FOLDING, DIMERISATION AND THE INTERACTION WITH $A\beta_{1-40}$ OF HUMAN CYSTATIN C

EMMA KEELEY

Submitted for the degree of Doctor of Philosophy

September 2007

Department of Molecular Biology and Biotechnology

University of Sheffield

PAGE NUMBERING AS

ORIGINAL

For Emmas Nanny,

Mrs P. Chalmers

ABSTRACT

Amyloid formation is a predominant feature of many human diseases including Alzheimers' disease, Parkinsons' disease, type II diabetes and Creutzfeldt-Jacob disease. The process of amyloidogenesis involves the self-assembly of soluble protein into insoluble fibrous material. Amyloidogenic proteins, which share no common sequence, structure or function, form amyloid fibres which have a common morphology. Neither a detailed structure of mature amyloid or the mechanism by which it forms is fully understood. The work presented in this thesis uses the cystatins as a model system for probing the mechanism of amyloid formation.

The formation of amyloid requires a refolding event as the protein involved refolds from its native structure to the cross-β structure common to amyloid. Characterisation of the folding pathway of human cystatin C indicates that it folds via a partially folded kinetic intermediate, in an analogous manner to chicken cystatin. Analysis of the early stages of amyloidogenesis in cystatins indicates domain-swapped dimers are the building block of cystatin amyloid. As is observed with chicken cystatin, the dimerisation of human cystatin C is a bimolecular process. m-value analysis indicates that the structure of the dimerisation transition state is very close to the structure of the unfolded state and is more unfolded than the kinetic intermediate identified in the human cystatin C folding pathway. No terameric species are observed in the amyloidogenesis of human cystatin C, supporting evidence that tetramers are an off-pathway intermediate in the amyloidogenesis of chicken cystatin. Following the formation of dimer, isomerisation of the proline conserved across the cystatins is required prior to the formation of amyloid fibres.

A preliminary study of the interaction between human cystatin C and Aβ shows that there is no interaction between monomeric hCC and monomeric Aβ_{1-40}. Given that hCC has been shown to inhibit Aβ amyloid formation, hCC must interact with one of the oligomeric species of Aβ that is populated during amyloidogenesis. Further experimentation is required to determine the exact nature of the interaction between hCC and Aβ.

ACKNOWLEDGEMENTS

There are many people who have helped me over the past four years, too many to name, but the fact you're reading this suggests you're one of them. Thank you! I doubt I'd be writing this now if it hadn't been for any one of you. There are, of course, a few names that can not go unmentioned...

First on the list is my supervisor Rosie Staniforth, who has been an absolute star through out this project. Thank you for always being available with advice and encouragement while you guided me though this project. From the days of Anna, Silva, Manal and Gareth through to Caz, Clare and Rob, the Cystatin group has been fantastic. Thanks not only for the help and advice you've all given me, but also for making the office and lab a fun, enjoyable place to work.

The rest of the NMR group have been pretty great too. Thanks to Jon Waltho, for not only standing in when Rosie wasn't around, but also for being equally available with advice when she has been around. Thanks to Jeremy, Andrea, Laz and Rich for taking the time to explain various aspects of NMR to me at a speed my brain could handle. Thanks too to Matt and Clare for always being around with sound advice whenever I've needed it. I'd have been lost without all the help I've received from the members of the lab, but I'd have been equally lost without the many friends I've made in the lab. Knowing I can always recruit someone for a pint and a laugh at the end of a tough day, or with any feeble excuse really, is great. Thanks!

There are many friends outside of the lab that I should thank, I'll not list them all, but thanks for always being top mates. Rachel and Tom deserve a special mention for putting up with living with me for the majority of PhD. As well as being really great mates, they've welcomed me into their home and looked after me when I really needed it. Cheers guys! Thanks too to my many buddies from SUSAC who have kept me in a good state of mind through countless dive trips and late night drinks in the Cobbie.

My family, as always, have been wonderful. Knowing your love and support is constantly behind me means I can take on any challenge life throws at me. Maa,

Paa, Nan, Egg, Paul, Nic, Alex, Alf, John and Pat - thank you. Your words of encouragement, great advice and even the gentle nagging are greatly appreciated. You've all been a great help, but I must specially mention my Nanny. I wouldn't be here without my Nan! For your words of wisdom, countless Sunday dinners, drops of medicine and your ability to always make me laugh, thank you from the bottom of my heart. Thanks also to Big G for the many time you've motivated me by silently standing at my shoulder and telling me to "come on, Katie".

CONTENTS

LIST OF FIGURES AND TABLES

ABBREVIATIONS

Aβ	Amyloid β peptide
APP	Aβ precursor protein
AD	Alzheimer's disease
AL	Amyloid light chain
ANF	Atrial natriuretic factor
ANS	1-anilino-8-naphthalene sulfonate
AUC	Analytical ultracentrifugation
BSE	Bovine spongiform encephalopathy
CAA	Cerebral amyloid angiopathy
cC	Chicken cystatin
CD	Circular dichroism
CJD	Creutzfeldt-Jakob disease
CNS	Central nervous system
CSF	Cerebrospinal fluid
ECL	Electrochemiluminescence
EM	Electron microscopy
ER	Endoplasmic reticulum
FGF-2	Fibroblast growth factor-2
GdnHCl	Guanidine hydrochloride
GSS	Gerstmann-Straussler-Schneiker syndrome
hCC	Human cystatin C
HCHWA-D	Hereditary Cerebral Haemorrhage with Amyloidosis - Dutch type
HCHWA-I	Hereditary Cerebral Haemorrhage with Amyloidosis - Icelandic type
HPLC	High pressure liquid chromatography
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl-β-D-galactosidase
K_i	Dissociation constant
LB	Luria Bertani media
LBK	Potassium buffered LB.
MRI	Magnetic Resonance Imaging
MS	Mass spectrometry
NMR	Nuclear magnetic resonance

OD_n	Optical density (at n nm)
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEI	Polyethyleneimine
PHF	Paired helical filaments
PMSF	Phenylmethylsulfonylfluoride
QLS	Quasi-elastic light scattering
SAP	Serum amyloid protein
sAPPα	Secreted Aβ precursor protein alpha
sAPPβ	Secreted Aβ precursor protein beta
SDS-PAGE	Sodium dodecyl sulfate-poly acrylamide gel electrophoresis
SEC	Size exclusion chromatography
SPECT	Single photon emission computed tomography
ssNMR	Solid-state NMR
TEM	Transmission electron microscopy
TFA	Trifluoroacetic acid
TTR	Transthyretin
WT	Wild type

CHAPTER ONE
INTRODUCTION

1.1   CYSTATINS

1.1.1   The Cystatins

Members of the cystatin superfamily are non-covalent competitive inhibitors of papain-like cysteine proteases.   Cystatins bind to cysteine proteases in an equimolar, tight, reversible manner forming a complex that lacks proteolytic activity.   Amongst other functions the cystatins protect host tissues from proteolysis by cysteine proteases of host, bacterial and viral origin.[1-4]   Although the cystatins share a common "hot-dog" fold, the superfamily has been divided into smaller families based on the localisation, size and complexity of the polypeptide chain.   Cystatins were originally categorised into three types, the stefins (type I), the cystatins (type II) and the kininogens (type III).[1, 4]   Figure 1.1 highlights the main structural differences between these subfamilies.

Stefins are mainly located intracellularly and are the lowest molecular weight subfamily.   Although they have the general cystatin fold, they lack disulphide bonds and carbohydrate residues.   The cystatins are generally slightly larger than the stefins and, as they are translated with a secretory peptide leader sequence, are considered to be extracellular.   Cystatins contain four cysteine residues that are involved in the formation of two characteristic disulphide bonds.   The cystatins are described as non-glycosylated although there are some exceptions to this rule.   For example, cystatin F and E/M contain functional N-glycosylation sites and approximately 20% of rat cystatin is N-glycosylated.   The kininogens are mainly intravascular proteins and consist of three cystatin-like domains.   Each domain contains the characteristic disulphide bonds found in type II cystatins as well as additional disulphide bonds.   Kininogens are glycosylated and have an additional polypeptide at the C-terminus containing the bradykinin sequence.[1, 5, 6]

More recently a group of important fetal proteins known as fetuins have been identified as an additional cystatin subfamily.   The fetuins are N- and O-glycosylated and phosphorylated.   The N-terminal contains two tandem type 2 cystatin domains and the C-terminal region contains a histidine-rich domain flanked by proline-rich domains.[7]

Figure 1.1 **A.** Line representation of the three main cystatin families. Loops represent the presence of a disulphide bond. For type 3 cystatins each segment represents a type 2 cystatin like domain.[3] **B.** An example of a type I cystatins, stefin A (PDB ID: 1gd3[8]) **C.** An example of a type II cystatins, chicken cystatin (PDB ID: 1cew[9])
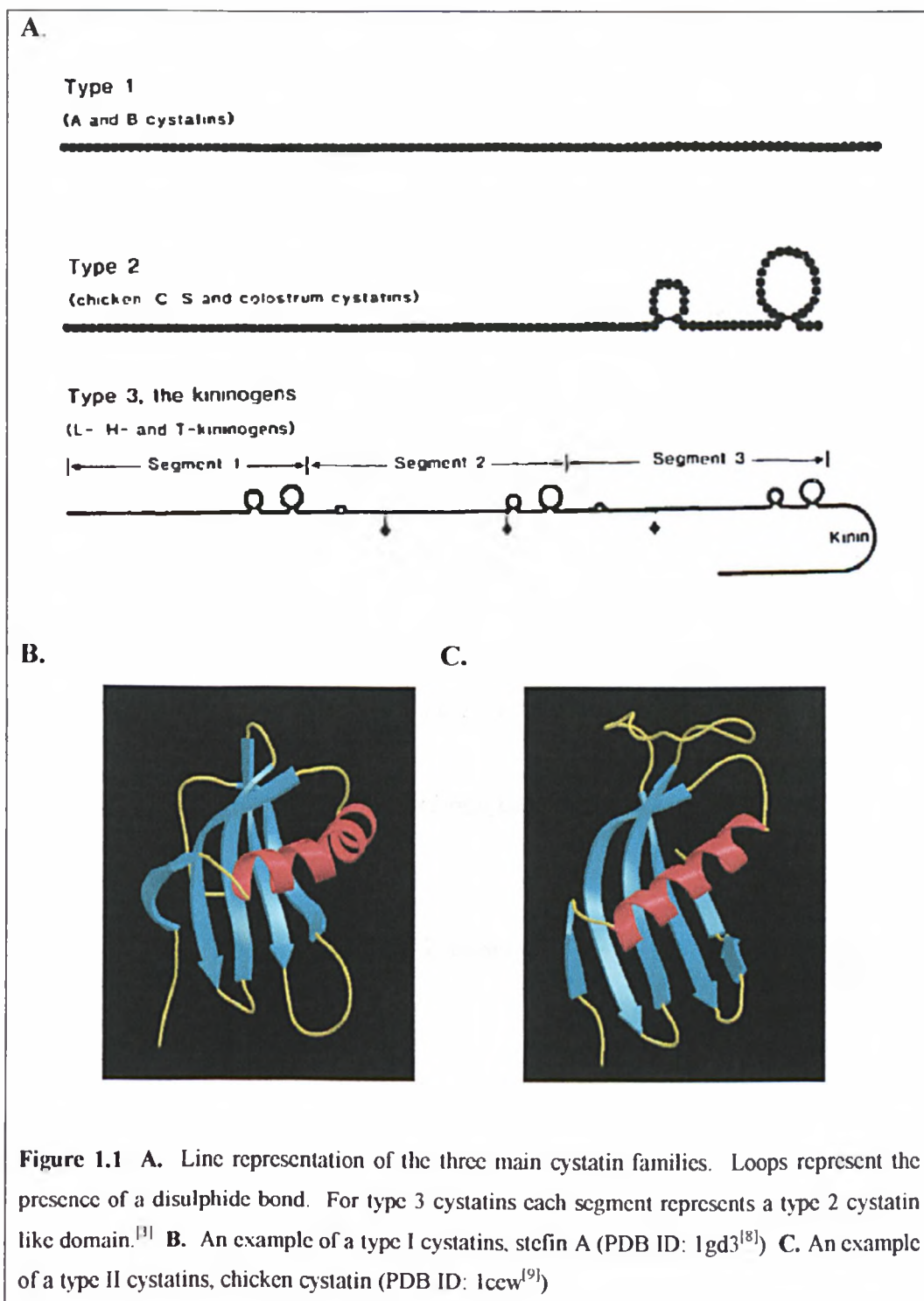
Human Cystatin C

Human cystatin C (hCC) is a typical member of the type II cystatin family.[1] It is encoded by a single copy gene located on chromosome 20 and is synthesised as a proform with a hydrophobic leader sequence.[10] Once the leader sequence is cleaved, a 120 residue protein remains with a calculated molecular weight of 13 359 Da.[11] As is typical of the type II cystatins, hCC is not glycosylated and contains two disulphide bonds towards the C terminus.[1] hCC synthesis is not

tissue specific, all nucleated cells constitutively express and secrete hCC.[5] For experimental use hCC can be either isolated from urine[12] or recombinant hCC can be synthesised and purified from *E. coli*.[13-18] hCC is present at a relatively high concentration in most body fluids, as summarised in Table 1.1[11]

	Concentration [nM]
Plasma	105 ± 30
CSF	545 ± 180
Urine	7.5 ± 0.5
Saliva	135 ± 70
Seminal plasma	3810 ± 600

Table 1.1 Typical concentration of hCC found in the main bodily fluids. Data taken from [11].

Through the mechanism of three dimensional domain swapping, human cystatin C has been shown to self associate under mildly denaturing conditions, such as the presence of denaturant, low pH or high temperature.[19, 20] Domain swapping is further discussed later in this chapter but, briefly, in this process a dimer is formed which has two monomer-like domains that are structurally similar to chicken cystatin. The main exception is the linker region that unfolds to form the open interface of the dimer.[19] Figure 1.2 shows the structure of domain swapped hCC and the positions of the open and closed interfaces.

Figure 1.2 Domain swapped hCC dimer, with each monomer coloured blue or green. The "closed interface" is formed between the swapped domains and is the main adhesive force. In addition an "open interface" is formed in the linker region which provides some additional stabilization to the dimer.[21] The red dot marks the position of the L68Q mutation. Image adapted from [19].
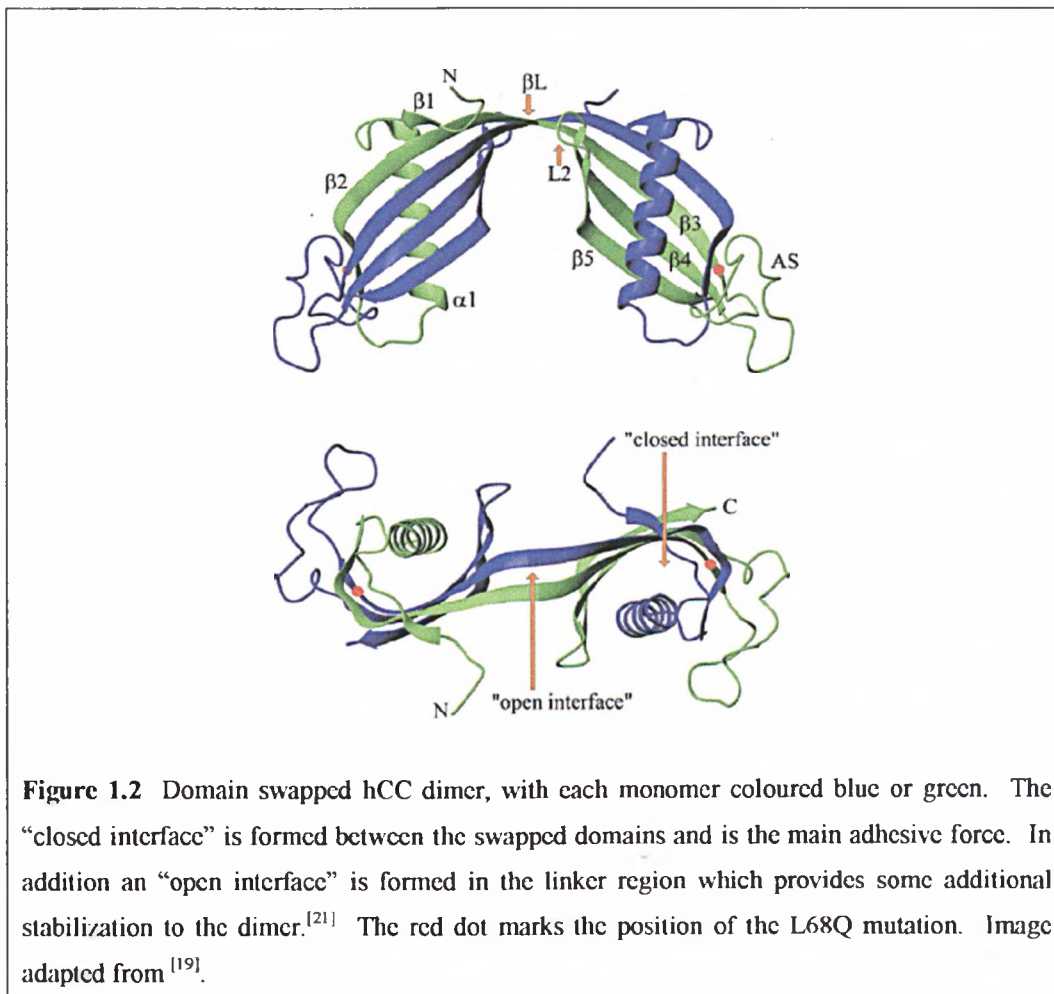
L68Q Variant hCC

hCC is associated with the disease hereditary cerebral haemorrhage with amyloidosis – Icelandic type (HCHWA-I), where it is present as a L68Q variant. HCHWA-I is discussed further later in the chapter. The L68Q mutation is caused by a point mutation changing the codon for residue 68 from CTG to CAG.[22] The L68Q mutation results in a reduced hCC concentration in the cerebrospinal fluid and in the deposition of hCC amyloid in the brain.[23] There is no difference in the ability of wild type or L68Q mutant to inhibit cathepsins. Both normal and mutant are expressed and cleared from the cell at a similar rate.[24] L68Q variant cystatin C has a ^{1}H-^{15}N heteronuclear single quantum coherence spectroscopy (HSQC) spectrum very similar to wild type cystatin C suggesting that they fold in a similar way.[23]

