW_1 data was 81 %. Attempts were made to improve this value by imposing a resolution limit, but even when this reached 2.15 Å, the resolution of AnCE_IW_2, the overall completeness was still less than 85 %. In contrast, the overall completeness for the AnCE_IW_2 data at this resolution was 100 % hence this dataset was used for further processing and structure solution.

5.3.3 Data Processing and Structure Solution

The four datasets, one for AnCE with each dipeptide, were processed using MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) and Aimless (Evans, 2006; Evans, 2007) in the space group *h3*, which is the space group normally seen for AnCE crystals grown in this condition (Akif *et al.,* 2010). Phases were calculated by molecular replacement using Phaser (McCoy *et al.,* 2007) with the native AnCE structure, PDB code 2X8Y (Akif *et al.,* 2010), as the search model. In each case there was one molecule per asymmetric unit in the crystal.

Following an initial round of rigid body refinement, performed using REFMAC (Murshudov *et al.,* 1997; Murshudov *et al.,* 2011), there was clear difference density at each of the three N-linked glycosylation sites that are present in the native AnCE structure, hence N-linked glycan chains were added to the models at Asn53, Asn196 and Asn311. The catalytic zinc ion was modelled into the active site difference density, co-ordinated by His367, His371 and Glu395. Following these adjustments a portion of difference density remained in the active site of each of the four structures. Dipeptide residues were fitted to this density as appropriate and this is discussed in the subsequent sections. Further rounds of refinement followed prior to structure validation using Molprobity (Chen *et al.,* 2010). Data processing statistics are detailed in Table 5.2.

	AnCE_IW	AnCE_VW	AnCE_YW	AnCE_WR
Beamline at DLS	i04	i04	i04	i04
Detector	Pilatus3 6M	Pilatus3 6M	Pilatus3 6M	Pilatus3 6M
Wavelength (Å)	0.980	0.980	0.980	0.980
Space group	*h3*	*h3*	*h3*	*h3*
Unit cell parameters:				
a (Å)	173.40	173.07	172.67	173.30
b (Å)	173.40	173.07	172.67	173.30
c (Å)	103.20	102.49	102.61	103.00
α (°)	90.00	90.00	90.00	90.00
β (°)	90.00	90.00	90.00	90.00
γ (°)	120.00	120.00	120.00	120.00
Molecules per asymmetric unit	1	1	1	1
Resolution range (Å)	86.69-2.15	86.54-2.20	86.33-1.72	86.64-1.71
$^{\#}R_{merge}$ (%)	9.5 (64.1)	8.2 (48.8)	5.7 (49.8)	6.1 (71.4)
$^{+}R_{pim}$ (%)	7.3 (48.8)	5.5 (32.4)	5.2 (46.0)	4.7 (46.2)
Mean *I/σ(I)*	1.0 (0.8)	9.4 (2.8)	7.9 (1.6)	11.2 (1.7)
Completeness (%)	100.0 (100.0)	99.3 (99.8)	98.4 (95.5)	100.0 (100.0)
Number of reflections:				
Total	299, 754	209, 712	307, 199	586, 652
Unique	62, 903	57, 669	119, 892	124, 816
Multiplicity	4.8 (4.6)	3.6 (3.7)	2.6 (2.2)	4.7 (3.9)
Wilson *B* factor (Å^2)	34.47	35.11	25.72	28.15
Average *B* factor (Å^2)				
Protein	34.00	34.63	24.41	26.66
Ligand	29.23	69.46	58.19	48.31
Zinc	27.06	30.27	21.24	23.67
Glycosylated sugars	59.58	58.15	46.37	47.60
Solvent	36.85	35.25	35.76	38.02
$^{\S}R_{cryst}$/$^{\#\#}R_{free}$ (%)	16.45/19.92	17.93/22.13	17.25/19.61	16.81/18.57
R.M.S deviation from ideal values				
Bond lengths (Å)	0.009	0.011	0.007	0.006
Bond angles (°)	1.275	1.435	1.114	1.103
Ramachandran plot statistics (%)				
Favoured	97.99	97.65	98.66	98.49
Disallowed	0.00	0.00	0.00	0.00

$^{\#}R_{merge} = \Sigma_{hkl}\Sigma_i|\boldsymbol{I_i}\,(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I_i}(hkl)$, where $\boldsymbol{I}_i$(hkl) is the intensity of *i*th measurement and <***I****(hkl)*> is the average of symmetry-related observations of a unique reflection.

$^{+}R_{pim} = \Sigma_{hkl}\sqrt{\frac{1}{n-1}}\Sigma_{i=1}^{n}|\boldsymbol{I}_i(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I}_i(hkl)$

$^{\S}R_{cryst} = \Sigma_h|F_o - F_c|/\Sigma_h F_o$, where F_o and F_c are observed and calculated structure factor amplitudes of reflection *h*, respectively.

$^{\#\#}R_{free}$ is equal to R_{cryst} for a randomly selected 5 % subset of reflections.

Table 5.2: X-ray Diffraction Data Collection, Processing and Refinement Statistics for AnCE with IW, VW, YW and WR. Outer shell values are shown in parentheses.

5.3.4 Crystal Structures of AnCE in Complex with the Dipeptides

Analysis of the overall topology of each structure indicates that there has been no large-scale movement of secondary structure elements compared to the native AnCE structure. This is confirmed by calculated RMSD values in the range 0.094 to 0.139 for each structure compared to native AnCE. Key features of AnCE seen in the structures include the long substrate binding channel that extends the length of the molecule and that is capped by an "N-terminal lid" formed by helices α1, α2 and α3. The active site is located at the centre of this channel, marked by the location of the catalytic zinc ion. It is here that difference density was observed in each of the dipeptide structures.

5.3.4.1 AnCE and IW

A total of 1200 diffraction images were collected from the crystal of AnCE with the dipeptide IW, which covered 180°. All 1200 of these images were used to process the data to a final resolution of 2.15 Å. Data processing statistics are summarised in Table 5.2.

Following initial rounds of model building and refinement, a quantity of difference density was visible in the AnCE active site. The IW peptide was placed into this density such that the isoleucine and tryptophan residues were occupying the S_1' and S_2' pockets respectively. As shown in Figure 5.3, all atoms of the dipeptide are well accommodated by the electron density at this position.

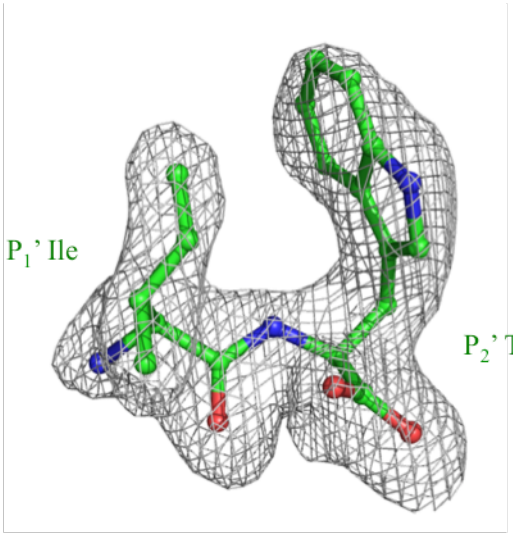


Figure 5.3: Fo-Fc Density Map Contoured at 3σ to Illustrate Binding of the IW Dipeptide to the AnCE Active Site. The IW residues are shown as green sticks and the Fo-Fc map as a grey mesh. Both residues are well accommodated by the density.

The mode of binding of the dipeptide is such that the C-terminal tryptophan side chain resides in a "hydrophobic pocket" that is formed primarily by Phe441, Phe551 and Phe363 (Figure 5.4). Further to this, the P_2' tryptophan residue also forms a number of interactions with the S_2' pocket that are well conserved in peptide binding to ACE homologues. The C-terminal carboxylate group interacts with the side chains of residues of Gln265, Lys495 and Tyr504. Additionally, the indole ring of the tryptophan side chain interacts with the side chain of Gln266.

Further contacts with the active site are made by the P_1' isoleucine. This residue contacts the side chains of His337 and His497 through its carbonyl oxygen, and the protein backbone at Ala338 via its amide nitrogen.

The dipeptide does not appear to interact directly with the catalytic zinc ion, rather it forms a water-mediated interaction from the P_1' amide nitrogen. Interestingly this nitrogen forms a direct interaction with Glu368, which has been implicated in the catalytic mechanism, where it is thought to have a role in activating the catalytic water molecule for nucleophilic attack (section 1.6.4) (Zhang *et al.,* 2013). The binding of IW to the active site is illustrated in Figure 5.4 and the interactions that occur between the dipeptide and the enzyme are listed in Table 5.3.

An important observation is that the position adopted by the tryptophan residue in the structure of AnCE with IW is almost identical to that of the tryptophan moiety of the C-domain selective inhibitor RXPA380 in the crystal structure with both AnCE and the C-domain of sACE (Akif *et al*., 2010; Corradi *et al.,* 2007). This indicates that the IW dipeptide is likely to bind to the C-domain in the same manner as observed here for AnCE.

IW atom	Interacting atom	Distance (Å)
P_2' Trp OXT	Water	2.79
P_2' Trp O	Lys495 NZ	2.60
P_2' Trp O	Tyr504 OH	2.68
P_2' Trp NE2	Gln266 OE1	3.34
P_1' Ile O	His497 NE2	3.11
P_1' Ile O	His337 NE2	2.75
P_1' Ile N	Ala338 O	3.18
P_1' Ile N	Glu368 OE2	2.71
P_1' Ile N	Water	2.65

Table 5.3: List of Interactions Formed by the IW Dipeptide in the AnCE Active Site.

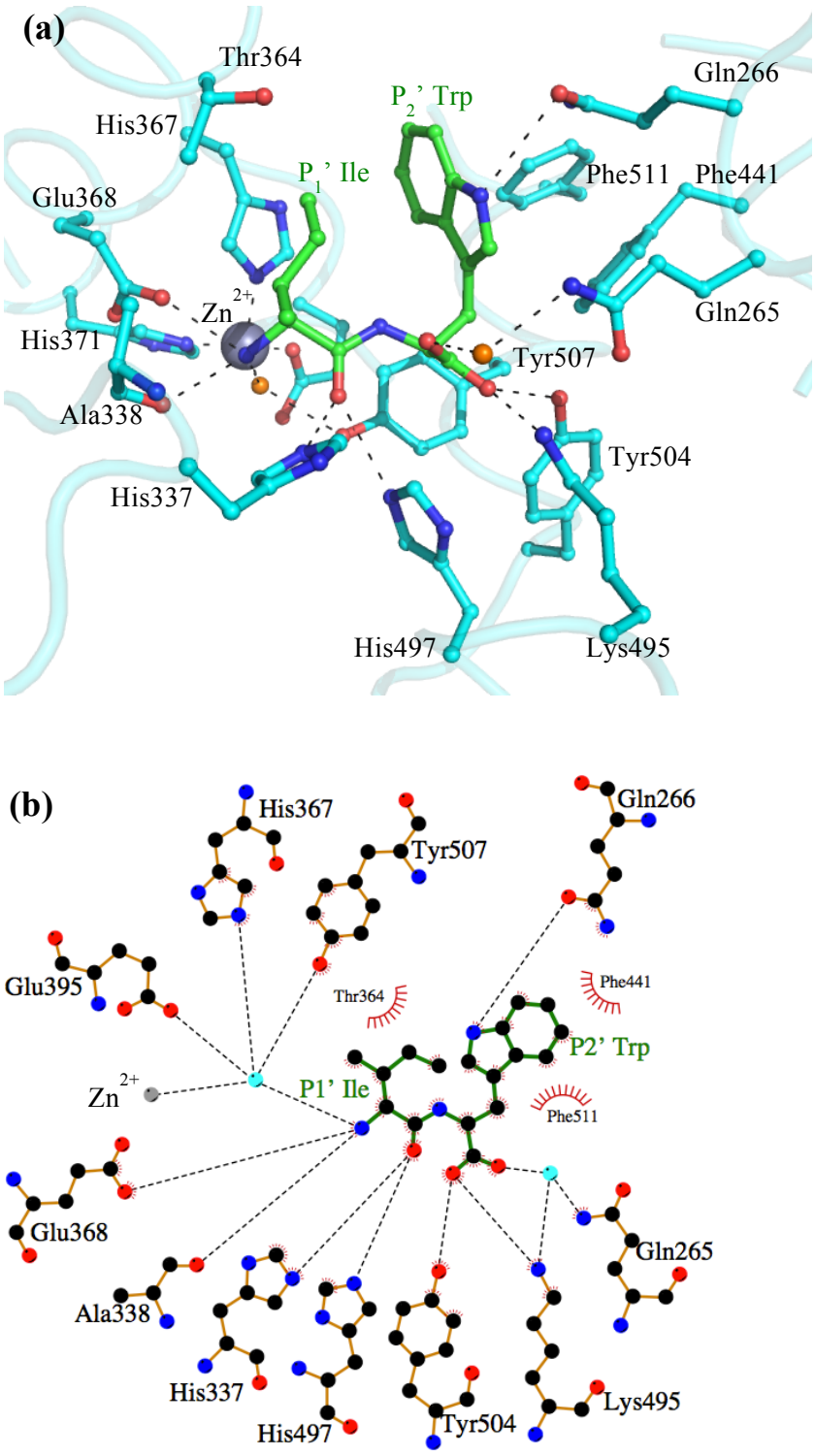


Figure 5.4: Binding of the Dipeptide IW to the AnCE Active Site. (a) The isoleucine and tryptophan residues occupying the S_1' and S_2' subsites, respectively, are shown as green sticks, with interacting protein residues shown as cyan sticks. This includes residues Thr364, Phe441 and Phe511, which make a critical contribution to the hydrophobic nature of the binding pocket. Water molecules involved in binding are shown as orange spheres and the catalytic zinc ion is a grey sphere. (b) Schematic representation of (a). Colouring as in (a) with the exception of water molecules, which are shown in turquoise, and hydrophobic residues, which are shown as red crescents.

5.3.4.2 AnCE and VW

A total of 900 diffraction images were collected from the crystal of AnCE with the dipeptide VW. These covered a total of 180° and reached a resolution of 1.90 Å. Following integration of all images performed using MOSFLM, attempts were made to scale the data using Aimless, but this resulted in exceedingly high R_{merge} values.

Visual inspection of the diffraction images indicated that for a subset, images 501-700, there were very few Bragg peaks at greater than 2 Å resolution. These images were therefore removed and further attempts made to scale the data.

Removing this subset of images did improve the R_{merge} value to some degree but not to an acceptable value. Ultimately, a resolution limit of 2.2 Å was imposed and this, combined with removing images 501-700, resulted in an overall R_{merge} value of 48.8 %. Importantly, the overall completeness remained at 99.3 %, indicating that removing the subset of images had not adversely affected the results. Full data processing statistics are detailed in Table 5.2.

Following initial refinement of the structure of AnCE with VW some difference density was clearly visible in the active site. As shown in Figure 5.5, this density was somewhat ambiguous. Although more extensive than that seen for IW (Figure 5.3), it was not so well defined, making placement of the dipeptide residues more difficult.

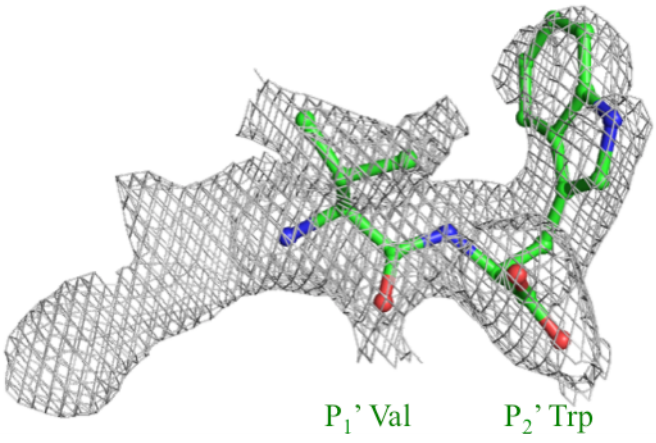


Figure 5.5: Fo-Fc Density Map Contoured at 3σ to Illustrate Binding of the VW Dipeptide to the AnCE Active Site. The IW residues are shown as green sticks and the Fo-Fc map as a grey mesh.

Initially, two alanine residues were placed into the density to occupy the S_2' and S_1' subsites. After refinement there was very clear density for the valine side chain at the P_1' position and some remaining density at the P_2' position. The valine and tryptophan side chains were therefore added to the residues at P_1' and P_2' respectively. As shown in Figure 5.5, the difference density very well accommodates the peptide backbone and the valine side chain and the large, bulky tryptophan side chain also fits reasonably well.

Although VW binds in approximately the same location as reported for IW, the P_1' residue adopts a slightly different position. The most obvious manifestation of this is that the valine nitrogen is able to interact directly with the catalytic zinc ion, rather than doing so through a water molecule, as seen in IW. This subtle movement means that the interaction with Glu368 that is seen with IW is not replicated here, as the isoleucine nitrogen is now more than 4 Å from the glutamate side chain. The interactions formed by the VW dipeptide in the AnCE active site are illustrated in Figure 5.6 and listed in Table 5.4.

VW atom	Interacting atom	Distance Å
P_2' Trp OXT	Water	2.74
P_2' Trp O	Lys495 NZ	2.65
P_2' Trp O	Gln265 NE2	3.23
P_2' Trp O	Tyr504 OH	2.45
P_2' Trp NE2	Gln266 OE1	3.05
P_1' Val O	His497 NE2	2.58
P_1' Val O	His337 NE2	2.84
P_1' Val N	Zn^{2+}	2.56

Table 5.4: List of Interactions Formed by the VW Dipeptide in the AnCE Active Site.

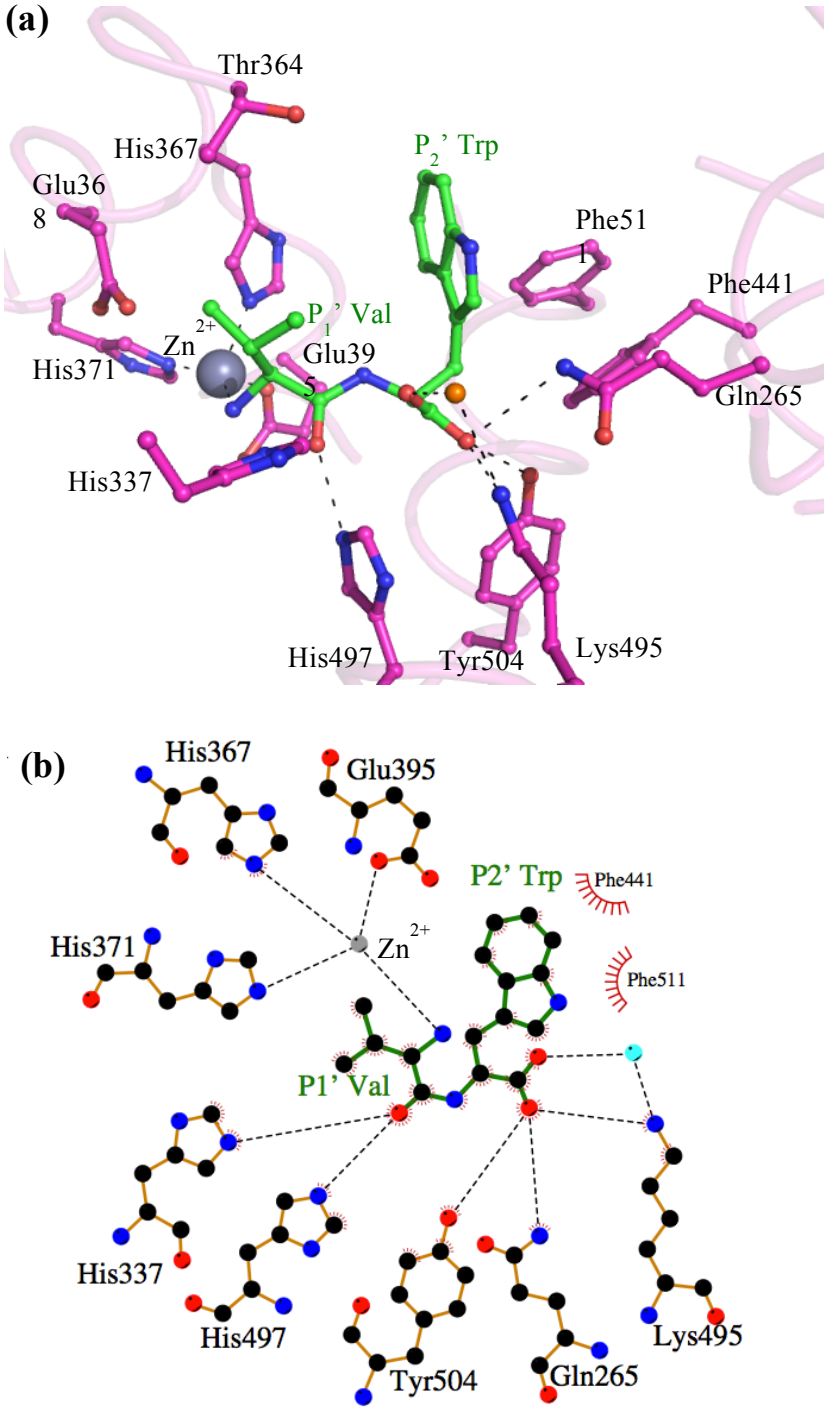


Figure 5.6: Binding of the Dipeptide VW to the AnCE Active Site. (a) The valine and tryptophan residues occupying the S_1' and S_2' subsites, respectively, are shown as green sticks, with interacting protein residues shown as magenta sticks. This includes residues Thr364, Phe441 and Phe511, which make a critical contribution to the hydrophobic nature of the binding pocket. Water molecules involved in binding are shown as orange spheres and the catalytic zinc ion is a grey sphere. (b) Schematic representation of binding. Colouring as in (a) with the exception of water molecules, which are shown in turquoise, and hydrophobic residues, which are shown as red crescents.

After placement of the dipeptide in the S_2' and S_1' subsites, some residual density was visible, extending past the zinc ion towards the S_1 pocket, Figure 5.5. This is quite a significant piece of density; furthermore it appears to be continuous with the portion of density that has been attributed to the VW dipeptide. In spite of this the density is not particularly well defined, hence it was challenging to build the model into it.

Based on the fact that the active site difference density is continuous, it is most likely due to the dipeptide in some way. With this in mind, an attempt was made to place the VW dipeptide in a second position in the active site, such that the valine residue remained in the S_1' site, but the tryptophan residue now resided in the S_1 pocket. This is illustrated in Figure 5.7, which shows that whilst this model does not provide a "perfect fit" for the residual density, it is the best interpretation of the data available. It is therefore possible that in some molecules the VW dipeptide may have bound in this alternative location.

Binding in this manner would mean that the C-terminal carboxylate of the tryptophan residue could interact with the catalytic zinc ion. Furthermore, the tryptophan nitrogen may be close enough to interact with the side chain of Glu368, a residue that has been implicated in catalysis. In AnCE Trp341 and Phe375 may play a role in providing a hydrophobic environment for the large tryptophan side chain.

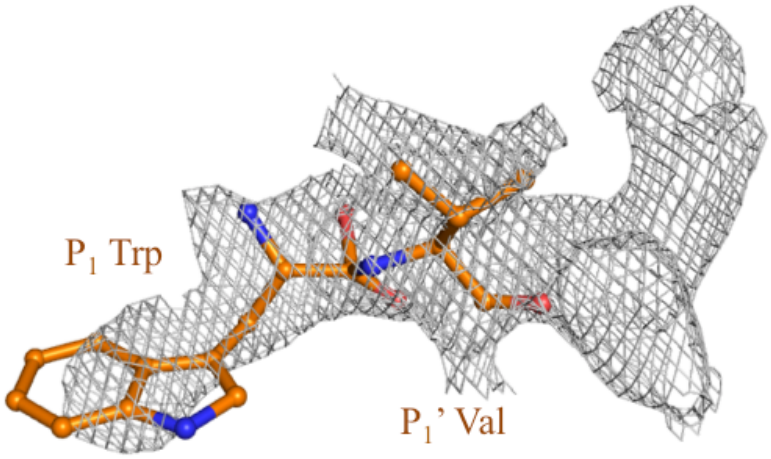


Figure 5.7: Placement of a Second VW Molecule at a Second Position in the AnCE Active to Fill the Remaining Difference Density. The VW molecule is shown as orange sticks, with the valine and tryptophan residues adopting the P_1' and P_1 positions respectively. The difference density is displayed as a Fo-Fc map contoured at 3σ, shown as a grey mesh.

5.3.4.3 AnCE and YW

900 diffraction images were collected from the crystal of AnCE with YW, covering 180°. The maximum resolution of the data was 1.72 Å. In order to achieve good statistics for data processing at this resolution a subset of images, 51-250, were excluded. This resulted in an overall R_{merge} value of 5.7 % and an overall completeness of 98.4 %. Full data processing statistics are presented in Table 5.2.

Following initial rounds of model building and refinement, some difference density was visible in the active site. This was located in a very similar position to that seen for the dipeptides IW and VW (sections 5.3.4.1 and 5.3.4.2, respectively). The YW dipeptide was modelled into this difference density such that the tryptophan and tyrosine residues occupied the S_2' and S_1' subsites, respectively. As illustrated in Figure 5.8, this density very well accommodates the dipeptide backbone, whilst the large tryptophan and tyrosine side chains fit adequately. Unlike in the structure with VW, there is no evidence of any further difference density in the active site channel, indicating that YW binds at a single location.

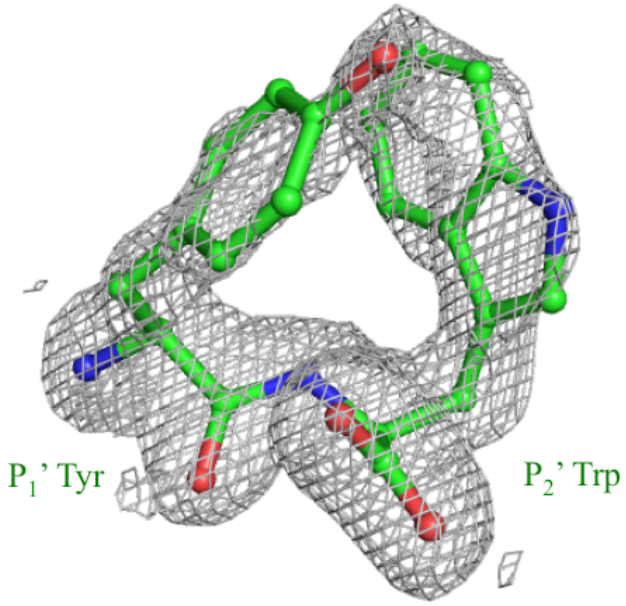


Figure 5.8: Fo-Fc Density Map Contoured at 3σ to Illustrate Binding of YW to the AnCE Active Site. The YW dipeptide is shown as green sticks in the Fo-Fc map, displayed as a grey mesh. The tryptophan and tyrosine residues adopt the P_2' and P_1' positions, respectively.

The YW dipeptide binds in such a way as to form very similar interactions with the active site residues as previously described for IW and VW. The C-terminus is held in place via conserved interactions between the P_2' tryptophan carboxylate group and the side chains of Gln265, Lys495 and Tyr504. The P_1' carbonyl group also contacts the side chains of His337 and His497.

Based on the interpretation of the active site difference density, the P_1' tyrosine nitrogen is not close enough to the active site zinc ion to interact with it directly, the interaction is instead mediated by a water molecule. This nitrogen does directly contact the protein backbone at Ala338 and also appears to interact with the side chain of Glu368, which has a proposed role in the catalytic mechanism (Zhang *et al.,* 2013). The interactions between YW and the active site are illustrated in Figure 5.9 and detailed in Table 5.5.

YW atom	Interacting atom	Distance (Å)
P_2' Trp OXT	Water	2.76
P_2' Trp O	Lys495 NZ	2.89
P_2' Trp O	Gln265 NE2	3.12
P_2' Trp O	Tyr504 OH	2.59
P_2' Trp NE1	Gln266 OE1	3.23
P_1' Tyr O	His337NE2	2.65
P_1' Tyr O	His497 NE2	2.95
P_1' Tyr N	Water	2.26
P_1' Tyr N	Glu368 OE2	2.71
P_1' Tyr N	Ala338 O	3.04

Table 5.5: List of Interactions Formed by the YW Dipeptide in the AnCE Active Site.