The L68Q mutant is stable between pH 7.5 and 9.0, but outside this range dimerisation occurs. At a concentration of 0.1 mg/ml hCC and under conditions where dimerisation is strongly favoured such as pH 6 or 1M NaCl, an equilibrium

is reached where approximately eighty percent of the variant cystatin is in the dimeric form. L68Q dimers will convert to monomers when dialysed against physiological buffer. Therefore, it is likely that an equilibrium between monomer and dimer is reached for L68Q under physiological conditions. The L68Q mutation destabilises the cystatin enough that the mildly denaturing conditions required for dimerisation occur with a small increase in temperature or pH change from physiological conditions. In contrast, wild type cystatin C tends not to dimerise under physiological conditions. Under conditions where both will dimerise, only the variant will form fibrils *in vitro*.[23]

L68Q is on the central β-strand on the concave face covered by the α-helix. The position of L68Q is marked on Figure 1.2 and 1.3. The substitution puts a hydrophilic residue into a hydrophobic environment and makes the van der Waals interactions prohibitively close, thus putting a repulsive force on the α-helix.[19] The energy difference corresponding to the unfolded state is lower as unfavourable solvent contacts in the newly exposed interface are reduced. The variant may be stabilised in the dimeric state by the additional stability provided by the open interface.[21]

1.1.2 Cystatin structure

The cystatin superfamily share a common structural motif composed of a single α-helix lying across a five stranded antiparallel β-sheet, often refered to as a "hot-dog" fold.[9] The monomeric structure of hCC has not been published, but the crystal structure of the dimer has been determined.[19] hCC has 41% sequence identity and 62.5% similarity with the better characterised chicken cystatin, making chicken cystatin a suitable molecular model for hCC.[21] Figure 1.3 shows the primary and secondary structure of hCC and the tertiary structure of chicken cystatin (cC) showing the numbering of the strands, helix and loops. The secondary structure of hCC shows an α-helix and five β-strands, with three β-bulges within the β-sheet and a unstructured highly mobile N-terminus.[25]
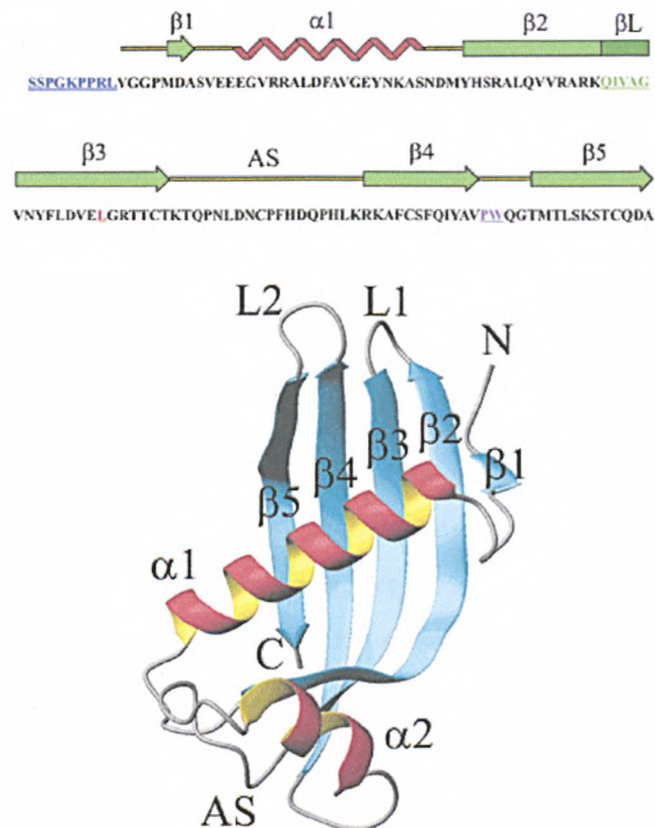
Figure 1.3   Primary and secondary structure of hCC and the tertiary structure of cC. Underlined in green, purple and blue in the primary structure are the regions involved in protease inhibition. In red in the primary sequence is the location of the L68Q mutation.

1.1.3   Biological role of Cystatins

Several biological functions have been proposed for the cystatins. In addition to their main function as protease inhibitors, cystatins have been identified as signalling molecules and are thought to play a role in the inflammation response.

C1 Cysteine Protease Inhibitory Activity

hCC is an effective, reversible inhibitor of papain-like cysteine proteases including cathepsin B, H, K, L and S as well as papain, ficin and dipeptidyl peptidase 1.[2, 11] hCC is arguably the most physiologically important extracellular human cystatin and is the controlling inhibitor for cathepsin B in almost all bodily fluids examined including CSF, seminal plasma and milk.[2, 12] Based on the dissociation constant (K_i), hCC is predicted to contribute to physiological inhibition of any cysteine proteases that pass from the lysosomal system to the extracellular fluid.[26] Table 1.2 shows K_i for common hCC-enzyme complexes.

Papain	Cathepsin B	Cathepsin H	Cathepsin L	Cathepsin S
0.000011	0.25	0.28	<0.005	<0.008

Table 1.2 The dissociation constant, K_i, of common cystatin-enzyme complexes [nM]. Data taken from [27].

Titration of papain with hCC shows a decreased enzymatic activity as a linear function of hCC concentration, suggesting an equimolar stoichiometry.[26] The mechanism of cysteine protease inhibition by cystatins is a one step process that is a simple, reversible, second order type.[2] The highly conserved region from Q53-G57 of chicken cystatin forms a tight beta-hairpin loop which binds to the cysteine proteinases. Similarly, a second hairpin loop made up of the conserved region P103–W104 along with the N-terminal segment both bind to the cysteine proteinase. These three regions form a wedge shape which is complimentary to the active site of papain as shown in Figure 1.4a. Although the active site is blocked, the cystatin remains too far from the reactive site of papain to be attacked by it.[9] The crystal structure of stefin A in complex with cathepsin H is shown in Figure 1.4.[28]

When hCC is exposed to an excess of papain, hCC is cleaved between residues 11 and 12 thus weakening its inhibitory activity. Truncation at G11-G12 decreases the affinity for papain by three orders of magnitude. This highlights the contribution of the N-terminal region, which contains the conserved G11, V10, L9 and R8, to cysteine protease binding. [2, 29, 30] The contribution of the N terminus to affinity for the peptidase varies with the different peptidases.[30]

Figure 1.4 A. A model of the monomeric structure of hCC with the region involved in the C1 protease inhibition site circled. The C13 protease inhibition site is less well defined; it is thought to involve the region around residue N39. Image adapted from [31]. **B.** Crystal structure of stefin A in complex with cathepsin H, images taken from [28].
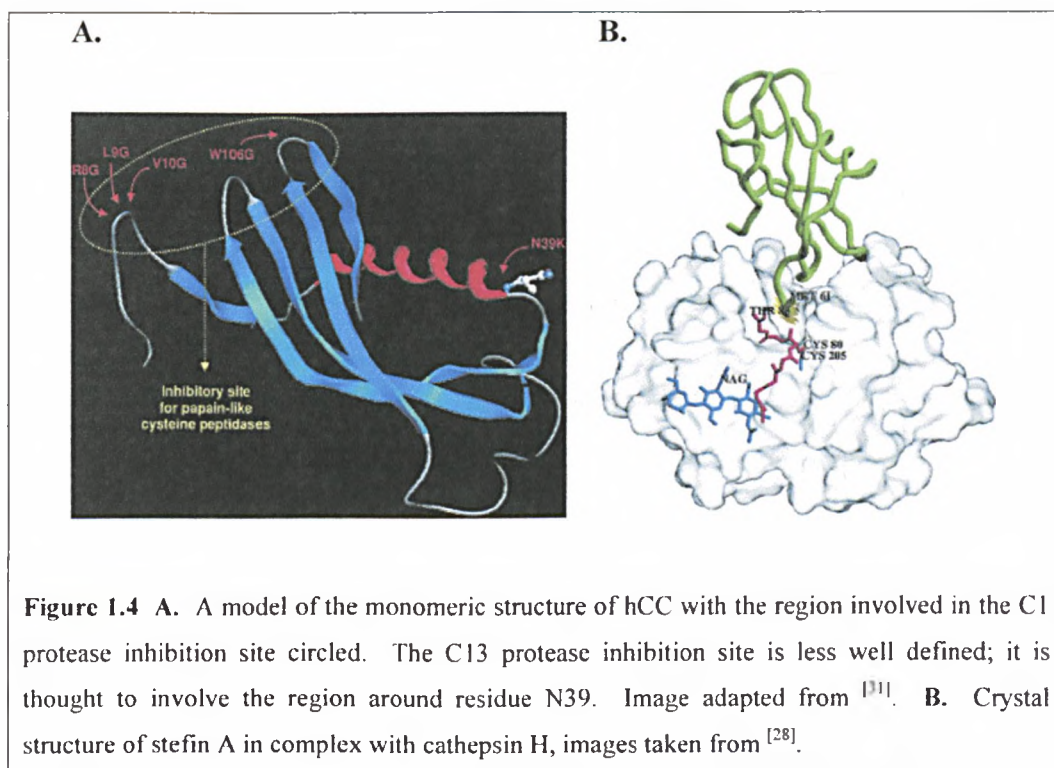
C13 Cysteine Protease Inhibitory Activity

hCC is unusual in that it has an additional legumain-inhibiting site. Legumain, a recently discovered lysosomal endopeptidase in mammals, is member of the C13 cysteine protease family. Legumain shows restricted hydrolysis of asparaginyl bonds, with unidentified factors preventing the hydrolysis of many of the asparaginyl bonds in the protein.[32] The legumain inhibitory interaction is different to the papain inhibitory site. Whereas dimerisation causes a loss of papain inhibition, legumain inhibition is still as active as the monomeric form.[31] It has been suggested that the loop around N39 followed by the conserved alpha helix is the site of interaction with legumain.[31] The backside loop (39-41) that is thought to be crucial for C13 protease inhibition is not perturbed by dimerisation, consistent with its preserved activity in the dimeric form.[19]

Paracrine/autocrine cofactor

Although hCC has not been shown to be glycosylated, it is worth noting that a cell signalling role has been suggested for a glycosylated form of rat cystatin C. The glycosylated form of rat cystatin C has been shown to act as a paracrine /autocrine co-factor for the mitogenic activity of fibroblast growth factor-2 (FGF-2) on neural stem cells. Furthermore, a combined delivery of FGF-2 and glycosylated cystatin C will stimulate neurogenesis.[33]

Inflammation

It has been proposed that cystatins play a role in the inflammation response following the observation that raised levels of protease and protease inhibitors are found at sites of inflammation.[4] Cystatins have been identified in several process that are involved in the inflammation response. Secretion of hCC by monocytes and macrophages is down-regulated by proinflammatory lipopolysaccharide and interferon γ.[34] The N-terminal tetrapeptide of hCC is thought to have an inhibitory effect on superoxide anion release and phagocytosis in human neutrophils.[35, 36] Leukocytes are involved in the initiation and maintenance of inflammation and must get from their usual location in the blood to the site of injury. hCC has been found to have an effect on polymorphonuclear leukocyte locomotion.[36]

1.1.4 Cystatins and disease

The ability of cystatins to inhibit exogenous cysteine proteases results in cystatins being identified as protective agents against several diseases. In contrast to this, cystatins have been identified as the causative agent in certain amyloid diseases. Further to this hCC has been shown to colocalise with amyloid β peptide (Aβ) amyloid deposits in several diseases including Alzheimers' disease. Cystatins have also been identified in several cancers where it is not known whether they play a protective or causative role.

Antimicrobial and antiviral activity

Cystatins exhibit antiviral and antimicrobial activity by inhibiting exogenous proteinases. Derivatives mimicking the proposed proteinase-binding centre of cystatin C, which irreversibly inhibit cysteine proteinases, have been found to inhibit specifically the growth of all strains of group A *Streptococci*.[37] Cystatin isolated from Horseshoe crab haemocytes show antimicrobial activity against Gram negative bacteria such as *S. typhimurium*, *E.coli* and *K. pneumoniae*.

Several studies identify cystatins as inhibitors of viral replication. hCC has been shown to inhibit *Herpes simplex* virus type I replication.[38] Chicken cystatin has been shown to alter intracellular proteolytic processing of poliovirus proteins.[39]

hCC is a powerful inhibitor of cysteine proteases of *Leishmania Mexicana,* a protozoan parasite.[10, 26] Both hCC and chicken cystatin are potent inhibitors of congopain and cruzipain. These are important cysteine peptidase virulence factors from the parasitic protozoa *Trypanosoma Congolese* and *Trypanosoma cruzi,* both of which cause major diseases in the developing world.

Cancer

Lysosomal cysteine proteases have been implicated in multiple steps of tumour progression, from immortilization and transformation through tumour invasion and angiogenesis to metastasis and drug resistance.[40-45] The involvement of lysosomal cysteine proteases in the development of tumours suggests that cystatins may play a protective role against tumour progression. However, the scenario is more complex as there is increasing evidence supporting a role for cystatins in the promotion and suppression of tumour growth, invasion and metastasis.[46] Stefin A has been shown to modulate growth and metastatic potential of human angiosarcomas, malignant fibrous histiocytomas and poorly differentiated ovarian carcinomas.[5] hCC and cathepsin B have been suggested as having a role in both benign and malignant ovarian cancers where a significantly high serum concentration of hCC is associated with malignancy.[47]

Increased levels of cathepsin B, but decreased levels of cathepsin B-hCC complexes have been observed in the sera of patients with lung cancer. This could be interpreted as a loss of control of hCC over serum cathepsin B during lung cancer. However, cystatin C, and its complex with cathepsin B, may accumulate in pleural effusions rather than in the sera of cancer patients.[48, 49]

Other cysteine proteases are overexpressed in tumours and some can compete with cathepsin B in complex formation with hCC. For example, in conditions where cathepsin L and cathepsin B are present, cathepsin L would displace cathepsin B from the hCC-cathepsin B complex.[50]

hCC has been identified as a novel TGF-β type II receptor antagonist that inhibits TGF-β binding and signalling in normal and cancer cells. There is evidence that TGF-β stimulation of initiating metastatic events, including decreased cell

polarization, reduced cell–cell contact and elevated cell invasion and migration, can be prevented by hCC treatment.[51]

Hereditary Cerebral Haemorrhage with Amyloidosis – Icelandic type

Studies on Icelandic families with HCHWA have identified the responsible gene as a single, dominant gene showing marked anticipation and almost complete penetration.[52] The L68Q variant of hCC has been shown to be the amyloid precursor protein HCHWA-I.[53] Amyloid material is found in the intima and/or outer wall of the small arteries and arterioles of the leptomeningea, cerebral cortex, basal ganglia, brainstem and cerebellum in all autopsied cases and asymptomatic deposits have occasionally been found in peripheral tissues. However, there is no association with neurite plaques or tangles. [22, 52, 54]

The mean age of onset for HCHWA-I has been reported as somewhere between 22.5-27.3 years, with 50% of fatalities occurring between the ages of 20 and 30. [52, 54, 55] HCHWA-I patients suddenly take ill with signs of central nervous system (CNS) damage including acute onset of headache, nausea, focal neurological signs, loss of consciousness, paralysis and sensory disturbances, but no hypertension. [52, 55] The first haemorrhage is usually fatal, if it is not the patient faces cognitive decline and dementia, with the dementia occurring in a stepwise fashion. A slow progressive dementia can also occur and the dementia may even precede stroke. [54, 55] The CSF level of hCC is much lower in HCHWA-I patients than in individuals with wild type protein, so much so that a 3.5 mg/l concentration in the CSF is a discrimination value that can diagnose HCHWA-I. [11, 54]

Alzheimers' disease

hCC has been found to co-localise with Aβ amyloid deposits in the brains of patients with AD, hereditary cerebral haemorrhage with amyloidosis - Dutch type (HCHWA-D) and sporadic cerebral amyloid angiopathy (CAA).[55-58] Co-localisation of Aβ and hCC also occurs in the muscular amyloid deposits found in cases of sporadic inclusion body myositis, the most common muscular disease of elderly.[59] When hCC is the main amyloidogenic protein, such as the cases of HCHWA-I, there is no evidence of co-immunostaining.[55, 58] The connection

between hCC and Alzheimers' disease is discussed in further detail later in this chapter (section 1.2.5) and in chapter 6.