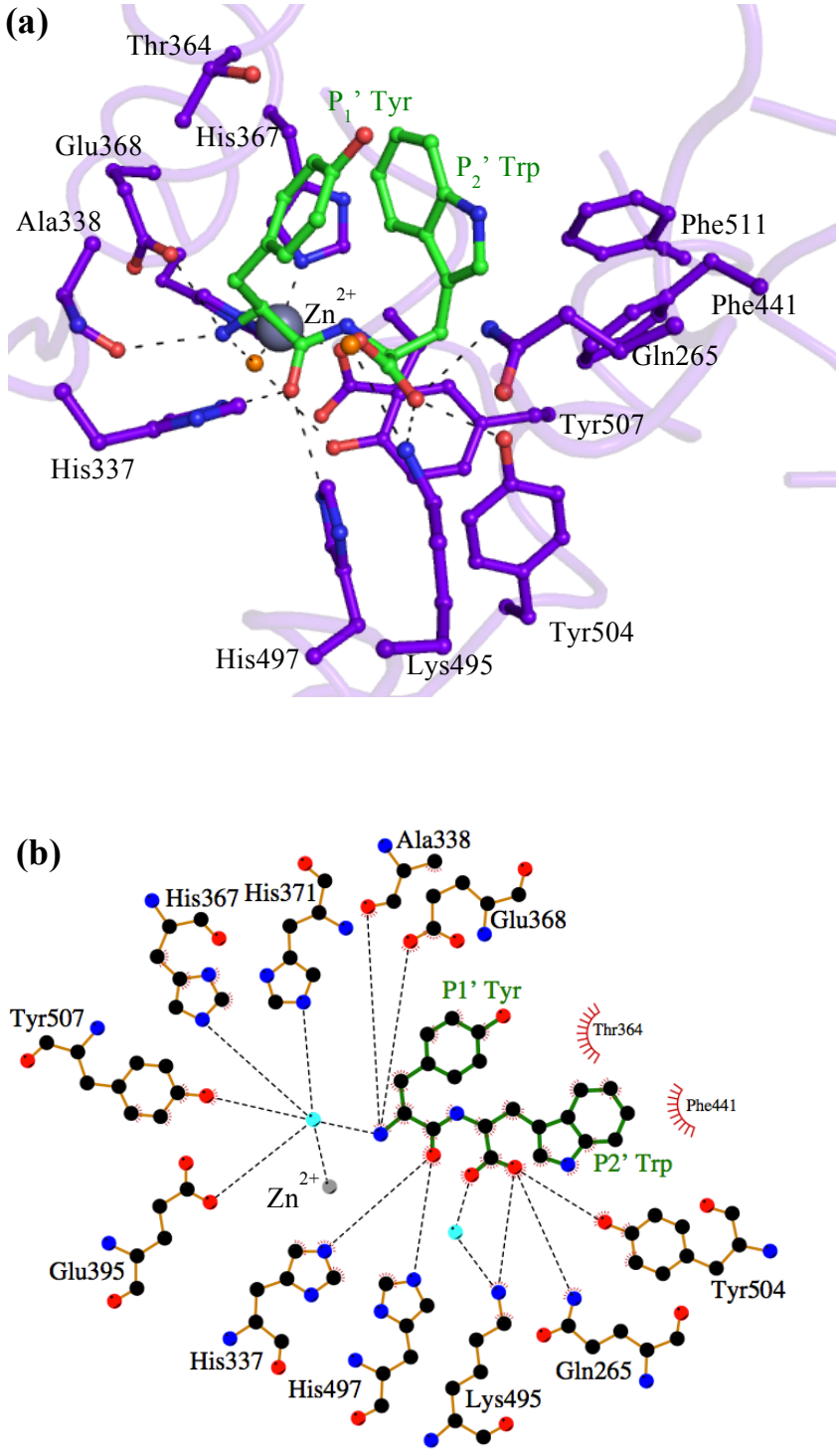


Figure 5.9: Binding of the YW Dipeptide to the AnCE Active Site. (a) The dipeptide is shown as green sticks with the tryptophan and tyrosine residues occupying the S_2' and S_1' subsites, respectively. The catalytic zinc ion is shown as a grey sphere and key water molecules involved in binding as orange spheres. Interacting AnCE residues are shown as purple sticks and this includes those that contribute to the hydrophobic nature of the binding site, such as Thr364, Phe441 and Phe551. (b) Schematic representation of dipeptide binding. The colour scheme is as described for (a), but with water molecules shown in turquoise and selected hydrophobic residues as red crescents.

5.3.4.4 AnCE and WR

900 diffraction images were collected for AnCE with WR, all of which were used to process the data to 1.71 Å resolution. Data processing statistics are detailed in Table 5.2.

Following initial cycles of refinement, some difference density remained in the active site. It was immediately clear however that this was far less extensive than that seen for AnCE with IW, VW and YW (Figures 5.3, 5.5 and 5.8) and that it could not accommodate the WR dipeptide.

The conclusion was hence drawn that the WR dipeptide had not bound to AnCE. A citrate ion from the crystallisation condition was placed in the active site density, Figure 5.10, as is seen in the native AnCE structure, PDB code 2X8Y (Akif *et al.,* 2010) when crystallised under the same conditions.

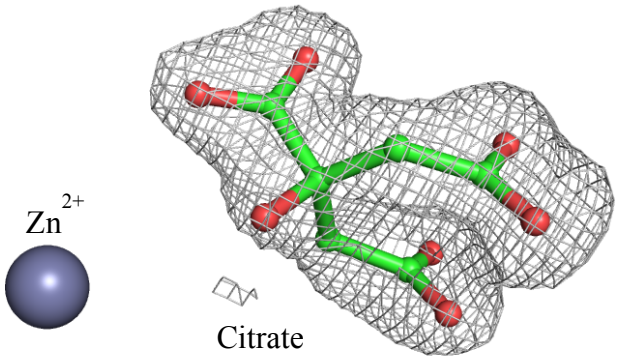


Figure 5.10: Fo-Fc Density Map Contoured at 3σ, Illustrating Binding of a Citrate Ion to the AnCE Active Site Following Co-crystallisation with the Dipeptide WR. The citrate ion is shown as green sticks in the difference density, which is displayed as a grey mesh. The active site zinc ion is shown as a grey sphere for reference. The citrate ion fits the observed density well and there is quite clearly insufficient difference density to accommodate the WR dipeptide.

5.4 Discussion

5.4.1 General Features of Dipeptide Binding to AnCE

The results obtained here illustrate that dipeptides with a C-terminal tryptophan residue function as competitive inhibitors of ACE by binding to the enzyme's active site. This is in accordance with kinetic data previously published by Lunow *et al*., 2015.

The three dipeptides with a C-terminal tryptophan residue that were studied (IW, VW and YW) all bind to the AnCE active site in the same location, such that the tryptophan residue occupies the S_2' subsite and the first residue the S_1' subsite. Binding in this

manner, the C-terminal carboxylate group is able to form interactions with conserved residues Gln265, Lys495 and Tyr504. Additionally, the carbonyl oxygen of the P_1' residue contacts the side chains of His337 and His497. These interactions are all commonly seen in peptide binding to angiotensin converting enzymes, as detailed in Chapters 3 and 4, and would be conserved in both the C- and N-domains of sACE.

Adopting this position also means that the indole nitrogen of the C-terminal tryptophan residue forms an interaction with Gln266, which is conserved in the structures of AnCE with IW, VW and YW. The importance of this interaction to dipeptide binding is highlighted by the fact that the Gln266 side chain in the VW structure exists in a different conformation to that in the structures with YW and IW. This is so that it can maintain this interaction with the tryptophan side chain, which is in a slightly different orientation, as illustrated in Figure 5.11.

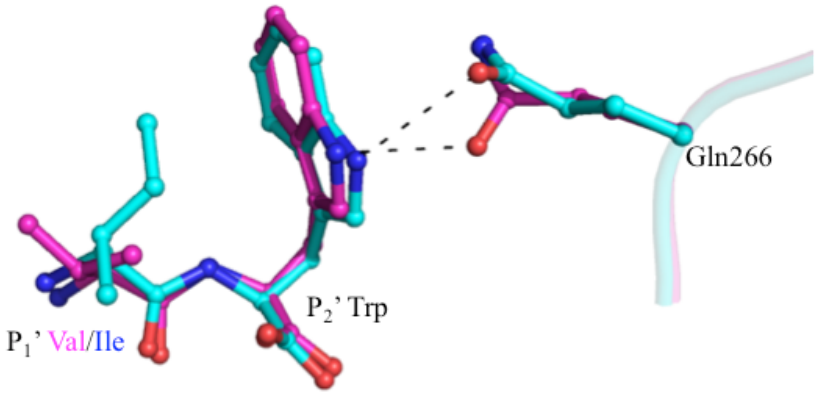


Figure 5.11: Interaction Between the Side Chain of Gln266 and the P_2' Dipeptide Tryptophan Residue. The Gln266 oxygen contacts the tryptophan indole nitrogen in the structures of AnCE with IW, VW and YW. The significance of this interaction is highlighted by the movement of the Gln266 side chain to accommodate the small variations in position of the tryptophan in the three structures. Here Gln266 is shown as cyan sticks interacting with the dipeptide IW, cyan sticks, and as magenta sticks, in a slightly different position, in order to interaction with VW, also magenta sticks.

Interestingly, Gln266 is conserved in neither the C-domain nor the N-domain, where Thr282 and Ser260 replace it, respectively. Neither threonine nor serine would be able to replicate this direct interaction with the tryptophan residue, although it is possible that a water-mediated interaction could occur.

Whilst this means that this interaction is not relevant in terms of understanding the domain selectivity displayed by these dipeptides, it is a good example of how subtle

changes in active site residues can impact on substrate and inhibitor binding. It would be interesting to investigate the effect of this interaction further by mutating the C-domain Thr282 to glutamine and observing the effect that this has on inhibition. Small changes such as this are also important when considering the potential for designing inhibitors to specifically target insect ACE homologues, as discussed in section 1.10.

5.4.2 Role of the C-terminal Tryptophan

A C-terminal tryptophan moiety has previously been identified as an important determinant of selective inhibition of the C-domain, and this has been exploited in the development of inhibitors such as RXPA380 and kAW (Georgiadis *et al.,* 2004; Redelinghuys *et al.,* 2006;).

Dipeptides with a C-terminal tryptophan moiety have also been shown to be potent and selective inhibitors of the C-domain of ACE (Lunow *et al.,* 2013). Much of the hydrophobic nature of the C-domain is conserved in AnCE; thus, the structural data obtained here can be used to understand the basis of this selectivity.

Comparison of the structures of AnCE with IW, VW and YW with those of AnCE with RXPA380 and the C-domain with kAW indicates that the C-terminal tryptophan moiety is found at the same location in all of the crystal structures, Figure 5.12.

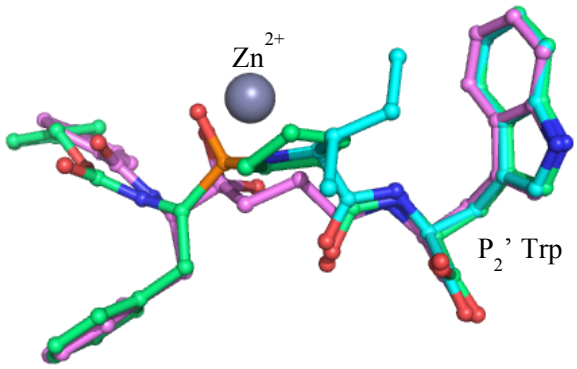


Figure 5.12: Comparison of Binding of the C-domain Selective Inhibitors IW, kAW and RXPA380 to ACE. The IW dipeptide from the structure with AnCE presented here is shown as cyan sticks with the zinc ion as a grey sphere for reference. Superposed onto this is RXPA380 from the structure with AnCE (PDB code 2X96 (Akif *et al.,* 2010), which is shown as green sticks, and kAW from the structure with the C-domain (PDB code 3BKL (Watermeyer *et al.,* 2008)), which is shown as pink sticks. The C-terminal P_2' tryptophan moiety is observed to bind at the same location in all three structures, indicating a conserved mode of binding.

This indicates that this is a conserved mode of binding and that the same factors that cause RXPA380 and kAW to be selective inhibitors of the C-domain are also responsible for this property of IW, VW and YW. The hydrophobic effect appears to be the principal driving force for tryptophan binding at this position, with AnCE residues Phe441, Phe511 and Phe363 contributing to a hydrophobic "pocket" for the tryptophan side chain.

Phe441 and Phe511 are both conserved in the N- and C-domains of ACE. Previous work has implicated two valine residues, Val379 and Val380, in the C-domain as being key determinants of selectivity. These are not conserved in the N-domain, being replaced by serine and threonine respectively.

In AnCE Val380 is, as in the N-domain, replaced by a threonine residue. Val379 is also not conserved, but the hydrophobic nature at this position is maintained because it is replaced by Phe363. This, combined with the structural data, suggests that Val379 perhaps has a more significant role than Val380 in tryptophan binding, because the presence of a threonine at the equivalent position to that of Val380 does not appear to have negatively impacted IW, VW or YW binding to AnCE.

It is important to take into consideration however that no kinetic data have yet been obtained for AnCE inhibition by these dipeptides. Although structural data indicate that IW, VW and YW are likely to be good AnCE inhibitors it is impossible to determine this from the structure alone. Furthermore, an additional interaction between the tryptophan side chain and Gln266 has been identified in AnCE that is likely to enhance binding, whilst the substitution of Val379 by Phe363 may increase the influence of hydrophobic interactions. These factors may well be sufficient to counter the negative effect of the Val380 threonine substitution, which may therefore have a much greater negative influence on binding to the N-domain.

5.4.3 Importance of the P_1' Residue

5.4.3.1 Role of an Aliphatic P_1' Residue in C-domain Inhibition

The data presented by Lunow *et al.,* 2015 indicate that whilst the C-terminal residue appears to be the key determinant of dipeptide domain selectivity, the P_1' residue does also have a role in this. An aliphatic residue at the P_1' position was observed to result in more effective inhibition of the C-domain, with IW and VW being more potent inhibitors than YW. The data presented here help to explain the role of the residue at this position.

Two C-domain valine residues that are not conserved in the N-domain have been established as key determinants of C-domain selectivity. One of these residues, Val380, is located in very close proximity to the position adopted by the P_1' residues of the dipeptides in the structures that are presented here.

In the N-domain, a threonine residue replaces Val380, and this is also the case in AnCE (Thr364), thus making this protein an ideal model for studying the role of this residue.

As illustrated in Figure 5.13, the P_1' isoleucine residue in the IW structure comes to within ~4 Å of Thr364. Assuming that the dipeptides bind in a similar position in the C-domain active site, then the hydrophobic interaction between Val380 and isoleucine or valine of the dipeptide is likely to be an important stabilising force for binding. The loss of this in the N-domain, combined with the previously outlined changes affecting the S_2' pocket, further illustrates why these dipeptides are much more effective inhibitors of the C-domain.

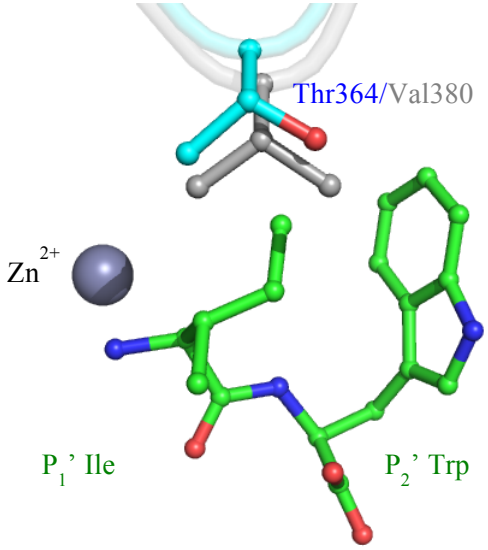


Figure 5.13: The Role of the P_1' Residue in Dipeptide Binding. The IW dipeptide from the structure with AnCE is shown as green sticks and the active site zinc ion as a grey sphere for reference. Val380 from the C-domain is shown as grey sticks. This residue is within 4 Å of the P_1' isoleucine residue, explaining the preference for an aliphatic P_1' residue for C-domain binding. In AnCE and the N-domain Val380 is replaced by threonine and AnCE Thr364 is shown as cyan sticks. Hence, the stabilising interaction present in the C-domain is unlikely to be conserved.

5.4.3.2 Effect of P_1' Tyrosine on YW Binding and Inhibition

The dipeptide YW is much less selective for the C-domain than either IW or VW. It is both a weaker inhibitor of the C-domain and a much stronger inhibitor of the N-domain than these two dipeptides (Lunow *et al.,* 2013).

Analysis of the structure of AnCE with YW indicates that the hydroxyl group of the P_1' tyrosine is ~3 Å from the side chain of Asp360, hence, there is the potential for a hydrogen bond interaction between these two residues, stabilising YW binding at this position. This is supported by the observation that the side chain of Asp360 exists in a different conformation in the structure with YW compared to with IW and VW, in order to facilitate this interaction, Figure 5.14.

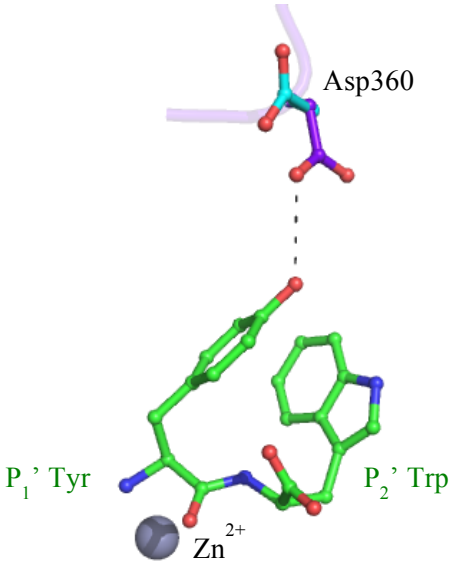


Figure 5.14: Interaction Between YW P_1' Tyrosine and AnCE Residue Asp360. The YW dipeptide from the structure with AnCE is shown as green sticks and the catalytic zinc ion as a grey sphere. The Asp360 side chain from the YW structure, purple sticks, adopts a different conformation to the equivalent residue in the structure with IW, cyan sticks, in order to form an interaction with the tyrosine side chain.

Asp360 is replaced by Glu376 in the C-domain. This means that the charge at this position is conserved, but the longer Glu376 side chain is likely to clash with the tyrosine residue, potentially explaining the decreased potency of YW as a C-domain inhibitor.

In contrast, Asp360 is conserved in the N-domain as Asp354, meaning that this interaction could be replicated, helping to explain why YW is a much better inhibitor of the N-domain than either IW or AW.

These observations indicate that, in addition to the well-characterised role of the P_2' moiety, the interactions formed by the P_1' inhibitor residue can also have an important role in binding to the active site and hence the potency of the inhibitor. This is an area that should be explored further, to investigate how these interactions might be further exploited in inhibitor design.

5.4.4 The Structure of AnCE with VW

The difference density found in the catalytic channels of both the YW and IW structures was very clear, particularly for IW, and placement of the dipeptides was quite straightforward. This was not the case for VW, where more extensive density was seen.

Initially the VW dipeptide was placed in the same position as YW and IW had been, meaning that all of the conserved interactions made by the C-terminus could be formed. These interactions, combined with the fact that the density for the valine side chain at P_1' is very well defined (Figure 5.5), mean that there is very little uncertainty surrounding this placement.

Positioned in this way, the peptide backbone at the P_1' position has moved by ~ 1 Å compared to the YW and IW structures. This brings the valine side chain very close to Glu368, conserved in both the N- and C-domains. Consequently, the side chain of this residue has rotated relative to the YW and IW structures, presumably to minimise any contacts, as illustrated in Figure 5.15. This also means that the VW dipeptide is able to directly contact the catalytic zinc ion, whereas YW and IW do so through a water molecule.

It is not immediately apparent by studying the structures why the VW dipeptide binds in a slightly different manner but it is clear from the difference density that this is the position that it is adopting. It might be the case that this is not the case in the C-domain, where the valine residue may reside closer to Val380, replaced by Thr364 in AnCE.

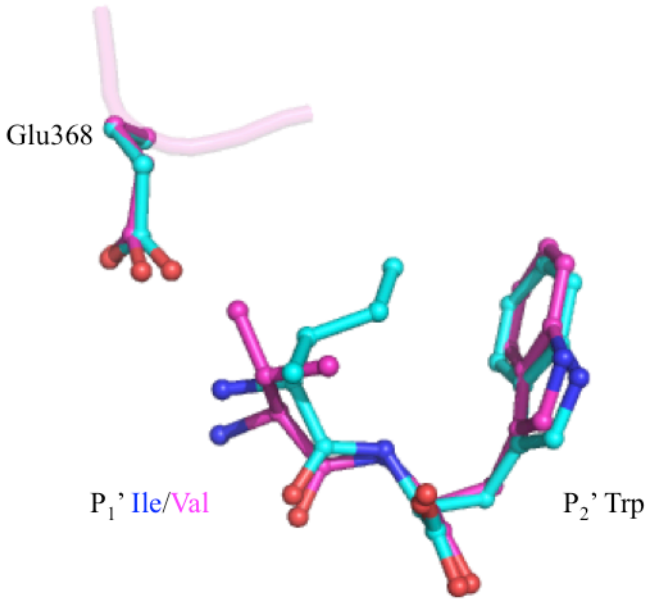


Figure 5.15: Positions of the Dipeptides VW and IW in the AnCE Active Site. The VW dipeptide is shown as magenta sticks and the IW dipeptide as cyan sticks. The P_2' tryptophan residues are found at the same position in both structures, but the peptide backbone at the P_1' position has moved by approximately 1 Å between the two structures. The side chain of Glu368 has also moved slightly between the two structures in order to avoid a clash with the valine side chain of VW. The Glu368 side chain from the structures with VW and IW are shown as cyan and magenta sticks, respectively.

What is perhaps more surprising than this movement is the residual density beyond the P_1' valine. This is a significant portion of density, indicating that it is likely to be due in someway to the dipeptide. It also appears to be continuous with the VW portion of the density, with no visible break.

Based on these observations, attempts were made to place the dipeptide in a second position. In Chapter 4, two of the datasets collected for AnCE with the inhibitor tynorphin (VVYPW) appear to show the PW dipeptide fragment bound at the P_2 and P_3 positions. Given the similarity between PW and VW, an attempt was made to place VW at the position occupied by PW in the tynorphin structures. However analysis of the difference density seen in this structure is not indicative of the dipeptide binding at the same position. The density is located much closer to the catalytic zinc ion and in fact appears continuous with the C-terminal portion.

Ultimately, the VW dipeptide was placed in the active site density in a second orientation. The valine remained in the S_1' position and the tryptophan was now occupying the S_1 site. The different density adequately accommodates this arrangement, as shown in Figure 5.7.

Binding in this manner, the VW dipeptide might co-ordinate the catalytic zinc ion through its carboxylate group. The tryptophan moiety may then be in a similar position to the phenylalanine moieties of ACE inhibitors captopril, lisinopril, enalprilat and RXPA380, stabilised by hydrophobic interactions with Trp341 and Val502. Although this would provide an explanation for the density seen in the active site, it is not clear why VW would bind in this manner, but not IW, which is so very similar. It would therefore be prudent to collect more data for this complex before drawing a firm conclusion.

5.4.5 AnCE and WR

In spite of the ambiguity surrounding the VW structure, it remains clear that the dipeptides IW, VW and YW are all present, in some manner, in the AnCE structures. This does not appear to be the case for WR. This is not entirely surprising, as WR is the only one of the dipeptides without a C-terminal tryptophan residue, and, analysis of the structures of IW, VW and YW has indicated that the bulky tryptophan side chain has a critical role in dipeptide binding to AnCE through the hydrophobic effect.

The data presented by Lunow *et al.,* 2015 indicate that WR is a very poor inhibitor of the C-domain, and the absence of WR in the AnCE structure strongly suggests that this is due to it having a very poor affinity for the active site. Considering the structures of the other dipeptides bound to AnCE provides some insight into the reasons for this.

Interactions between the C-terminus of peptides and the S_2' subsite have been shown to be important conserved features of binding. It is therefore highly likely that WR would bind to the ACE active site with the C-terminal arginine residue at the P_2' position and the tryptophan residue at P_1'. This would result in the large, positively charged arginine side chain projecting into a very hydrophobic pocket. Specifically, the arginine side chain is likely to be within 4 Å of two conserved phenylalanine residues: Phe441 and Phe511 (using AnCE numbering).

In the AnCE active site, the presence of the large, hydrophobic, Phe363 residue is likely to further impede binding. In the C-domain the valine residue at this position is also likely to have a similar, although possibly less pronounced effect.

WR is a slightly better inhibitor of the N-domain than the C-domain. This is likely a result of Phe363 being replaced by Ser357, which would help to reduce the

hydrophobicity of the S_2' subsite somewhat. However the fact that WR is not a very potent N-domain inhibitor indicates that this region of the active site remains sufficiently hydrophobic such that insertion of a large polar side chain is not entropically favourable. This means that whereas addition of a hydrophobic moiety at the P_2' position of potential inhibitors can help to enhance C-domain inhibition, large polar P_2' groups are unlikely to be a feature of successful N-domain inhibitors.

5.5 Conclusions

The data presented here provide, for the first time, an insight into the structural basis of ACE inhibition by dipeptides. The tryptophan containing dipeptides IW, VW and YW are all confirmed to be competitive inhibitors that bind to the active site of the enzyme.

Key binding features include interactions formed by the C-terminal carboxylate group with Gln265, Lys495 and Tyr504, which are conserved features of peptide binding to angiotensin converting enzymes. Only one of the dipeptides, VW, appears to directly interact with the catalytic zinc ion, indicating that this is not a pre-requisite.

A C-terminal tryptophan moiety has been shown to result in selective inhibition of the C-domain by these dipeptides. The data obtained here indicate that this is likely to be due to the increased hydrophobicity of the S_2' pocket of the C-domain compared to the N-domain, with the dipeptide tryptophan residues adopting the same positions on binding as the C-terminal tryptophan moieties of the inhibitors RXPA380 and kAW. Furthermore, an additional interaction with AnCE Gln266 was identified, that is conserved neither in the N- nor the C-domain. This highlights the importance of small substitutions in the active site pockets.

In addition to the role of the P_2' position, the importance of the P_1' residue has also been identified. An aliphatic residue at P_1' improves C-domain inhibition and the structural data here indicate that this is due to the presence of a valine residue that is not conserved in the N-domain.

An interaction has also been identified between a P_1' tyrosine and AnCE residue Glu368. This is conserved in the N-domain, but not in the C-domain. This feature should be the subject of further study, as it may be possible to exploit it in the design of N-domain specific inhibitors.

Chapter 6: A New Structure of AnCE in a Different Crystal Form

6.1 Introduction

6.1.1 Existing AnCE Crystal Structures

AnCE, an angiotensin converting enzyme homologue from *D. melanogaster*, has been used extensively as a model to study substrate and inhibitor binding to ACE. This includes the work presented in Chapters 3, 4 and 5 of this thesis.

The crystal structure of AnCE was first reported by Kim *et al.,* in 2003 (Kim *et al.,* 2003). This structure, referred to as the "form I" structure was determined at 2.40 Å resolution using crystals in the $P2_1$ space group. In 2010 a second, higher resolution (1.90 Å), AnCE structure was reported. This is referred to as the "form II" structure and was obtained from crystals in the *h3* space group (Akif *et al.,* 2010). It is the form II structure that has been used in all of the ensuing complex structures (Akif *et al.,* 2010; Akif *et al.,* 2011; Akif *et al.,* 2012; Masuyer *et al.,* 2014) and in the work presented in Chapters 3, 4 and 5.

Crystals for the form II structure are typically grown using the hanging drop vapour diffusion technique in a condition that contains 1.2-1.3 M sodium citrate, 0.1 M HEPES, pH 7.5. The condition gives rise to reproducible, diffraction quality crystals that consistently appear within 7-10 days of incubation at 21 ^{o}C. A comprehensive description of the AnCE crystallisation process, including pictures of crystals and an example diffraction image is given in section 2.7.4.

6.1.2 Requirement for a New AnCE Crystallisation Condition

The form II crystal structure has been used with great success to study binding of a wide range of inhibitors to AnCE. However it has recently become apparent that crystals grown using this condition may not always be ideal.

All of the ACE inhibitors studied thus far have been competitive inhibitors, which form strong interactions with the active site, frequently through co-ordination of the catalytic zinc ion. As part of the process of developing a new generation of inhibitors, the need for which is discussed in some detail in section 1.5.1, attention has moved towards novel compounds, which bear very little resemblance to current inhibitors.

Flavonoids are one example of this. These are low molecular weight polyphenol compounds isolated from plant extracts, which have been observed to display some ACE inhibitory activity (Guerrero *et al*., 2012; Loizzo *et al.,* 2007; Nyman *et al.,* 1998;

Park *et al.,* 2003). These are particularly interesting, because their presence in plant extracts means that they are bioavailable through diet.

Luteolin, Figure 6.1, is the most potent flavonoid ACE inhibitor identified thus far, with an IC_{50} value of 23 μM reported (Guerrero *et al.,* 2012). Attempts to co-crystallise AnCE with luteolin in order to understand more about its inhibitory actions have so far proved unsuccessful. In each case crystals appear in the standard condition but the inhibitor is not present in the active site.

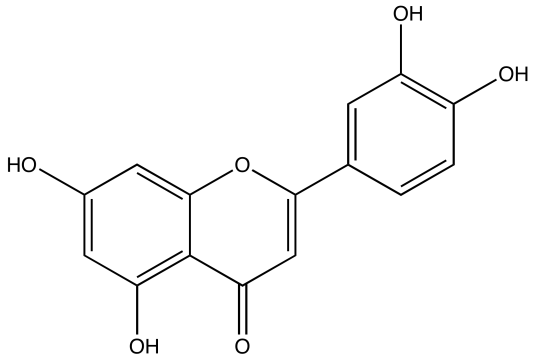


Figure 6.1: Chemical Structure of the Flavonoid Luteolin, which Inhibits ACE with an IC_{50} of 23 μM (Guerrero *et al.,* 2012).

The condition used for AnCE crystallisation contains a high concentration of sodium citrate (1.2-1.3 M) and in the absence of luteolin, or any other inhibitor, a citrate ion is seen co-ordinating the catalytic zinc ion. It is possible that the affinity of luteolin, and other flavonoids, for the AnCE active site is insufficient to displace the citrate ion, hence why co-crystals have not formed. It would therefore be beneficial to obtain AnCE crystals from a condition that does not contain citrate, as this may be more suitable for co-crystallisation with new, weaker, inhibitors.

In addition to this, as discussed in Chapter 3, it has not so far been possible to obtain crystals for AnCE with some of the BPPs using the current condition. The hypothesis is that binding of these peptides is causing a movement of secondary structure elements, similar to that seen in the structure of tACE with BPPb (Masuyer *et al.,* 2012), and that this is having an effect on the crystal contacts that form between the protein molecules. In any new crystallisation condition that is identified it is probable that different crystal contacts would form and that these may not be disrupted by BPP binding.

6.1.3 Aims

The aim of the work presented in this chapter is to identify a new condition in which AnCE can crystallise and to use these crystals to determine the structure of AnCE in a new crystal form. Ultimately, the aim is then to use this new condition to obtain crystals of AnCE with new inhibitors that have not co-crystallised in the standard condition. By studying the mechanisms of binding of these compounds it may then be possible to learn more about how different features of the ACE active site could be exploited in inhibitor design.

6.2 Methods

6.2.1 AnCE Expression and Purification

Recombinant AnCE was expressed by the methylotropic yeast *P. pastoris* and purified using a two step protocol consisting of HIC followed by SEC, as described in section 2.7. Purity of the protein was confirmed by SDS PAGE analysis.

6.2.2 AnCE Crystallisation

Following purification AnCE was dialysed to completeness into 5 mM HEPES pH 7.5, 0.1 mM PMSF. Dialysis was performed at 4 ^{o}C using 12-14 kDA MWCO visking tubing (Medicell Membranes Ltd, London). The dialysed sample was then concentrated to 15 mg ml^{-1} by centrifugation at 3000 g using 30 kDa molecular weight cut-off Amicon Ultra-15 Centrifugal Filter Units (Merck Millipore) and a Heraeus centrifuge (Thermo Fisher Scientific).