1.2   AMYLOID

1.2.1   History

Rudolph Virchow first introduced the term amyloid in 1854 to describe waxy macroscopic tissue abnormalities found in cerebral tissue that stained positively with an iodine staining reaction. The staining property meant that deposits were originally identified as being starch-like and therefore given the name amyloid, derived from the greek amylon meaning starch. By 1859 Friedrich and Kekule had demonstrated that the deposits actually had a high protein content and lacked carbohydrate.[27]

Today, several major incurable diseases are characterised by the deposition of amyloid including Alzheimer's disease, Parkinson's disease, Huntington's disease, type II diabetes and the transmissible spongiform encephalopathies. Many of the diseases associated with amyloid deposition are late-onset diseases that are prevalent in the developed world. As the population surviving until old age expands, the medical importance of understanding the mechanism of amyloidogenesis increases. In the 150 years since Virchow studied amyloid there have been many advances in our knowledge of amyloid structure and function and its role in disease. However, amyloid remains an important area of research as there are still many unanswered questions regarding the mechanism of amyloidogenesis, its significance in disease and the development of treatments for amyloidosis.

1.2.2   Structure

Although amyloid precursor proteins are distinct in their amino acid sequence and native fold, they aggregate into mature amyloid fibres that share a similar structure, implying a common mechanism of fibril formation. [60, 61] Amyloid is identified by staining with specific dyes, electron microscopy (EM) and a characteristic X-ray fibre diffraction pattern. Although amyloid is generally defined as being extracellular, intracellular structures with the same core structure have been observed. For example, Lewy bodies found in Parkinson's diseases are intracellular deposits of α-synuclein.[62] Amyloid-like fibrils can also be formed

in vitro from proteins unconnected to amyloidoses.[63] In fact, it has been suggested that nearly all proteins have the ability to form amyloid, provided they are put under suitable conditions.[62] The overall morphology of amyloid depends on the conditions in which amyloidogenesis takes place and different morphologies can be observed in the same preparation.[64] Fibril insolubility and inability to form crystals prevents structural studies by solution nuclear magnetic resonance (NMR) and X-ray crystallography. Instead, models of amyloid have been developed based on structural information on the amyloid fibre ascertained from several techniques which are described in the following section.[65]

Dye binding

Amyloid is characterised by its binding properties with certain dyes, most notably congo red and thioflavin T. Congo red staining is one of the key diagnostic tests for amyloid. Fibres appear pink-orange after congo red staining when viewed by light microscopy and turn an apple-green colour under polarised light due to birefringence. Birefringence is the decomposition of light into two rays when it passes through certain types of material depending on the polarisation of the light. As this effect only occurs if the structure is anisotropic, the birefringence associated with congo red stained amyloid indicates an ordered submicroscopic structure.[27, 62, 66, 67] Figure 1.5 shows the binding of congo red to amyloid deposits and the characteristic apple-green birefringence under polarized light.
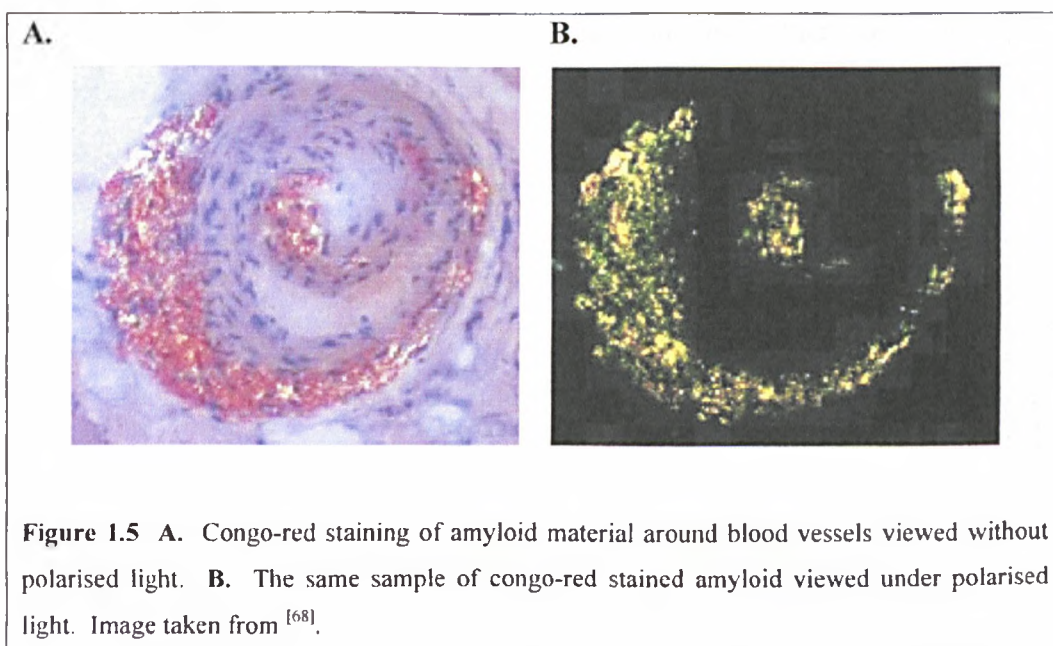
Figure 1.5 A. Congo-red staining of amyloid material around blood vessels viewed without polarised light. **B.** The same sample of congo-red stained amyloid viewed under polarised light. Image taken from [68].

Thioflavin T is a benzothiazole dye that exhibits enhanced fluorescence upon binding to amyloid fibres. The free dye has an excitation maximum of 385 nm and emission of 445 nm, binding to amyloid gives rise to a new excitation maximum at 450 nm and enhanced emission at 482 nm. Monitoring the shift in fluorescence that is observed upon staining amyloid with Thioflavin T can be used to follow aggregation in solution.[27, 62, 66, 67, 69]

Electron microscopy

Amyloid is also identified by certain features that are seen by electron microscopy. Amyloid fibres are uniform, straight, rigid and unbranched. A typical fibre diameter is between 60-130 Å and length is usually around 1000-16000 Å, although length depends heavily on solution conditions.[27, 62] Figure 1.6 shows an example of a typical image of amyloid fibres produced by EM. EM imaging has also identified the 25-35 Å diameter filamentous subunit structure of amyloid fibres known as protofilaments.[27] Protofilaments assemble themselves into pairs or larger groupings by coiling around each other with a long range twist.[64] By averaging multiple EM images of cross sections of fibres under cryogenic conditions, the substructure of amyloid from several precursor proteins has been identified. Amyloid formed from insulin, a fragment of $A\beta$, $A\beta_{1-40}$ and the SH3 domain of phosphatidylinositol-39-kinase have been visualized in this way.[64, 70-72] Cryo-EM images of $A\beta_{11-25}$ fibres at high resolution show striations running across the fibril corresponding to 4.7 Å. This direct visualisation of the β-sheet supports the cross-β model and indicates that more than one β-sheet must be arranged so that β-stands are in direct register perpendicular to the fibre axis.[65, 72]
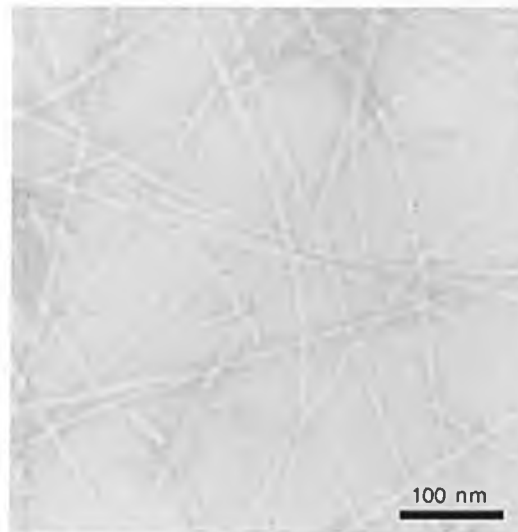
Figure 1.6 Electron microscopy (EM) of negatively stained islet amyloid polypeptide (IAPP) amyloid fibrils. Image taken from [62].

X-ray fibre diffraction

Amyloid fibres are characterised by their X-ray fibre diffraction pattern which shows an ordered, repeating β-sheet conformation perpendicular to the fibre axis known as a "cross-β" structure. X-ray diffraction data provides information on the distances between different features of the fibre. Crystalline order in fibres is usually much lower in directions perpendicular to the fibre axis than in those parallel to the axis therefore equatorial reflections are much weaker and broader.[63] A strong reflection at 4.7 Å in the meridian direction of the X-ray diffraction pattern reflects the hydrogen bonding distance between β-strands perpendicular to the fibre axis. A more diffuse reflection on the equator 10-11 Å represents the distance between the β-sheets. Higher order reflections in the meridian direction reflect the helical twist of the β-strands.[21, 63, 65] Figure 1.7 shows a typical X-ray diffraction pattern of amyloid fibres and highlights the reflections mentioned above.
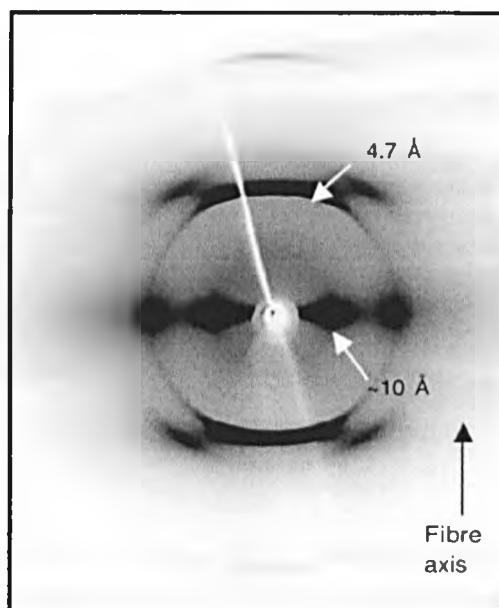
Figure 1.7 X-ray fibre diffraction pattern from aligned IAPP amyloid fibrils, showing the positions of the 4.7 Å meridional and approx. 10 Å equatorial reflections in a cross-β pattern. Image taken from [62].

Solid state NMR

Solid state NMR (ssNMR) data is able to show the distance between selectively labelled residues and is very accurate for distances less than 6 Å.[73] The constraints produced from ssNMR experiments can provide information on the secondary, tertiary and quaternary structure of the amyloid fibre.[74]

Each of the techniques mentioned above have limitations in the information they can provide about the structure of amyloid. However, by combining the constraints produced by ssNMR with the results of other experiments a model has been produced of the Aβ protofilament. The structure, shown in Figure 1.8, agrees with most of the data already gatherered on Aβ fibrils.[75, 76]

Figure 1.8 Structural model for $A\beta_{1-40}$ fibrils, consistent with ssNMR constraints on the molecular conformation and intermolecular distances and incorporating the cross-β motif common to all amyloid fibrils. Residues 1–8 are considered fully disordered and are omitted. Image adapted from [75]
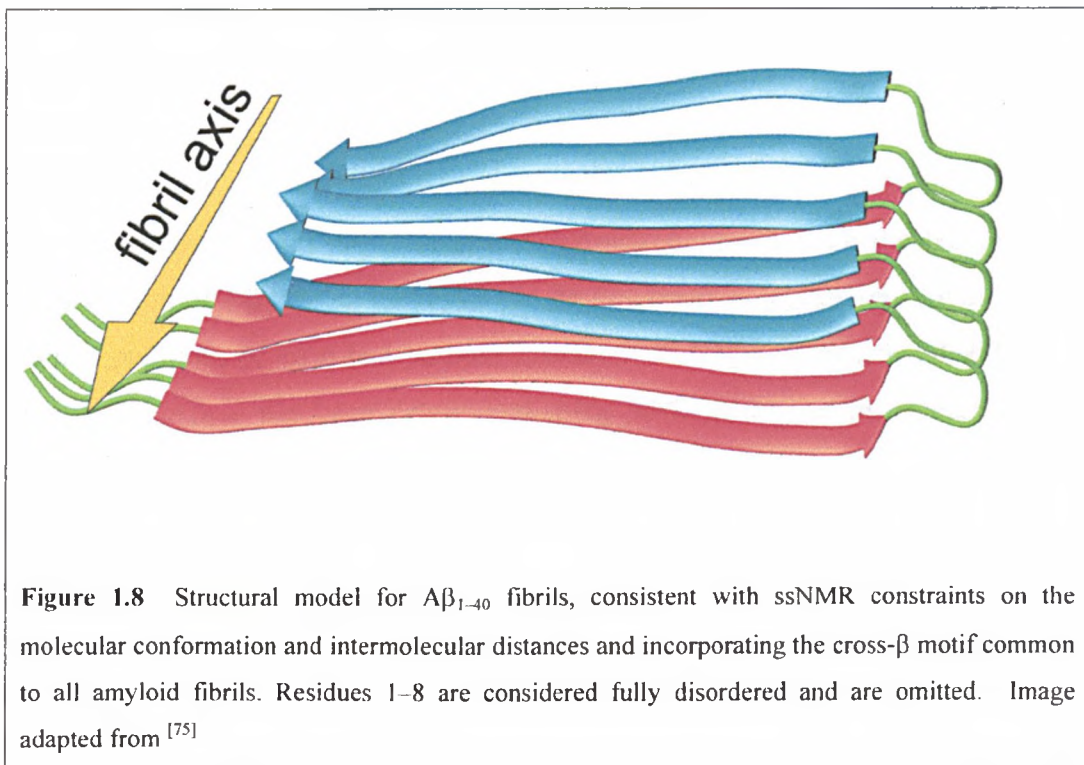
Similarly, the combination of information from several techniques has enabled model structures of other amyloid fibres to be produced, for example transthyretin[77], β2-microglobulin fragment[78] or $A\beta$.[70, 75, 76, 79]

1.2.3 Aβ and amyloid

Aβ peptide

$A\beta$ is the amyloidogenic protein associated with Alzheimers' disease. It has been shown to be derived from $A\beta$ precursor protein (APP), a cell surface receptor for which there is limited knowledge of its function. APP is cleaved by α-secretase into secreted APPα (sAPPα) and c83 or by β-secretase into secreted APPβ (sAPPβ) and c99. c83 or c99 can then be cleaved by γ-secretase into p3 or $A\beta$ respectively. sAPPα has a protective function as it reduces the effects of molecules such as $A\beta$ and glutamate. Cell death may result either from an increase of $A\beta$ or a decrease in sAPPα. Cleavage of APP is thought to occur at the cell surface and involve plasma membrane invaginations.[80]

Neuronal and non-neuronal cells have been shown to process APP differently. Neuronal cells secrete $A\beta_{1-40}$ through the trans-Golgi, but $A\beta_{42}$ is found in the endoplasmic reticulum (ER) and nuclear envelope requiring cell death before it is released. Non-neuronal cells however, produce both at the cell membrane and

secrete them. The majority of Aβ produced is Aβ$_{1-40}$ with only 5-20% being the more amyloidogenic Aβ$_{1-42}$.[81, 82]