Crystallisation experiments were performed using a Phoenix crystallisation robot (Art Robbins Instruments). Experiments were prepared in Intelli-plate 96 crystallisation plates (Hampton Research) using the sitting drop vapour diffusion method (described in section 2.6.2.3). 50 µl of the appropriate reservoir solutions were transferred, using the crystallisation robot, into each of the 96 wells of the crystallisation plates. 300 nl sitting drops were then assembled containing a 1:1 ratio of AnCE:reservoir solution. Plates were sealed using ClearVue sheets (Molecular Dimensions Ltd., Newmarket, UK) and incubated at 21 ^{o}C. All crystallisation experiments were observed regularly for any signs of crystal formation.

In excess of one thousand different crystallisation conditions were screened using the following commercially available, pre-prepared screens: Clear Strategy™ Screen I HT-96, Clear Strategy™ Screen II HT-96, Heavy + Light HT-96, JCSG-*plus*™ HT-96, MemGold™ HT-96, MIDAS™ HT-96, Morpheus® HT-96, PACT *premier*™ HT-96,

The PGA Screen™ HT-96, ProPlex™ HT-96, Structure Screen 1 + 2 HT-96 and Stura FootPrint Combination HT-96 (all purchased from Molecular Dimensions Ltd., Newmarket, UK).

6.2.3 Crystal Harvesting and X-ray Diffraction Data Collection

Crystals grown in a drop containing 0.1 M MMT (Malic acid, MES monohydrate, Tris) buffer pH 4.0, 25 % w/v PEG 1500 were mounted directly from the drop into litholoops (Molecular Dimensions Ltd., Newmarket, UK) and flash frozen in liquid nitrogen for transportation to the synchrotron.

All X-ray diffraction data were collected at 100 K on beamline I24 at DLS and captured using a Pilatus3 6M detector.

6.2.4 Diffraction Data Processing

Raw diffraction images were processed using MOSFLM (Battye *et al*., 2011; Leslie and Powell, 2007) prior to scaling using Aimless Evans, 2006; Evans, 2007), part of the CCP4 suite (Winn *et al*., 2011). Initial phases were calculated by molecular replacement using Phaser (McCoy *et al.,* 2007) with the form II AnCE structure, PDB code 2X8Y (Akif *et al.,* 2010) as the search model.

The resulting model was subjected to an initial round of rigid body refinement using REFMAC (Murshudov *et al.,* 1997; Murshudov *et al.,* 2011). Following this, iterative rounds of model building and refinement were performed using *C*oot (Emsley *et al.,* 2011) and REFMAC, respectively, with a subset of 5 % of reflections separated for use as the R_{free} set for cross-validation.

Model building included adding water molecules to the structure where Fo-Fc peaks in electron density exceeded 3σ and where potential hydrogen bonds could be formed. Ligands were added to the active site as appropriate based on interpretation of the difference density.

Validation of the final model was performed using Molprobity (Chen *et al.,* 2010). LigPlot^{+} (Laskowski and Swindells, 2011) was used to study interactions of ligands with the AnCE active site and PDBePISA (Krissinel and Henrick, 2007; Krissinel, 2010) to probe crystal contacts between molecules. The molecular graphics program PyMol (The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC) was used to produce all of the structural figures.

6.3 Results and Discussion

6.3.1 AnCE Crystallisation

Following incubation at 21 ^{o}C for approximately 3 months, crystals were observed in a condition containing 0.1 MMT buffer pH 4.0 and 25 % w/v PEG 1500. An example of these crystals is shown in Figure 6.2.

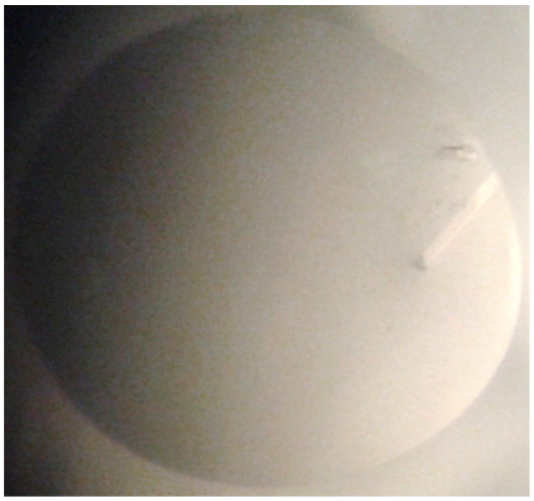


Figure 6.2: Crystals of AnCE Grown in a Condition Containing 0.1 M MMT Buffer pH 4.0 and 25 % w/v PEG 1500. Crystals are shown in a 300 nl drop.

6.3.2 X-ray Diffraction Data Collection

Crystals grown in the "new" condition were mounted directly into litholoops for data collection at DLS. A total of 250 diffraction images were collected to 1.85 Å resolution, an example of a diffraction image is shown in Figure 6.3.

Although diffraction experiments were performed at 100 K, no cryoprotecting solutions were used. This is because the crystallisation condition contains 25 % w/v PEG and PEGs have been reported to act as a cryoprotectant in some cases (Berejnov *et al.,* 2006).

The crystal was largely stable during data collection, however there was some evidence of radiation damage occurring towards the end of the experiment. The quality of later diffraction images was visibly worse than earlier images, with a clear reduction in resolution. For this reason, only images 1-200 (inclusive) were used in data processing.

There is also some evidence of ice rings on the diffraction images at approximately 3.7 Å, 2.2 Å and 1.9 Å (Chapman and Somasundaram, 2010; Dowell and Rinfret, 1960) (Figure 6.3). The presence of these ice rings did not cause any difficulty during data processing.

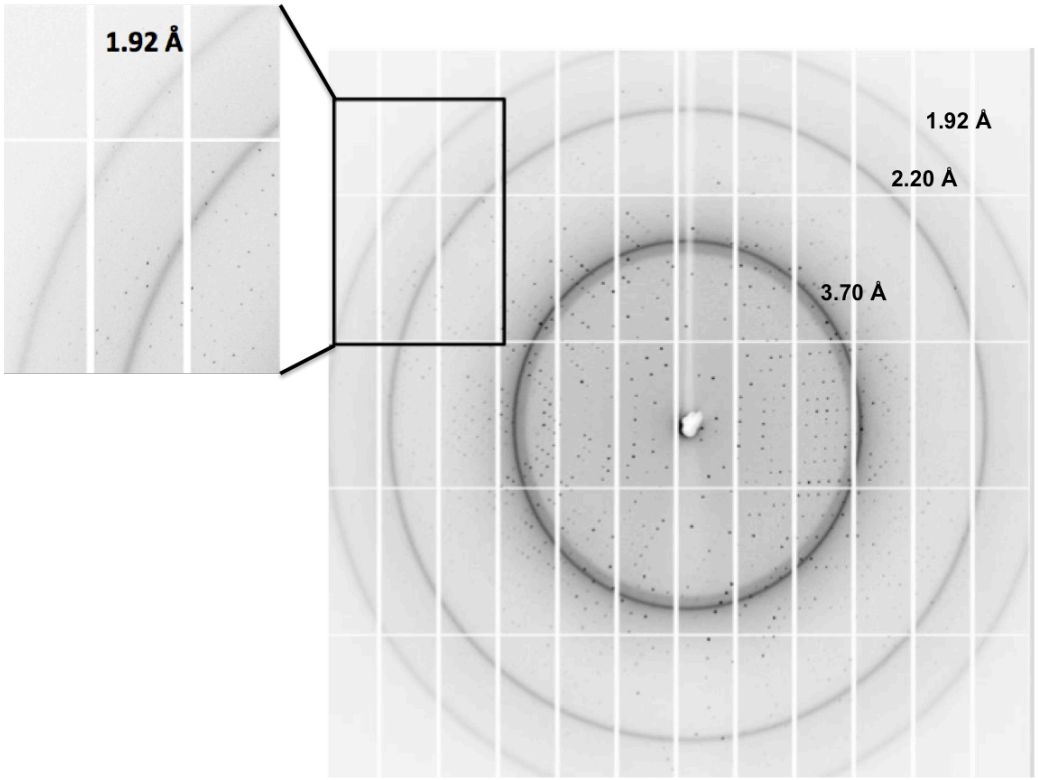


Figure 6.3: X-ray Diffraction Image Collected from an AnCE Crystal Grown in the New Condition. Ice rings are visible 3.7 Å, 2.2 Å and 1.9 Å. Diffraction beyond 1.9 Å is highlighted in the inset figure.

6.3.3 Data Processing and Structure Solution

Integration and scaling of diffraction images 1-200 using MOSFLM and Aimless indicated that the crystals used for the data collection belonged to the space group $P2_12_12_1$. This is in contrast to both the form I and form II AnCE structures, which were determined using crystals from the $P2_1$ and *h3* space groups, respectively (Akif *et al.,* 2010; Kim *et al.,* 2003). There was one molecule per asymmetric unit in the crystal. Data processing statistics are detailed in Table 6.2.

Phases were calculated by molecular replacement performed using Phaser. The form II AnCE structure, PDB code 2X8Y (Akif *et al.,* 2010) was used as the search model. A single phaser solution was found and the molecular replacement statistics, detailed in Table 6.1, indicate that this is almost certainly the correct solution. In particular, the TFZ score (translation function z-score) of 11.3 is significantly higher than 8, which is normally indicative of a correct solution (Oeffner *et al.,* 2013).

Parameter	RFZ	TFZ	PAK	LLG
Value	6.4	11.3	0	15, 534

Table 6.1: Molecular Replacement Statistics for the Form III AnCE Structure. Statistics obtained following molecular replacement performed using Phaser with the form II AnCE structure (PDB code 2X8Y) as the search model. RFZ = rotation function z-score, TFZ = translation function Zzscore, PAK = packing clashes and LLG = log likelihood gain.

After molecular replacement, clear electron density was visible for AnCE residues 18-614 (20-614 in the form II structure), the highly conserved catalytic zinc ion (co-ordinated by His367, His371 and Glu395) and for N-linked glycans, which were added to the model at Asn53 and Asn311.

Following initial rounds of model building and refinement clear difference density was visible in the catalytic channel and this was filled by a malate ion from the crystallisation solution. Final rounds of refinement were then performed, the statistics from which are shown in Table 6.2. The R_{cryst} value (16.31 %) is acceptable for a structure of this resolution (1.85 Å) and indicates that the model corresponds well to the diffraction data. Similarly, the difference between the R_{cryst} and R_{free} values is 2.3 %,

which indicates that overfitting of the model has not occurred during refinement (section 2.6.3.4.1) (Kleywegt and Brünger, 1996). This model is now referred to as the "form III" AnCE structure.

Beamline at DLS	i24
Detector	Pilatus3 6M
Wavelength (Å)	1.000
Space group	$P2_12_12_1$
Unit cell parameters: **a, b, c (Å)** **α, β, γ (°)**	 86.12, 94.90, 99.03 90.00, 90.00, 90.00
Molecules per asymmetric unit	1
Resolution range (Å)	68.62-1.85
$^{\#}R_{merge}$ (%)	6.1 (58.8)
$^{+}R_{pim}$ (%)	5.0 (49.2)
Mean *I/σ(I)*	10.6 (2.2)
Completeness (%)	97.8 (89.5)
Number of reflections: **Total** **Unique**	 219, 797 68, 121
Multiplicity	3.2 (2.2)
Wilson *B* factor (Å^2)	25.0
Average *B* factor (Å^2) **Protein** **Tris** **Malate** **Zinc** **Glycosylated sugars** **Solvent**	 24.35 41.90 28.33 21.00 49.35 34.73
$^{\S}R_{cryst}$/$^{\#\#}R_{free}$ (%)	16.31/18.56
R.M.S deviation from ideal values **Bond lengths (Å)** **Bond angles (°)**	 0.007 1.280
Ramachandran plot statistics (%) **Favoured** **Disallowed**	 99.16 0.00
PDB code	5A2R

$^{\#}$ $R_{merge} = \Sigma_{hkl}\Sigma_i|\boldsymbol{I_i}\,(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I_i}(hkl)$, where $\boldsymbol{I}_i$(hkl) is the intensity of *i*th measurement and <***I****(hkl)*> is the average of symmetry-related observations of a unique reflection.

$^{+}R_{pim} = \Sigma_{hkl}\sqrt{\frac{1}{n-1}}\Sigma_{i=1}^{n}|\boldsymbol{I}_i(hkl) - \langle \boldsymbol{I}(hkl)\rangle|/\Sigma_{hkl}\Sigma_i \boldsymbol{I}_i(hkl)$

$^{\S}R_{cryst} = \Sigma_h|F_o - F_c|/\Sigma_h F_o$, where F_o and F_c are observed and calculated structure factor amplitudes of reflection *h*, respectively.

$^{\#\#}R_{free}$ is equal to R_{cryst} for a randomly selected 5 % subset of reflections.

Table 6.2: X-ray Diffraction Data Collection, Processing and Refinement Statistics for the Form III AnCE Crystal Structure. Outer shell values are shown in parentheses.

6.3.4 Analysis of the Form III Structure

6.3.4.1 Overall Structure

The overall fold of the form III AnCE structure is illustrated in Figure 6.4. It adopts the characteristic, predominantly helical fold found in structures of angiotensin converting enzymes, consisting of twenty-one α-helices, nine short 3_{10} helices and only two anti-parallel β-strands. Key ACE structural features are conserved, including the long, central substrate binding channel with the active site at its centre, and the lid that controls access to this, which is formed by helices α1, α2 and α3.

Visual inspection of the structure indicates that the overall topology is identical to that of both the form I and form II AnCE structures, with no gross movement of structural elements. This is confirmed by RMSD values of 0.27 Å and 0.33 Å for the form I and form II structures, respectively.

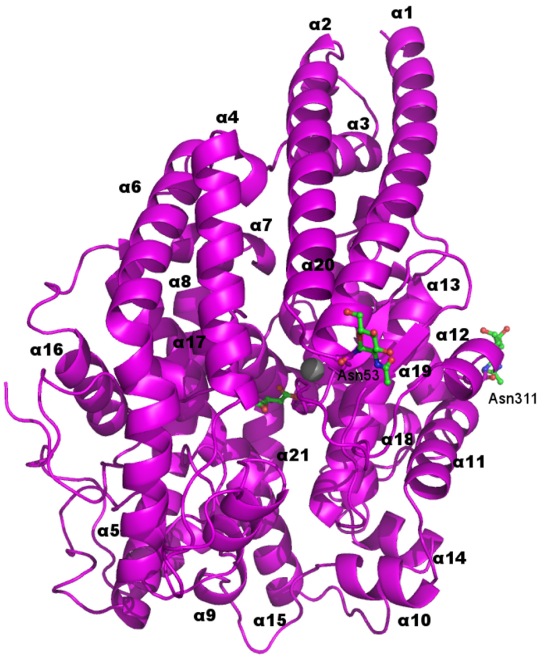


Figure 6.4: The Overall Appearance of the Form III AnCE Structure (PDB code 5A2R). Cartoon representation of secondary structure elements highlights the dominance of α-helices, which are labelled α1-21. There are only two short β-strands. The overall topology is identical to that of the form I and form II structures. Key features include the long substrate binding channel with the active site at its centre, highlighted by the location of the catalytic zinc ion, grey sphere, and a malate ion from the crystallisation condition, green sticks. This channel is capped by a lid formed by helices α1-3. N-linked glycosylation is visible at Asn53 and Asn311 and is displayed as green sticks.

Given that it is the form II structure that has been used in all of the inhibitor binding studies reported in the last five years, this will be the main focus of comparisons made during the analysis of the new AnCE structure.

6.3.4.2 The Active Site

Following initial rounds of refinement there remained a small portion of difference density in the active site. Based on the size and shape of the density and the components of the purification buffers and crystallisation solutions, a malate ion, presumably arising from the MMT buffer, was modelled into this density. As illustrated in Figure 6.5 the malate ion fits the density well and there was no residual density remaining following further refinement.

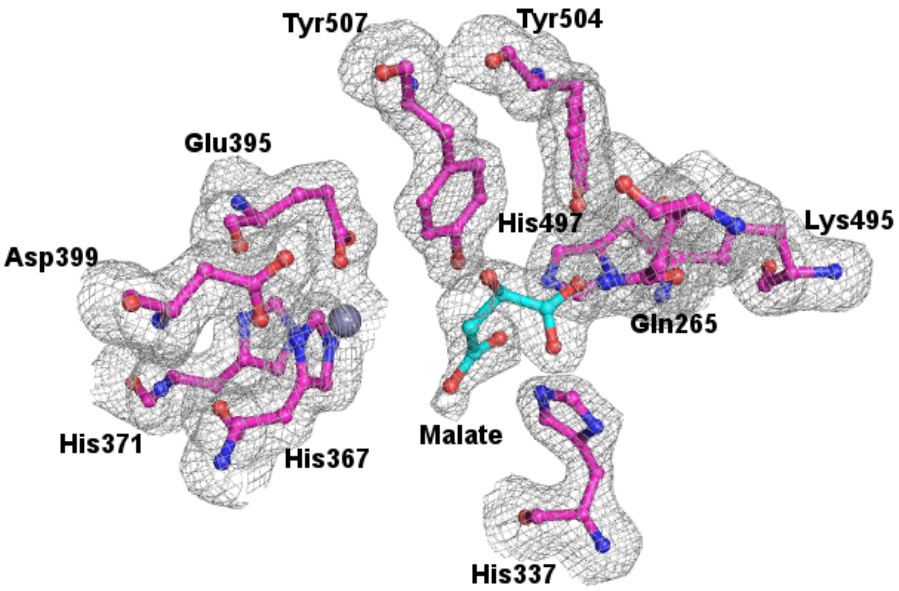


Figure 6.5: Electron Density Map for Malate Binding to the AnCE Active Site. Electron density map is shown as a grey mesh contoured at 1σ around key active site residues (pink sticks), the catalytic zinc ion (grey sphere) and the malate ion from the crystallisation condition (cyan sticks). The malate ion fits the observed density well and there is no residual density remaining.

The manner in which the malate ion binds to the active site is not dissimilar to that displayed by the C-terminal portion of the peptide and dipeptide inhibitors considered in Chapters 3, 4 and 5. One of the carboxylate groups interacts with the side chains of active site residues Gln265, Lys495 and Tyr504, whilst the other contacts the nitrogen atoms of His337 and His497. This is illustrated in Figures 6.5 and 6.6.

6.3.4.2.1 Comparison with the Form II Active Site

As outlined in section 6.1.2, in the form II AnCE structure, in the absence of a ligand, a citrate ion is observed to co-ordinate the catalytic zinc ion. This has been identified as a problem in terms of studying the binding of new, weaker inhibitors, which are unable to displace this ion.

Analysis of the form III structure indicates that the malate ion binds to the AnCE active site in a very similar manner to the citrate ion in the form II structure. Consequently, there is no observable movement of active site residues.

Both ions interact with Gln265, Lys495 and Tyr504 through a carboxylate group. However whereas the second malate carboxylate contacts His337 and His497, the two remaining citrate carboxylate groups partake in water-mediated interactions with Tyr507, Asp399 and Glu395, the latter of which is also involved in zinc ion co-ordination.

Crucially, the citrate ion is also observed to interact with the catalytic zinc ion via a water molecule. Whilst this water molecule is conserved in the new AnCE structure, the position of the malate ion indicates that it would be unable to replicate this interaction. This suggests that the malate ion is not held so strongly in the active site and hence might be more easily displaced by weaker inhibitors. A summary of these interactions in both the form II and form III structures is presented in Figure 6.6 and summarised in Tables 6.3.

Citrate atom	**Interacting atoms**	**Distance (Å)**
OA1	Water	2.61
OA2	Y504 OH	2.87
OA2	Q265 NE2	3.28
OA2	K495 NZ	3.22
OHB	Water	2.86
OG2	Water	3.20
OB1	Water	2.96
Malate atom	**Interacting atoms**	**Distance (Å)**
O1	Y504 OH	2.52
O1	Q265 NE2	3.02
O1	K495 NZ	2.70
O2	Water	2.79
O4	H497 NE2	2.95
O4	H337 NE2	2.67

Table 6.3: List of Interactions Formed by the Citrate and Malate Ions in the Active Sites of the Form II and III AnCE Structures.

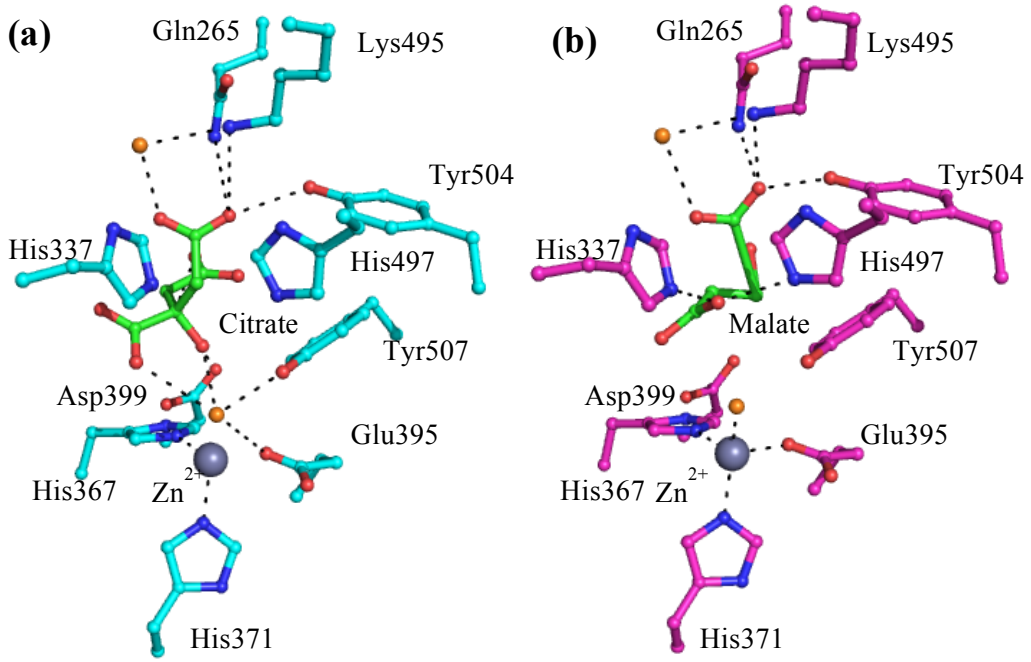


Figure 6.6: Binding of Citrate and Malate Ions to the Form II (PDB code 2X8Y) and Form III AnCE Structures (5A2R). (a) Key active site residues from the form II AnCE structure that interact with the citrate ion are shown as cyan sticks with the citrate ion as green sticks. The catalytic zinc ion is shown as a grey sphere and a water molecule that mediates an interaction between the citrate and zinc ions as an orange sphere. (b) The same active site residues are shown for the form III structure this time in magenta, with the malate ion in green. The catalytic zinc ion is a grey sphere. The water molecule from (a) is conserved and shown as an orange sphere, but the orientation of the malate ion means that there is no water-mediated interaction between the malate and the zinc.

6.3.4.3 Glycosylation Profiles of the Form II and Form III Structures

Glycosylation is a common feature of angiotensin converting enzyme structures, and has been shown to be essential for the stability and function of human sACE (Gordon *et al.,* 2003).

Whilst no glycosylation was reported for the form I AnCE structure, three N-linked glycosylation sites, at Asn53, Asn196 and Asn311, are clearly visible on the form II structure. Interestingly, although there is very clear difference density for N-linked glycans at Asn53 and Asn311 on the form III structure, there is no evidence of N-linked glycosylation at Asn196. This is in spite of the fact that both the form II and form III structures were obtained from protein expressed using *P. pastoris*. Analysis of the two structures indicates that this observed difference is almost certainly due to differences in crystal packing.

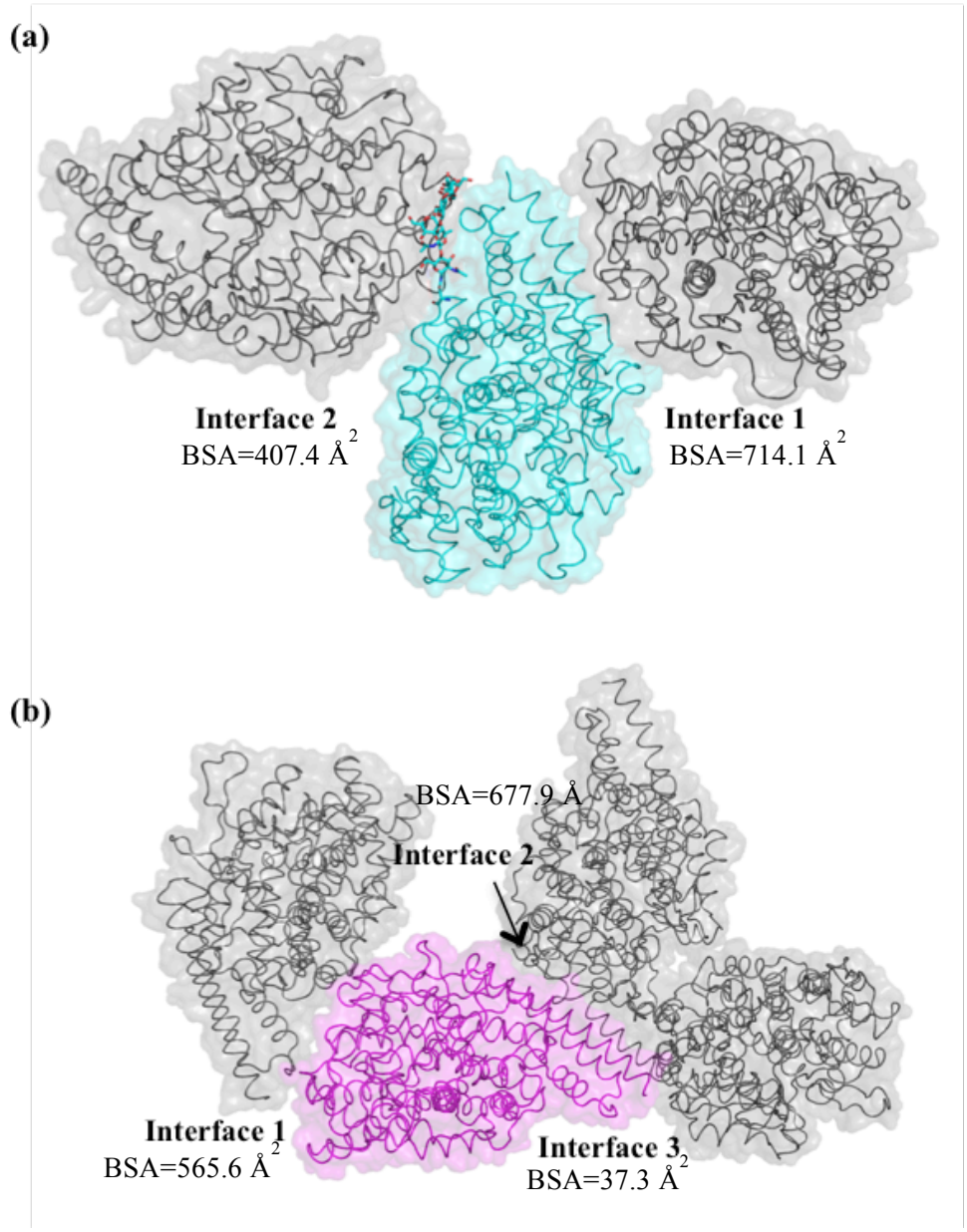


Figure 6.7: Crystal Contacts in the Form II (PDB code 2X8Y) and Form III (PDB code 5A2R) AnCE Structures. (a) In the form II AnCE structure that there are two different interfaces between symmetry molecules. To illustrate this, three AnCE molecules are shown as cartoon and surface representations. The central molecule is shown in cyan and the two symmetry molecules in grey. The two interfaces are labelled as interface 1 and interface 2. The N-linked glycosylation at Asn196 is shown as cyan sticks and is an integral part of interface 2. (b) The three different interfaces between symmetry molecules in the new structure. The interfaces are labelled 1-3. The molecules are shown as surface and cartoons with the central molecule in magenta and the three symmetry molecules in grey. For all interfaces the buried surface area (BSA) was calculated using PDBePISA and is labelled on the figures (Krissinel and Henrick, 2007; Krissinel, 2010).

In the form II structure, each molecule contacts four symmetry related molecules, meaning that there are two different types of interface in the crystal, as shown in Figure 6.7. The first of these, interface one, involves the N-terminal lid of one molecule (helices α1 and α2) and a region of the second molecule containing parts of α10, α11, α14 and α18. Interface two is considerably larger, and involves α3, α6 and α8 of the first molecule and α5, α8 and α12 of the second.

Each molecule in the form III structure contacts six symmetry related molecules, which results in three different interfaces, as shown in Figure 6.7. The first interface involves helices α5, α8 and α9 of one molecule and helices α3, α8, α18 and α20 of the neighbouring molecule. The second interface involves the N-terminal lid of one molecule and helices α5 and α8 of a second molecule. As illustrated in Figure 6.7, this interface is quite similar to interface one of the form II structure. The final interface in the form III structure is much smaller than the other interfaces; Lys20 of α1 of one molecule contacts four residues of α18 belonging to a second molecule.

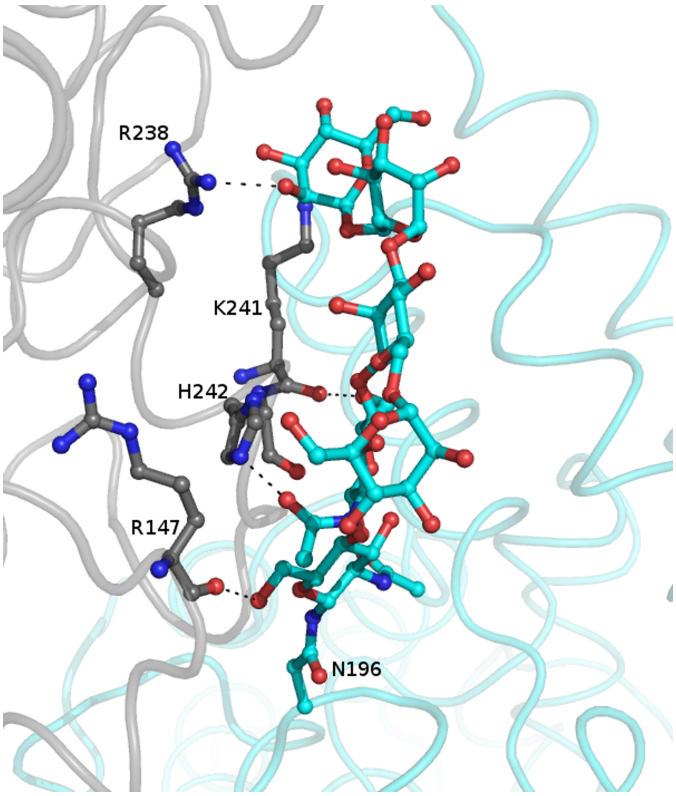


Figure 6.8: **Stabilisation of N-linked Glycosylation at Asn196 by Crystal Contacts in the Form II AnCE Structure.** The glycosylation is shown as blue sticks with hydrogen bonding residues from the symmetry molecule in grey. Hydrogen bonds are shown as black dashes.