Structure of Aβ

Extrinsic factors such as pH, temperature and solvent have a marked effect on the conformation of Aβ.[80] In water Aβ can be described as random coil, with only small populations of local non-random structure.[83] A three-dimensional solution structure of Aβ$_{1-40}$ has been determined using NMR spectroscopy at pH 5.1 in aqueous sodium dodecyl sulfate (SDS) micelles. In this environment the peptide is unstructured between residues 1 and 14 which are mainly polar and likely solvated by water. The rest of the protein adopts an alpha-helical conformation between residues 15 and 36 with a kink or hinge at 25-27. This largely hydrophobic region is likely solvated by SDS.[84]

Aβ amyloid

Fibrillogenesis of Aβ is highly concentration dependent, consistent with a nucleation dependent mechanism, and proceeds via a partially folded intermediate.[85] Fibrillisation of Aβ involves a transient population of α-helical structures, while β-sheet content has been linked to fibrillogenesis and is related to neurotoxicity.[85, 86] The nucleus for fibrillisation is likely to be oligomeric Aβ.[87]

Examination of the formation of Aβ amyloid by EM shows the formation of short, wavy protofibrils at early time points, followed by mature amyloid fibres at later time points.[88] Transient formation of protofibrils of several morphologies such as spheres, rings and short filaments have been observed in *in vitro* preparations of Aβ fibrils.[89] Protofibrils may be on- or off-pathway intermediates in the formation of Aβ amyloid. Hydrogen exchange data suggests that the β-sheet elements comprising the amyloid fibril are already present in protofibrils, but are expanded into some adjacent residues upon the formation of mature amyloid.[90]

Negative staining TEM of mature Aβ amyloid shows a mixture of straight and twisted fibres, with an average diameter of approximately 70 Å.[91] Platinum shadowing TEM of the fibres reveals the periodicity and the crossover structure of the fibrils associated with a fibril supertwist.[70]

Image processing of electron micrographs has shown that amyloid fibrils are composed of protofilaments wound around one another.[88] The environment in which amyloid formation takes place can greatly affect the macromolecular morphology of the fibrils, often showing variation in the number or twists of protofilaments.[70, 74, 88]

Cross-sections of Aβ amyloid indicate that fibrils consist of several protofilaments.[92] The fibril cross section comprises three core regions of higher density, each being approximately 8nm long and 3.5nm wide, that may represent the individual protofilaments.[70] Within the cross section the core regions are joined by end to end contacts. The size of the core region fits with the U-turn like structure described for Aβ (see below). Depending on whether the central core contains a separate protofilaments, the cross section can be fitted to two or three peptide units in cross section.[70] Scanning TEM suggests that each protofilaments contains two β-sheet layers.[88] Figure 1.9 shows diagrammatically how the Aβ fibre is formed from several protofilaments.

Figure 1.9 A representation of the components of the Aβ amyloid fibre. Top diagrams represent the end on view of the fibre, while the lower images are the axial view of the fibre. Image adapted from [72]
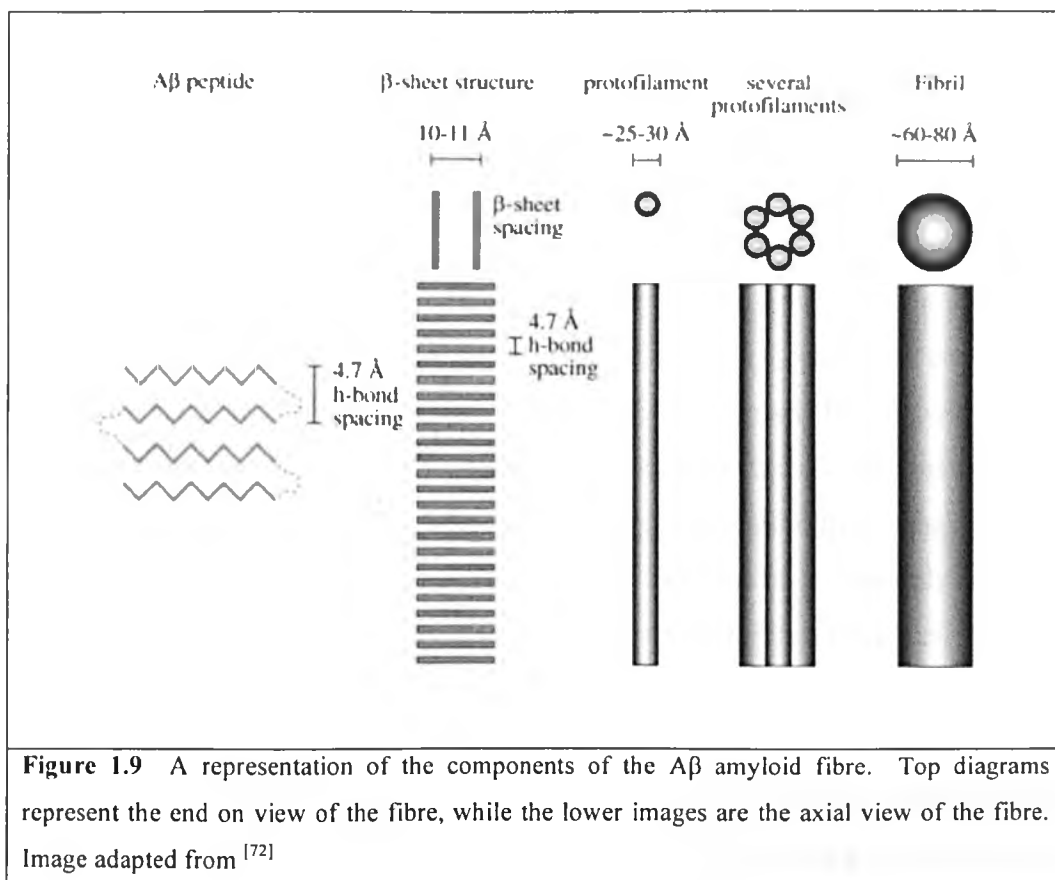
Proline and cysteine mutant scanning of Aβ in amyloid fibres has identified regions of Aβ that are likely to be a turn as well as unstructured regions when Aβ becomes incorporated into amyloid. The thermodynamic stabilities of amyloid fibrils formed from these mutants were used to characterize the susceptibility of different residue positions of the Aβ sequence to proline or cysteine substitution. Both experiments suggest an unstructured N and C termini and two regions between 16-19 and 31-34 that form β-sheet, with the residues in between forming a turn region.[93, 94]

A 3D structure of Aβ$_{1-42}$ fibrils has been modelled using hydrogen-bonding constraints from quenched hydrogen/deuterium-exchange NMR, side-chain packing constraints from pairwise mutagenesis studies, and parallel, in-register beta-sheet arrangement from solid-state NMR studies. Residues 18-42 form a β-strand-turn-β-strand motif that contains two intermolecular, parallel, in-register beta-sheets that are formed by residues 18-26 (β1) and 31-42 (β2). At least two molecules of Aβ$_{1-42}$ are required to achieve the repeating structure of a protofilament. Intermolecular side-chain contacts are formed between the odd-numbered residues of strand β1 of the nth molecule and the even-numbered residues of strand β2 of the (n - 1)th molecule. This interaction pattern leads to partially unpaired beta-strands at the fibrillar ends, which explains the sequence selectivity, the cooperativity, and the apparent unidirectionality of Aβ fibril growth.[79]

Models of Aβ$_{1-40}$ fibres, such as that shown in figure 1.8, based on constraints from solid state NMR spectroscopy, x-ray fibre diffraction and EM show a similar structure to Aβ$_{1-42}$. Residues 12–24 and 30–40 adopt β-strand conformations and form parallel β-sheets. Residues 25–29 contain a turn region that brings the two β-sheets in contact through sidechain-sidechain interactions. The only charged sidechains in the core are those of D23 and K28, which form salt bridges. [75, 95]

1.2.4 Amyloid and disease

Many diseases, known as amyloidoses, are associated with the deposition of amyloid. Table 1.3 shows a summary of these diseases and the amyloidogenic protein associated with the disease. Each clinical syndrome is associated with a specific protein, which is normally soluble, but in the disease state undergoes a

conformational change that causes it to aggregate into insoluble amyloid fibres.[21, 27, 54] A range of other components such as proteins and carbohydrates are also incorporated in the deposits *in vivo*. However, self assembly into amyloid will occur *in vitro* without these other components being present.[60]

Most amyloid is deposited extracellularly and may be either isolated to a single organ or found systemically.[27, 60] The quantity of the deposit can vary greatly from almost undetectable levels in certain neurodegenerative diseases to kilograms found in some systemic diseases.[60] Most mutations in amyloidogenic proteins are associated with enhanced fibril formation due to destabilising the native fold, thus increasing the steady state concentration of the partially unfolded species.

Amyloidogenic protein	Native structure	Disease	Location of amyloid deposits
α-synuclein	Unfolded	Parkinson's disease	Brain
Amyloid β and fragments of	Unfolded	Alzheimer's disease Cerebrovascular amyloidosis Down's syndrome	Brain
Apolipoprotein Al variants	α-helical	Hereditary systemic amyloidosis Familial visceral amyloid	Eyes
Atrial natriuretic factor (ANF)	Unfolded	Atrial amyloidosis	Heart
β2 – Microglobulin	β-sheet	Haemodialysis-related amyloidosis	Musculosceletal system, gastrointestinal and urogenital tracts, tongue, heart.
BRI protein variants	Unfolded	Familial British dementia Familial Dutch dementia	Brain
Calcitonin	Unfolded	Medullary carcinoma of the thyroid	Thyroid
Cystatin C variant	α/β	Hereditary cystatin C amyloid angiopathy (HCCAA)	Brain
Fibrinogen Aα-chain variants and fragments of	Unknown	Hereditary renal amyloidosis	Kidney
Gelsolin variants	α/β	Hereditary systemic amyloidosis	A number of visceral organs and tissues
Huntingtin	α-helical / unfolded	Huntingdon's disease	Brain
Immunoglobulin light chain variable domains	β-sheet	Systemic amyloid light chain (AL) amyloidosis Nodular amyloidosis	Virtually all tissues and organs
Insulin	α-helical (mostly)	Injection-localised amyloidosis	Site of injection
Islet amyloid polypeptide (IAPP, or amylin)	Unfolded	Type II diabetes	Pancreas
Lysozyme variants	α + β	Hereditary systemic amyloidosis	A number of visceral organs and tissues
Prion protein and fragments of	N-terminal unfolded, C-terminal mostly α-helical	Creutzfeldt-Jakob disease (CJD) Gerstmann-Straussler-Schneiker syndrome (GSS) Fatal familiar insomnia Kuru Bovine spongiform encephalopathy (BSE)	Brain
Prolactin	α-helical	Ageing pituitary-gland amyloidosis	Pituitary
Serum amyloid A and fragments of	α-helical	Reactive systemic amyloid A amyloidosis Chronic inflammatory disease	Bladder, stomach, thyroid, gastrointestinal tract, kidney, liver, spleen
Tau protein	Unfolded	Frototemporal dementia Alzheimer's disease Pick's disease Progressive supranuclear palsy	Brain
Transthyretin (TTR) (wild type (WT) and many variants)	β-sheet (mostly)	Senile systemic amyloidosis (wild-type TTR) Familial amyloid polyneuropathy (type I)	Virtually all tissues and organs

Table 1.3 Summary of amyloidogenic proteins and the related amyloidoses.

Toxicity

There is a correlation between amyloid deposition and disease indicating that fibre formation is pathogenic.[61] However, it is not known whether it is the fibres themselves or an intermediate in fibre formation that is the toxic species. There is evidence that soluble oligomeric intermediates are the toxic species rather than insoluble fibrils.[82, 96] A protofibril only fraction has been shown to be neurotoxic.[82] Amyloid has been suggested as a protective measure as early prefibrillar aggregates associated with neurodegenerative diseases can be highly damaging to the cell while mature fibrils are relatively benign.[82] In experiments with non-disease associated proteins under amyloid forming conditioned, early species in aggregation were found to be cytotoxic. This evidence has been used to suggest a common mechanism of pathogenicity in protein folding diseases which is related to the structure of the early aggregates.[97]

Alzheimers' disease

Alzheimers' disease (AD) is the most common form of dementia in the elderly, affecting over 4 million patients in the US alone.[81, 82] The risk of being diagnosed with AD increases with age from a less than 0.05% chance for those under sixty, to over 20% chance in those over eighty. In the developed world the survival rate is approximately six years after diagnosis.[81] Although there are specific criteria used to diagnose AD, at present diagnosis can only be confirmed by autopsy. Hopefully magnetic resonance imaging (MRI) and/or single photon emission computed tomography (SPECT) imaging may allow diagnosis before autopsy and identify presymptomatic patients in the future.

Over 90% of the cases of AD are sporadic, the remainder are familial cases, which usually involve a mutation that increases the amount of $A\beta_{1-42}$, or the propensity of it to form fibres.[82] Missense mutations in presenilin-1 gene are the most common cause of early onset familial AD due to an increase in extracellular and brain concentration of $A\beta_{1-42}$. Polymorphisms in presenelin-1 are also associated with late onset AD and CAA.[98, 99] Down's syndrome patients in their forties nearly always develop AD, most likely due to the trisomy of chromosome 21 where the APP gene is located.[82]

AD is characterised by a progressive loss of cognitive function as well as the presence of amyloid plaques and neuro fibrillar tangles.[82] The pyramidal neurons of cortical layers III and V and often layer II are the neurons most susceptible to loss in AD.[57] Thinning of the cortical gyri, widening of the sulci and enlarging of the ventricles result in an overall loss of grey matter. Aβ amyloid plaques in neurophil and cerebral vessel walls are a hallmark of the AD brain, although Aβ deposits are also present in unaffected individuals. [82, 85, 100]

Aβ can bind to membrane lipids, which may have a detrimental effect on cell function. Aβ disrupts membranes with acidic phospholipids which are present in higher numbers at slightly acidic pH (such as that found in the endosomes).[101] There is significant evidence of Aβ-GM1 ganglioside interactions being involved in Aβ aggregation and may even act as a seed for aggregation.[102] The ability of Aβ to disrupt endosomal and plasma membranes may be one of the causes of Aβ toxicity.[101] Aβ is toxic to vascular smooth muscle cells and there is good correlation between the toxic effect of Aβ and the ability to bind membranes. Oligomers have an increase toxicity and increased binding of lipids. Membrane binding is influenced by the cholesterol concentration of the membrane, with reduced cholesterol content being associated with reduced Aβ binding.[103]

Chronic inflammation plays an important role in AD pathogenesis. There is evidence of Aβ being an inflammatory stimuli to microglia resulting in neuronal loss and cognitive decline. Cathepsin B has been identified as one of over five hundred genes transcriptionally induced by Aβ_{1-42} in murine microglia cells. A high percentage of these genes encode enzymes, particularly hydrolases such as cathepsin B and L. Inhibiting cathepsin B by various means leads to a diminished toxic effect on primary neurons. Inhibiting only extracellular cathepsin B was sufficient to abolish the neurotoxic effects.[104]

Cerebral amyloid angiopathies (CAA)

In cases of CAA, amyloid is deposited in the walls of blood vessels of the CNS.[54] Several proteins are known to form cerebrovascular amyloid including Aβ, hCC, prion protein, transthyretin, gelsolin and ABri.[98] Amyloid deposits affect the leptomeningeal and cortical arteries and arterioles most frequently. Initially amyloid fibrils are deposited in the abluminal aspect of the basal lamina

around smooth muscle cells, then gradually spread towards the internal elastic lamina of arteries and endothelium of arterioles. In capillaries, fibrils are found within basal lamina with larger deposits extending into adjacent neurophils. This results in a loss of smooth muscle cells and the appearance of vascular changes. Mutations in different amyloidogenic proteins have been associated with severe CAA. Polymorphisms in several risk factors have also been reported that are associated with either an increase in risk of CAA or CAA related haemorrhage.[54, 98]

A range of elderly patients with either AD, Downs syndrome, intracranial haemorrhage, cerebral infarction or elderly with no evidence of neurological disorders were all found to have some degree of cerebrovascular amyloid deposits in the leptomeningeal and cortical vessels.[105] The majority of CAA cases are sporadic and are found in elderly individuals with and without AD. 46% of those over the age of seventy have some CAA. CAA associated lobar cerebral haemorrhage accounts for 12-15% of all cerebral haemorrhages in the elderly.[54] Hypertension is not believed to be important factor in CAA. Changes due to CAA can co-exist with atherosclerotic or arteriosclerotic cerebrovascular disease.[106] The usual location of CAA related haemorrhage is the cerebral cortex, most commonly in lateral ganglia, and involves underlying white matter. Haemorrhage is usually accompanied with a high frequency of meningeal signs and symptoms such as headache, vomiting and nuchal rigidity, which are due to rupture into the subarachnoid space.[107]