Importantly, this results in the N-linked glycosylation at Asn196 being an integral component of interface two in the form II structure. Hydrogen bonds are formed between the glycan chains and residues Arg147, Arg238, Lys241 and His242 of the symmetry-related molecule, Figure 6.8. The extent of these interactions with the symmetry related molecule means that the large glycan chain should be held in a single conformation; hence, it is visible in the crystal structure.

In contrast, none of the interfaces between symmetry molecules in the form III structure include the region surrounding Asn196. The large glycan chain is therefore likely to be highly mobile and exist in many different conformations throughout the molecule, explaining why no electron density is visible for it in the new structure.

6.3.4.4 Further Differences Between the Form II and Form III AnCE Structures Relating to Crystal Contacts

There are some additional, albeit more subtle, differences between the two structures in the region surrounding Asn196. As shown in Figure 6.9, the region encompassing the tops of helices α2 and α3 has moved by approximately 1.3 Å between the two structures.

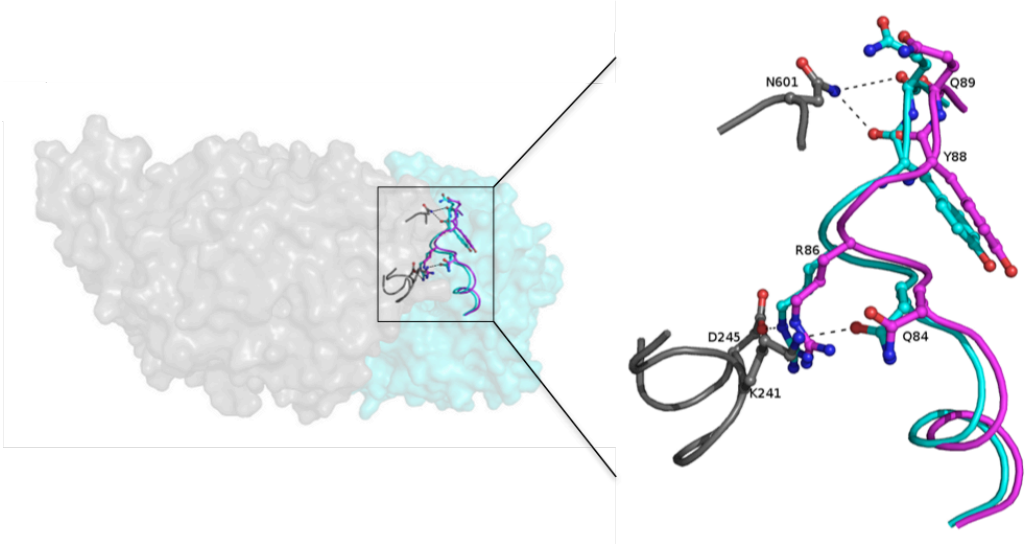


Figure 6.9: Movement of α2 due to Differences in Crystal Contacts in the Form II (PDB code 2X8Y) and form III (PDB code 5A2R) AnCE structures. Surface representation of interface 2 in the form II structure shows one molecule in cyan and the symmetry molecule in grey. Close-up view of the interface shows the interacting residues from each molecule of the interface in cyan and grey respectively. Additionally, corresponding residues from the form III structures have been superposed and are shown in magenta to illustrate the movement of the protein backbone in this region. Hydrogen bonding interactions with Asn601, Asp245 and Lys241 of the symmetry molecule in the form II structure pull part of α2 and the loop between α2 and α3 towards the symmetry molecule, resulting in movement of the peptide backbone by up to 1.3 Å compared to the new structure.

In the form II structure, this region has been pulled closer to the symmetry related molecule by hydrogen bonding interactions between the main chains of Tyr88 and Gln89 and the side chain of Asn601 from the symmetry molecule, and between the side chains of Gln84 and Arg86 and the side chains of Lys241 and Asp245, respectively, of the related molecule.

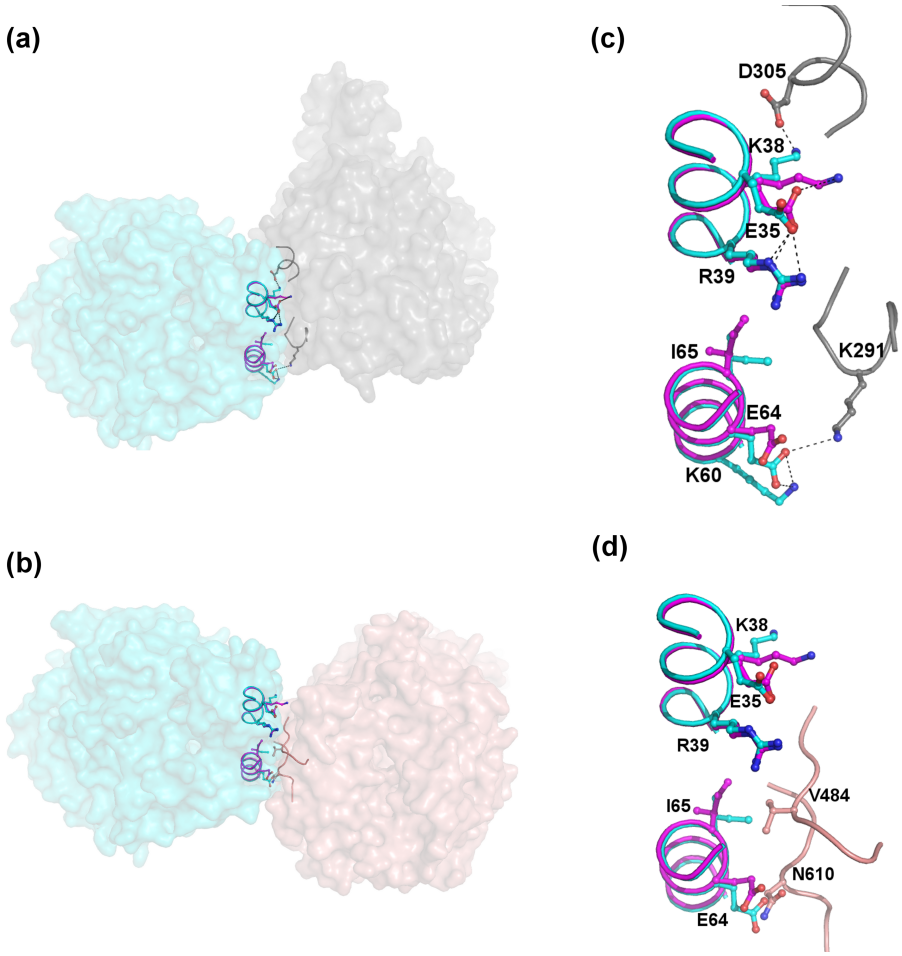


Figure 6.10: Comparison of Positions Adopted by Side-Chains at Interface 2 in the Form II Structure and Interface I in the New Structure. (a) Surface representation of interface 2 of the form II structure with AnCE shown in cyan and the symmetry molecule in grey. Key interacting residues at this interface are shown as sticks with a close-up view in (c). For comparison, equivalent residues from the new structure are shown in (c) as magenta sticks. The Lys38 side chain moves by approximately 5 Å in the two structures to hydrogen bond with Asp305 of the symmetry molecule in the form II structure and Glu35 of the same molecule in the new structure. (b) Surface representation of interface 1 in the new structure, which overlaps significantly with interface 2 in the form II structure. The AnCE surface is shown in cyan with the symmetry molecule in coral. Key interacting residues are shown as sticks with a close-up view in (d), where symmetry molecule residues are in coral and AnCE residues from the new structure in magenta with residues from the original structure shown in cyan for comparison. To avoid a clash with Asn610 of the symmetry molecule, Glu64 in the new structure adopts a different conformation to in the form II structure where it forms a hydrogen bond with Lys291 of the symmetry molecule. Similarly, to avoid a clash with Val484 of the symmetry molecule Ile65 exists as a different rotamer in the two structures.

Figure 6.7 illustrates that there is quite a significant overlap between interface 1 of the form II structure and interface 2 of the form III structure. In the latter, residue Lys38 forms a hydrogen bond with neighbouring residue Glu35, Figure 6.10. In the form II structure, Lys38 has moved by approximately 5 Å so as to interact with Asp305 of the symmetry related molecule. Consequently, the side chain of Glu35 adopts a slightly different conformation and forms an additional, second interaction with Arg39, within the molecule (Figure 6.10).

Remaining at this interface, Glu64 of the form II structure is observed to form hydrogen bonds with two lysine residues: Lys60 from within the molecule and Lys291 of the symmetry molecule. In the new structure, the proximity of the symmetry molecule in this region means that Glu64 has adopted a different conformation so that it avoids a clash with Asn610. Consequently, the interaction with Lys60 is lost, Figure 6.10. Furthermore, Ile65 is observed to exist as different rotamers in the two structures, most probably to avoid a clash with Val484 in the form III structure, Figure 6.10.

6.4 Conclusions and Future Directions

In this chapter a new crystallisation condition has been identified that gives rise to diffraction quality crystals of the *D. melanogaster* ACE homologue, AnCE. The new structure of AnCE, the form III structure is reported in a different space group to the two previous structures. Some minor differences compared to previous AnCE structures have been identified, including the absence of one of the three N-linked glycosylation sites, which are the result of differences in crystal contacts between symmetry related molecules.

The overall fold of the structure is otherwise identical to that reported previously, and the positions of active site residues are unchanged. This indicates that this is a valid structure and that this condition could be used to produce crystals for future structural studies.

Crucially, the citrate ion found to co-ordinate the active site zinc ion in the form II structure is absent from the new structure. There is a malate ion, originating from the crystallisation condition, in the active site, however it does not appear to interact with the zinc ion and hence may not be so tightly bound.

The presence of the active site citrate ion has been identified as a problem in studying the structural basis of inhibition by new, potential inhibitor precursors such as

flavonoids, which may not have sufficient affinity for the active site to displace the citrate. The data presented here indicate that the new condition may be more suitable for this role.

Future work will now involve co-crystallisation experiments with a range of potential inhibitors, including luteolin (Figure 6.1) and other flavonoids, as well as some of the BPPs for which crystals were not obtained from the original condition (Chapter 3).

Chapter 7: Recombinant Expression of an ACE Homologue from *Anopheles gambiae* by *Pichia pastoris*

7.1 Introduction

Due to its role as a key vector in the transmission of the *Plasmodium falciparum* parasite, a major causative agent in malaria, the mosquito *Anopheles gambiae* is an exceedingly important organism to study.

More recently, it has also become apparent that *A. gambiae* may be an important tool for studying angiotensin converting enzymes. Whilst the existence of insect ACE homologues has been known for two decades (Lamango and Isaac, 1994) the *A. gambiae* genome codes for ten ACE-like enzymes (AnoACEs). This is the highest number seen in any insect genome to date and may be indicative of the gene products having a wide range of important roles in mosquito physiology. Studying these enzymes could therefore help to elucidate some of the roles of ACE-like enzymes not only in insects but also in mammals, outside of blood pressure regulation.

Given the significant evidence that exists for insect ACE homologues having a role in reproduction (section 1.9.1.1) (Ekbote *et al.,* 2003a; Ekbote *et al.,* 2003b; Hurst *et al.,* 2003; Macours and Hens, 2004; Vercruysse *et al.,* 2004; Wijffels *et al.,* 1996), and the role of *A. gambiae* as a vector for transmission of *P. falciparum*, there is also the long-term possibility of using AnoACE inhibition in insecticides as a method for controlling the transmission of malaria.

A complete introduction to the *A. gambiae* ACE homologues is presented in section 1.10. Currently, work on these enzymes has been limited to a single genomics study and some analysis of gene expression patterns (Burnham *et al.,* 2005). Whilst this has provided a good foundation, a lot more work is required in order to understand the functions and physiological roles of these enzymes.

7.1.1 Aims and Approaches

As discussed in section 1.10, the aim of this work is to begin by focusing on a single one of these enzymes: AnoACE3, and to work towards protein expression and purification. The longer-term goal would then be to produce sufficient quantities of pure protein to enter crystallisation trials and ultimately to produce a crystal structure. As has been the case for ACE homologues from humans and *D. melanogaster* this could then be used to study the structural basis of substrate and inhibitor binding. Pure protein could also be used in activity assays to learn more about substrate specificity, complementing any crystallographic data.

Due to their extensive glycosylation, recombinant expression of both the human and *D. melanogaster* ACE homologues has only been achieved using eukaryotic systems; human ACE is expressed using Chinese hamster ovary (CHO) cells and *D. melanogaster* AnCE using *P. pastoris*.

The sequence alignment for AnoACE3 and AnCE presented in section 1.10 indicates that the three N-linked glycosylation sites present in the AnCE crystal structure are likely to be conserved in AnoACE3. This, combined with the high degree of overall sequence similarity shared by AnoACE3 with AnCE (65 %) and each domain of human sACE (52 %) indicates that AnoACE3 is likely to undergo quite extensive post-translational modification and that this will be important for its function and stability.

Given that AnCE is routinely expressed using *P. pastoris* producing high yields of active protein, this system was chosen as the starting point for AnoACE3 expression work. The primary advantage of using *P. pastoris* over higher eukaryotic systems such as the CHO cells used for expression of human ACE is that it is much more cost effective. It requires no specialist equipment, as the yeast can be grown in shake flasks in much the same way as *E. coli*. Genetic manipulation is also much more straightforward and there is the potential to achieve high protein yields in a much shorter time frame.

7.2 Methods

The *AnoACE3* coding sequence cloned into the pPICZα vector was kindly provided by Prof. Elywn Isaac, University of Leeds and, prior to this work beginning, was transformed into the GS115 *P. pastoris* expression strain by Dr Akif Mohd. Key features of the pPICZα vector are outlined in section 2.3.1 and a vector map is included in Appendix A.

7.2.1 Mut Phenotype Determination

To determine the methanol utilisation (Mut) phenotype of the transformants, cells were streaked onto minimal dextrose medium plus histidine (MDH: 1.34 % YNB, 4×10^{-5} % biotin, 2 % dextrose, 4×10^{-2} % histidine) and minimal methanol medium plus histidine (MMH: 1.34 % YNB, 4×10^{-5} % biotin, 0.5 % dextrose, 4×10^{-2} % histidine) agar plates. Plates were incubated at 30 ^{o}C for 48 hours. After 48 hours growth on each plate was observed. Transformants observed to have grown equally well on both plates are

determined to be of the Mut^+ phenotype whilst those that show significantly poorer growth on the MMH plates are of the Mut^S phenotype.

7.2.2 Mut^+ Expression Protocol

GS115 cells transformed with the pPICZα/AnoACE3 construct previously confirmed to be of the Mut^+ phenotype were grown on YPD agar plates at 30 ^oC for 48 hours so that discreet colonies appeared.

20 ml BMGY media prepared in a 250 ml baffled flash was inoculated with a single one of these colonies and grown at 30 ^oC with orbital shaking at 250 rpm until log-phase growth was reached. This was determined by measuring the optical density at 600 nm and typically took 16-8 hours. At this point the culture was harvested by centrifugation at 3000 g, 18 ^oC for 5 minutes using the Beckman Coulter^{TM} Avanti J-25 centrifuge with the 16.250 rotor. The supernatant was discarded and the cell pellet resuspended in 200 ml BMMY media for induction. The 200 ml culture was returned to the incubator at 30 ^oC, 250 rpm for 48-72 hours, exact time dependent on the aim of the experiment. Every 24 hours HPLC grade methanol was added to a final concentration of 0.5 % to maintain the concentration required for induction, compensating for losses through evaporation and consumption.

At the end of the expression period the culture was harvested by centrifugation at 3000 g as described above. Samples of the supernatant and cell pellets were analysed by SDS PAGE following the general protocol outlined in section 2.5.2 to detect any secreted or intracellular expression.

7.2.3 Mut^S Expression Protocol

Where there was ambiguity in the results of the Mut phenotype testing expression trials were also conducted following the protocol for cells with the Mut^S phenotype. As for Mut^+ expression, this protocol begins with growing individual colonies on YPD agar plates.

50 ml BMGY media prepared in a 500 ml flask was inoculated with an individual colony of GS115 cells transformed with the pPICZα/AnoACE3 construct and incubated at 30 ^oC with orbital shaking at 250 rpm until log phase was reached. This was determined by measuring the optical density at 600 nm and typically took 16-18 hours. At this point cells were harvested by spinning at 3000 g, 18 ^oC as detailed above. The

supernatant was discarded and the cell pellet resuspended in 10 ml BMMY media in a 100 ml baffled flask.

The 100 ml culture was incubated at 30 ^{o}C, 250 rpm for a further 48-96 hours. The length of induction was varied between experiments in order to investigate the optimal time period. Every 24 hours 100 % HPLC grade methanol was added to the culture to a final concentration of 0.5 % in order to maintain the concentration required for induction, compensating for losses through evaporation and consumption.

At the end of the expression the culture was harvested by centrifugation at 3000 g, 18 ^{o}C as described previously. Samples of the cell pellet and expression supernatant were analysed by SDS PAGE to identify any intracellular or secreted expression of AnoACE3.

7.2.4 Optimisation of Expression

Following the Mut^{+} and Mut^{S} expression protocols outlined above failed to yield any detectable AnoACE3, as discussed in sections 7.3.2 and 7.3.3. To try and overcome this problem the protocols were altered and expression trials were performed following the general methodology outlined above but varying methanol concentration (0.5 - 3 %), growth temperature (28 - 30 ^{o}C), induction time (24 - 96 hours) and media type (BMGY/BMMY and MMH/MM).

7.2.5 Ammonium Sulphate Precipitation of Expression Samples

Owing to the continued lack of detectable AnoACE3 expression, ammonium sulphate precipitation was performed on expression supernatants prior to SDS PAGE analysis. Supernatants were treated with an equal volume of 3 M ammonium sulphate, 20 mM Tris pH 8.0 and incubated for 10 minutes at room temperature. Ammonium sulphate treated samples were spun down at 4000 g, 4 ^{o}C for 10 minutes using the Beckman Coulter^{TM} Avanti J-25 centrifuge and the 16.250 rotor. Both the pellet and supernatant were subsequently analysed by SDS PAGE to detect any AnoACE3.

7.3 Results and Discussion

7.3.1 Mut Phenotype Determination

The preliminary work completed by Dr Akif Mohd had produced four positive transformants of GS115 cells with the pPICZα/AnoACE3 construct. All four of these were screened to determine the Mut phenotype by comparing growth on MMH and MDH agar plates. Although this is very much a subjective analysis it was very clear that

one of the transformants had grown equally well on both plates and was hence of the Mut^+ phenotype. This result indicates that the *AOX1* gene remains intact so these cells should grow well using methanol as a metabolite.

The results for the other three transformants were not quite so clear, but appeared to indicate poorer growth on the MMH compared to the MDH agar plate. This indicates that the *AOX1* gene has been lost and the cells are not able to use methanol as a metabolite as efficiently. They are of the Mut^S phenotype.

The Mut^+ expression protocol facilitates the growth of larger expression cultures much more easily than the Mut^S protocol. Expression trials were therefore initiated using the Mut^+ transformant in the first instance.

7.3.2 Expression of AnoACE3

Due to the presence of the α-factor in the pPICZα vector, any AnoACE3 that is expressed should be secreted from the cells into the culture medium. However, upon analysis by SDS PAGE there was no evidence of any protein expression close to the expected molecular weight of AnoACE3 (73.6 kDa). An example of a typical expression gel is shown in Figure 7.1. This gel contains samples of the culture media taken before induction and at regular intervals up to 50 hours post-induction. At no point is a protein close to the expected molecular weight of AnoACE3 visible. There are trace amounts of some smaller species, however if anything AnoACE3 should appear larger than expected due to its probable glycosylation.

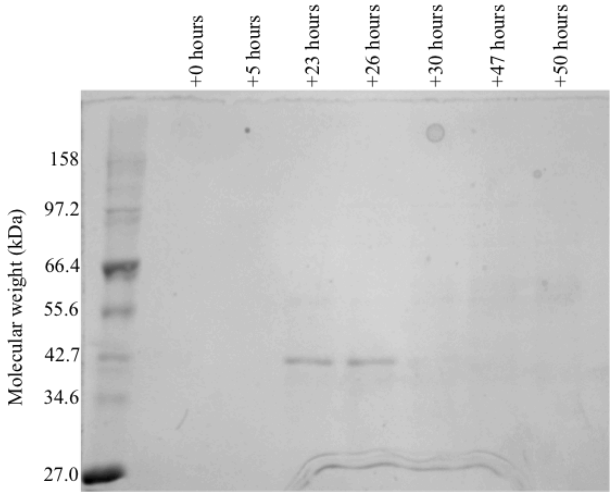


Figure 7.1: Analysis of AnoACE3 Expression by SDS-PAGE. Samples of the supernatant taken at 0 hours (pre-induction), +5, +23, +26, +30, +47 and +50 hours following induction were analysed. At no point on the gel is there any evidence of a species close to the expected molecular weight, 73.6 kDa, of AnoACE3. There are only trace amounts of some much smaller species at +23 and +26 hours.

It is possible that AnoACE3 was being expressed, but poorly. The large culture volume would have compounded this problem by diluting the protein, potentially to a point that it was not visible on the gel. To overcome this samples were concentrated prior to analysis, however no expression was detected.

As detailed in the section 7.2.4, many attempts were made to achieve AnoACE3 expression including altering the growth temperature from 30 ^{o}C to 28 ^{o}C, varying the concentration of methanol used for induction between 0.5 and 3 %, testing for expression at time points up to 96 hours and changing the expression medium from BMGY/BMMY to MMH/MM. The reason for the change in medium is that unlike BMGY/BMMY, MMH/MM media is not buffered. It is possible that AnoACE3 was being expressed but degraded by proteases active at the neutral pH maintained by the buffered media. Removing this buffering may have inactivated the proteases so that any AnoACE3 was no longer being degraded. Unfortunately none of the above approaches proved successful.

Another possibility for the lack of detectable AnoACE3 is that the protein is being produced but retained in the cells rather than being secreted. To check for this cell pellets from the various expression trials were routinely subjected to SDS PAGE analysis but there was never any evidence of AnoACE3 within the cells.

Expression trials were also performed using the transformants deemed to be of the Mut^{S} phenotype. The variations outlined above were also applied to the Mut^{S} expression protocol however again there was no evidence of AnoACE3 expression.

As a final attempt to concentrate the sample so as to detect any weak AnoACE3 expression, the expression media was treated with ammonium sulphate. Both the supernatant and pellet from this treatment were analysed by SDS PAGE but no AnoACE3 was detected.

At this stage the decision was taken to cease attempts to express AnoACE3 using *P. pastoris* and move to alternative expression systems. This is the basis of the work described in Chapter 8.

7.3.3 Possible Reasons for the Failure to Express AnoACE3

There is always the possibility with recombinant protein expression that the target protein will prove to be toxic to the cells in some way. This seems very unlikely in the case of AnoACE3 however, given that the closely related homologue, AnCE, is expressed well by *P. pastoris*. Furthermore the expression protocols that were trialled all involved the cultures reaching high optical density values prior to the induction of expression. This technique can often enable *P. pastoris* to produce proteins that are toxic to it (Ahmad *et al.,* 2014); hence it is very unlikely that this was the problem here.

It can also be the case that the target protein is being expressed, but is either itself unstable or being degraded by yeast proteases present in the culture. To try and determine whether this was the problem with AnoACE3, expression trials were performed using a non-buffered media. The rationale for this being that in the absence of a buffer to maintain the neutral pH, culture conditions may be such that some of the yeast proteases are no longer active and so unable to act on any AnoACE3. Unfortunately this change did not have any impact on detectable expression levels. Another possibility would have been to add a protease inhibitor such as PMSF to the culture, but given that AnoACE3 is itself a protease this would have been somewhat counter-productive.

In some ways the lack of detectable AnoACE3 expression is surprising, given that the closely related homologue, AnCE, has been successfully expressed using this system. Taking this into consideration, perhaps the most likely explanation for the lack of expression is due to incorrect post-translational processing.

As explained in section 1.10, whilst the core peptidase domain is very well conserved between AnCE and AnoACE3, AnoACE3 also has an additional N-terminal sequence of unknown function not found in AnCE. Furthermore, although both proteins have a conserved N-terminal signal sequence, this is omitted from the AnCE expression construct but still present in the AnoACE3 construct.

It is quite possible that these differences in the N-terminal region may in some way be responsible for the lack of AnoACE3 expression in *P. pastoris*. For example, further glycosylation sites might be present that are essential for protein stability. Although *P. pastoris* is able to glycosylate proteins the nature of the glycans added is not the same as in the native system. In *P. pastoris* these glycans have a very high mannose content, which may not be ideal.

Furthermore, the human ACE enzymes also have O-linked glycosylation sites. Whilst this has not been an issue with AnCE it is possible that the additional N-terminal region of AnoACE displays O-linked glycosylation. To investigate this further a glycosylation prediction was performed on the full AnoACE3 sequence and this did identify a potential additional N-linked glycosylation site and also an O-linked glycosyation site in the N-terminal region (Gupta and Brunak, 2004).

Another issue that may arise due to the different nature of the N-termini of AnoACE3 and AnCE again relates to post-translational processing of the expressed protein. Both constructs contain the α-factor signal sequence at their N-termini, which should target them for secreted expression.

As part of the secretion pathway the α-factor signal sequence is cleaved from the target protein by the enzyme Kex2. Cleavage of this sequence then allows the protein to move from the Golgi apparatus to vesicles for secretion. There is increasing evidence to suggest that the nature of the N-terminus of the target protein can have a profound effect on the ability of Kex2 to remove the α-factor, thus impacting on the secretion of the target protein (Yang *et al.,* 2013). It is quite possible that the differing nature of the N-termini of AnCE and AnoACE3 has therefore affected this process, resulting in the lack of detectable AnoACE3 expression.

7.4 Conclusions

AnoACE3 is an ACE homologue from *A. gambiae* that shares 54 % sequence similarity with the *D. melanogaster* enzyme AnCE. Given that AnCE has been successfully expressed using the methylotropic yeast *P. pastoris*, recombinant expression of AnoACE3 using the same system was attempted.

In spite of significant efforts no AnoACE3 expression by *P. pastoris* was detected. The reasons for this are unknown, however it is possible that it may be related to the unique nature of the N-terminus of the AnoACE3 construct, which may impact the post-translational processing of the protein.

As a result of this, work towards the expression of AnoACE3 in higher eukaryotic systems has been initiated and this is presented in Chapter 8.

Chapter 8: Recombinant Expression of ACE Homologues from *Drosophila melanogaster* and *Anopheles gambiae* in Mammalian and Insect Systems

8.1 Introduction

Due to the lack of success seen when trying to express AnoACE3 using the *Pichia pastoris* expression system, the decision was made to begin creating new constructs for expression in mammalian and insect systems using HEK293T and *Sf*9 cells respectively. This work was completed with support from the group of Prof. Ray Owens at the Oxford Protein Production Facility (OPPF), Harwell Science and Innovation Campus, Didcot, Oxon, UK.

All work with AnoACE3 was carried out concurrently with the two *Drosophila melanogaster* ACE homologues: AnCE and ACER (introduced in section 1.9). Previous efforts to express ACER using *Pichia pastoris* have also proved unsuccessful hence moving to mammalian and insect systems at the same time as for AnoACE3 was a logical step. The reasons for also attempting this with AnCE were twofold. Primarily as a positive control but also because producing AnCE in an insect system when it is itself an insect protein may in the longer term be interesting for comparison with the protein produced from yeast.

8.2 Methods

8.2.1 Construct Design

Graphical illustrations of the predicted domain structures of AnoACE3 and ACER are shown in Figure 8.1. ACER and AnoACE3 are both predicted to be secreted proteins as illustrated by the predicted signal sequences at their N-termini. This is in contrast to mammalian ACE which is membrane bound, but comparable to AnCE which has a 17 residue secretion signal sequence at its N-terminus, also illustrated in Figure 8.1. Like AnCE, the ACER signal sequence appears to be located at the N-terminus of the protein and is comprised of the first 22 residues.

AnoACE3 appears to be slightly different. The signal sequence is predicted to be located directly upstream of the conserved peptidase domain, consisting of residues 60-87. There is then an “undefined” 60 residue sequence at the very N-terminus of the protein which is conserved in neither AnCE nor ACER. At this point the function of this region is unknown.

In order to maximise the likelihood of successfully expressing AnoACE3 a series of constructs, AnoACE_1 to AnoACE_6, was created. These contained either the full sequence (residues 1-698), the 17 residue signal sequence onwards (61-698) or only the

conserved peptidase domain (88-698). The constructs and their key features are summarised in Table 8.1.

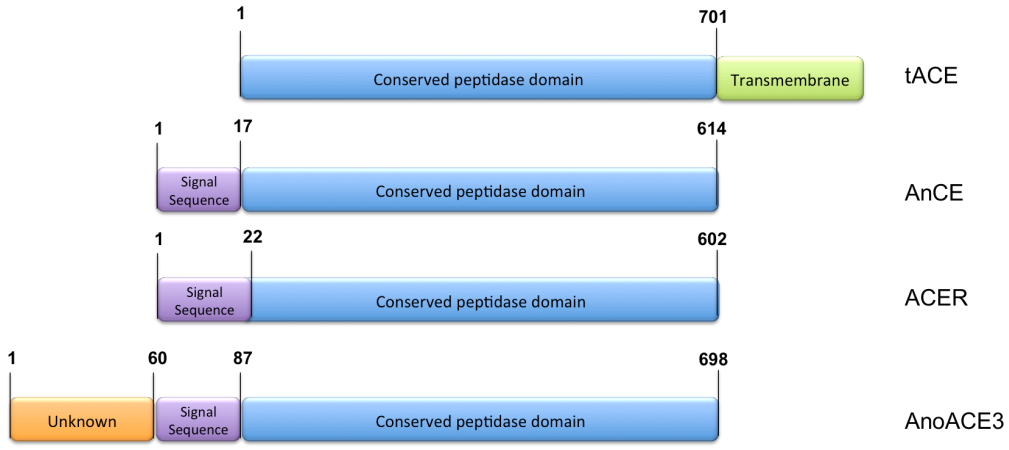


Figure 8.1: Domain Organisation of tACE, AnCE, ACER and AnoACE3. All four proteins share a conserved peptidase domain shown in blue. tACE and sACE (not shown in this picture) both have a membrane binding domain, shown in green, at their C-termini, which is present in neither AnCE, ACER nor AnoACE3. Instead AnCE and AnoACE3 both have a 17 residue signal sequence and ACER has a 22 residue signal sequence, shown in purple, at their N-termini, meaning that they are secreted rather than membrane bound. AnoACE3 also has an additional 60 residue region at the N-terminus of unknown function. This is displayed in orange.

All 6 constructs were designed to be cloned into the pOPIN vector suite using the "In-Fusion®" technique (Clontech Laboratories, Inc.). This requires the insert to have 15 base pair sequences at the 5' and 3' ends that are complementary to the vector sequence. These additional residues are indicated in green on the primer sequences in Tables 8.1, 8.2 and 8.3.

AnoACE_1 and AnoACE_2, both containing the predicted signal sequence, were designed for cloning into the pOPINEneo vector. Their native signal sequences will be used to drive secreted expression in both *Sf*9 and HEK293T cells. The synthesised proteins will also contain a C-terminal eight-residue histidine tag to aid purification and detection by western blot. For all constructs the decision was made to locate the poly-histidine tags at the C-termini due to the variable nature of the N-terminus in different constructs and due to the presence of the N-terminal lid in other ACE homologues. The pOPINEneo vector map is shown in Appendix A.

Construct name	Features	Protein sequence	Size (bp)/kDa	Vector	Forward primer	Reverse primer
AnoACE_1	"Unknown" region Native signal sequence Conserved peptidase domain C-terminal His-tag	1-698	2090/81	pOPINEneo	AGGAGATATACCATG CGATTTGGAATCA T_m = 64.2 °C	GTGATGGTGATGTTT GGCCGCCATAAAAT T_m = 67.8 °C
AnoACE_2	Native signal sequence Conserved peptidase domain C-terminal His-tag	60-698	1913/75	pOPINEneo	AGGAGATATACCATG ATGGCAAGTAATA T_m = 65.3 °C	GTGATGGTGATGTTT GGCCGCCATAAAAT T_m = 67.8 °C
AnoACE_3	μ-phosphatase signal sequence Conserved peptidase domain C-terminal His-tag	87-698	1835/72	pOPINTTGneo	GCGTAGCTGAAACCG GCCGTCGTTCAA T_m = 71.8 °C	GTGATGGTGATGTTT GGCCGCCATAAAAT T_m = 67.8 °C
AnoACE_4	HBM signal sequence Conserved peptidase domain C-terminal His-tag	87-698	1835/72	pOPIN-HBM	TCTTACATCTATGCGC GTCGTTCAACGG T_m = 68.2 °C	GTGATGGTGATGTTT GGCCGCCATAAAAT T_m = 67.8 °C
AnoACE_5	HA signal sequence Conserved peptidase domain C-terminal His-tag	87-698	1835/72	pOPIN-HA	TGCCTGGTGTTCGCGC GTCGTTCAACGG T_m = 72.5 °C	GTGATGGTGATGTTT GGCCGCCATAAAAT T_m = 67.8 °C
AnoACE_6	Ac64 signal sequence Conserved peptidase domain C-terminal His-tag	87-698	1835/72	pOPIN-Ac64	CATTCTGCCTTTGCGC GTCGTTCAACGG T_m = 70.1 °C	GTGATGGTGATGTTT GGCCGCCATAAAAT T_m = 67.8 °C

Table 8.1: AnoACE Construct and Primer Details. The regions of the primers shown in green indicate the 15 bp sequences complementary to the vector required for In-Fusion® cloning. The molecular weights given in kDa are for the mature proteins so do not include the signal sequences.

Construct name	Features	Protein sequence	Size (bp)/kDa	Vector	Forward primer	Reverse primer
ACER_1	μ-phosphatase signal sequence Conserved peptidase domain C-terminal His-tag	31-602	1799/71	pOPINTTGneo	GCGTAGCTGAAACCG GCGTGCTGGAGG T_m = 72.8 °C	GTGATGGTGATGTTT GGTGCTAAACGTT T_m = 67.7 °C
ACER_2	HBM signal sequence Conserved peptidase domain C-terminal His-tag	31-602	1799/71	pOPIN-HBM	TCTTACATCTATGCGG TGCTGGAGCGC T_m = 68.0 °C	GTGATGGTGATGTTT GGTGCTAAACGTT T_m = 67.7 °C
ACER_3	HA signal sequence Conserved peptidase domain C-terminal His-tag	31-602	1799/71	pOPIN-HA	TGCCTGGTGTTCGCGG TGCTGGAGGCG T_m = 73.6 °C	GTGATGGTGATGTTT GGTGCTAAACGTT T_m = 67.7 °C
ACER_4	Ac64 signal sequence Conserved peptidase domain C-terminal His-tag	31-602	1799/71	pOPIN-Ac64	CATTCTGCCTTTGCGG TGCTGGAGGCG T_m = 73.1 °C	GTGATGGTGATGTTT GGTGCTAAACGTT T_m = 67.7 °C

Table 8.2: ACER Construct and Primer Details. The regions of the primers shown in green indicate the 15 bp sequences complementary to the vector required for In-Fusion® cloning. The molecular weights given in kDa are for the mature proteins so do not include the signal sequences.

Construct name	Features	Protein sequence	Size (bp)/kDa	Vector	Forward primer	Reverse primer
AnCE_1	μ-phosphatase signal sequence Conserved peptidase domain C-terminal His-tag	18-614	1796/70	pOPINTTGneo	GCGTAGCTGAAACCG GCGCGCTGGTCA T_m = 73.1 °C	GTGATGGTGATGTTT TGATGAGACGCAT T_m = 65.4 °C
AnCE_2	HBM signal sequence Conserved peptidase domain C-terminal His-tag	18-614	1796/70	pOPIN-HBM	TCTTACATCTATGCGG CGCTGGTCAAGG T_m = 69.2°C	GTGATGGTGATGTTT TGATGAGACGCAT T_m = 65.4 °C
AnCE_3	HA signal sequence Conserved peptidase domain C-terminal His-tag	18-614	1796/70	pOPIN-HA	TGCCTGGTGTTCGCGG CGCTGGTCAAGG T_m = 73.9 °C	GTGATGGTGATGTTT TGATGAGACGCAT T_m = 65.4 °C
AnCE_4	Ac64 signal sequence Conserved peptidase domain C-terminal His-tag	18-614	1796/70	pOPIN-Ac64	CATTCTGCCTTTGCGG CGCTGGTCAAGG T_m = 71.3 °C	GTGATGGTGATGTTT TGATGAGACGCAT T_m = 65.4 °C

Table 8.3: AnCE Construct and Primer Details. The regions of the primers shown in green indicate the 15 bp sequences complementary to the vector required for In-Fusion® cloning. The molecular weights given in kDa are for the mature proteins so do not include the signal sequences.

Constructs AnoACE_3, AnoACE_4, AnoACE_5 and AnoACE_6 were designed for cloning into the pOPINTTGneo, pOPIN-HBM, pOPIN-HA and pOPIN-Ac64 vectors respectively. Due to the fact that these four constructs only contain the conserved peptidase domain, these vectors were chosen to provide the proteins with the μ-phosphatase, HBM, HA and Ac64 signal sequences respectively. These four constructs all include a C-terminal histidine tag. Vector maps are included in Appendix A.

The AnCE and ACER constructs that have previously been used for *P. pastoris* expression do not contain the native signal sequences, hence the AnCE and ACER constructs created for the insect and mammalian work also only contain the conserved peptidase domain. Four constructs were designed for each protein: AnCE_1, AnCE_2, AnCE_3 and AnCE_4 and ACER_1, ACER_2, ACER_3 and ACER_4, for cloning into the pOPINTTGneo, pOPIN-HBM, pOPIN-HA and pOPIN-Ac64 vectors respectively. Construct details are summarised in Tables 8.2 (ACER) and 8.3 (AnCE).

8.2.2 PCR Amplification of AnoACE3, ACER and AnCE Coding Sequences

The inserts for the new AnoACE, AnCE and ACER constructs were created by PCR amplification using the *P. pastoris* vector DNA as the template. Successful PCR amplification was confirmed by analysis of products on an agarose gel following the method outlined in section 2.5.1.

8.2.2.1 AnoACE3

The PCR amplification of the AnoACE constructs was performed as outlined in Tables 8.4 and 8.5, which detail the components of the reaction mixture and the thermal cycling parameters respectively. All PCR reactions were performed using KOD hot start DNA polymerase (Merck Millipore) using a PTC-100 Thermal Cycler (MJ Research). The components of the reaction mixture and the thermal cycling parameters were based upon the manufacturer's instructions. In particular the annealing temperature of 64 ^{o}C was chosen as it is the lowest primer T_m and the extension time of 40 seconds was selected based on the rate of 20 seconds per kilobase (quoted by the manufacturer) and product sizes ranging from 1835 to 2090 bp.

8.2.2.2 ACER

The general protocol for amplification of the ACER coding sequences was as described for AnoACE3 with slight alterations to the parameters to suit the specific properties of the ACER primers, as detailed in Tables 8.4 and 8.5.

Component	Volume per reaction (µl)
10 × buffer for KOD hot start polymerase	5
$MgSO_4$ (25 mM)	3
dNTPs (2 mM each)	5
Forward primer (10 µM)	1.5
Reverse primer (10 µM)	1.5
Template DNA (15 ng μl^{-1})	3
KOD hot start DNA polymerase (1 U μl^{-1})	1
HPLC grade water	5

Table 8.4: Components of the AnoACE3 and ACER Polymerase Chain Reaction.

Step	Temperature and time (AnoACE3)	Temperature and time (ACER)
Polymerase activation	95 ^oC for 2 minutes	95 ^oC for 2 minutes
Denaturation	95 ^oC for 20 seconds	95 ^oC for 20 seconds
Annealing	64 ^oC for 10 seconds	68 ^oC for 10 seconds
Extension	70 ^oC for 40 seconds	70 ^oC for 40 seconds
Repeat denaturation to extension for a total of 30 cycles.		
Final extension	72 ^oC for 2 minutes	72 ^oC for 2 minutes

Table 8.5: PCR Thermal Cycling Conditions for AnoACE3 and ACER.

8.2.2.3 AnCE

Due to the fact that the original AnCE constructs were received from our collaborator (Prof. Elwyn Isaac) the only available starting material was the GS115 transformant used for AnCE expression as detailed in section 2.7.

In order to obtain template DNA for the PCR amplifications a total genomic DNA extraction was performed based on the protocol presented by Lõoke *et al.,* 2011.

2 ml of YPD media were inoculated with a single colony of GS115 cells transformed with the pPIC9/AnCE construct and grown for 24 hours at 30 ^{o}C with orbital agitation at 250 rpm. The culture was harvested by centrifugation at 4600 g, 4 ^{o}C for 10 minutes using a Heraeus centrifuge (Thermo Fisher Scientific, MA, USA) and the supernatant discarded. The cell pellet was resuspended in 100 µl of 200 mM lithium acetate, 1 % SDS and incubated for 5 minutes at 70 ^{o}C. 300 µl of 70 % ethanol were added to the lysate and the suspension resuspended vigorously before spinning down at 15, 000g for 3 minutes at room temperature using a microcentrifuge (Centrifuge 5415 D, Eppendorf). The resulting pellet was washed with 70 % ethanol prior to resuspension in 100 µl of HPLC grade water and a final centrifugation step at 15, 000 g for 3 minutes. The supernatant was used as the PCR template.

PCR parameters for the AnCE constructs were determined in the same way as for AnoACE and ACER. Here the annealing temperature was increased slightly to 65 ^{o}C to match the lowest primer T_m. Additionally, betaine was included in the reaction mixture to improve specificity given that genomic DNA was being used as the template. Similarly, an additional 5 cycles were performed, bringing the total to 35. Full details of reaction components and thermal cycling parameters are included in Tables 8.6 and 8.7 respectively.

8.2.3 DpnI Treatment of PCR Products

In order to remove any remaining template DNA all of the AnoACE3, AnCE and ACER PCR products were treated with the restriction enzyme DpnI (Promega, WI, USA). The reaction components listed in Table 8.8 were added directly to the PCR mixtures after thermal cycling and then incubated at 37 ^{o}C for 5 hours. A "PCR clean-up" was then performed on each reaction using the Wizard® SV gel and PCR clean-up system (Promega, WI, USA) in accordance with the manufacturer's instructions. The concentration of the purified DNA was determined by measuring the absorbance at 260 nm using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, MA, USA).

Component	Volume per reaction (µl)
10 × buffer for KOD hot start polymerase	5
$MgSO_4$ (25 mM)	3
dNTPs (2 mM each)	5
Forward primer (10 µM)	1.5
Reverse primer (10 µM)	1.5
Template DNA	3
Betaine (3 M)	3
KOD hot start DNA polymerase (1 U μl^{-1})	1
HPLC grade water	5

Table 8.6: Components of the AnCE Polymerase Chain Reactions

Step	Temperature and time
Polymerase activation	95 ^{o}C for 2 minutes
Denaturation	95 ^{o}C for 20 seconds
Annealing	65 ^{o}C for 10 seconds
Extension	70 ^{o}C for 40 seconds
Repeat denaturation to extension for a total of 35 cycles.	
Final extension	72 ^{o}C for 2 minutes

Table 8.7: PCR Thermal Cycling Conditions for AnCE.

Component	Volume (µl)
DpnI (10 u μl^{-1})	0.5
Acetylated BSA (10 µg μl^{-1})	0.2
10 × Promega buffer B	2.5
HPLC grade water	2.8
PCR product	19

Table 8.8: Preparation of the Reaction Mixture for the DpnI Treatment of the AnoACE3, ACER and AnCE PCR Products.

8.2.4 In-Fusion® Reactions

All In-Fusion® reactions were performed using In-Fusion® HD EcoDry™ Cloning kits (Clontech Laboratories, Inc.). These consist of lyophilised pellets of the In-Fusion® HD enzyme premix. 250 ng of each purified PCR product was added to the pellets in a 96-well PCR plate. HPLC grade water was added to each well to a final volume of 10 µl. The plate was sealed and incubated at 42 ^{o}C for 30 minutes in a Veriti® Thermal Cycler. After 30 minutes 40 µl TE (Tris EDTA) buffer was added to each well to stop the reactions.

8.2.5 Transformation of *E. coli*

Products of the In-Fusion® reactions were immediately transformed into One Shot® OmniMAX™ 2 T1^{R} chemically competent *E. coli* (Life Technologies) using a heat shock method.

The competent cells were purchased in pre-prepared 50 µl aliquots. 4 µl of the diluted In-Fusion® reactions were each added to an aliquot and incubated on ice for 30 minutes. The cells were then "heat shocked" at 42 ^{o}C for 30 seconds and incubated on ice for a further 2 minutes. 300 µl SOC media were added to each aliquot prior to incubation at 37 ^{o}C for 1 hour.

After incubation 25 µl of each transformation was plated onto 1 ml LB-agar supplemented with 50 µg ml^{-1} carbenicillin, 0.02 % X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside) (in DMF) and 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside). The addition of X-Gal and IPTG is to enable "blue-white screening" of colonies.

8.2.6 Plasmid DNA Preparation

Individual white colonies were picked for each transformation and incubated overnight at 37 ^{o}C with orbital agitation at 600 rpm in 1.2 ml power broth supplemented with 50 µg ml^{-1} carbenicillin.

Plasmid DNA was extracted from each culture using a QIAprep Spin Miniprep Kit (Qiagen) in accordance with the manufacturer's protocol. Plasmid DNA concentrations were determined by measuring the absorbance at 260 nm using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, MA, USA).

8.2.7 PCR Verification of Constructs

The presence of the appropriate insert in the purified plasmid DNA was determined using PCR. The reactions were prepared as detailed in Table 8.9 and the thermal cycling parameters are listed in Table 8.10. The primers used are the same as those that were used for the original PCR amplification (8.2.2). PCR products were analysed on an agarose gel to confirm presence of the correct insert following the general protocol given in section 2.5.1.

Component	Volume per reaction (µl)
Phusion Flash® High Fidelity PCR Mix (Thermo Fisher Scientific)	12.5
Forward primer (10 µM)	1.5
Reverse primer (10 µM)	1.5
Template DNA	1.5
HPLC grade water	8

Table 8.9: List of Components for the Polymerase Chain Reaction to Confirm the Presence of the AnoACE3, ACER and AnCE Inserts in the pOPIN Vectors After InFusion® Reactions.

Step	Temperature and time
Polymerase activation	94 ^{o}C for 2 minutes
Denaturation	94 ^{o}C for 30 seconds
Annealing	60 ^{o}C for 30 seconds
Extension	68 ^{o}C for 2 minutes
Repeat denaturation to extension for a total of 30 cycles.	
Final extension	68 ^{o}C for 4 minutes

Table 8.10: Thermal Cycling Parameters for the Polymerase Reaction to Confirm the Presence of the AnoACE3, ACER and AnCE Inserts in the pOPIN Vectors After InFusion® Reactions.

8.2.8 Trial Transfections of HEK293T and *Sf*9 Cells Performed at OPPF, UK

8.2.8.1 Transient Transfection of HEK293T Cells

24 hours prior to transfection, 1 ml of HEK293T cells was plated at a density of 1.5-2.0 × 10^5 cells ml^{-1} in individual wells of a 24-well tissue culture plate. Cells were incubated for 24 hours at 37 ^{o}C in a 5 % CO_2/95 % air environment.

2 µl 1.33 mg ml^{-1} GeneJuice® transfection reagent (Merck Millipore) and 1 µg purified plasmid DNA were added to 60 µl serum free DMEM, mixed thoroughly and incubated at room temperature for 10 minutes. During this time the medium was removed from each well of the 24-well plate and replaced with 1 ml of DMEM supplemented with 2 % FBS. After incubation the DNA/GeneJuice cocktails were added to the appropriate wells and incubated at 37 ^{o}C in 5 % CO_2/95 % air for 3 days.

After 3 days secreted protein expression was assessed by analysis of the media by western blot following the general protocol outlined in section 2.5.3. A His Tag peroxidase conjugated antibody (R & D Systems, Oxon, UK) was used to detect the presence of proteins with a polyhistidine tag.

8.2.8.2 Transfection of Sf9 Cells

500 µl aliquots of *Sf*9 cells at a density of 5 × 10^5 cells ml^{-1} in Sf-900™ ii SFM (serum free media) (Invitrogen) were prepared in a 24-well plate and incubated at room temperature for 30 minutes in order to attach to the plate surface.

250 ng Ian Jones bacmid, 50 µl Sf-900™ ii SFM and 0.75 µl FuGeneHD transfection reagent (Promega) were combined with 250 ng of plasmid DNA and incubated for 30 minutes at room temperature.

After the incubation periods the transfection mixtures were added to the appropriate cells and incubated for 5 days at 27 ^{o}C. After 5 days expression was assessed by western blot performed by Dr Joanne Nettleship at the OPPF.

8.2.9 Trial Transfections of HEK293T Cells Performed at University of Bath

Having seen some evidence of expression when working at the OPPF, the aim was now to try and express two of these constructs: AnoACE_2 and ACER_3 using HEK293T cells at the University of Bath. This was attempted using two different transfection reagents: GeneJuice® (Merck Millipore) and FuGeneHD (Promega).

8.2.9.1 Transfections using GeneJuice®

Cells were plated at a density of 1×10^5 cells ml^{-1} in 1 ml of DMEM supplemented with 10 % FBS in a 24 –well tissue culture plate and incubated for 24 hours at 37 ^{o}C in a 5 % CO_2/95 % air environment.

1 µg purified plasmid DNA and 2 µl 1.33 mg ml^{-1} GeneJuice® transfection reagent (Merck Millipore) were added to 60 µl serum free DMEM, mixed thoroughly and incubated at room temperature for 10 minutes.

The media on the previously prepared cells was removed and replaced with 1 ml DMEM supplemented with 2 % FBS. The transfection mixture was added to the appropriate well of the plate and incubated at 37 ^{o}C in 5 % CO_2/95 % air for 72 hours.

Expression was detected by western blots performed on media samples following the protocol outlined in section 2.5.3 using a His Tag peroxidase conjugated antibody (R & D Systems, Oxon, UK) to detect proteins with polyhistidine tags.

8.2.9.2 Transfections using FuGeneHD

Cells were plated at a density of 5×10^5 cells ml^{-1} in 1 ml of DMEM supplemented with 10 % FBS in a 24 –well tissue culture plate and incubated for 24 hours at 37 ^{o}C in a 5 % CO_2/95 % air environment.

After 24 hours the media on the cells was removed and replaced with 1 ml of fresh DMEM supplemented with 10 % FBS.

45 µl of serum free DMEM and 1 µg of the appropriate plasmid DNA were mixed together and 3 µl FuGeneHD (Promega) transfection reagent added. The transfection mixture was incubated at room temperature for 10 minutes before being added to the appropriate cells. Transfected cells were incubated at 37 ^{o}C in a 5 % CO_2/95 % air environment for 72 hours. After this time expression was detected by western blot following the protocol outlined in section 2.5.3 using a His Tag peroxidase conjugated antibody (R & D Systems) to detect proteins with polyhistidine tags.

8.3 Results

8.3.1 PCR Amplification of AnoACE, AnCE and ACER Coding Sequences

Successful amplification of the AnoACE, ACER and AnCE coding sequences for the new constructs was confirmed by agarose gel electrophoresis. Gel images are shown in Figure 8.2 and clearly indicate products at the expected sizes as detailed in Tables 8.1, 8.2 and 8.3.

8.3.2 Transformation into *E. coli* and PCR Verification

Following the In-Fusion® reactions to introduce the PCR products into the appropriate pOPIN vector, the constructs were transformed into One Shot® OmniMAX™ 2 T1^{R} chemically competent *E. coli* using a heat shock method. Successful transformations were identified by the growth of white colonies on LB-agar plates supplemented with carbenicillin, X-Gal and IPTG. There were multiple white colonies for each of the constructs, indicating that all of the In-Fusion® reactions and transformations had been successful.

This was further verified by PCR using isolated plasmid DNA as the template and the original cloning primers for each construct. Analysis of the PCR products by gel electrophoresis indicated products at the expected size for each of the constructs, confirming successful fusion and transformation.

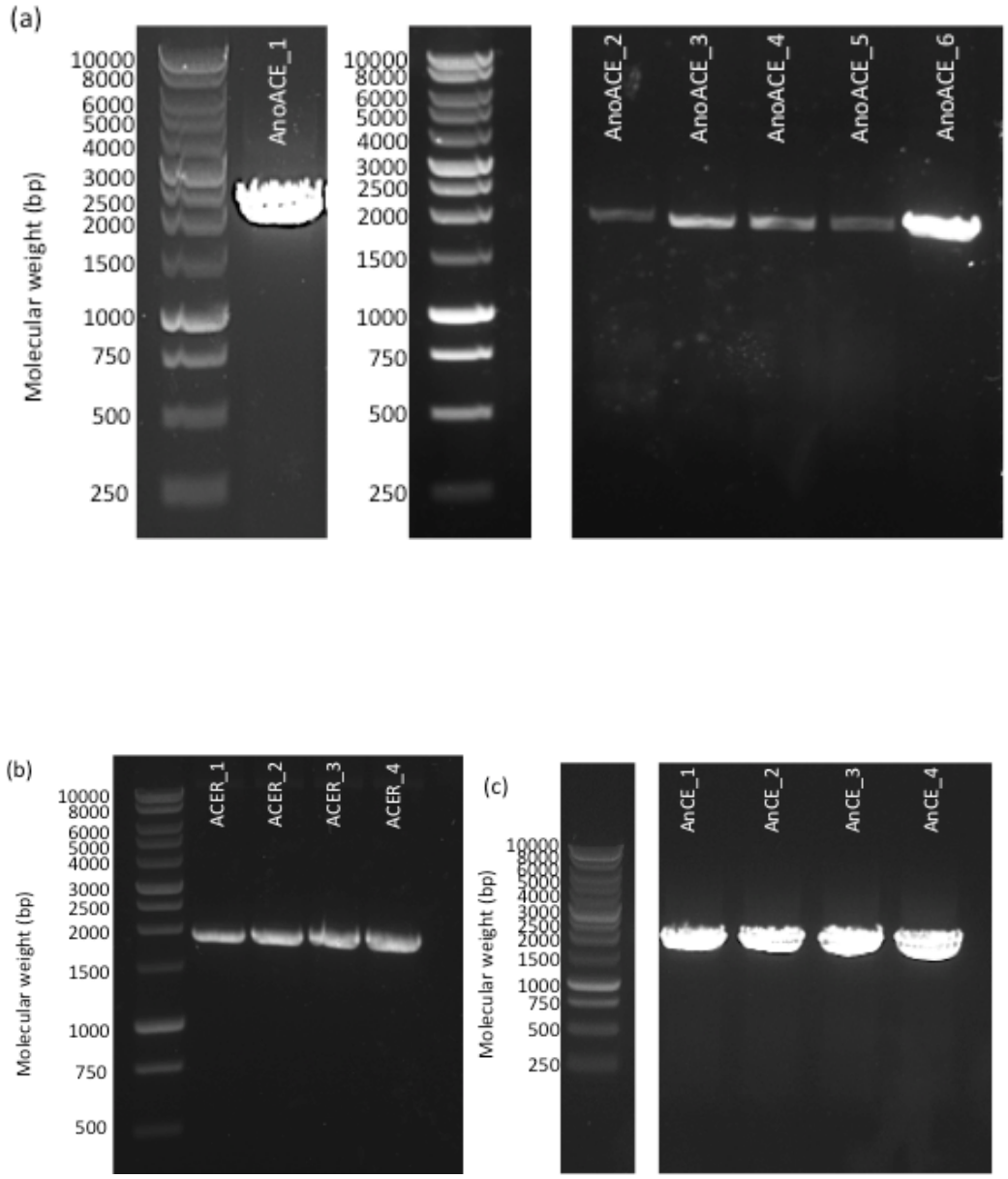


Figure 8.2: Analysis of the AnoACE3 (a), ACER (b) and AnCE (c) PCR Products by Agarose Gel Electrophoresis. Figure illustrates that all of the PCR amplifications for the new constructs were successful, with products visible at the expected sizes for each construct as indicated by the values in Tables 1.1, 1.2 and 1.3.

8.3.3 Trial Transfections Performed at OPPF, UK

Expression of the ACE homologues by the HEK293T and *Sf*9 cells was detected by western blotting using an antibody recognising the polyhistidine tags included in all of the constructs. The two western blots are shown in Figures 8.3 and 8.4 and a summary of the results is presented in Table 8.11.

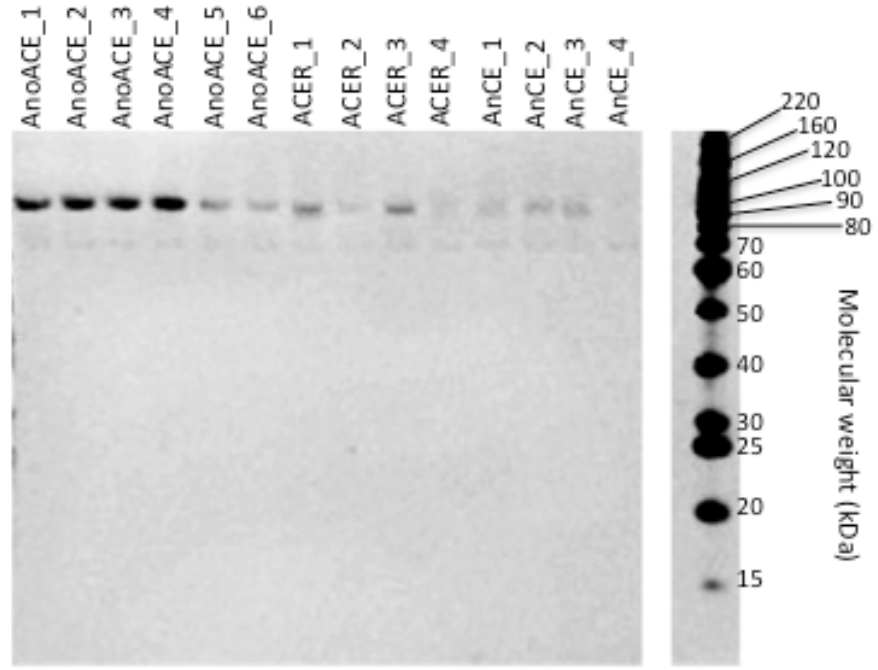


Figure 8.3: Analysis of Secreted AnoACE3, ACER and AnCE Expression by HEK293T Cells by Western Blotting with an Anti-Histidine Tag Antibody.

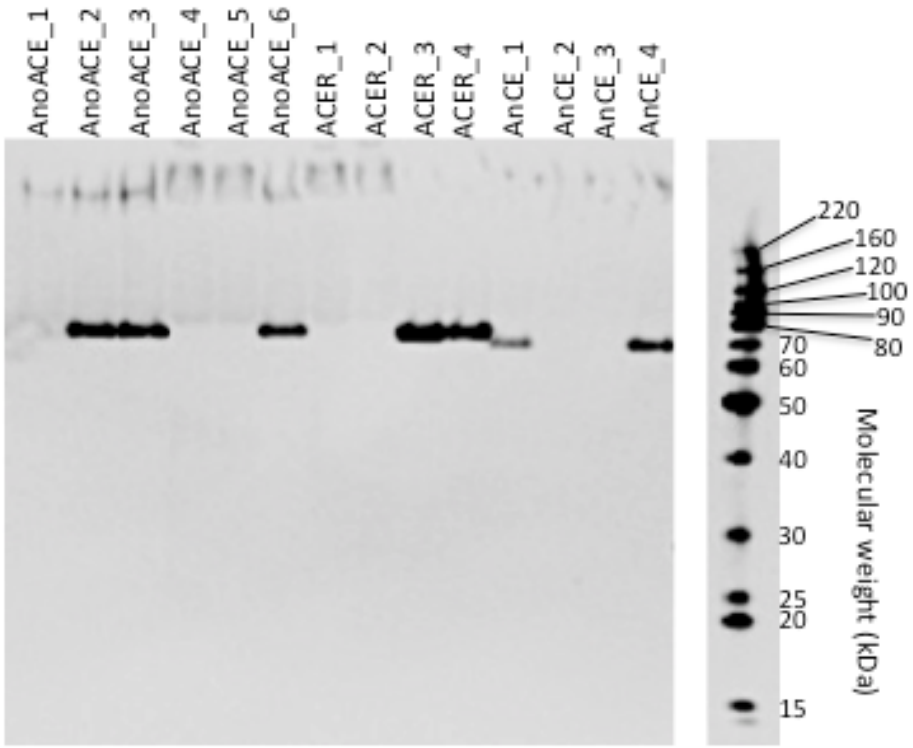


Figure 8.4: Analysis of Secreted AnoACE3, ACER and AnCE Expression by *Sf*9 Cells by Western Blotting with an Anti-Histidine Tag Antibody.