Alzheimers' Disease related Cerebral Amyloid Angiopathy

The incidence of both CAA and AD increase with age and AD increases the frequency and severity of CAA. Vascular amyloid deposits in the leptomeninges and cortices are found in over 80% of all AD cases, with over 25% of these having moderate or severe CAA.[54, 81, 108] The mean age of onset of dementia is significantly greater in subjects with high levels of CAA than those with low level, but the duration of AD is not influenced by CAA. The frequency of CAA is also age dependant with severity increasing with age. Severity of CAA has been shown to correlate with atherosclerosis and arteriosclerosis in cerebral vessels.[108]

Hereditary Cerebral Amyloid Angiopathy

Hereditary CAA has been described in Icelandic and Dutch families showing autosomal, dominant inheritance of recurrent haemorrhage. In Icelandic families the amyloidogenic protein is hCC, whereas in Dutch families it is Aβ.[55] Although carriers occasionally reach old age, it is not common as a fatal haemorrhage usually occurs in the first few decades of life. Haemorrhage lesions in these patients are most common in the basal ganglia region but are not infrequent in other locations. Occasionally areas of infarction are also found. Hyalinisation and thickening of walls of small arteries and arterioles occurs in the meninges and within brain tissue itself. Similar changes are also seen in the spinal cord. Commonly the intima is separated from the rest of the wall by a definitive space, empty except for delicate strands and the occasional macrophage. An abnormal immune response has been reported and a reduced number of monocytes in the blood.[109]

Human cystatin C and Aβ amyloid

There is a significant correlation between a coding polymorphism in the hCC gene (CST3) and the age of onset of AD. There is also some evidence of a correlation between the CST3 gene and ApoE in increasing the risk of AD.[110] Three polymorphic sites have been identified in the promoter region of hCC, homozygosity in one of these alleles (A/A) is significantly associated with late-onset AD, whereas heterozygosity is not.[111]

Co-localisation of Aβ and hCC has been studied by immunostaining tissue sections with specific antibodies to Aβ or hCC in brain sections from several Aβ amyloid diseases. Immunostained AD brain sections showed co-localisation of Aβ and hCC within arterial walls, but rarely in senile plaques or amyloidotic arteriolar walls. [56-58] Strong anti-hCC and anti-Aβ staining occurs in the pyramidal neurons mainly in the third and fourth layers of the cerebral cortex, the areas associated with AD.[56] hCC staining is sometimes observed within the cytoplasm of pericytes, but predominantly stains the adventitial component of arteriolar walls. However, Aβ staining is usually more diffuse throughout the vessel wall.[58] hCC is deposited less frequently than Aβ, with approximately one fifth of the deposits that are immunoreactive to Aβ co-staining with anti-hCC antibody. Where Aβ immunoreactivity was observed throughout the blood vessel,

cystatin C was found mainly in the outer media to the adventitia of larger cortical and leptomeningeal vessels. [56, 105] hCC associated with Aβ is not an intrinsic part of the amyloid fibril.[98]
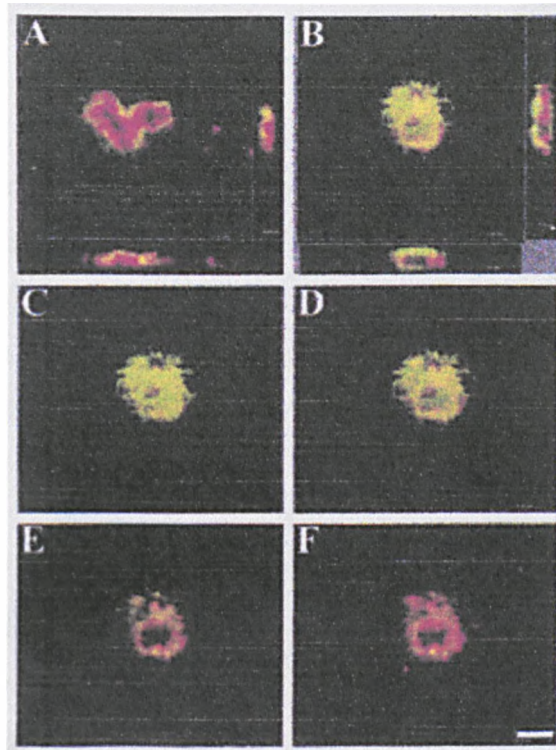
Figure 1.10 Laser confocal microscopy showing distict layering of the plaque. Green represents antibody binding to Aβ, whilst red represent antibody binding to hCC. A and B are images of two different plaques while C-D are different optical sections of B. Imaging shows a immunonegative plaque core, surrounded by a layer of hCC that is enclosed by a layer of amyloid. Image taken from [112]

As shown in Figure 1.10, laser confocal microscopy shows distinct layering to the plaque. An immunonegative core is surrounded by a layer of hCC and a top layer of Aβ.[112] Immunoprecipitation assay indicate that hCC binds to full length APP and secreted APP. Deletion mutants of APP localise the binding site to the Aβ region of APP. Analysis of the association between hCC and Aβ suggests a specific, saturable high affinity 1:1 binding between hCC and Aβ. The association of hCC and Aβ results in a concentration dependent inhibition of Aβ fibril formation. Co-incubation of hCC with monomeric Aβ significantly reduces formation of Aβ oligomers and protofibrils, but hCC is not able to dissolve Aβ oligomers.[113, 114]

The colocalisation of hCC and Aβ is not limited to AD. In patients with HCHWA-D a majority of patients co-stained for hCC in blood vessels that stained for Aβ. A third of patients with sporadic Aβ CAA co-stained for hCC.[55] In cases of sporadic inclusion body myositis, the most common muscular disease of the elderly, pathology includes vacuolar degeneration of muscle fibres, intrafibre Aβ deposits and focal cytoplasmic paired helical filaments containing phosphorylated tau, both of which are congophilic. 80-90% of the vacuolated fibres show numerous, well defined plaque-like inclusions that are strongly immunoreactive with anti-hCC antibody. The hCC either co-localised or was adjacent to Aβ immunoreactive inclusions.[59] When hCC is the main amyloidogenic protein, such as the cases of HCHWA-I, no co-immunostaining has been reported, deposits immunostain only with anti-hCC despite using several anti-Aβ.[55, 58]

1.3   FOLDING AND AMYLOID

1.3.1 Folding intermediates and amyloid

In order to understand the process of amyloid formation, it is important to consider what causes amyloidogenic proteins to form amyloid. It is believed that the initiation of amyloid formation is dependant on the formation of a partially unfolded intermediate.[87, 115] This is not surprising considering that amyloidogenic proteins of diverse structures all rearrange into a cross-β structure in the amyloid fibre. For proteins that are predominantly β-sheet in the native state this may require relatively minor structural changes. However, a considerable proportion of amyloidogenic proteins have a high α-helical content (as shown in table 1.3). In such cases it is clear that the amyloidogenic protein must break native contacts prior to aggregation.

Further evidence supports the role of partially folded intermediates in the formation of amyloid. Many mutations that are associated with amyloidosis, including the L68Q mutation in hCC, reduce the stability of the protein relative to the wild-type protein.[116] Whether the mutation destabilised the native state or stabilises the unfolded or partially folded state, the result is a reduction in protein in the native state and an increase in partially folded intermediates and unfolded protein.[117] In addition to this, amyloid formation is often induced in vitro by

incubating in conditions that destabilise the native state. It has been found that many proteins that are not associated with the formation of amyloid *in vivo* will form amyloid *in vitro* if placed under suitably denaturing conditions.[123, 163]

1.3.2 Models of amyloid formation

The structural studies of amyloid mentioned earlier in the chapter provide us with valuable information about the end point of amyloidogenesis, but provide little information on the mechanism of amyloidogenesis. Kinetic experiments are able to provide information on the process of amylodogenesis. The formation of amyloid can be followed using a fibril sensitive probe such as thioflavin T or congo red or by following changes in secondary structure.

Multiple models of the common core structure of amyloid fibres have been proposed. These models can be grouped into three classes; refolding models, gain of function models and natively disordered models. Although each model class proves some insight into how amyloid fibres form, no single model is yet to provide a full explanation of amyloid behaviour.[118]

Refolding model

Refolding models propose that the fibrillogenic protein exists in two distinct states, the native state and the fibrillar state. In order to convert between these two states the protein must unfold then refold in the alternative conformation. It has been suggested that backbone interactions, which are available to all protein sequences, dominate the fibrillar state.[119] The insulin protofilament model is an example of such a model. Although the disulphide bonds in the native structure are maintained, insulin has to completely unfold from its native α-helical structure to form the cross-β structure of the amyloid fibre.[64]

Natively disordered models

Natively disordered models are based on proteins where the native state is unstructured. They generally describe only segments of proteins so the full protein may have structured regions in the native state but the region responsible for fibril formation is unstructured. An example of this type of model is the expanded polyglutamine region of huntingtin protein. Several models of the

huntingtin amyloid fibre have been proposed, each one proposes the polyQ region as a nucleation site for amyloid formation.[120-122]

Gain of function models

Gain of function models propose that fibre formation is driven by a conformational change in a limited region of the native structure of the amyloidogenic protein which exposes an interaction surface that would otherwise be inaccessible in the native state. This group of models can be further divided into four groups; direct stacking, cross-β spine, 3D domain swapping and 3D domain swapping with cross-β spine.[118]

Direct stacking models

Direct stacking models require a small conformational change to the native state that allows fibril formation via the stacking of subunits. For example, in the transthyretin model, a conformational change displaces the terminal β-strand of the β-sheet exposing the penultimate strand that can form novel β-sheet interactions between transthyretin subunits. Transthyretin subunits stack up head-to-head and tail-to-tail producing the repeating β-sheets of the amyloid fibre.[77, 123, 124]

Cross-β spine model

The cross-β spine model proposes that short segments of the native protein change conformation to form one or more β-strands of a cross-β spine. The atomic structure of a seven-residue peptide of Sup35 which forms amyloid-like fibres has been determined. The structure reveals a double β-sheet bound by a steric zipper, with each sheet formed from parallel segments stacked in register.[125] A cross-β spine model of β2-microglobulin (β2m) has been proposed where residues 83-97 form a β-hairpin in the fibril that stack into antiparallel β-sheets. Two sheets pack together to form the cross-β spine.[126] A direct stacking model has also been proposed for β2m in which the N and C terminal strands are displaced and form the core from native-like β2m.[127]

3D domain swapping models

Eisenberg was the first to observe the process known as three-dimensional domain swapping after studying diphtheria toxin.[21, 86] In 3D domain swapping two or

more subunits exchange domains. The domain can be any part of the protein, from a short segment of secondary structure to a large complete functional globular domain. A flexible linker or hinge region is needed to allow partial unfolding, this region is the only region where the structure is different to the monomer. The main adhesive force is the "closed interface" between the swapped domains. The linker region forms the "open interface" which provides some additional stabilisation to the dimer.[21] hCC and human prion protein are two examples of domain swapped proteins that are also known to form amyloid.[19, 128] 3D domain swapping and the cystatins are discussed in detail later in the chapter.

Models of domain swapping can be based on a closed ended swap, where the swap is reciprocal and there are no exposed domains. Models have also been proposed where an open ended "runaway" domain swap occurs. In such models there are exposed domains at each terminus of the domain swapped oligomer that allow propagation of the oligomer. Figure 1.11 shows a representation of 3D domain swapping and possible mechanisms of propagation in the amyloid fibre.
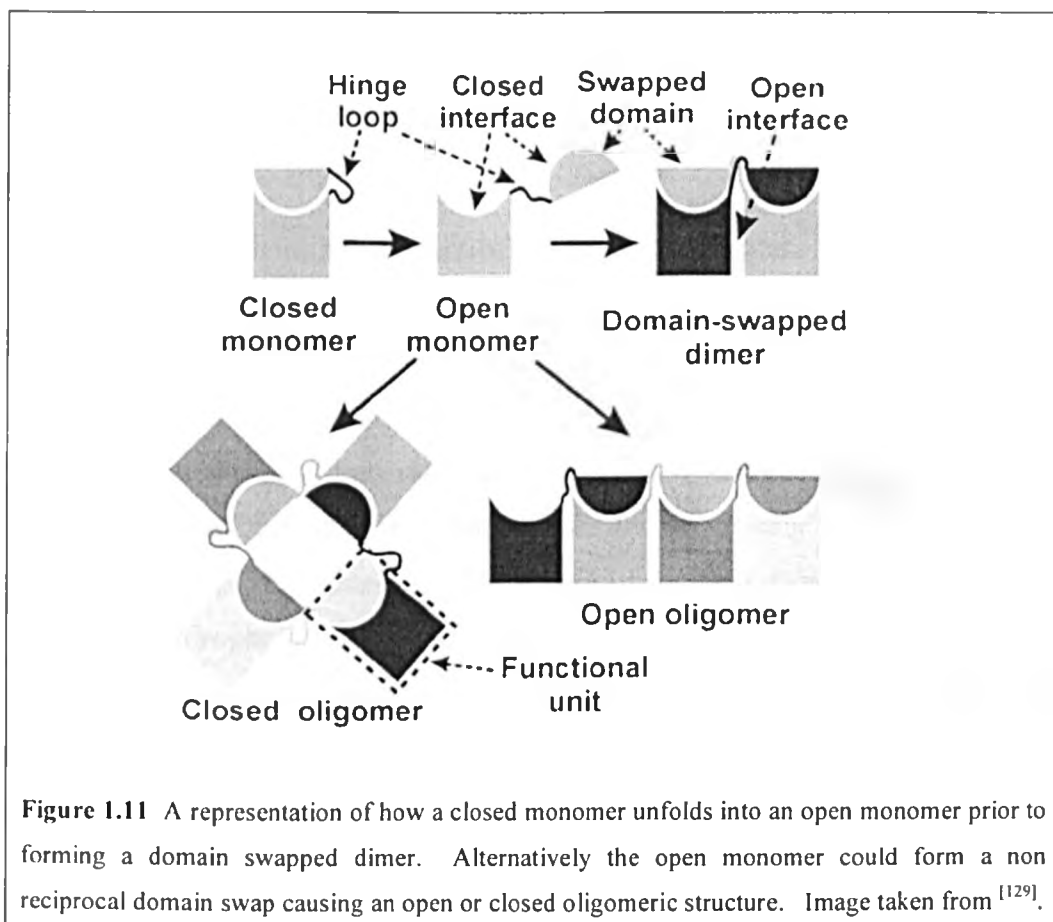
Figure 1.11 A representation of how a closed monomer unfolds into an open monomer prior to forming a domain swapped dimer. Alternatively the open monomer could form a non reciprocal domain swap causing an open or closed oligomeric structure. Image taken from [129].