Construct	Expression in HEK293T Cells	Expression in *Sf*9 Cells
AnoACE_1	Strong	In well
AnoACE_2	Strong	Strong
AnoACE_3	Strong	Strong
AnoACE_4	Strong	Weak/in well
AnoACE_5	Moderate	Weak/in well
AnoACE_6	Moderate	Strong
ACER_1	Strong	Weak/in well
ACER_2	Weak	Weak/in well
ACER_3	Strong	Strong
ACER_4	Weak	Strong
AnCE_1	Weak	Moderate
AnCE_2	Moderate	In well
AnCE_3	Moderate	In well
AnCE_4	None	Strong

Table 8.11: Initial Results of Expression Trials in HEK293T and *Sf*9 Cells. Expression levels (where present) are graded as strong, moderate or weak. "In well" refers to the samples that have stayed at the top of the gel in Figure 8.4.

Western blot analysis of the HEK293T media 72 hours after transfection indicates that all but one of the constructs, AnCE_4, was successfully expressed. The level of expression seen varies between the different constructs. AnoACE_1, AnoACE_2, AnoACE_3 and AnoACE_4 appear to have expressed particularly well, whilst only trace expression is seen for ACER_2, ACER_4 and AnCE_1.

The histidine tagged proteins all appear at around the 120 kDa marker on the western blot. This is considerably larger than the expected molecular weights listed in Tables 8.1, 8.2 and 8.3 (~ 70 kDa) and is likely due to the extensive glycosylation that is a characteristic of ACE homologues.The western blot results indicate that expression of the different constructs was much more variable in the *Sf*9 cells compared to the HEK293T cells. There are high levels of AnoACE_2, AnoACE_3, AnoACE_6, ACER_3, ACER_4 and AnCE_4 whilst a moderate amount of AnCE_1 has been detected. The remaining samples give very little, if any, signal at the expected molecular

weight. In a number of cases there is quite a clear band right at the very top of the gel. This may indicate aggregation of some description that has prevented the protein from migrating through the gel. The blot was repeated to rule out any problems with sample preparation and the same result was achieved.

Interestingly, whilst the products of the HEK293T cells migrated with much larger apparent molecular weights than expected, the proteins produced by the *Sf*9 cells appear close to the predicted size of 70 kDa.

8.3.4 Trial Transfections Performed at the University of Bath

Following the successful results achieved with the support of Prof. Ray Owen's group at the OPPF, work towards the expression of some of these constructs at the University of Bath was initiated. To begin with two constructs were selected: AnoACE_2 and ACER_3 and transfection and expression trials using HEK293T cells were performed using two transfection reagents: GeneJuice® (Merck Millipore) and FuGeneHD (Promega). Transfections were also performed simultaneously with a control protein known to show good secreted expression from HEK293T cells.

Expression of target proteins was detected by western blot using an antibody recognising the polyhistidine tag. These blots are shown in Figures 8.5 and 8.6.

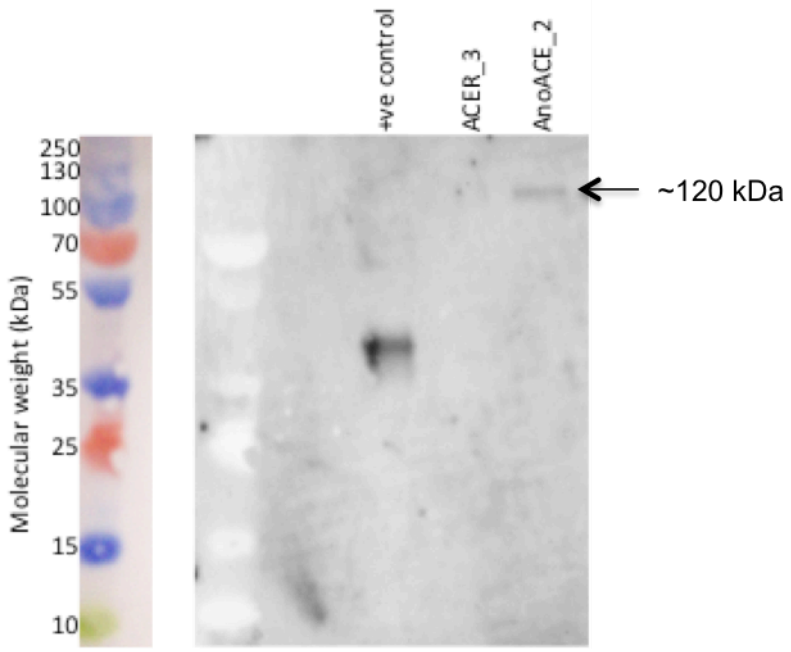


Figure 8.5: Analysis of Secreted ACER_3 and AnoACE_2 Expression by HEK293T Cells Following Transfection. Transfection Performed using GeneJuice® Transfection Reagent.

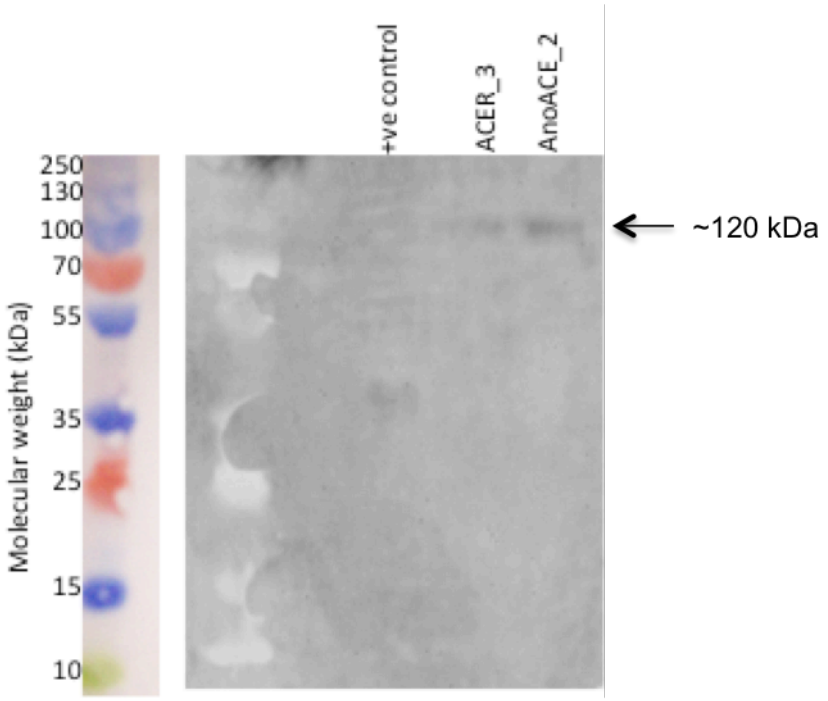


Figure 8.6: Analysis of Secreted ACER_3 and AnoACE_2 Expression by HEK293T Cells Following Transfection. Transfection Performed using FuGeneHD Transfection Reagent.

The western blots indicate that, as expected, the control protein was expressed well with both transfection reagents. There was also some expression of AnoACE_2 with both reagents, but ACER_3 expression is only seen when FuGeneHD is used.

Both AnoACE_2 and ACER_3 appear on the western blots to be 100-130 kDa in size. This is significantly larger than their predicted molecular weights (~70 kDa) but is consistent with the results of the previous expression trials (section 8.3.3, Figures 8.3 and 8.4).

8.4 Discussion

8.4.1 Trials Transfections Performed at the OPPF

Overall the transfections and subsequent expression trials performed at the OPPF were very successful. All 14 constructs showed some expression in either the HEK293T of *Sf*9 cells and many were expressed well in both systems. Whilst AnCE and ACER have previously been expressed using recombinant techniques in *P. pastoris*, this is the first time that expression of AnoACE3, or any of the *Anopheles gambiae* ACE homologues, has been achieved.

8.4.1.1 Comparison of Expression in HEK293T and *Sf*9 Cells

Based on the western blot in Figure 8.3, the best expression in HEK293T cells was seen for AnoACE_1, AnoACE_2, AnoACE_3 and AnoACE_4. Moderate expression of the other constructs was detected with the exception of AnCE_4, which did not appear to show any expression.

Apart from the signal sequence, AnCE_4 is identical to the other three AnCE constructs. Given that these did all show some level of expression in HEK293T cells the lack of AnCE_4 expression is not indicative of expression being incompatible with the HEK293T system, for example due to incorrect post-translational processing or the absence of key chaperones. Furthermore, the AnCE_4 construct displayed very good expression in *Sf*9 cells (Figure 8.4). This indicates that there were no undetected problems with the cloning as the same vector preparations were used for both transfections.

The most probable reason for the lack of detectable expression of the AnCE_4 construct in HEK293T cells is the signal sequence. The AnCE_4 construct uses the Ac64 signal sequence. Interestingly, this is also used for the ACER_4 and AnoACE_6 constructs, both of which show considerably weaker expression in HEK293T cells compared to the other constructs. The Ac64 signal sequence is derived from the gp64 protein of a nuclear polyhedrosis virus infecting *Autographa californica*, a type of moth. This signal sequence is therefore most suitable for use with the Baculovirus expression system in insect cells; hence the absence of AnCE_4 expression in HEK293T cells is neither unexpected nor a cause for concern.

It follows then that the AnoACE_6, ACER_4 and AnCE_4 constructs were amongst those that showed the best expression in the *Sf*9 cells. Many of the other constructs did not show such good expression in the *Sf*9 cells, but again this is most likely due to the signal sequences not being ideal for use in insect systems.

8.4.1.2 Comparison of Expression of the Different AnoACE Constructs

The most interesting expression profiles in many ways are those of AnoACE_1 and AnoACE_2. Neither of these constructs has an additional signal sequence introduced during cloning. The AnoACE_2 construct has the N-terminal signal sequence that is conserved in other secreted ACE homologues, including AnCE and ACER. Perhaps unsurprisingly given that it is an insect protein, the AnoACE_2 construct shows very good secreted expression in the Sf9 cells. The AnoACE_2 expression detected in the

HEK293T cells is also amongst the strongest of all of the constructs, indicating that this signal sequence is able to target the protein for secretion in the HEK293T cells.

The AnoACE_1 construct contains an additional sequence at its N-terminus, upstream of the signal sequence, which is not conserved in other ACE homologues. The function of this sequence is entirely unknown. Interestingly, whilst the AnoACE_1 construct shows good expression in the HEK293T cells, the only evidence of AnoACE_1 expression in the *Sf*9 cells is a small amount of protein that has remained in the wells of the gel in Figure 8.4. The fact that the sample has not been able to migrate down the gel suggests that it has aggregated. Even taking this into consideration, the amount of protein present is small compared to AnoACE_2.

Given that the S*f*9 cells are probably more representative of the *A. gambiae* system than the HEK293T cells, this raises very interesting questions as to the biological role of the N-terminal region. The very limited data available at the moment make it impossible to speculate as to what this may be but it is certainly an area that should be investigated further.

The most important next step in future work though would be to repeat the *Sf*9 expression trials, which, unfortunately time did not permit in this work. It is important to remember that these are the results of a single experiment and any number of factors could have caused anomalous results. It may ultimately transpire that the AnoACE_1 construct does in fact express very well in *Sf*9 cells.

8.4.1.3 Glycosylation Profiles of the Expressed Proteins

The western blots in Figures 8.3 and 8.4 show that the products of the HEK293T expression trials are much larger than those from the *Sf*9 cells. Whilst the *Sf*9 products migrate close to the expected molecular weight of approximately 70 kDa, the HEK293T products appear at close to double the expected size. This is almost certainly due to differences in the glycosylation profiles of the two systems.

This is not unexpected, as the differences in the N-glycosylation pathways in insect and mammalian systems have been well documented. Although insect cells are able to add N-linked glycans to proteins these are not processed beyond the initial high mannose and paucimannose structures. In contrast mammalian systems are able to add much larger, more complex glycan structures terminating in galactose or sialic acid residues (Harrison and Jarvis, 2006; Kost *et al.,* 2005; Nettleship *et al.,* 2010).

The different glycosylation profiles produced by the two systems may prove to be important for future work. The glycosylation profile could affect the activity of the enzymes, which would be an important consideration in any biochemical characterisation. It would also almost certainly have an impact on the crystallisation process. Generally speaking extensive glycosylation makes crystallisation very difficult as it increases heterogeneity. Based on these results it would perhaps be sensible to begin work on this using protein expressed in *Sf*9 cells, where the glycosylation appears to be less extensive. This is also likely to more closely resemble the glycosylation profile that the enzymes would have when expressed in their native systems.

Ultimately, whichever system is chosen, it may well still prove necessary to use glycosidases to reduce some of the glycan chains or even to express mutant forms of the protein with any non-essential glycosylation sites removed.

8.4.2 Trial Transfections Performed at the University of Bath

The long-term aim of this work is to establish conditions under which useful quantities of AnoACE, ACER and AnCE can be expressed for use in crystallisation trials and other experimental work. Given that good expression has been seen in both the mammalian and insect system, the expression system of choice for this aim would probably be *Sf*9 cells, as the potential protein yields are higher than for the HEK293T cells. However due to time constraints, the decision was made to begin work in Bath towards expression in HEK293T cells, as the transfection process is considerably faster.

Two constructs were initially chosen for this: AnoACE_2 and ACER_3. This decision was based on the yields seen in the expression trials performed using HEK293T cells at the OPPF.

This work is still in its early stages, thus the results are only preliminary. As shown in Figures 8.5 and 8.6 there is evidence that expression of both of these proteins in HEK293T cells has been achieved. Based on the western blots the yield is very low, however these are small-scale trials (culture volume 1 ml). Furthermore, the expressed proteins are being secreted from the cells into the media, thus being greatly diluted.

The western blots in Figures 8.5 and 8.6 indicate that whilst some level of AnoACE_2 expression was achieved with both of the transfection reagents tested, ACER_3 expression was only detectable using FuGeneHD. These results were only based on a single experiment, so it would be premature to conclude that FuGeneHD is the most

suitable transfection reagent, but it will be important to consider this observation as the work progresses.

Having established some level of expression the next stages of this work will be to scale-up and perform the transfections on a larger scale. This will most likely involve some work to optimise the transfection process so as to obtain the best possible yields. Following this, the aim will be to purify the proteins and prepare them for crystallisation trials.

8.5 Conclusions

Expression of three insect ACE homologues: AnCE and ACER from *D. melanogaster* and AnoACE3 from *A. gambiae* has been achieved using both mammalian HEK293T and insect *Sf*9 cells. Although AnCE expression has previously been achieved using the yeast *P. pastoris*, this is the first time that expression of an ACE homologue from *A. gambiae* has been achieved in any system. It is therefore an important step forward.

Work has now been initiated to optimise and scale-up the expression process in HEK293T cells with the aim of producing enough protein to begin crystallisation trials. It would also be desirable to begin this process in *Sf*9 cells. This is because this cell line should ultimately be able to produce higher yields of the target proteins. Furthermore the glycosylation of the proteins produced by the *Sf*9 cells appears to be less extensive than of those from the HEK293T cells, meaning that proteins produced in this system may be more amenable to crystallisation.

Chapter 9: General Discussion

9.1 Angiotensin Converting Enzymes and ACE Inhibitors

The role of angiotensin converting enzyme in mammalian blood pressure regulation has been the subject of extensive study for over 60 years, since the initial discovery of the enzyme in 1954 (Skeggs *et al.,* 1954; Acharya *et al.,* 2003). As an integral component of the renin-angiotensin aldosterone system, ACE functions to increase blood pressure, and, as such, has been the target of anti-hypertensive agents, ACE inhibitors, since the 1980s (Patchett *et al.,* 1980).

Surprisingly, the development of ACE inhibitors took place long before the first ACE structure was reported, thus, these drugs were designed with no structural information. Currently ACE inhibitors are amongst the most widely prescribed drugs in the World, and are an effective treatment for hypertension and other cardiovascular ailments. In spite of this however many patients suffer from adverse side effects that range from a persistent dry cough to angioedema, which can be fatal (Bas *et al.,* 2007).

It is now widely accepted that ACE has a whole host of physiological functions outside of blood pressure homeostasis. It is the inhibition of the function of ACE in these other processes that appears to be the cause of many of the adverse effects pertaining from ACE inhibition.

The domain organisation of ACE appears to be central to these additional roles. ACE consists of two homologous domains, arranged in tandem, each of which possesses a functional peptidase active site (Soubrier *et al.,* 1988). These two domains, known as the N- and C-domains, display different substrate specificity. The C-domain has been shown to be primarily responsible for blood pressure homeostasis *in vivo*, whilst the N-domain is able to hydrolyse novel substrates such as GnRH and AcSDKP more efficiently (Georgiadis *et al.,* 2003).

With this in mind, significant efforts are currently being made to develop new inhibitors that specifically target only one of the two domains of ACE. Selective inhibition of the C-domain could target the hypertensive function of ACE, whilst leaving the role of the N-domain in other physiological processes uninhibited, thus reducing or eliminating many of the common side effects arising from treatment with ACE inhibitors.

Similarly, selectively targeting the role of the N-domain outside of the RAAS might be an equally exciting prospect. For example, inhibiting hydrolysis of AcSDKP by the N-

domain has been identified as a potential method for protecting haematopoietic stem cells during some aggressive chemotherapy regimes (Bogden *et al.,* 1991).

The aim of much of the work presented in this thesis was to investigate the structural basis of ACE inhibition and how the interaction of different inhibitors with the active site could be exploited in the design of new, domain selective, ACE inhibitors.

The enzyme AnCE from *D. melanogaster* was the primary tool for doing this. AnCE is a close homologue of ACE, sharing 60 % sequence identity with each of the two catalytic domains. It has previously been used successfully as a model for studying the structural basis of inhibitor binding (Akif *et al.,* 2010).

The main reason for using AnCE is that it is much more amenable to X-ray crystallography studies than human ACE. The challenges associated with working with ACE are highlighted by the fact that almost fifty years elapsed between the initial discovery of this enzyme in 1954 (Skeggs *et al.,* 1954) and the report of the first crystal structure in 2003 (Natesh *et al.,* 2003). The primary reason for this is the extensive glycosylation that is required for the correct processing and function of this enzyme. This makes ACE unsuitable for recombinant expression using traditional prokaryotic systems such as *E. coli*, and makes crystallisation challenging due to the heterogeneity introduced by the glycan chains.

In contrast, whilst AnCE still undergoes some post-translational processing, it can be expressed using the methylotropic yeast *P. pastoris* and appears much more amenable to crystallisation; a condition has been identified which typically yields diffraction quality crystals within 7-10 days. (Akif *et al.,* 2010).

9.2 Bradykinin Potentiating Peptides

Bradykinin potentiating peptides (BPPs), isolated from the venom of the pit viper *B. jararaca,* are potent inhibitors of ACE that formed the basis for the development of the very first inhibitors (Cushman and Ondetti, 1999).

These peptides are characterised by a C-terminal IPP motif that is thought to be central to ACE inhibition; the proline residue at the penultimate position protects the scissile peptide bond from cleavage.

More recently, a number of BPPs have been characterised as selective inhibitors of the C-domain (Cotton *et al.,* 2002) and interactions between the P_2 residue of the peptide and the S_2 subsite have been implicated in this.

In Chapter 3, the crystal structures of AnCE with two BPPs, BPP9a and BPPn, are presented. BPP9a is a selective inhibitor of the C-domain, and the structure with AnCE re-enforces the proposed role of the P_2 residue in this. The P_2 glutamine of BPP9a comes into close proximity of the side chain of AnCE residue Thr387. In the N-domain this is replaced by Arg381, the long, positively charged, side chain of which is highly likely to clash with the P_2 moiety (Figure 3.9).

The structural data presented in Chapter 3 indicate that the P6 position of the peptide may also have an important role in ACE inhibition. The previously reported structure of tACE with BPPb (Masuyer *et al.* 2012) suggests that there is an interaction between the P_6 proline of BPPb and tACE residue Trp59. In the structure of AnCE with BPP9a (Chapter 3) only the C-terminal portion of the peptide is visible, and this is thought to be due to a lack of stabilising interactions with the active site to hold the N-terminal portion of the peptide in a single conformation.

Comparison of the tACE and AnCE structures indicates that Thr40 in AnCE replaces Trp59 in tACE. This substitution means that the interaction with the P_6 tryptophan would not be conserved, which could lead to increased flexibility of this portion of the peptide, resulting in it not being visible in the crystal structure.

Based solely on the structural data presented in this thesis it is not possible to quantify the importance of any interaction between the protein and peptides at the P_6 position, however it would be interesting to expand this study to consider the effects of this interaction on the potency of BPPs as ACE inhibitors. This could include synthesis of new peptides, identical to BPP9a save for the P_6 position, and investigating their potency as inhibitors of both the N- and C-domains.

The second structure presented in Chapter 3 is of AnCE in complex with a novel peptide, BPPn. BPPn was designed to investigate the importance of the conserved P_3 proline residue, which has been mutated to arginine.

Only the conserved IPP motif is visible in the active site of the structure with BPPn (Figure 3.6). It is possible that this is due to cleavage of the peptide by AnCE, however

the fact that there is some residual density beyond the P_1 isoleucine indicates that it is perhaps more likely that the remainder of the peptide is not visible due to its flexibility.

With this in mind, an interaction between the conserved P_3 proline residue, mutated to arginine in BPPn, and a conserved tryptophan residue found in both domains of ACE and in AnCE has been proposed as having an important role in stabilising the peptide in the active site (Figure 3.10).

The structural data alone however are insufficient to determine whether this has an effect on the potency of the peptide as an ACE inhibitor. Further work should now be carried out to quantify the inhibitory activity of BPPn.

If it does transpire that this mutation has had a deleterious effect on inhibitor potency, then maintaining and indeed improving this interaction, for example by addition of a P_3 tryptophan moiety, would be an important factor to consider in future inhibitor design.

9.3 Spinorphin and Tynorphin

Spinorphin and tynorphin are both peptides that have been identified as ACE inhibitors. Spinorphin is a natural peptide derived from the enzymatic cleavage of the β-chain of haemoglobin whilst tynorphin is a related, truncated peptide that is not such a potent ACE inhibitor (Nishimura and Hazato, 1993; Yamamoto *et al.,* 2000).

Interestingly, the sequences of spinorphin and tynorphin bear very little resemblance to the BPPs, thus the basis of their action as ACE inhibitors has not been clear.

The aim of the work presented in Chapter 4 was to use AnCE as a model to study the binding of spinorphin and tynorphin to ACE and hence to learn more about their actions as ACE inhibitors.

To this end a number of sets of diffraction data were collected for the two complexes, which were used to produce six crystal structures, three for AnCE with each peptide.

As discussed in section 4.4, interpretation of the electron density maps for these structures was not always straightforward. In all cases clear residual difference density was visible in the active site that was thought to be due to peptide binding, but it was not possible to model the full-length peptides into the active sites of any of the structures.

The conclusion was drawn that AnCE was cleaving both spinorphin and tynorphin into multiple fragments, and it was these that were present in the active site of the crystal structures.

What is not clear from this study alone is whether this cleavage is occurring on a timescale that is biologically relevant, or whether it is an artefact from crystallisation, where AnCE and the inhibitors are held at high concentrations in a constrained conformation for an extended period of time.

This needs to be investigated further. The most appropriate method for this is likely to be incubating spinorphin and tynorphin with ACE for different time periods and using a combination of HPLC and mass spectrometry to monitor the persistence of the full-length peptides and to identify any smaller fragments that are produced.

9.4 Dipeptide Inhibitors

The structures obtained of AnCE with spinorphin and tynorphin indicate that much smaller peptide fragments might be responsible for the observed inhibitory activity of these peptides. This correlates well with recent reports of the inhibitory actions of small dipeptides towards ACE.

Tryptophan-containing dipeptides have recently been characterised as ACE inhibitors and, perhaps more importantly, dipeptides with a C-terminal tryptophan residue have been shown to be selective inhibitors of the C-domain (Lunow *et al*., 2013; Lunow *et al.,* 2015).

In Chapter 5, the crystal structures of AnCE bound to three such inhibitory dipeptides are presented. The data indicate that these dipeptides all bind to the AnCE active site and as such are expected to behave as competitive inhibitors. The mode of binding is the same as seen with other peptide inhibitors such that conserved interactions between the C-terminus of the dipeptide and the active site are made.

The C-terminal tryptophan residue occupies the S_2' hydrophobic pocket and this is likely to be the main driving force for domain selectivity, as the hydrophobicity of this environment is not retained to such an extent in the N-domain. Interestingly, an additional interaction was identified between the tryptophan and AnCE residue Gln266 that would neither be conserved in the N- nor the C-domain.

The importance of the P_2' residue in domain selectivity has been reported previously. Based on the structures presented in Chapter 5, it appears that the P_1' residue may also have an important role. An aliphatic residue at this position favours C-domain inhibition due to a stabilising interaction with a valine residue that is not conserved in the N-domain. Conversely, an interaction was identified between the side chain of the P_1' tyrosine residue and a glutamate reside in AnCE. This interaction would not be conserved in the C-domain but could be in the N-domain. This is an important feature that could potentially be used in the development of N-domain selective inhibitors.

Overall the work with the dipeptides and AnCE illustrates how very subtle differences confined to a relatively small area of the active site can have a profound effect on inhibitor potency. It is important to consider this alongside the other interactions along the catalytic channel that have been identified through the studies on BPPs when considering the design and development of new inhibitors.

9.5 A New AnCE Crystal Structure

All of the structural data presented in Chapters 3, 4 and 5 were obtained using AnCE crystals grown from the same condition: 1.2-1.3 M sodium citrate, 0.1 M HEPES, pH7.5. Whilst this crystallisation condition has been used very successfully on these occasions, a number of circumstances have arisen where it has not been ideal.

Part of the focus on developing new ACE inhibitors is to consider natural compounds that display some inhibitory activity and to consider how features of these could be used in inhibitor design. Structural studies are therefore essential to understand the basis of their interaction with ACE.

In some cases this has proved to be very challenging, and crystallisation experiments have repeatedly failed to yield co-crystals of AnCE and the inhibitor. In all cases only a citrate ion, arising from the crystallisation condition, is present in the active site. It is thought that the weaker inhibitors are unable to displace the citrate ion, hampering attempts at structure-based studies.

In an attempt to overcome this a new crystallisation condition has been identified and using crystals grown in this condition a new structure of AnCE in a different crystal form has been elucidated. Careful analysis of this structure, presented in Chapter 6, indicates that it is largely identical to existing AnCE structures and as such crystals obtained from this condition can confidently be used in future experiments.

Whilst there is no citrate ion in the active site of the new structure, a malate ion, also thought to arise from the crystallisation condition, is present. Unlike the citrate ion, the malate ion does not appear to interact with the catalytic zinc ion (Figure 6.6). This suggests that it may not be so tightly bound and hence could be more easily displaced by weaker inhibitors. This could make this condition more suitable for studying the binding of novel inhibitors, and work has already been initiated to try and achieve this.

9.6 ACE Homologues from *A. gambiae*

In addition to AnCE, many other insect ACE homologues have been identified. ACE homologues have been identified in every insect genome sequenced to date, indicating that they must have important biological roles (Isaac *et al.,* 2007). Given that insects do not have a circulatory system resembling that found in mammals, studying the roles of insect ACE-like enzymes could help to better our understanding of the roles of ACE outside of the RAAS.

Furthermore, there is a significant amount of evidence to suggest that insect ACE-like enzymes have an important conserved role in reproduction. With this in mind, the prospect of developing ACE inhibitors that specifically target insect enzymes and using these in insecticides has been considered.

The ACE homologues from the mosquito *A. gambiae* (AnoACEs) are particularly interesting for a number of reasons. Firstly, the genome appears to code for ten different ACE-like enzymes, the highest number found in any organism thus far. In addition to this, one of these enzymes, AnoACE9, is predicted to consist of two catalytic domains arranged in tandem, as seen in sACE, which indicates that this arrangement has been selected for on more than one occasion, hence must be advantageous in some way (Burnham *et al.,* 2005). Finally, the role of *A. gambiae* as a vector for the transmission of *P. falciparum*, a causative agent of malaria, makes targeting AnoACEs in insecticides an exciting long term prospect.

The aim of the work involving AnoACEs that is presented in this thesis was to focus on one of these enzymes, AnoACE3, and to work towards its expression and purification. The longer term aim of this being to use the purified enzyme in structural and activity based studies to learn more about its biological functions.

AnoACE3 shares 54 % sequence identity with AnCE and all three of the N-linked glycosylation sites present in AnCE are predicted to be conserved. First attempts to

express AnoACE3 were made using the methylotropic yeast *P. pastoris*, which has been used successfully for the recombinant expression of AnCE.

As detailed in Chapter 7, in spite of significant efforts, no detectable AnoACE3 expression was achieved. There are many possible reasons why this may have been the case, but perhaps the most likely explanation is that the unique nature of the AnoACE3 N-terminus impeded expression in some way.

AnoACE3 has a unique sequence at its N-terminus that is not conserved in AnCE, the function of which is entirely unknown. It is possible that this region undergoes extensive post-translational modifications, which could not be reproduced in *P. pastoris*, that are in someway essential for stability.

Both the AnCE and AnoACE3 constructs used for *P. pastoris* expression contain an N-terminal α-factor signal sequence that targets the proteins for secreted expression. This sequence must be cleaved from the protein by the enzyme Kex2 so that the protein can move from the Golgi apparatus into vesicles for secretion. The different nature of the N-terminus of the AnoACE3 construct may have impeded this process, leading any protein that was expressed to be "stuck" at this stage of the pathway.

Ultimately, the decision was made to try and express AnoACE3 using higher eukaryotic systems (Chapter 8). This work was completed concurrently with AnCE and a second *D. melanogaster* ACE homologue, ACER.

Expression trials were performed using both HEK293T and S*f*9 cells and there was evidence of expression of all three proteins using both systems. This is an important step forward and is the first time that recombinant expression of an ACE homologue from *A. gambiae* has been achieved.