3D domain swapping with cross-β spine model

A 3D domain swapping with cross-β spine model has been proposed for a polyglutamine insertion mutant of RNAse A. The model incorporates a runaway domain swap of the C-terminal strand and the cross-β spine is formed from the polyglutamine insertion. The polyglutamine spine is comprised of 2 β-sheets forming a steric zipper. The remainder of the protein remains in its native state and packs around the outside of the cross-β spine. Although this model incorporates a runaway domain swap, the data does not exclude fibril formation by direct stacking of domain swapped dimers or oligomers.[118, 130]

1.3.3 3D domain swapping

Domain swapped oligomers are made up of subunits that are identical in structure to the monomer with the exception of the hinge loop and the secondary interface. These conformational differences can either favour or disfavour domain swapping and will greatly influence the equilibrium between monomer and domain-swapped forms.[131] Studies on the mechanism of domain swapping reveal that domain swapping is closely linked with the folding reaction. Generally there is a large kinetic barrier separating the interconversion between monomer and domain-swapped species. For interconversion to occur many native interactions, often in the hydrophobic core, must be disrupted to be replaced with by identical interactions with another protein chain.[131]

The more independent the structure of the swapping elements the later during the folding pathway they can associate. In many cases the swapping elements are highly intertwined and not independently stable. Consequently association will occur before too much of the structure is present. [131] In the case of barnase, the structures that domain swap can only fold partially in an independent manner. Therefore it has been proposed that association must occur at the latest at an intermediate stage of folding.[132] In the case of suc1 the domain swap occurs in the central hydrophobic core. The interactions that connect the domain swapped region are formed very early in the folding reaction, in order to break these interactions substantial unfolding is required. A model has been proposed in which a suc1 molecule can fold from the denatured state to either monomer or a domain-swapped dimer.[133]

Domain swapping and the cystatins

Members of both of the main subfamilies of cystatins have been shown to form dimers via 3D domain swapping.[134] In the ^{1}H-^{15}N HSQC spectrum of the hCC dimer sixty-nine percent of the residues show no significant change in chemical shift relative to the monomer. The changes that do occur are found to be in residues close to the protease binding site. The three regions that show changes in chemical shift are R8-G11 in the N terminus, Q55-N61 in the first hairpin loop and Y102-L112 in the second hairpin loop. [25] The dimer is symmetrical as there are no extra signals in the a ^{1}H-^{15}N HSQC spectrum of the dimer compared to the monomer.[25] A minimum requirement for domain-swapping is the unfolding of the monomer so that the α-helix is separated from the β-sheet and β2 and β3 are torn apart. When domain swapping occurs the open interface forms two new long antiparallel β-strands running from β2-β3 incorporating L1 into the β sheet.[19, 21, 134] The β bulges that cause the sheet to curve slightly are preserved in the dimer.[21]

The structure of the transition state en route to dimer is closer to the unfolded state than the folded state showing that a considerable disturbance of structure is required for transition from monomer to dimer.[134] The reaction is limited by a bimolecular process, most likely limited by the association of predominantly unfolded proteins.[134] The hinge region must account for any favourable energy changes, the most likely source of stability is alleviation of a distortion in the turn connecting strands two and three of the monomers. End to end distance in the crystallographic dimer structure (60-80 Å) closely matches the width of the amyloid fibrils. [134]

1.3.4 Further aggregation

The kinetics of fibril formation has a lag phase which disappears, or is reduced, if a seed or preformed nucleus is present in the supersaturated solution.[86] Models of fibril growth based on the principles of nucleated processes suggest that the first phase of aggregation is the formation of oligomers as a result of non-specific interactions or domain swapping.[86] Large amorphous aggregates are formed early in amyloid formation and may provide a high local concentration to facilitate nucleation.[61] The earliest species detectable by EM and AFM are

beadlike amorphous aggregates or micelles, which anneal to form chain-like protofibrils, then proceed to form annular species and/ or amyloid.[60, 82] Protofibrils are thought to be on-pathway intermediates in the formation of amyloid. Aβ protofibril elongation involves both the incorporation of monomers and the association of immature protofibrils, these dissociate via the release of monomers only.[61]

The structural similarity that defines amyloid fibres is principally at the level of β-sheet folding within the protofilament, while the different types vary in the assembly of their protofilaments.[135] While protofilaments may assemble into fibres with variable packing, the protofilament interactions are maintained along the length of the fibril. This suggests that protofilaments adopt a series of interchain interactions during their assembly that propagate during fibrillar growth.[64]

Further aggregation of cystatins

The crystal structures of hCC revealed interesting packing interactions between dimers that may be significant for further aggregation, specifically four dimers arranged in a crystallographic tetrad.[21] Prolonged incubation of chicken cystatin under destabilizing, but predenaturing conditions leads to further self association into dimers and tetramers, but no trimeric or pentameric species have been reported. Non-reducing sodium dodecyl sulfate-poly acrylamide gel electrophoresis (SDS-PAGE) and mass spectrometry (MS) shows no intermolecular disulphide bonds are present. The overall cystatin fold can be observed in the tetramer using far-UV-circular dichroism (far-UV-CD). Only marginal differences in tryptophan residues are seen by near-UV-CD. 1-anilino-8-naphthalene sulfonate (ANS) binding shows an increased solvent exposure of protein core in tetramer.[136]

The guanidine hydrochloride (GdnHCl) dependence of the tetramerization reaction indicates that monomer to dimer rate is limited by a major unfolding event whereas tetramer formation is limited by a slow conformational rearrangement within the dimer. The hydrodynamic radius of the tetramer (approximately 46 Å) is appropriate for dimensions of the amyloid fibril of cystatins (7-13 nm) suggesting it could be the building block of the fibril.[136]

Studies on the tetramer of stefin B indicate that two domain swapped dimers form tetramer by a process that occurs with a *trans-* to *cis-*proline isomerisation in the loop between strands 4 and 5. This isomerisation shifts the position of the loop sufficiently to allow the formation of an interface between the two domain swapped dimers involving the loop and the N-terminus. This is not the first time a proline isomerisation has been suggested as a key step in the amyloidogenesis pathway. Studies of β-2-microglobulin have identified a folding intermediate containing a non native trans-proline isomer as a direct precursor of fibril elongation. Structural analysis of this intermediate shows that it is highly native like in structure but contains perturbations of the edge strand which increase the likelihood of self aggregation.[137, 138]

1.3.5 *In vivo* factors

Amyloid deposits *in vivo* have also been found to contain accessory molecules, such as serum amyloid P, α_1-antichymotrypsin and apolipoprotein E, in addition to the main amyloidogenic protein.[27] However, amyloid fibres that appear identical to their *in vivo* counterparts can be produced *in vitro* without the requirement for accessory molecules. Whilst not essential for amyloid formation the role of these accessory molecules could be analogous to the molecular chaperones found in protein folding.[54]

Serum Amyloid-P

Serum amyloid-P (SAP) has been shown to bind all amyloid like fibres formed *in vitro* and is found ubiquitously associated with amyloid deposits *in vivo*. *In vitro* SAP stabilises amyloid deposits from phagocytic and proteolytic degradation. SAP requires the presence of calcium, at a concentration of 0.1 mg/ml SAP pentamer binds every 50 apoC-II monomer in the fibril. The effect of SAP on fibril sedimentation suggests a higher order reaction than a one-to-one binding. In the presence of SAP EM analysis shows a distinct "clumping" of fibrils without changing the morphology of the fibres. In the absence of SAP fibre-fibre interactions are weak and more susceptible to shear.[139]

α₁-antichymotrypsin

The 45KDa serine protease inhibitor, known as α_1-antichymotrypsin, has been identified through screening for amyloid and paired helical filaments (PHF) binding. α_1-antichymotrypsin staining of aging and AD brains shows plaques and vascular deposits almost identical to the anti-amyloid pattern. Colloidal gold labelled secondary antibodies localised α_1-antichymotrypsin to Aβ in neuritic plaque cores from AD showing that α_1-antichymotrypsin is an integral part or very tightly bound to the amyloid.[140]

Apolipoprotein E (ApoE)

ApoE is a circulating secretory protein which, although synthesised primarily in the liver, is synthesised locally by astrocytes and microglia in the brain. There is an absence of most other plasma apolipoproteins in the brain, making ApoE the major apolipoprotein of the cerebrospinal fluid (CSF). Brain ApoE is believed to play a role in the redistribution of lipid and cholesterol during membrane repair and synaptic plasticity as well as transport of ApoE containing lipids in the CSF. [81, 141, 142] ApoE has been shown to bind all amyloid like fibres formed *in vitro* and is found ubiquitously associated with amyloid deposits *in vivo*.[139]

There are three common alleles of ApoE, ε2 (8%), ε3 (78%) and ε4 (14%). ApoEε4 has been suggested as a risk factor for AD, AD associated CAA and sporadic CAA.[54, 143-145] ApoEε4 has been shown to increase the risk of AD in younger age groups, while ApoEε2 confers a reduced risk.[110, 141] In AD brains the number and size of Aβ deposits increases in ApoEε4 carriers. It is also associated with amyloid burden in other brain disorders and in normal aging. Transgenic mice studies show an increased ApoE mRNA with aging and a great elevation in insoluble ApoE.[81, 141] ApoEε4 has also been reported as increasing the risk of CAA associated haemorrhage independently of AD.[146] The severity of CAA has been linked with a significant increase in the ApoEε4 genotype reporting that single ApoEε4 allele increased the odds of moderate or severe CAA by almost 3 fold and two copies of the allele increased the odds by over 13 fold.[144]

OVERVIEW OF THESIS

The overall aim of this project is to characterise the mechanisms that lead to the formation of cystatin amyloid fibres. cC has proved a useful model in studying cystatin amyloid, but relatively little information regarding the human counterpart hCC is known. Studying the folding and oligomerisation of hCC and relating it to cC adds to the developing model of cystatin amyloid formation. A second aim to this project is to characterise the interaction between hCC and Aβ.

The initial stages of this project were devoted to the development of a protocol to produce recombinant hCC. The strategy used in the development of this protocol is described in detail in Chapter 3. The characterisation of the folding pathway of hCC and comparison with the folding pathway of cC is the basis of Chapter 4. Chapter 5 discusses the oligomerisation observed in the early stages of cystatin amyloid formation. Chapter 6 describes the use of NMR to study the interaction between hCC and $A\beta_{1-40}$.

CHAPTER TWO
MATERIAL AND METHODS

2.1   INTRODUCTION

This chapter includes the details of the common experimental procedures used throughout the work presented in this thesis. Further details of the material and methods that are relevant to specific chapters are found at the end of each chapter.

2.2   BUFFERS AND REAGENTS

Unless otherwise stated all reagents were purchased from BDH or Sigma-Aldrich. Deionised water from a Sartorius 611VF water purifying system was used throughout all experiment. Phosphate and Tris-HCl buffer was made according to the protocol in Sambrook et al.[147]

2.3   DNA MANIPULATION

2.3.1   Expression vectors

Wild type hCC

Wild type hCC cloned into the pIN-III-ompA expression system was provided by Dr, R.A. Staniforth. Expression and purification of hCC is discussed further in chapter 3.

P103A chicken cystatin

Recombinant cC cloned into the pIN-III-ompA periplasmic expression system was provided by E.A. Auerswald (Ludwig-Maximilians-University, Munich, Germany).[17] Expression was carried out in *E. coli* TG1 strain for which an efficient purification protocol had been established.[148] Mutagenesis on the wild type plasmid was carried out by Dr. M. El-baghdady to produce the P103A mutant.

2.3.2   Plasmid extraction

Plasmid DNA was extracted from 5ml overnight growths using QIAprep Spin Miniprep kits (Qiagen), according to the protocol provided by the manufacturer. Plasmid was eluted using 10 mM Tris-HCl pH 8.5, quantified and stored at -20°C.

2.3.3 Evaluating DNA concentration

The concentration of plasmid and oligonucleotide solutions was calculated using absorbance at 260 nm where an absorbance reading of 1 is equivalent to a DNA concentration of 50 μg/ml. The purity of plasmid preparations was assessed by comparing the ratio of absorbance at 260 nm and 280 nm assuming a pure DNA sample has a ratio of approximately 2. All absorbance readings were taken on a Varian Cary spectrophotometer.

2.3.4 Agarose gel electrophoresis

1% Agarose gel

A 1% agarose gel was prepared by placing 0.5 g electrophoresis grade agarose in 50 ml 1x TAE buffer and gently heating in a loosely stoppered conical flask until all the agarose had dissolved. The solution was poured into the taped gel tray of a Biorad mini sub DNA cell, the well comb inserted and allowed to cool. Once the gel has set the tape was removed from the tray prior to it being placed in the gel tank and submerged in 1x TAE buffer. The apparatus was ran at 50 V for approximately 1 hour. Gels were stained in 1x TAE containing 0.5 μg/ml ethidium bromide for approximately 1 hour. Gels were viewed on a transilluminator and photographed using a Polaroid instant camera.

DNA molecular weight standards

GeneRuler 1 kb DNA ladder (Fermentas) was prepared and used according to the manufacturers protocol. The ladder contains 14 fragments of 10000, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500, 1000, 750, 500 and 250 base pairs.

2.3.5 Competent cells

Preparation of competent cells

The *E. coli* strain XL10 Gold was routinely used for plasmid production and XL1 Blue were used during mutagenesis experiments. TG1, Rosetta, Origami and Rosetta-gami strains of *E.coli* were used during the optimisation of the expression of human cystatin C which is discussed further in chapter three. The TG1 strain was used for expression of chicken cystatin. Supercompetent XL1 Blue cells were purchased from Stratagene, competent Rosetta and Rosetta-gami cells were purchased from Novagen, all other strains were made competent using the

following technique. In order to maintain their phenotype Rosetta require 34 µg/ml chloramphenicol, Origami require 12.5 µg/ml tetracycline and 15 µg/ml kanomycin, Rosetta-gami have to combined antibiotic requirements of Rosetta and Origami cells. XL1 Blue, XL10 Gold and TG1 cells were purchased from Stratagene. Rosetta, Origami and Rosetta-gami cells were purchased from Novagen.

Glycerol stocks of cells stored at -80°C were streaked onto Luria Bertani media-agar (LB-agar) containing any antibiotics required to maintain the strain and incubated at 37°C overnight. A single colony was picked to inoculate 5 ml of Luria Bertani media (LB) which was then incubated overnight at 37°C, shaking at 250 rpm. 10 ml of LB was inoculated with 200 µl of the overnight growth and left shaking at 37°C until the optical density at 600nm (OD_{600}) reached 0.6. Cells were centrifuged at 1663 g for 10 minutes to pellet the cells. The supernatant was discarded and the cells gently resuspended in 3.3 ml RF1 buffer. After incubation on ice for 30 minutes the cells were again pelleted by centrifugation. The supernatant was discarded and the cell pellet resuspended in 1 ml RF2 buffer. Following incubation on ice for a further 30 minutes cells were divided into 200 µl aliquots and either transformed immediately or frozen in liquid nitrogen prior to storage at -80°C.

RF1 buffer

30 mM KCH_3CO_2, 100 mM RbCl, 10 mM CaCl, 50 mM MnCl4, 15% glycerol, pH 5.8

RF2 buffer

10 mM MOPS, 10 mM RbCl, 75 mM CaCl2, 15% glycerol, pH 6.5

2.3.6 Transformations

0.5 µl of plasmid DNA at an approximate concentration of 50 µg/ml was put into a chilled 14 ml polypropylene tube. 100 µl of competent cells were added and the tube and gently mixed prior to being left on ice for 30 minutes. Tubes were placed at 42°C for 90 seconds then returned to ice for a further 30 seconds. After adding 900 µl of warmed LB tubes were incubated at 37°C for 1 hour shaken at

250 rpm. Where cells were purchased a supercompetent or competent cells transformations were carried out following the protocol provided by the supplier. 20 µl was streaked and 20 µl and 100 µl cells plated onto selective LB-agar plates and incubated at 37°C overnight.

2.4   Growth media and solutions

2.4.1   LB media (Luria Bertani media)

Per litre of water,

Bacto- tryptone	10 g	(Difco)
Bacto- yeast extract	5 g	(Difco)
NaCl	10 g	

The solution was sterilised by autoclaving at 121°C for 20 minutes. LB-agar was produced by adding 15 g/l of agar to the LB prior to autoclaving.

2.4.2   M9 minimal media

Per litre of water,

Na_2HPO_4	6 g
KH_2PO_4	3 g
NaCl	0.5 g

The media solution was adjusted to pH 7.4, then made up to 1 l before being sterilised by autoclaving at 121°C for 20minutes. The following were added to the above solution after autoclaving:

trace elements[†]	650 µl	
glucose[*]	3 g	
1 mg ml^{-1} thiamine[*]	1 ml	
0.5 mg ml^{-1} ($^{15}NH_4)_2SO_4$[*]	2 ml	(Cambridge Isotope Laboratories)
1M $MgSO_4$[†]	1 ml	
1 M $CaCl_2$[†^]	0.1ml	

[†] autoclaved at 121°C for 20 mins, [*] 0.2 µm syringe filter sterilised, [^]Added last. The flask was swirled immediately to disperse precipitate. The preparation was abandoned if precipitate did not disperse.

Trace elements

Per 100 ml of water:

$CaCl_2.2H_2O$	550 mg

$MnSO_4.H_2O$	140 mg
$CuSO_4.5H_2O$	40 mg
$ZnSO_4.7H_2O$	220 mg
$CoCl_2.6H_2O$	45 mg
$Na_2MoO_4.2H_2O$	26 mg
H_3Bo_4	40 mg
KI	26 mg

The above were added to 70 ml of water and the solution adjusted to pH 8.0 before adding:

EDTA   500 mg

The solution was adjusted to pH 8.0 before adding:

$FeSO_4.7H_2O$ 375 mg

The solution was then made up to 100 ml with water before autoclaving at 121°C for 20 mins.