The work on the AnoACEs is in its very early stages and there is a lot of scope to build on the results that have been achieved. The immediate task is now to scale-up the expression. Ideally this should be performed in both the HEK293T and the S*f*9 cells. This is because, based on the molecular weights of the species produced (Figures 8.3 and 8.4), the glycosylation profiles of the enzymes produced by the two cell lines seems to be quite different. It would be really interesting to see if this has any effect on the activity and stability of the enzymes.

Once pure protein has been obtained it should then be possible to conduct a series of activity assays to try and determine what some of the substrates of AnoACE3 are. A good place to start here would be considering the activity of the enzyme towards angiotensin I and bradykinin, which are both well characterised as substrates of ACE and AnCE. Studying the action of known ACE inhibitors towards AnoACE could also provide more information about the biochemical properties of this enzyme.

Structural data would be highly useful to complement this data, providing insights into the reasons for the observed substrate and inhibitor binding preferences. This work may be somewhat challenging due to the extensive post-translational modifications evident by the sizes of the expression products seen in Figures 8.3 and 8.4. It may be necessary to use glycosidases to cut back the glycan chains to some extent and possibly, as was the case for sACE, to mutate any glycosylation sites that are not essential for stability and activity (Anthony *et al.,* 2010; Gordon *et al.,* 2003).

9.7 Concluding Remarks

In conclusion, the work presented here includes some important progress in the understanding of angiotensin converting enzymes from mammals and insects. By using the *D. melanogaster* ACE homologue AnCE as a model in structural studies to investigate the binding of different inhibitors, a number of interactions with the active site have been identified that could potentially be used in the development of new inhibitors, for which there is clearly a need.

In addition to this, the structural basis of inhibition by small dipeptide fragments has been investigated and illustrates how it may be possible to create potent, domain specific inhibitors, by enhancing interactions with a relatively small area of the active site.

Finally, recombinant expression of an ACE homologue from *A. gambiae* has been reported for the first time. Future work should now involve both structure and activity based studies on this enzyme to learn more about the function of ACE homologues in *A. gambiae*.

References

Acharya, K.R., Sturrock, E.D., Riordan, J.F., Ehlers, M.R.W., 2003. ACE revisited: a new target for structure-based drug design. *Nature Reviews Drug Discovery* 2, 891-902.

Acharya, K.R., Sturrock, E.D., 2003. Deglycosylation, processing and crystallization of human testis angiotensin-converting enzyme. *Biochemical Journal* 371, 437-442.

Ahmad, M., Hirz, M., Pichler, H., Schwab, H., 2014. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology* 98, 5301-5317.

Amunts, A., Brown, A., Bai, X-C., Llácer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H., Ramakrishnan, V., 2014. Structure of the yeast mitochondrial large ribosomal subunit. *Science* 343, 1485-1489.

Akif, M., Georgiadis, D., Mahajan, A., Dive, V., Sturrock, E.D., Isaac, R.E., Acharya, K.R., 2010. High-resolution crystal structures of *Drosophila melanogaster* angiotensin-converting enzyme in complex with novel inhibitors and antihypertensive drugs. *Journal of molecular biology* 400, 502-517.

Akif, M., Ntai, I., Sturrock, E.D., Isaac, R.E., Bachmann, B.O., Acharya, K.R., 2010. Crystal structure of a phosphonotripeptide K-26 in complex with angiotensin converting enzyme homologue (AnCE) from *Drosophila melanogaster*. *Biochemical and Biophysical Research Communications* 398, 532-536.

Akif, M., Masuyer, G., Schwager, S.L.U., Bhuyan, B.J., Mugesh, G., Isaac, R.E., Sturrock, E.D., Acharya, K.R., 2011. Structural characterization of angiotensin I-converting enzyme in complex with a selenium analogue of captopril. *FEBS Journal* 278, 3644-3650.

Akif, M., Masuyer, G., Bingham, R.J., Sturrock, E.D., Elwyn, R.E., Acharya, K.R., 2012. Structural basis of peptide recognition by the angiotensin-1 converting enzyme homologue AnCE from *Drosophila melanogaster*. *FEBS Journal* 279, 4525-4534.

Almquist, R.G., Chao, W-R., Ellis, M.E., Johnson, H.L., 1980. Synthesis and biological activity of a ketomethylene analogue of a tripeptide inhibitor of angiotensin converting enzyme. *Journal of Medicinal Chemistry* 23, 1392-1398.

Anthony, C.S., Corradi, H.R., Schwager, S.L.U., Redelinghuys, P., Georgiadis, D., Dive, V., Acharya, K.R., 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. *The Journal of Biological Chemistry* 285, 35685-35693.

Anthony, C.S., Masuyer, G., Sturrock, E.D., Acharya, K.R., 2012. Structure based drug design of angiotensin-I converting enzyme inhibitors. *Current Medicinal Chemistry* 19, 845-855.

Azizi, M., Rousseau, A., Ezan, E., Guyene, T-T., Michelet, S., Grognet, J-M., Lenfant, M., Corvol, P., Ménard, J., 1996. Acute angiotensin converting enzyme inhibition increases the plasma level of the natural stem cell regulator *N*-acetyl-seryl-aspartyl-lysyl-proline. *Journal of Clinical Investigations* 97, 839-844.

Bas, M., Adams, V., Suvorava, T., Niehues, T., Hoffmann, T.K., Kojda, G., 2007. Nonallergic angioedema: role of bradykinin. *Allergy* 62, 842-856.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D* 67, 271-281.

Berejnov, V., Husseini, N.S., Alsaied, O.A., Thorne, R.E., 2006. Effect of cryoprotectant concentration and cooling rate on vitrification of aqueous solutions. *Journal of Applied Crystallography* 39, 244-251.

Bernstein, K.E., Shen, X.Z., Gonzalez-Villalobos, R.A., Billet, S., Okwan-Duodu, D.O., Ong, F.S., Fuchs, S., 2011. Different *in vivo* functions of the two catalytic domains of angiotensin-converting enzyme (ACE). *Current Opinion in Pharmacology* 11, 105-111.

Bezerra, G. A., Dobrovetsky, E., Viertlmayr, R., Dong, A., Binter, A., Abramic, M., Macheroux, P., Dhe-Paganon, S., Gruber, K., 2012. Entropy-driven binding of opioid peptides induces a large domain motion in human dipeptidyl peptidase III. *Proceedings of the National Academy of Science USA* 109, 6525-6530.

Bingham, R.J., Dive, V., Philips, S.E.V., Shirras, A.D., Isaac, R.E., 2005. Structural diversity of angiotensin-converting enzyme insights from structure-activity comparisons of two *Drosophila* enzymes. *FEBS Journal* 273, 362-373.

Bogden, A.E., Carde, P., de Paillette, E.D., Moreau, J.P., Tubiana, M., Frindel, E., 1991. Amelioration of chemotherapy-induced toxicity by cotreatment with AcSDKP, a tetrapeptide inhibitor of haematopoietic stem cell proliferation. *Annals of the New York Academy of Sciences* 628, 126-139.

Bond, C. S., Schüttelkopf, A. W., 2009. *ALINE*: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallographica Section D* 65, 510-512.

Bonnet, D., Lemoine, F.M., Khoury, E., Pradelles, P., Najman, A., Guigon, M., 1992. Reversible inhibitory effects and absence of toxicity of the tetrapeptide acetyl-N-Ser-Asp-Lys-Pro (AcSDKP) in human long-term bone marrow culture. *Experimental Hematology* 20, 1165-1169.

Bragg, W.H., Bragg, W.L., 1913. The reflection of x-rays by crystals. *Proceedings of the Royal Society of London* 88, 428-438.

Brantl, V., Gramsch, C., Lottspeich, F., Mertz, R., Jaeger, K-H., Herz, A., 1986. Novel opioid peptides derived from haemoglobin: haemorphins. *European Journal of Pharmacology* 123, 309-310.

Brew, K., Dinakarpandian, D., Nagase, H., 2000. Tissue inhibitors of metalloproteinases: evolution, structure and function. *Biochimica et Biophysica Acta* 1477, 267-283.

Brünger, A.T., 1992. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-475.

Brünger, A.T., 1993. Assessment of phase accuracy by cross validation: the free *R* value. Methods and applications. *Acta Crystallographica Section D* 49, 24-36.

Burnham, S., Smith, J.A., Lee, A.J., Isaac, R.E., Shirras, A.D., 2005. The angiotensin-converting enzyme (ACE) gene family of *Anopheles gambiae*. *BMC Genomics* 6, 172-185.

Byers, L.D., Wolfenden, R., 1973. Binding of the by-product analogue benzylsuccinic carboxypeptidase A. *Biochemistry* 12, 2070-2078.

Caldwell, P.R.B., Seegal, B.C., Hsu, K.C., Das, M., Soffer, R.L., 1976. Angiotensin-converting enzyme: vascular endothelial localisation. *Science* 191, 1050-1051.

Carhan, A., Tang, K., Shirras, C.A., Shirras, A.D., Isaac, R.E., 2011. Loss of angiotensin-converting enzyme-related (ACER) peptidase disrupts night-time sleep in adult *Drosophila melanogaster*. *The Journal of Experimental Biology* 214, 680-686.

Chapman, M.S., Somasundaram, T., 2010. De-icing: recovery of diffraction intensities in the presence of ice rings. *Acta Crystallographica Section D* 66, 741-744.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* 66, 12-21.

Coates, D.R., Isaac, R.E., Cotton, J., Siviter, R., Williams, T.A., Shirras, A., Corvol, P., Dive, V., 2000. Functional conservation of the active sites of human and *Drosophila* angiotensin I-converting enzyme. *Biochemistry* 39, 8963-8969.

Corradi, H.R., Schwager, S.L.U., Nchinda, A.T., Sturrock, E.D., 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. *Journal of Molecular Biology* 357, 964-974.

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

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

Corvol, P., Williams, T.A., Soubrier, F., 1995. Peptidyl dipeptidase A: angiotensin I-converting enzyme. *Methods in Enzymology* 248, 283-305.

Cotton, J., Hayashi, M.A.F., Cuniasse, P., Vazeux, G., Ianzer, D., De Camargo, A.C.M., Dive, V., 2002. Selective inhibition of the C-domain of angiotensin I converting enzyme by bradykinin potentiating peptides. *Biochemistry* 41, 6065-6071.

Cushman, D.W., Cheung, H.S., Sabo, E.F., Ondetti, M.A., 1977. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry* 16, 5484-5491.

Cushman, D.W., Ondetti, M.A., 1999. Design of angiotensin converting enzyme inhibitors. *Nature Medicine* 5, 1110-1112.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35, 375-383.

Deddish, P.A., Jackman, H.L., Skidgel, R.A., Erdös, E.G., 1997. Differences in the hydrolysis of encephalin congeners by the two domains of angiotensin converting enzyme. *Biochemical Pharmacology* 53, 1459-1463.

Deddish, P.A., Marcic, B., Jackman, H.L., Wang, H-Z., Skidgel, R.A., Erdös, E.G., 1998. N-domain specific substrate and C-domain inhibitors of angiotensin-converting enzyme: angiotensin-(1-7) and keto-ACE. *Hypertension* 31, 912-917.

Diederichs, K., Karplus, P.A., 1997. Improved *R*-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology* 4, 269-275.

Dive, V., Cotton, J., Yiotakis, A., Michaud, A., Vassiliou, S., Jiracek, J., Vazux, G., Chauvet, M., Cuniasse, P., Corvol, P., 1999. RXP407, a phosphinic peptide is a potent inhibitor of angiotensin I converting enzyme able to differentiate between its two active sites. *Proceedings of the National Academy of Science USA* 96, 4330-4335.

Dowell, L.G., Rinfret, A.P., 1960. Low-temperature forms of ice as studied by x-ray diffraction. *Nature* 188, 1144-1148.

Ehlers, M.R.W., Fox, E.A., Strydom, D.J., Riordan, J.F., 1989. Molecular cloning of human testicular angiotensin-converting enzyme: the testis isozyme is identical to the C-terminal half of endothelial angiotensin-converting enzyme. *Proceedings of the National Academy of Sciences USA* 86, 7741-7745.

Ellis, S.B., Brust, P.F., Koutz, P.J., Waters, A.F., Harpold, M.M., Gingeras, T.R., 1985. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast, *Pichia pastoris*. *Molecular and Cellular Biology* 5, 1111-1121.

Ekbote, U., Coates, D., 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 Letters* 455, 219-222.

Ekbote, U., Looker, M., Isaac, R.E., 2003a. ACE inhibitors reduce fecundity in the mosquito, *Anopheles stephensi*. *Comparative Biochemistry and Physiology Part B* 134, 593-598.

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

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of C*oot*. *Acta Crystallographica Section D* 66, 486-501.

Evans, P.R., 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D* 62, 72-82.

Evans, P.R., 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D* 67, 282-292.

Ferreira, S.H., 1965. A bradykinin-potentiating factor (BPF) present in the venom of *Bothrops jararaca. British Journal of Pharmacology* 24, 163-169.

Ferreira, S.H., Greene, L.J., Alabaster, V.A., Bakhle, Y.S., Vane, J.R., 1970. Activity of various fractions of bradykinin potentiating factor against angiotensin I converting enzyme. *Nature* 225, 379-380.

Fischer, N., Neumann, P., Konevega, A.L., Bock, L.V., Ficner, R., Rodnina, M.V., Stark, H., 2015. Structure of the *E. coli* ribosome-EF-Tu complex at < 3 Å resolution by C_S-corrected cryo-EM. *Nature* 520, 567-570.

Fruitier-Arnaudin, I., Cohen, M., Bordenave, S., Sannier, F., Piot, J-M., 2002. Comparative effects of angiotensin IV and two haemorphins on angiotensin-converting enzyme activity. *Peptides* 23, 1465-1470.

Fuchs, S., Xiao, H.D., Cole, J.M., Adams, J.W., Frenzel, K., Michaud, A., Zhao, H., Keshelavat, G., Capecchi, M.R., Corvol, P., Bernstein, K.E., 2004. Role of the N-terminal catalytic domain of angiotensin-converting enzyme investigated by targeted inactivation in mice. *The Journal of Biological Chemistry* 279, 15946-15953.

Fuchs, S., Frenzel, K., Hubert, C., Lyng, R., Muller, L., Michaud, A., Xiao, H.D., Adams, J.W., Capecchi, M.R., Corvol, P., Shur, B.D., Bernstein, K.E., 2005. Male fertility is dependent on dipeptidase activity of testis ACE. *Nature Medicine* 11, 1140-1142.

Fuchs, S., Xiao, H.D., Hubert, C., Michaud, A., Campbell, D.J., Adams, J.W., Capecchi, M.R., Corvol, P., Bernstein, K.E., 2008. Angiotensin-converting enzyme C-terminal catalytic domain is the main site of angiotensin I cleavage in vivo. *Hypertension* 51, 267-274.

Garman, E.F., Mitchell, E.P., 1996. Glycerol concentrations required for cryoprotection of 50 typical protein crystallisation solutions. *Journal of Applied Crystallography* 29, 584-587.

Georgiadis, D., Beau, F., Czarny, B., Cotton, J., Yiotakis, A., 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. *Circulation Research* 93, 148-154.

Georgiadis, D., Cunlasse, P., Cotton, J., Yiotakis, A., Dive, V., 2004. Structural determinants of RXPA380, a potent and highly selective inhibitor of the angiotensin-converting enzyme C-domain. *Biochemistry* 43, 8048-8054.

Gonzalez-Villalobos, R.A., Shen, X.Z., Bernstein, E.A., Janjulia, T., Taylor, B., Giani, J.F., Blackwell, W-L.B., Shah, K.H., Shi, P.D., Fuchs, S., Bernstein, K.E., 2013. Rediscovering ACE: novel insights into the many roles of the angiotensin-converting enzyme. *Journal of Molecular Medicine* 91, 1143-1154.

Gordon, K., Redelinghuys, P., Schwager, S.U., Ehlers, M.E., Papageorgiou, A.C., Natesh, R., Acharya, K.R., Sturrock, E.D., 2003. Deglycosylation, processing and crystallisation of human testis angiotensin-converting enzyme. *Biochemical Journal* 371, 437-442.

Green, D.W., Ingram, V.M., Perutz, M.F., 1954. The structure of haemoglobin. 4. Sign determination by the isomorphous replacement method. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical sciences* 225, 287-307.

Guerrero, L., Castillo, J., Quiñones, M., Garcia-Vallvé, S., Arola, L., Pujadas, G., Muguerza, B., 2012. Inhibition of angiotensin-converting enzyme activity by flavonoids: structure-activity relationship studies. *PLOS ONE* 7, 1-11.

Gupta, R., Brunak, S., 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pacific Symposium on Biocomputing* 7, 310-322.

Hardy, L.W., Malikayil, A., 2003. The impact of structure-guided drug design on clinical agents. *Current Drug Discovery,* 15-20.

Harrison, R.L., Jarvis, D.L., 2006. Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. *Advances in Virus Research* 68, 159-161.

Hendrickson, W.A., Smith. J.L., Phizackerley, R.P., Merritt, E.A., 1988. Crystallographic structure analysis of lamprey haemoglobin from anomalous dispersion of synchrotron radiation. *Proteins* 4, 77-88.

Hendrickson, W.A., Horton, J.R., LeMaster, D.M., 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal* 9, 1665-1672.

Hirose, S., Kawamura, Y., Yokota, K., Kuroita, T., Natsume, T., Komiya, K., Tsutsumi, T., Suwa, Y., Isogai, T., Goshima, N., Noguchi, T., 2011. Statistical analysis of features associated with protein expression/solubility in an *in vivo Escherichia coli* expression system and a wheat germ cell-free expression system. *The Journal of Biochemistry* 150, 73-81.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., Salzberg, S.L., Loftus, B., Yandell, M., Majoros, W.H., Rusch, D.B., Lai, Z., Kraft, C.L., Abril, J.F., Anthouard, V., Arensburger, P., Atkinson, P.W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G.K., Chrystal, M.A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C.A., Flanigan, M., Grundschober-Freimoser, A., Friedi, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M.E., Hladun, S.L., Hogan, J.R., Hong, Y.S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J.J., Lobo, N.F., Lopez, J.R., Malek, J.A., McIntosh, T.C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S.D., O'Brochta, D.A., Pfannkoch, C., Qi, R., Regier, M.A., Remington, K., Shao, H., Sharakhova, M.V., Sitter, C.D., Shetty, J., Smith, T.J., Strong, R., Sun, J., Thomasova, D., Ton, L.Q., Topalis, P., Tu, Z., Unger, M.F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K.J., Wortman, J.R., Wu, M., Yao, A., Zdobnov, E.M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S.C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C.W., Louis, C.,

Kalush, F., Mural, R.J., Myers, E.W., Adams, M.D., Smith, H.O., Broder, S., Gardner, M.J., Fraser, C.M., Birney, E., Bork, P., Brey, P.T., Venter, J.C., Weissenbach, J., Kafatos, F.C., Collins, F.H., Hoffman, S.L., 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129-149.

Hooper, N.M., Keen, J., Pappin, D.J.C., Turner, A.J., 1987. Pig kidney angiotensin converting enzyme: purification and characterisation of amphipathic and hydrophilic forms of the enzyme establishes *C*-terminal anchorage to the plasma membrane. *Biochemical Journal* 247, 85-93.

Hooper, N.M., Turner, A.J., 1987. Isolation of two differentially glycosylated forms of peptidyl-dipeptidase A (angiotensin converting enzyme) from pig brain: a re-evaluation of their role in neuropeptide metabolism. *Biochemical Journal* 241, 625-633.

Houard, X., Williams, T.A., Michaud, A., Dani, P., Isaac, R.E., Shirras, A.D., Coates, D., Corvol, P., 1998. The *Drosophila melanogaster*-related angiotensin-I-converting enzymes Acer and Ance. *European Journal of Biochemistry* 257, 599-606.

Hubert, C., Houot, A-M., Corvol, P., Soubrier, F., 1991. Structure of the angiotensin I-converting enzyme gene: two alternate promoters correspond to evolutionary steps of a duplicated gene. *The Journal of Biological Chemistry* 266, 15377-15383.

Hurst, D., Rylett, C.M., Isaac, R.E., Shirras, A.D., 2003. The *Drosophila* angiotensin-converting enzyme homologue Ance is required for spermiogenesis. *Developmental Biology* 254, 238-247.

Isaac, R.E., Schoofs, L., Williams, T.A., Corvol, P., Veelaert, D., Sajid, M., Coates, D., 1998. Toward a role for angiotensin-converting enzyme in insects. *Annals New York Academy of Sciences* 839, 288-292.

Isaac, R.E., Ekbote, U., Coates, D., Shirras, A.D., 1999. Insect angiotensin-converting enzyme. A processing enzyme with broad substrate specificity and a role in reproduction. *Annals New York Academy of Sciences* 897, 342-347.

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

Ishimoto, H., Lark, A., Kitamoto, T., 2012. Factors that differentially affect daytime and nighttimes sleep in *Drosophila melanogaster*. *Frontiers in Neurology* 3, 1-5.

Israili, Z. H., Hall, W. D., 1992. Cough and angioneurotic edema associated with angiotensin-converting enzyme inhibitor therapy. A review of the literature and pathophysiology. *Annals of Internal Medicine* 117, 234-242.

Jaspard, E., Wei, L., Alhenc-Gelas, F., 1993. Differences in the properties and enzymatic specificities of the two active sites of angiotensin I-converting enzyme (kininase II). *The Journal of Biological Chemistry* 268, 9496-9503.

Junot, C., Gonzales, M-F., Ezan, E., Cotton, J., Vazeux, G., Michaud, A., Azizi, M., Vassiliou, S., Yiotakis, A., Corvol, P., Dive, V., 2001. RXP 407, a selective inhibitor of the N-domain of angiotensin I-converting enzyme, blocks in vivo the degradation of hemoregulatory peptide acetyl-Ser-Asp-Lys-Pro with no effect on angiotensin I hydrolysis. *The Journal of Pharmacology and Experimental Therapeutics* 297, 606-611.

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

Kleywegt, G.J., Jones, T.A., 1995. Where freedom is given, liberties are taken. *Structure* 3, 535-540.

Kleywegt, G.J., Brünger, A.T., 1996. Checking your imagination: applications of the free R value. *Structure* 4, 897-904.

Kohlstedt, K., Shoghi, F., Muller-Esterl, W., Busse, R., Fleming, I., 2002. CK2 phosphorylates the angiotensin-converting enzyme and regulates its retention in the endothelial cell plasma membrane. *Circulation Research* 91, 749-756.

Kohlstedt, K., Busse, R., Fleming, I., 2005. Signaling via the angiotensin-converting enzyme enhances the expression of cyclooxygenase-2 in endothelial cells. *Hypertension* 45, 126-132.

Kost, T.A., Condreay, J.P., Jarvis, D.L., 2005. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nature Biotechnology* 23, 567-575.

Kouz, P.J., Davis, G.R., Stillman, C., Barringer, K., Cregg, J.M., Thill, G., 1989. Structural comparison of the *Pichia pastoris* alcohol oxidase genes. *Yeast* 5, 167-177.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 37, 774-797.

Krissinel, E., 2010. Crystal contacts as nature's docking solutions. *Journal of Computational Chemistry* 31, 133-143.

Kuoppala, A., Lindstedt, K.A., Sarrinen, J., Kovanen, P.T., Kokkonen, J.O., 2000. Inactivation of bradykinin by angiotensin-converting enzyme and by carboxypeptidase N in human plasma. *American Journal of Physiology: Heart and Circulatory Physiology* 278, 1069-1074.

Lamango, N.S., Isaac, R.E., 1994. Identification and properties of a peptidyl dipeptidase in the housefly, *Musca domestica*, that resembles mammalian angiotensin-converting enzyme. *Biochemical Journal* 299, 651-657.

Lamango, N.S., Sajid, M., Isaac, R.E., 1996. The endopeptidase activity and the activation by Cl^- of angiotensin-converting enzyme is evolutionarily conserved: purification and properties of an angiotensin-converting enzyme from the housefly, *Musca domestica*. *Biochemical Journal* 314, 639-646.

Lantz, I., Glämsta, E-L., Talbäck, L., Nyberg, F., 1991. Haemorphins derived from haemoglobin have an inhibitory action on angiotensin converting enzyme activity. *FEBS Journal* 287, 39-41.

Laskowski, R.A., Swindells, M.B., 2011. Ligplot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modelling* 51, 2778-2786.

Lenfant, M., Wdzieczak-Bakalat, J., Guittet, E., Promef, J-C., Sotty, D., Emillia, F., 1989. Inhibitor of haematopoietic pluripotent stem cell proliferation: purification and determination of its structure. *Proceedings of the National Academy of Sciences USA* 86, 779-782.

Leslie, A.G.W., Powell, H.R., 2007. Processing diffraction data with mosflm. *Evolving Methods for Macromolecular Crystallography* 245, 41-51.

Li, P., Xiao, H.D., Xu, J., Ong, F.S., Kwon, M., Roman, J., Gal, A., Bernstein, K.E., Fuchs, S., 2010. Angiotensin-converting enzyme N-terminal inactivation alleviates bleomycin-induced lung injury. *American Journal of Pathology* 177, 1113-1121.

Liao, F-T., Chang, C-Y., Su, M-T., Kuo, W-C., 2014. Necessity of angiotensin-converting enzyme-related gene for cardiac functions and longevity of *Drosophila melanogaster* assessed by optical coherence tomography. *Journal of Biomedical Optics* 19, 1-6.

Loizzo, M.R., Said, A., Tundis, R., Rashed, K., Statti, G.A., Hufner, A., Menichini, F., 2007. Inhibition of angiotensin converting enzyme (ACE) by flavonoids isolated from *Ailanthus excels* (Roxb) (Simaroubaceae). *Phytotherapy Research* 21, 32-36.

Lõoke, M., Kristjuhan, K., Kristjuhan, A., 2011. Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques* 50, 325-328.

Lunow, D., Kaiser, S., Brückner, S., Gotsch, A., Henle, T., 2013. Selective release of ACE-inhibiting tryptophan-containing dipeptides from food proteins by enzymatic hydrolysis. *European Food Research and Technology* 237, 27-37.

Lunow, D., Kaiser, S., Rückriemen, J., Pohl, C., Henle, T., 2015. Tryptophan-containing dipeptides are C-domain selective inhibitors of angiotensin converting enzyme. *Food Chemistry* 166, 596-602.

Macours, N., Hens, K., 2004. Zinc-metalloproteases in insects: ACE and ECE. *Insect Biochemistry and Molecular Biology* 34, 501-510.

Marczak, E.D., Usui, H., Fujita, H., Yang, Y., Yokoo, M., Lipkowski, A.W., Yoshikawa, M., 2003. New antihypertensive peptides isolated from rapeseed. *Peptides* 24, 791-798.

Martin, M., Wellner, A., Ossowski, I., Henle, T., 2008. Identification and quantification of inhibitors for angiotensin-converting enzyme in hypoallergenic infant milk formulas. *Journal of Agricultural and Food Chemistry* 56, 6333-6338.

Masuyer, G., Schwager, S.L.U, Sturrock, E.D., Isaac, R.E., Acharya, K.R., 2012. Molecular recognition and regulation of human angiotensin-I converting enzyme (ACE) activity by natural inhibitory peptides. *Scientific Reports* 2, 1-10.

Masuyer, G., Akif, M., Czarny, B., Beau, F., Schwager, S.L.U., Sturrock, E.D., Isaac, R.E., Dive, V., Acharya, K.R., 2014. Crystal structures of highly specific phosphinic tripeptide enantiomers in complex with the angiotensin-I converting enzyme. *FEBS Journal* 281, 943-956.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. *Phaser* crystallographic software. *Journal of Applied Crystallography* 40, 658-674.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D* 53, 240-255.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* 67, 355-367.

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

Nchinda, A. T., Chibale, K., Redelinghuys, P., Sturrock, E.D., 2006. Synthesis of novel keto-ACE analogues as domain-selective angiotensin I-converting enzyme inhibitors. *Bioorganic and Medicinal Chemistry Letters* 16, 4612-4615.

Nettleship, J.E., Assenberg, R., Diprose, J.M., Rahman-Huq, N., Owens, R.J., 2010. Recent advances in the production of proteins in insect and mammalian cells for structural biology. *Journal of Structural Biology* 172, 55-65.

Nilsson, A.B.M., Nitescu, N., Chen, Y., Guron, G.S., Marcussen, N., Matejka, G.L., Friberg, P., 2000. IGF-I treatment attenuates renal abnormalities induced by neonatal ACE inhibition. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology* 279, 1050-1060.

Nishimura, K., Hazato, T., 1993. Isolation and identification of an endogenous inhibitor of encephalin-degrading enzymes from bovine spinal cord. *Biochemical and Biophysical Research Communications* 194, 713-719.

Nussberger, J., Cugno, M., Amstutz, C., Cicardi, M., Pellacani, A., Agostoni, A., 1998. Plasma bradykinin in angio-edema. *Lancet* 351, 1693-1697.

Nyman, U., Joshi, P., Madsen, L.B., Pedersen, T.B., Pinstrup, M., Rajasekharan, S., George, V., Pushpangadan, P., 1998. Ethnomedical information and in vitro screening for angiotensin-converting enzyme inhibition of plants utilised as traditional medicines in Gujarat, Rajasthan and Kerala (India). *Journal of Ethnopharmacology* 60, 247-263.

Oeffner, R.D., Bunkóczi, G., McCoy, A.J., Read, R.J., 2013. Improved estimates of coordinate error for molecular replacement. *Acta Crystallographica Section D* 69, 2209-2215.

Ondetti, M.A., Williams, N.J., Sabo, E.F., Plusec, J., Weaver, E.R., Kocy, O., 1971. Angiotensin-converting enzyme inhibitors from the venom of *Bothrops jararaca*. Isolation, elucidation of structure, and synthesis. *Biochemistry* 19, 4033-4039.

Overlack, A., 1996. ACE inhibitor-induced cough and bronchospasm incidence, mechanisms and management. *Drug Safety* 15, 72-78.

Patchett, A.A., Harris, E., Tristram, E.W., Wyvratt, M.J., Wu, M.T., Taub, D., Peterson, E.R., Ikeler, T.J., ten Broeke, J., Payne, L.G., Ondeyka, D.L., Thorsett, E.D., Greenlee, W.J., Lohr, N.S., Hoffsommer, R.D., Joshua, H., Ruyle, W.V., Rothrock, J.W., Aster, S.D., Maycock, A.L., Robinson, F.M., Hirschmann, R., 1980. A new class of angiotensin-converting enzyme inhibitors. *Nature* 288, 280-283.