2.4.3   Antibiotic solutions

Ampicillin   (Melford)

A 1000x stock solution was produced by dissolving in water at a concentration of 100mg ml^{-1}, then filter sterilised with a 0.2 μm syringe filter. 1 ml aliquots of the stock solution were stored at -20°C until required when they were gently thawed and added to growth media at a final concentration of 100 μg ml^{-1}.

Chloramphenicol

A 1000x stock solution was produced by dissolving in ethanol at a concentration of 34 mg ml^{-1}. 10 ml aliquots of the stock solution were stored at -20°C until required when they were added to growth media at a final concentration of 34 μg ml^{-1}.

Tetracyclin

A 400x stock solution was produced by dissolving in ethanol at a concentration of 5 mg ml^{-1}. 10ml aliquots of the stock solution were stored at -20°C until required when they were added to growth media at a final concentration of 12.5 μg ml^{-1}.

Kanomycin

A 1000x stock solution was produced by dissolving in water at a concentration of 15 mg ml^{-1}, then filter sterilised with a 0.2 μm syringe filter. 1 ml aliquots of the stock solution were stored at -20°C until required when they were gently thawed and added to growth media at a final concentration of 15 μg ml^{-1}.

2.4.4 Isopropyl-β-D-galactosidase (IPTG) (Melford)

A 1000x stock solution was produced by dissolving in water at a concentration of 120 mg ml^{-1}, then filter sterilised with a 0.2 μm syringe filter. 1 ml aliquots of the stock solution were stored at -20°C until required when they were gently thawed and added to growth media at a final concentration of 120 μg ml^{-1}.

2.5 PROTEIN EXPRESSION AND PURIFICATION

2.5.1 Human cystatin C

The development and optimisation of the protocol for expression and purification of wild-type human cystatin C is described in full in chapter three.

2.5.2 ^{15}N-labelled P103A chicken cystatin

Protein expression and extraction

The OmpA signal sequence in the chicken cystatin plasmid targets the expressed protein to the periplasm from where protein can be extracted using osmotic shock. Competent *E.coli* strain TG1 were transformed with the pIN-III-OmpA chicken cystatin plasmid and plated onto LB-agar plates containing 100 μg ml^{-1} ampicillin. A single colony was selected from the plate and used to inoculate 5 ml M9 media containing 100 μg ml^{-1} ampicillin which was then grown over night at 37°C shaking at 250 rpm. 2 ml/l of overnight growth was added to fresh M9-ampicillin and grown at 37°C, shaking at 250 rpm until the optical density at 600 nm is 0.5, this takes approximately 6 hours. Cells were induced with a final concentration of 120 μg ml^{-1} IPTG using a 1000x stock solution and left in growth conditions overnight. Cells were harvested by centrifuging at 4402 g for 40 minutes. The growth media was discarded and cells were resuspended in 35 ml 20% sucrose, 25 mM Tris-HCl pH 8.0 per litre of original growth media. Cells were left to stand for 10 minutes before being centrifuged at 48384 g for 15 minutes at 20°C. The sucrose solution was discarded and cells were resuspended

on ice in an equal volume of ice cold 2 mM EDTA, pH 8.0. Cells were left to stand for a further 10 minutes before being centrifuged at 48384 g for 15 minutes at 4°C. The supernatant was recovered and 1 complete protease inhibitor cocktail tablet (Roche diagnostics) per 50 ml was added.

Protein purification

The protocol established by Anastasi *et al*[148] describes the use of a Cm-Papain-Sepharose affinity column to purify chicken cystatin from ovomucin. This protocol was adapted so that chicken cystatin could be purified from the periplasmic extract. A papain affinity column was equilibrated at room temperature with binding buffer, 50 mM potassium phosphate buffer pH 6.5, 0.5 M KCl. The buffer concentration, pH and conductivity of the periplasmic extract were adjusted so that they matched the binding buffer by adding the necessary amount of K_2HPO_4 and KH_2PO_4 stock solutions and solid KCl. The adjusted periplasmic extract was then loaded onto the top of the papain affinity column at a rate of 2 ml/min and the breakthrough eluent collected. The column was washed with the binding buffer until the optical density at 280 nm of the eluent was less than 0.1. Chicken cystatin was then eluted from the column by washing it with the eluting buffer, 0.5 M KCl, 50 mM potassium phosphate buffer pH 11.5, 2 mM sodium azide, until the OD_{280nm} of the eluent was less than 0.05. The affinity column was re-equilibrated in binding buffer for storage. The pH of the cystatin containing fraction was adjusted to a neutral pH prior to concentrating it to 5-10 mls using an Amicon ultrafiltration device.

The chicken cystatin was further purified by size exclusion chromatography (SEC) using superdex-75 resin. The column was equilibrated with the eluting buffer used for the papain affinity column prior to loading the chicken cystatin. Eluting buffer was run through the column at a rate of 3 ml/min and 2.5 min fractions were collected. The pH of fractions thought to contain chicken cystatin was neutralised, then the presence of chicken cystatin confirmed by SDS-PAGE. Once identified chicken cystatin fractions were pooled and buffer exchanged to 10 mM potassium phosphate buffer pH 7.0, 2 mM sodium azide by dialysis.

2.6 PROTEIN PROCEDURES

2.6.1 SDS-PAGE

Buffers

4x upper gel buffer	0.5 M Tris/HCl, 0.4% (w/v) SDS, pH 6.8
4x lower gel buffer	1.5 M Tris/HCl, 0.4% (w/v) SDS, pH 8.8
5x running buffer	25 mM Tris/HCl, 250 mM glycine, 0.1% (w/v) SDS, pH 8.3
1x loading buffer	50 mM Tris/HCl pH 6.8, 100 mM DTT, 2% (w/v) SDS, 0.1% (w/v) Bromophenol blue, 10% (v/v) glycerol
Stain	0.25% (w/v) Coomassie brilliant blue R250, 10% (v/v) acetic acid, 45% (v/v) methanol, 45% water
De-stain	10% (v/v) acetic acid, 45% (v/v) methanol, 45% water.

16 % resolving gel

4x lower buffer	2.5 ml
40 % Bis acrylamide	4 ml (Biorad)
dH$_2$0	3.5 ml
10 % (w/v) APS	100 µl
TEMED	10 µl (Biorad)

4 % stacking gel

4x upper buffer	2.5 ml
40 % Bis acrylamide	1.1 ml (Biorad)
dH$_2$0	6.4 ml
10 % (w/v) APS	110 µl
TEMED	11 µl (Biorad)

SDS-PAGE molecular weight markers

Sigma, low molecular weight range:

Bovine serum albumin (BSA)	66 000 Da
Ovalbumin (chicken egg)	45 000 Da

Glyceraldehyde-3-phosphate dehydrogenase	36 000 Da
Carbonic anhydrase (bovine erythrocytes)	29 000 Da
Trypsinogen (bovine pancreas)	24 000 Da
Trypsin inhibitor (soybean)	20 000 Da
α-lactalbumin (bovine milk)	14 200 Da
Aprotinin	6 500 Da

Biorad, broad range molecular weight standards:

Myosin	200 000 Da
β-galactosidase	116 250 Da
Phosphorylase b	97 400 Da
Serum albumin	66 200 Da
Ovalbumin	45 000 Da
Carbonic anhydrase	31 000 Da
Trypsin inhibitor	21 500 Da
Lysozyme	14 400 Da
Aprotinin	6 500 Da

SDS-PAGE was carried out using a Bio-Rad mini-protean II apparatus. A 16% polyacrylamide resolving gel was prepared and overlain with a 4% polyacrylamide stacking gel containing 10 large or 15 small wells. Protein samples were prepared with a 1:1 ratio of protein solution:SDS loading buffer, of which 5 – 20 µl was loaded onto the gel. Gels were electrophoresed at a constant voltage of 200V for 50 – 60 minutes. Gels were stained with agitation at room temperature for a minimum of 1 hour before de-staining. Rapid de-staining was carried out by boiling the gels in water for 15 – 20 minutes, changing the water at intervals, followed by shaking at room temperature in de-stain for a minimum of 1 hour. The de-stained gels were placed on a light box and photographed using a Polaroid instant camera or scanned and stored as jpeg files using a HP scanjet 3670.

2.6.2   Protein concentration determination

The concentration of protein solutions was calculated using absorbance at 280 nm and the Beer-Lambert law:

(Eq 2.1.1) $\qquad A = \varepsilon c l$

Where A = Absorbance, c = Concentration (M), l = pathlength (cm), ε = molar extinction coefficient ($M^{-1}cm^{-1}$). The molar extinction coefficient of human cystatin C is 11050, P103A chicken cystatin is 11745 and Abeta$_{1-40}$ is 1490. Each were calculated from their sequences using the "ProtParam" tool available on the expasy website (http://au.expasy.org/tools/protparam.html). All absorbance readings were taken on a Varian Cary spectrophotometer.

2.6.3 Protein concentration and buffer exchange

Protein solutions were concentrated using an Amicon ultrafiltration device in conjunction with a Millipore ultrafiltration membrane with a 10 000 Da molecular weight cut off. When the volume of the protein solution was less than 10 ml a Vivaspin centrifugal concentrator with a 5 000 Da molecular weight cut off was used (Vivascience). If the volume of the protein solution was less than 0.5 ml a Microcon centrifugal concentrator with a 10 000 Da molecular weight cut off was used (Millipore). Both the ultrafiltrtion device and the centrifugal concentrators were used according to the manufacturers instructions. The preferred method of buffer exchange was dialysis at 4°C against the desired buffer using Spectra/Por molecular porous membrane tubing, 6-8 KDa molecular weight cut off (SPECTRUM). When small volumes of protein solution required buffer exchange several rounds of concentrating and diluting in the required buffer were carried out until the original buffer was diluted to an insignificant concentration.

2.6.4 Analytical Size exclusion chromatography

The purity and molecular weight of cystatin preps was determined by SEC-high pressure liquid chromatography (HPLC) using a Shodex KW-800 HPLC column and KWG guard column (Shodex, Japan) with a Perkin Elmer Series 200 HPLC instrument. The column was equilibrated with 50 mM sodium phosphate buffer pH 7.0, 2 mM sodium azide at a flow rate of 1 ml/min. The sample injection volume was 20 µl and the elution time was determined by recording absorbance at 280 nm. The column was regularly calibrated using standard proteins and a calibration graph was produced by plotting the elution time against the logarithm of the molecular weight of the standards. Cystatins show a retarded elution time relative to their molecular weight suggesting that there is some interaction with

the matrix in the column. SEC-HPLC was also used to follow decrease in monomer and appearance of oligomers under dimerising condition, further details are found in the material and methods section of the relevant chapters.

2.7 SPECTROSCOPIC TECHNIQUES

2.7.1 Fluorescence

Tryptophan fluorescence measurements were taken on either a Varian Cary Eclipse or a Shimadzu RF-5301PC fluorescence spectrometer. The excitation wavelength was set at 280 nm and emission spectra were recorded between 300-420 nm. Slit widths of 5 nm were used for both excitation and emission. Further details are available in chapter 4.2.2

2.7.2 Circular Dicroism (CD)

CD measurements were taken on a Jasco J-810 Spectropolarimeter. For equilibrium unfolding experiments the protein concentration was 20 μM. Temperature was maintained at 20°C and a 0.1 cm pathlength cuvette was used throughout the experiment. Spectra were recorded over the range of 190-300 nm. Further details are available in chapter 4.2.1

2.7.3 Stopped flow

An Applied Photophysics SX-18MV stopped flow reaction analyser was used to follow the change in fluorescence caused by cystatin folding or unfolding. All reactions were carried out at 25°C in 10 mM sodium phosphate buffer pH 8.0 and varying concentrations of GdnHCl. The monochromator slit widths were set at 5 mm and a wavelength 280 nm was used. Spectra were fitted to a single exponential to provide determine the amplitude change and the observed rate. Further details are available in chapter 4.2.3

2.7.4 EM

Carbon coated copper grids were glow-discharged with 3 x 15 second pulses. 2 μl of sample was adsorbed onto a grid for 1 minute then the edge blotted on filter paper. This was followed by three washes, two in water and one in 2 % (w/v) uranyl formate, pH 4.4 with blotting between washes. Grids were held in a drop of uranyl formate for 20 seconds then removed by blotting and any remaining fluid was removed by gentle vacuum. Micrographs were recorded on a Philips

CM100 microscope operating at an acceleration voltage of 100 kV equipped with a 1024x1024 CCD camera. Further details can be found in chapter 5.2.3

2.7.5 NMR

NMR spectra were recorded on Bruker DRX spectrometers operating at proton frequencies of 500, 600 or 800 MHz controlled using XWinNMR (Bruker). NMR data was processed using Felix (Accelrys). Further details are available in chapter 6.

CHAPTER THREE
PURIFICATION AND CHARACTERISATION OF HUMAN CYSTATIN C EXPRESSED IN *E. COLI* pIN-III-OMPA2 EXPRESSION SYSTEM

3.1   INTRODUCTION

A number of expression systems for hCC are reported in the literature. A low level of hCC expression (0.05 mg hCC/l culture) was achieved when using an *E. coli* cytoplasmic expression system, probably due to a high level of proteolysis of hCC in the cytoplasm or the reducing environment of the cytoplasm disfavouring disulphide bond formation. The periplasm provides a different environment to the cytoplasm and Abrahamson and co-workers produce a substantial amount of hCC using a periplasmic expression system.[16]   In this system, transport of recombinant proteins to the periplasm is achieved by fusing the desired gene to the signal sequence of the *E. coli* outer membrane protein A (ompA), which in their case is under the control of the major rightward promoter of phage λ.[16-18] With this system expression is temperature induced as the promoter is repressed at temperatures below 35°C.[16] Dimerisation and further aggregation of L68Q hCC is highly temperature dependant, raising the temperature from 37°C to 40°C results in a 150% increase in the dimerisation rate.[14] Therefore, an expression system that is not temperature induced expression system would be more suitable for the expression of L68Q hCC.

The pIN-III-ompA periplasmic expression system has successfully been used to produce a chicken analogue of hCC with a high yield.[17] The system is IPTG induced as pIN-III-ompA is under the control of the lipoprotein promoter and the lactose promoter and operator. To create an IPTG inducible expression system hCC was cloned into the pIN-III-ompA vector.

With an ompA expression system production of recombinant protein should be confined to the periplasm. Cell lysis is therefore not required as periplasmic extraction is sufficient to release recombinant protein. Extraction from the periplasm can be achieved using osmotic shock. Periplasmic extraction is a favourable method of protein extraction as the concentration of bacterial proteins

in the periplasm is lower than in the cytoplasm, therefore reducing the amount of purification stages required to produce pure recombinant protein. In this chapter the development of a pIN-III-ompA2-hCC periplasmic expression system is described. The final protocol for the production of hCC shown in section 3.4 yields approximately 4 mg of monomeric hCC per litre of cell culture.

3.2   MATERIALS AND METHODS

3.2.1   Manipulation of the pIN-III-ompA2-hcc plasmid

The pIN-III-ompA2-linker-hCC plasmid was produced by Dr. R.A. Staniforth and has a short linker sequence inserted between the ompA2 signal sequence and the hCC gene to provide the natural ompA cleavage site and the required restriction site to allow cloning of the hCC gene into the pIN-III-ompA2 plasmid. Briefly, an IMAGE consortium clone of hCC in the pBluescript SK vector was supplied by American Type Culture Collection and used to amplify the hCC gene with a short sequence containing an EcoR1 restriction site immediately upstream of the hCC gene and a BamH1 restriction site immediately downstream of the hCC gene. The amplified sequence and the pIN-III-ompA2 were cut with EcoR1 and BamH1 restriction enzymes and ligated together. Sequence analysis confirmed the correct insertion of hCC into the pIN-III-ompA2 vector. Figure 3.1 shows the pIN-III-ompA2-linker-hCC plasmid map and the DNA sequence of the region containing the hCC insert.

Expression of the pIN-III-ompA2-linker-hCC plasmid results in production of a modified hCC with an additional five amino acids at the N-terminus. Although it is not its natural cleavage site, it has been shown that *E. coli* signal peptidase is able to cleave between the C-terminal alanine of the ompA sequence and the N-terminal serine of mature hCC.[16, 18] The pIN-III-ompA2-linker-hCC plasmid was therefore manipulated to produce a plasmid with no linker sequence, pIN-III-ompA2-hCC (Figure 3.1).

In order to remove the linker sequence two primers were designed that only partially annealed to the plasmid. Primer 1 is complementary to the start of the hCC gene and has a 5' sequence that is complementary to the end of the ompA2 signal sequence. Primer 3 is complementary to the start of the non-coding strand of the ompA2 signal sequence and has a 5' sequence that is complementary to the

3' end of the non-coding strand of the hCC gene. Two other primers were designed "2" and "4", that bind to the downstream of the BamH1 site and upstream of the Xba1 site respectively (Figure 3.2A).