Pandyarajan, V., Weiss, M.A., 2012. Design of non-standard insulin analogs for the treatment of diabetes mellitus. *Current Diabetes Reports* 12, 697-704.

Park, P-J., Je, J-Y., Kim, S-K., 2003. Angiotensin I converting enzyme (ACE) inhibitory activity of hetero-chitooligosaccharides prepared from partially different deacetylated chitosans. *Journal of Agricultural and Food Chemistry* 51, 4930-4934.

Paul, M., Mehr, A.P., Kreutz, R., 2006. Physiology of local renin-angiotensin systems. *Physiological Reviews* 86, 747-803.

Perutz, M.F., Rossman, M.G., Cullis, A.F., Muirhead, H., Will, G., North, A.C.T., 1960. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by x-ray analysis. *Nature* 185, 416-422.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* 7, 95-99.

Redelinghuys, P., Nchinda, A.T., Chibale, K., Sturrock, E.D., 2006. Novel ketomethylene inhibitors of angiotensin I-converting enzyme (ACE): inhibition and molecular modelling. *Biological Chemistry* 387, 461-466.

Rieger, K-J., Saez-Servent, N., Papet, M-P., Wdzieczak-Bakala, J., Morgat, J-L., Thierry, J., Voelter, W., Lenfant, M., 1993. Involvement of human plasma angiotensin I-converting enzyme in the degradation of the haemoregulatory peptide *N*-acetyl-seryl-aspartyl-lysyl-proline. *Biochemical Journal* 296, 373-378.

Ripka, J.E., Ryan, J.W., Valido, F.A., Chung, A.Y.K., Peterson, C.M., Urry, R.L., 1993. N-glycosylation of forms of angiotensin converting enzyme from four mammalian species. *Biochemical and Biophysical Research Communications* 196, 503-508.

Rossman, M.G., Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* 15, 24-32.

Rousseau, A., Michaud, A., Chauvet, M-T., Lenfant, M., Corvol, P., 1995. The heaemoregulatory peptide N-acetyl-Ser-Asp-Lys-Pro is a natural and specific substrate of the N-terminal active site of human angiotensin-converting enyme. *Journal of Biological Chemistry* 270, 3656-3661.

Rylett, C.M., Walker, M.J., Howell, G.J., Shirras, A.D., Isaac, R.E., 2007. Male accessory glands of *Drosophila melanogaster* make a secreted angiotensin I-converting enzyme (ANCE), suggesting a role for the peptide-processing enzyme in seminal fluid. *The Journal of Experimental Biology* 210, 3601-3606.

Sakai, K., Sigmund, C.D., 2005. Molecular evidence of tissue renin-angiotensin systems: a focus on the brain. *Current Hypertension Reports* 7, 135-140.

Schechter, I., Berger, A., 1967. On the size of the active site in proteases. *Biochemical and Biophysical Research Communications* 27, 157-162.

Schmid, M.F., Herriot, N.M., 1976. Structure of carboxypeptidase B at 2.8 Å resolution. *Journal of Molecular Biology* 103, 175-190.

Schoofs, L., Veelaert, D., De Loof, A., Huybrechts, R., Isaac, E., 1998. Immunocytochemical distribution of angiotensin I-converting enzyme-like immunoreactivity in the brain and testis of insects. *Brain Research* 785, 215-227.

Sibony, M., Segretain, D., Gasc, J-M., 1994. Angiotensin-converting enzyme in murine testis: step-specific expression of the germinal isoform during spermiogenesis. *Biology of Reproduction* 50, 1015-1026.

Siviter , R.J., Nachman, R.J., Dani, M.P., Keen, J.N., Shirras, A.D., Isaac, R.E., 2002. Peptidyl dipeptidases (Ance and Acer) of *Drosophila melanogaster*: major differences in the substrate specificity of two homologs of human angiotensin I-converting enzyme. *Peptides* 23, 2025-2034.

Skeggs, L.T., Marsh, W.H., Kahn, J.R., Shumway, N.P., 1954. The existence of two forms of hypertension. *The Journal of Experimental Medicine* 100, 363-370.

Skeggs, L.T., Kahn, J.R., Shumway, N.P., 1956. The purification of hypertension II. *The Journal of Experimental Medicine* 103, 301-307.

Skidgel, R. A., Erdös, E. G., 1985. Novel activity of human angiotensin I converting enzyme: release of the NH_2- and COOH-terminal tripeptides from the lutenizing hormone-releasing hormone. *Proceedings of the National Academy of Science USA* 82, 1025-1029.

Soubrier, F., Alhenc-Gelas, F., Hubert, C., Allegrini, J., John, M., Tregear, G., Corvol, P., 1988. Two putative active centers in human angiotensin I-converting enzyme revealed by molecular cloning. *Proceedings of the National Academy of Science USA* 85, 9386-9390.

Steckelings, U. M., Artuc, M., Wollschläger, T., Wiestutz, S., Henz, B. M., 2001. Angiotensin converting enzyme inhibitors as inducers of adverse cutaneous reactions. *Acta Dermato-Venereologica* 81, 321-325.

Stoll, V., Stewart, K.D., Maring, C.J., Muchmore, S., Giranda, V., Gu, Y-G.Y., Wang, G., Chen, Y., Sun, M., Zhao, C., Kennedy, A.L., Madigan, D.L., Xu, Y., Saldivar, A., Kati, W., Laver, G., Sowin, T., Sham, H.L., Greer, J., Kempf, D., 2003. Influenza neuraminidase inhibitors: structure-based design of a novel inhibitor series. *Biochemistry* 42, 718-727.

Sturrock, E.D., Danilov, S.M., Riordan, J.F., 1992. Limited proteolysis of human kidney angiotensin-converting enzyme and generation of catalytically active N- and C-terminal domains. *Biochemical and Biophysical Research Communications* 236, 16-19.

Sturrock, E.D., Natesh, R., van Rooyen, J.M., Acharya, K.R., 2004. Structure of angiotensin I-converting enzyme. *Cellular and Molecular Life Sciences* 61, 2677-2686.

Taylor, C.A.M., Coates, D., Shirras, A.D., 1996. The *Acer* gene of *Drosophila* codes for an angiotensin-converting enzyme homologue. *Gene* 181, 191-197.

Tatei, K., Cai, H., Ip, Y.T., Levine, M., 1995. Race: a drosophila homologue of the angiotensin converting enzyme. *Mechanisms of Development* 51, 157-168.

Tickle, I.J., Laskowski, R.A., Moss, D.S., 1998. R_{free} and the R_{free} ratio. I. Derivation of ecpected values of cross-validation residuals used in macromolecular least-squared refinement. *Acta Crystallographica Section D* 54, 547-557.

Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C., Gingeras, T.R., 1987. Expression of the *lacZ* gene from two methanol regulated promoters in *Pichia pastoris*. *Nucleic Acids Research* 15, 3859-3876.

Turner, S.R., Strohbach, J.W., Tommasi, R.A., Aristoff, P.A., Johnson, P.D., Skulnick, H.I., Dolak, L.A., Seest, E.P., Tomich, P.K., Bohanon, M.J., Horng, M-M., Lynn, J.C., Chong, K-T., Hinshaw, R.R., Watenpaugh, K.D., Janakiraman, M.N., Thaisrivongs, S., 1998. Tipranavir (PNU-140690): A potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulphonamide class. *Journal of Medicinal Chemistry* 41, 3467-3476.

Vegter, S., de Jong-van den Berg, L.T.W., 2010. Misdiagnosis and mistreatment of a common side-effect – angiotensin-converting enzyme inhibitor-induced cough. *British Journal of Clinical Pharmacology* 69, 200-203.

Vercruysse, L., Gelman, D., Raes, E., Hooghe, B., Vermeirssen, V., Van Camp, J., Smagghe, G., 2004. The angiotensin converting enzyme inhibitor captopril reduces oviposition and ecdysteroid levels in Lepidoptera. *Archives of Insect Biochemistry and Physiology* 57, 123-132.

Voronov, S., Zueva, N., Orlov, V., Arutyunyan, A., Kost, O., 2002. Temperature-induced selective death of the C-domain with angiotensin-converting enzyme molecule. *FEBS Letters* 522, 77-82.

Watermeyer, J.M., Kröger, W.L., O'Neill, H.G., Sewell, B.T., Sturrock, E.S., 2008. Probing the basis of domain-dependent inhibition using novel ketone inhibitors of angiotensin-converting enzyme. *Biochemistry* 47, 5942-5950.

Wei, L., Alhenc-Gelas, F., Corvol, P., Clauser, E., 1991. The two homologous domains of human angiotensin converting enzyme are both catalytically active. *Journal of Biological Chemistry* 266, 9002-9008.

Wei, L., Alhenc-Gelas, F., Corvol, P., 1992. The two homologous domains of human angiotensin converting enzyme interact differently with competitive inhibitors. *Journal of Biological Chemistry* 267, 13393-13405.

Wei, C., Hase, N., Inoue, Y., Bradley, E.W., Yahiro, E., Li, M., Naqvi, N., Powell, P.C., Shi, K., Takahashi, Y., Saku, K., Urata, H., Dell'Italia, L.J., Hussain, A., 2010. Mast cell chymase limits the cardiac efficacy of Ang I-converting enzyme inhibitor therapy in rodents. *Journal of Clinical Investigation* 120, 1229-1239.

Weiss, M.S., Hilgenfield, R., 1997. On the use of the merging *R* factor as a quality indicator for x-ray data. *Journal of Applied Crystallography* 30, 203-205.

Weiss, M.S., 2000. Global indicators of x-ray data quality. *Journal of Applied Crystallography* 34, 130-135.

Wijffels, G., Fitzgerald, C., Gough, J., Riding, G., Elvin, C., Kemp, D., Willadsen, P., 1996. Cloning and characterisation of angiotensin-converting enzyme from the dipteran species, *Haematobia irritans exigua*, and its expression in the maturing male reproductive system. *European Journal of Biochemistry* 237, 414-423.

Williams, T.A., Corvol, P., Soubrier, F., 1994. Identification of two active site residues in human angiotensin I-converting enzyme. *The Journal of Biological Chemistry* 269, 29430-29434.

Williams, T.A., Michaud, A., Houard, X., Chauvet, M., Soubrier, F., Corvol, P., 1996. *Drosophila melanogaster* Angiotensin I-Converting enzyme expressed in *Pichia pastoris* resembles the C-domain of the mammalian homologue and does not require glycosylation for secretion and enzymatic activity. *Biochemical Journal* 318, 125-131.

Williams, S.P., Kuyper, L.F., Pearce, K.H., 2005. Recent applications of protein crystallography and structure-guided drug design. *Current Opinion in Chemical Biology* 9, 371-380.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* 67, 235-242.

Winter, G., 2010. *xia2*: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* 43, 186-190.

Yamamoto, Y., Hashimoto, J., Shimamura, M., Yamaguchi, T., Hazato, T., 2000. Characterisation of tynorphin, a potent endogenous inhibitor of dipeptidyl peptidase III. *Peptides* 21, 503-508.

Yang, H.Y.T., Erdös, E.G., Levin, Y., 1971. Characterisation of a dipeptide hydrolase (kininase II: angiotensin I converting enzyme). *The Journal of Pharmacology and Experimental Therapeutics* 177, 291-300.

Yang, S., Kuang, Y., Li, H., Liu, Y., Hui, X., Li, P., Jiang, Z., Zhou, Y., Wang,Y., Xu, A., Li, S., Liu, P., Wu, D., 2013. Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimising Kex2 P1' site. *PLoS ONE* 8, 1-11.

Yoshida, S., Naganawa, H., Aoyagi, T., 1991. Leuhistin, a new inhibitor of aminopeptidase M, produced by *Bacilllus laterosporus* BMI156-14F1. *The Journal of Antibiotics* 44, 579-581.

Yu, X., Ge, P., Jiang, J., Atanasov, I., Zhou, Z.H., 2011. Atomic model of CPV reveals the mechanism used by this single-shelled virus to economically caryy out functions conserved in multishelled reoviruses. *Structure* 19, 652-661.

Zaman, M. A., Oparil, S., Calhoun, D. A., 2002. Drugs targeting the renin-angiotensin-aldosterone system. *Nature Reviews Drug Discovery* 1, 621-636.

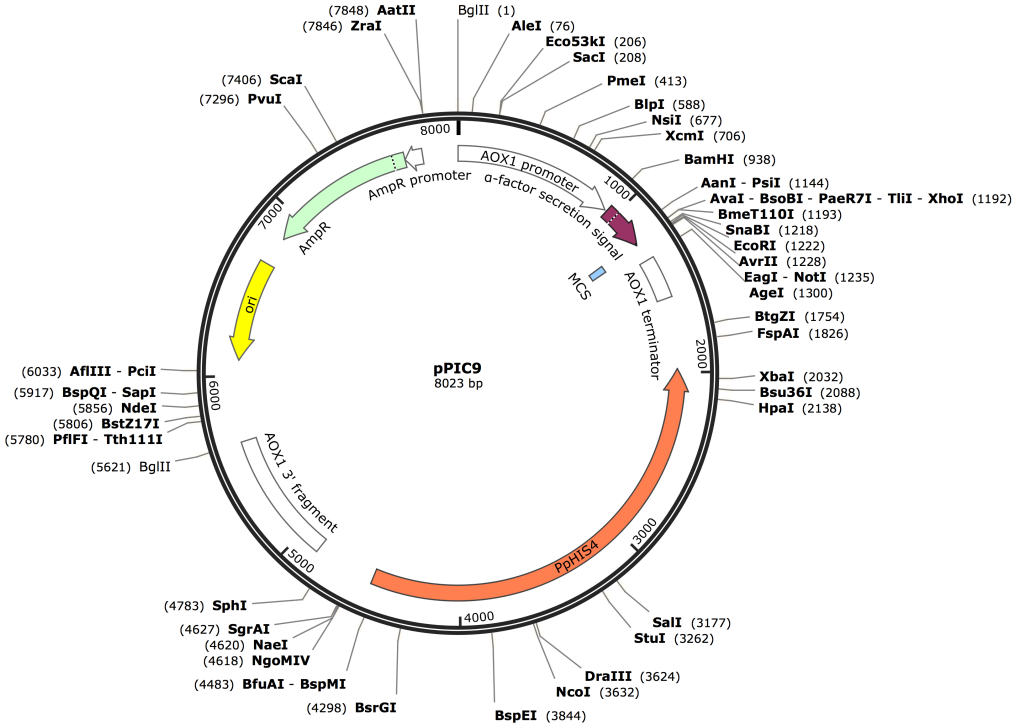
Zhang, C., Wu, S., Xu, D., 2013. Catalytic mechanism of angiotensin-converting enzyme and effects of the chloride ion. *The Journal of Physical Chemistry* 117, 6635-6645.

Zhao, Q., Sannier, F., Garreau, I., Guillochon, D., Piot, J.M., 1994. Inhibition and inhibition kinetics of angiotensin converting enzymes activity by haemorphins, isolated from a peptic bovine haemoglobin hydrolysate. *Biochemical and Biophysical Research Communications* 204, 216-223.

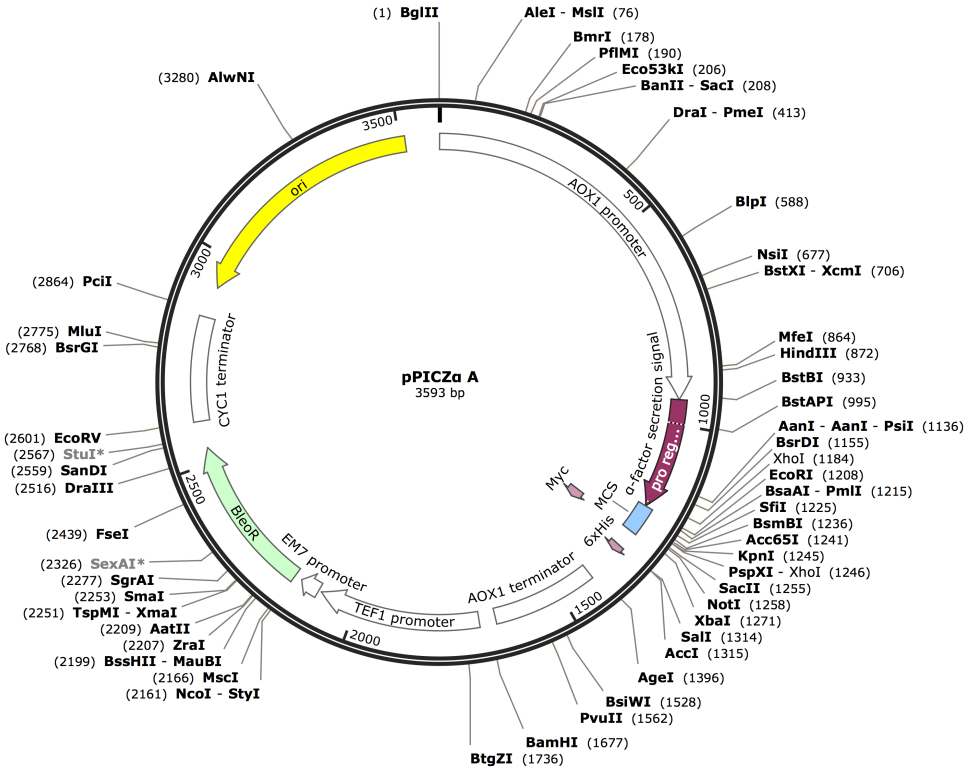
Zhao, Q., Piot, J.M., 1997. Investigation of inhibition angiotensin-converting enzyme (ACE) activity and opioid activity of two haemorphins, LVV-haemorphin-5 and VV-haemorphin-5, isolated from a defined peptic hydrolysate of bovine haemoglobin. *Neuropeptides* 2, 147-153.

Appendix A: Vector Maps

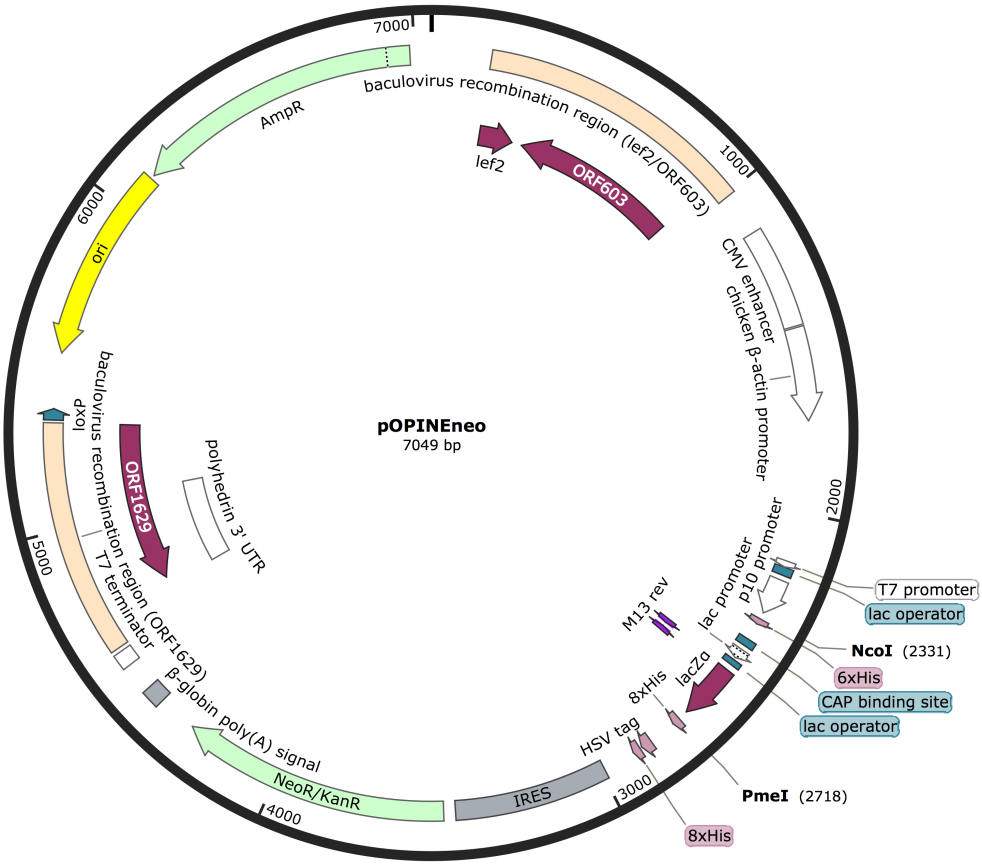
A1: pPIC9 Vector Map



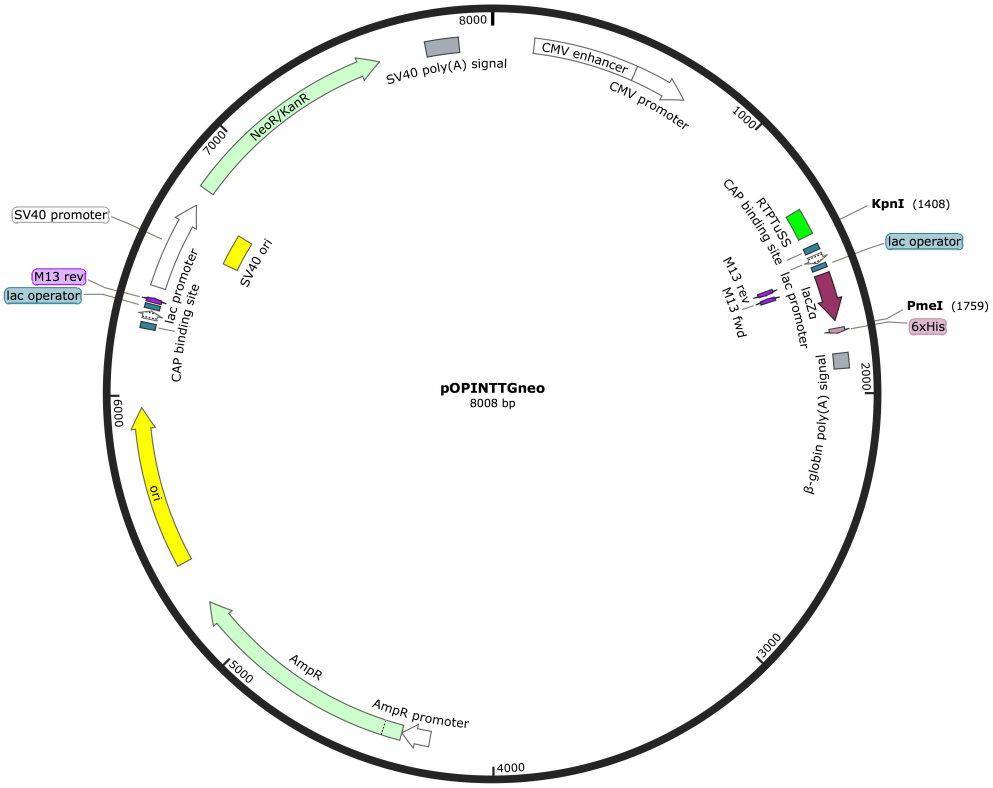
A2: pPICZα Vector Map



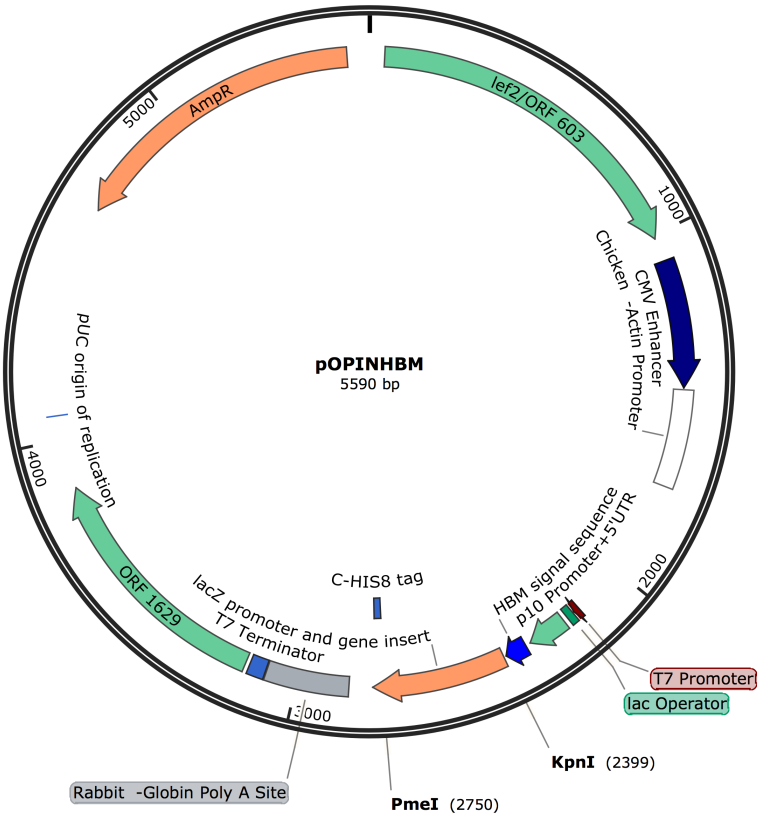
A3: pPOPINEneo Vector Map



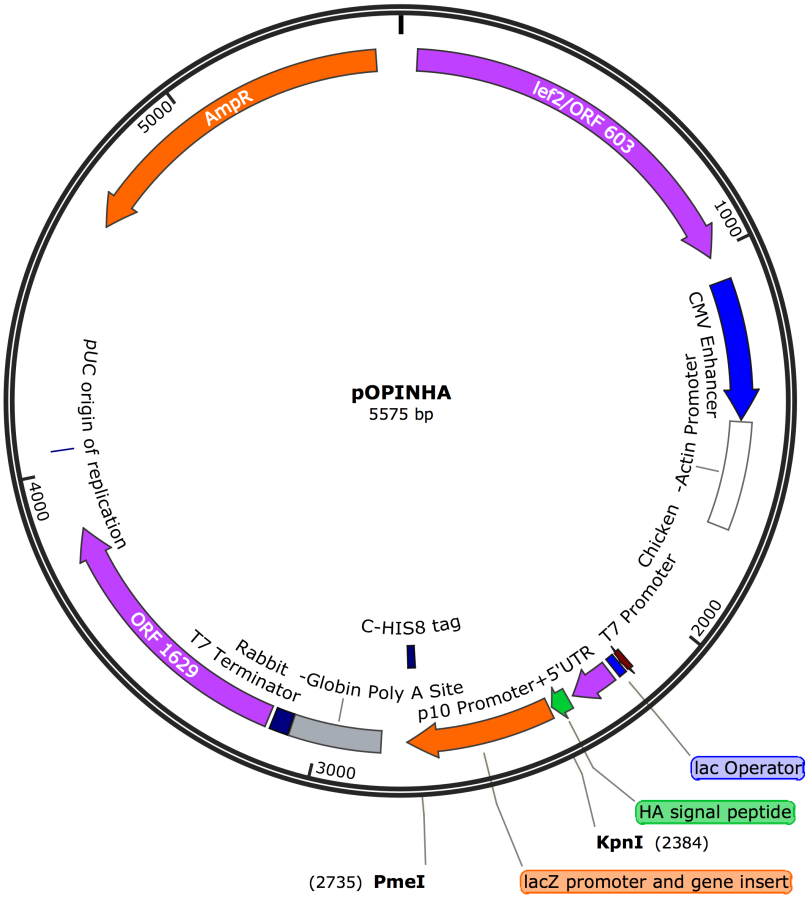
A4: pOPINTTGneo Vector Map



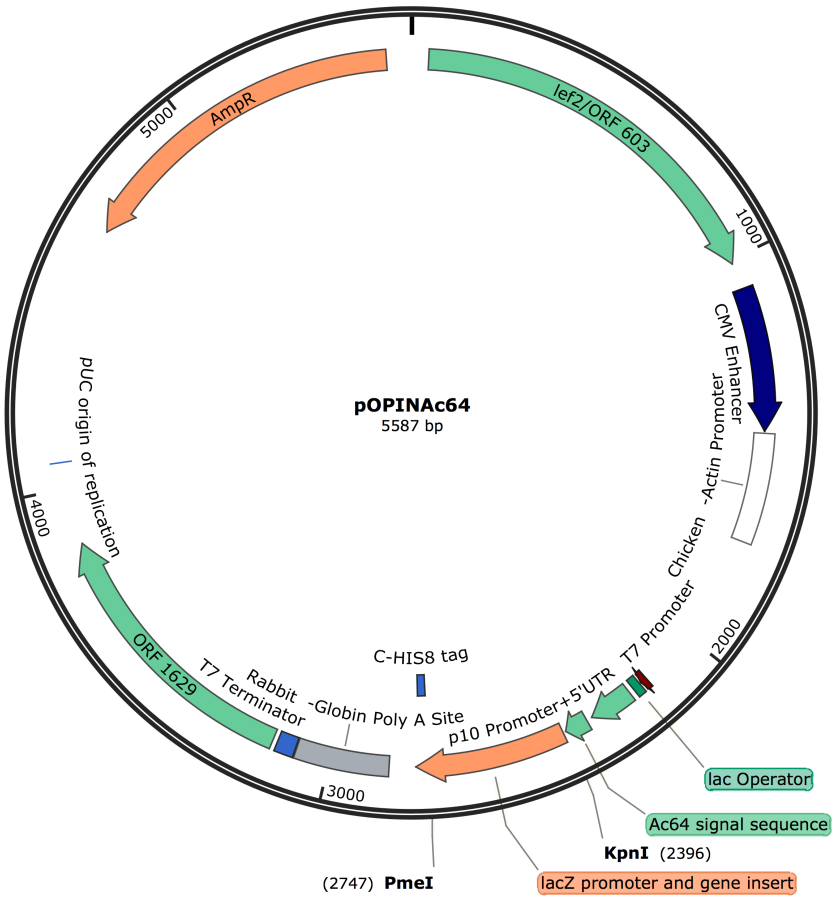
A5: pOPIN-HBM Vector Map



A6: pOPIN-HA Vector Map



A7: pOPIN-Ac64 Vector Map



Appendix B: List of Publications

Harrison, C., Acharya, K.R., 2014. ACE for all - a molecular perspective. *Journal of cell communication and signalling* 8, 195-210.

Harrison, C., Acharya, K.R., 2015. A new high-resolution crystal structure of the *Drosophila melanogaster* angiotensin converting enzyme homologue, AnCE. *FEBS Open Bio* 5, 661-667.