CTTT**GAGCGA** **ACGATCAAAA** **ATAAGTGCCT** TCCCATCAAA AAAATATTCT
CAACATAAAA AACTTTGTGT AATACTTGTA ACGCTACATG GAGATTAACT
CAATCTAGCT AGAGAGGCTT TACACTTTAT GCTTCCGGCT CGTATAATGT
GTGGAATTGT GAGCGGATAA CAATTTCACA CAGGAAACAG CTATGACCAT
GATTACGGAT TCACTGGAAC TCTAGATAAC GAGGGCAAAA AATGAAAAAG
ACAGCTATCG CGATTGCAGT GGCACTGGCT GGTTTC**GCTA** **CCGTAGCGCA**
GGCCGCTGAA TTCCATATG**T** **CCAGTCCCGG** **CAAGCCGCC**G CGCCTGGTGG
GAGGCCCCAT GGACGCCAGC GTGGAGGAGG AGGGTGTGCG GCGTGCACTG
GACTTTGCCG TCGGCGAGTA CAACAAAGCC AGCAACGACA TGTACCACAG
CCGCGCGCTG CAGGTGGTGC GCGCCCGCAA GCAGATCGTA GCTGGGGTGA
ACTACTTCTT GGACGTGGAG CTGGGCCGAA CCACGTGTAC CAAGACCCAG
CCCAACTTGG ACAACTGCCC CTTCCATGAC CAGCCACATC TGAAAAGGAA
AGCATTCTGC TCTTTCCAGA TCTACGCTGT GCCTTGGCAG GGCACAATGA
CCTTGTCGAA ATCCACCTGT CAGGACGCCT AGGGATCCGG CTGAGCAACG
ACGTGAACGC AATGCGTTCC GACGTTCAGG CTGCTAAAGA TGACGCAGCT
CGTGCTAACC A**GCGTCTGGA** **CAACATGGCT** **ACTAAATACC**

Figure 3.1. A. Plasmid map of the pIN-III-ompA2-linker-hCC plasmid provided by Dr. R.A. Staniforth. B. Plasmid map of pIN-III-ompA2-hCC plasmid produced following the removal of the linker sequence. C. The DNA sequence of the region containing the hCC insert. The sequence shows part of the pIN-III vector (black), the ompA signal sequence (grey), the linker sequence (pink) and the mature hCC gene (blue). Xba1 (orange), BamH1 (green) and EcoR1 (pink) restriction sites are underlined. Bold regions show the binding position of primers used to amplify the hCC sequence and the ompA2 sequence.

Figure 3.2. In all of the above diagrams the key features are represented by the following colours, hCC (blue), linker (pink), ompA2 (grey), pIN-III (black). A. Annealing positions of primers 1-4. B. The main polymerase chain reaction (PCR) product from PCR of the hCC gene consisting of the full hCC gene with the C-terminal end of the ompA2 sequence. C. The main PCR product from the PCR of the ompA2 sequence consisting of the full ompA2 sequence and the N-terminal end of the hCC gene. D-E. Unwanted annealing of ssDNA during the third PCR, strands either reanneal to their original sequence (D-E) or anneal to the opposite PCR product in a manner than prevents 3' extension. G. Desired annealing of ssDNA that allows amplification of the ompA2-hCC sequence beyond the Xba1 and BamH1 sites.

The hCC gene was amplified using primers 1 and 2 producing the coding strand of the hCC gene with the final 18 bases of the ompA2 signal sequenced attached at the 5' end (Figure 3.2B). The ompA2 signal sequence was then amplified using primers 3 and 4 producing the non-coding strand of the ompA2 signal sequence with the first 20 bases of the non-coding strand of the hCC gene attached at the 5' end (Figure 3.2C). The two purified PCR products were used in a third PCR where no additional primers were included. The four DNA strands may anneal in four different conformations (Figure 3.2D-G). When they anneal as in Figure 3.2G the annealing region is sufficient to act as internal primers and allow 3'-extension. The final PCR product contained the Xba1 site, the ompA2 signal sequence, no linker sequence, the hCC gene and the BamH1 site.

The original plasmid and the PCR product was cut at the Xba1 and BamH1 restriction sites, purified and ligated together. The EcoR1 site in the linker sequence provided a useful means of testing for successful removal of the linker sequence. BamH1 single digests and BamH1 and EcoR1 double digest were used to identify plasmid that cut once with BamH1 and once with BamH1 and EcoR1. Successful ligations were transformed into XL-1 Blue supercompetent cells and plasmids sequenced to confirm removal of the linker sequence.

Primers

Primer 1: 5'-gct acc gta gcg cag gcc tcc agt ccc ggc aag ccg cc-3'

Primer 2: 5'-cgg tat tta gta gcc atg ttg tcc aga cgc-3'

Primer 3: 5'-gg cgg ctt gcc ggg act gga ggc ctg cgc tac ggt agc-3'

Primer 4: 5'-g agc gaa cga tca aaa ata agt gcc t-3'

Primers were analysed by the DNA calculator tool available on the Sigma-Genosys website (http://www.sigma-genosys.com/calc/DNACalc.asp). No secondary structure or primer dimer were predicted for primers 2 and 4, a moderate chance of secondary structure was predicted for primers 1 and 2.

PCR of human cystatin C gene (PCR 1)

Distilled H₂0   40.6 µl

10x Pfu reaction buffer   5.0 µl

dNTPs (25 mM of each)	0.4 µl	(Fermentas)
hCC plasmid (100 ng/µl)	1.0 µl	
Primer 1 (100 ng/µl)	1.0 µl	
Primer 2 (100 ng/µl)	1.0 µl	
PfuTurbo DNA polymerase (2.5 U/µl)	1.0 µl	(Stratagene)

No of cycles	Temperature (°C)	Time (mins)
1	95	2
30	95	0.5
	60	0.5
	72	0.5
1	72	10

PCR product was analysed on a 1% agarose gel (see chapter 2.3.4) and extracted using a QIAquick gel extraction kit according to the manufacturers' instructions.

PCR of ompA2 signal sequence (PCR 2)

Distilled H₂0	40.6 µl	
10x Pfu reaction buffer	5.0 µl	
dNTPs (25 mM of each)	0.4 µl	(Fermentas)
hCC plasmid (100 ng/µl)	1.0 µl	
Primer 3 (100 ng/µl)	1.0 µl	
Primer 4 (100 ng/µl)	1.0 µl	
PfuTurbo DNA polymerase (2.5 U/µl)	1.0 µl	

No of cycles	Temperature (°C)	Time (mins)
1	95	2
30	95	0.5
	55	0.5
	72	0.5
1	72	10

PCR product was analysed on a 1% agarose gel (see chapter 2.3.4) and extracted using a QIAquick gel extraction kit according to the manufacturers' instructions.

PCR of hCC and ompA2 (PCR 3)

dH$_2$O	41.6 µl
10x Pfu reaction buffer	5.0 µl
dNTPs (25 mM of each)	0.4 µl (Fermentas)
hCC PCR product (100 ng/µl)	1.0 µl
ompA2 PCR product (100 ng/µl)	1.0 µl
PfuTurbo DNA polymerase (2.5 U/µl)	1.0 µl

No of cycles	Temperature (°C)	Time (mins)
1	95	2
30	95	0.5
	87	0.5
	72	1
1	72	10

PCR product was analysed on a 1% agarose gel (see chapter 2.3.4) and extracted using a QIAquick gel extraction kit according to the manufacturers' instructions.

Xba1 and BamH1 double digest

10x buffer	2 µl
BSA	0.2 µl
DNA (1 µg/µl)	1 µl
BamH1	0.5 µl (Promega)
Xba1	0.5 µl (Promega)
dH$_2$O	15.8 µl

The pIN-III plasmid and the ompA2-hCC PCR product were cut with Xba1 and BamH1. In each reaction samples were centrifuged briefly then incubated at 37°C for 1 hour. Digested plasmid was analysed on a 1% agarose gel (see chapter 2.3.4) and extracted using a QIAquick gel extraction kit according to the manufacturers' instructions.

Ligation

Plasmid	2.1 µl
Insert	0.4 µl

10xligase buffer	1 µl	
T4 DNA ligase	0.3 µl	(Promega)
dH₂O	6.2 µl	

Reactions were centrifuged briefly then incubated at 4°C overnight.

BamH1 single digest

10x buffer	2 µl
BSA	0.2 µl
DNA (1 µg/µl)	1 µl
BamH1	0.5 µl
Distilled H₂O	16.3 µl

Ligated plasmid was digested with BamH1 to determine whether the linker sequence had been successfully removed. Original plasmid was also cut as a positive control. Digest were centrifuged for a few seconds then left at 37°C for 1 hour.

EcoR1 and BamH1 double digest

10x buffer	2 µl
BSA	0.2 µl
DNA (1 µg/µl)	1 µl
BamH1	0.5 µl
EcoR1	0.5 µl
Distilled H₂O	15.8 µl

Ligated plasmid was digested with BamH1 and EcoR1 to determine whether the linker sequence had been successfully removed. Original plasmid was also cut as a positive control. Digest were centrifuged for a few seconds then left at 37°C for 1 hour.

Successful ligations were transformed into XL1-blue supercompetent cells (Stratagene) according to manufacturers' instruction. Plasmid was extracted from the transformed supercompetent cells according to the method in chapter 2.3.6. The plasmids were sequenced by Lark technologies.

3.2.2 Western blot analysis

Western blot analysis is an immunological technique that provides a method of probing cell fractions for the presence of a specific protein. Initially proteins were separated using gel electrophoresis, then electrophoretically transferred onto a nitrocellulose membrane. The membrane was first blocked with non-fat milk powder to prevent non-specific binding to the membrane. The membrane was then probed with a primary antibody specific to hCC, rabbit anti-human cystatin C (Upstate biotechnology). The membrane was thoroughly rinsed to remove any primary antibody that had bound non-specifically. The membrane was then probed with horseradish peroxidase conjugated pig anti-rabbit secondary antibody, this bound to the species-specific portion of the primary antibody. When the membrane was exposed to a chemiluminescent agent the HRP conjugated to the secondary antibody reacted with the agent producing localised luminescence. An image of the membrane showing the points of luminescence was visualised by exposing the membrane to light sensitive film.

Tris-glycine transfer buffer (Towbin buffer)

Per litre,

Glycine	14.4 g
Tris	3 g
Methanol	200 ml

The buffer was made up to 1 l with water, then sterilised by autoclaving at 121°C for 20 minutes.

Phosphate buffered saline-Tween 20 wash buffer

Per litre,

NaCl	8 g
KCl	0.2 g
Na_2HPO_4	1.44 g
KH_2PO_4	0.24 g
Tween 20	1 ml

The pH of the buffer was adjusted to pH 7.4 using HCl, before making up to 1 l with water.

Cell fractions

Growth media pre-induction and 3 hours post-induction were diluted with equal volumes of SDS-PAGE loading buffer. Whole cell fraction pre-induction and 3 hours post-induction prepared by pelleting 1 ml of culture in a microcentrifuge, resuspending in 100 μl loading buffer and boiling for 10 minutes. Periplasmic extract prepared as described in section 3.4.1. 1 ml of culture was sonicated for 3x 30 secs then pelleted in a microcentrifuge. The cell pellet was prepared as the whole cell fraction. hCC provided by Dr. M. Abrahamson was used as a positive control.

Electrophoretic transfer

SDS-PAGE was carried out on two identical gels according to the method described in chapter 2.6.1. The gel used in the transfer was not stained, the second gel was stained so that it could be used as a reference. The components of a Biorad mini-trans blot cell were soaked in tris-glycine transfer buffer prior to being loaded into the cassette according to the manufacturers' instructions. To ensure good gel-membrane contact, care was taken to remove any trapped air bubbles from the cassette prior to loading into the cell. The buffer tank was filled with tris-glycine transfer buffer cooled to 4°C. Electrotransfer was achieved by applying a constant voltage of 30 V overnight at 4°C.

Protein detection

Following electrophoretic transfer the membrane was transferred into a blocking solution of phosphate buffered saline (PBS)-Tween 20 buffer containing 50 g/l of non-fat milk powder and gently agitated for 1 hour. The membrane was rinsed in PBS-Tween 20 buffer for 1 minute, then incubated in PBS-Tween 20-milk with 1 μl/ml anti-cystatin for 1 hour during which time it was gently agitated. PBS-Tween 20 was used to wash the membrane for 1 minute twice, followed by 15 minutes once, then 5 minutes three times. Incubate the membrane for 1 hour with gently agitation in a 1:1000 dilution of HRP conjugated pig anti-rabbit IgG in PBS-Tween 20-milk buffer. Unbound HRP conjugated pig anti-rabbit IgG was removed by washing in PBS-Tween 20 for 1 minute twice, followed by 15 minutes once, then 5 minutes three times. The membrane was exposed to an electrochemiluminescence (ECL) solution according to the manufacturers'

instructions, then dried and sealed in a plastic sleave. The membrane was exposed to photographic film for 1 minute, 30 and 10 seconds. The film was developed according to manufacturers instructions.

3.2.3 Optimising the expression of hCC

The pIN-III-ompA2-cC plasmid expresses chicken cystatin well in TG1 cells grown at 37°C in LB media yielding 10-15 mg chicken cystatin/l of cell culture. Initial expression trials indicated that when the pIN-III-ompA2-hCC is induced under these conditions there is only a very low level of hCC expression (Figure 3.6A).

Phosphate buffered Luria Bertani media

Per litre,

Tryptone	10 g	(Difco)
Yeast extract	5 g	(Difco)
1M KCl	100 mls	
1M Buffer*	100 mls	

* For pH 7.0 use 1M MOPS (3-(N-morpholino)propanesulphonic acid), for pH 9.1 use 1M AMPSO (3-[1,1-dimethyl-2hydroxyethyl)amino]-2-hydroxypropanesulfonic acid). pH was adjusted using potassium hydroxide then the volume made up to 1 l with water. The media was sterilised by autoclaving at 121°C for 20 minutes.

3.2.4 Expression trials

2 ml/l overnight starter culture was used to inoculate 50 ml of the appropriate growth media in a 250 ml conical flask. All trials contained 100 μg/ml ampicilin and any additional antibiotics required to maintain the *E.coli* strain. Cultures were grown to an OD_{600} of 0.5, then induced with 120 μg/ml IPTG. 1 ml samples were taken prior to induction with IPTG, then at hourly intervals and finally after overnight growth. Samples were prepared for SDS-PAGE immediately by centrifuging for 5 minutes in a desktop microcentrifuge, resuspending the cell pellet in 100 μl of loading buffer, followed by boiling for 10 minutes. Samples were electrophoresed according to the method in chapter 2.6.1.

Enhancing expression of rare E. coli tRNAs

The expression of recombinant proteins in *E. coli* is decreased when the codon use in the recombinant gene differs from the codon usage in the host cell. This occurs because high-levels of expression of genes that require rare tRNAs causes a depletion in the internal tRNA pool. The result of this is an increase in expression time which may result in RNA being degraded before it is translated. Several host cell lines are commercially available which have enhanced expression of the codons that are rarely used by *E. coli*, these include the Rosetta strains (Novagen) and the CodonPlus stains (Stratagene). Both strains are derived from BL21 and enhance the expression of the rare *E. coli* tRNAs genes that decode the AGG, AGA, AUA, CUA, CCC. Rosetta cells contain an additional tRNA gene that decodes GGA and Rosetta 2 one that decodes CGG. The presence of these codons in the mature hCC gene sequence is highlighted in Figure 3.3.

TCC	AGT	CCC	GGC	AAG	CCG	CCG	CGC	CTG	GTG	GGA
Ser	Ser	Pro	Gly	Lys	Pro	Pro	Arg	Leu	Val	Gly
12										
GGC	CCC	ATG	GAC	GCC	AGC	GTG	GAG	GAG	GAG	GGT
Gly	Pro	Met	Asp	Ala	Ser	Val	Glu	Glu	Glu	Gly
23										
GTG	▇	CGT	GCA	CTG	GAC	TTT	GCC	GTC	GGC	GAG
Val	Arg	Arg	Ala	Leu	Asp	Phe	Ala	Val	Gly	Glu
34										
TAC	AAC	AAA	GCC	AGC	AAC	GAC	ATG	TAC	CAC	AGC
Tyr	Asn	Lys	Ala	Ser	Asn	Asp	Met	Tyr	His	Ser
45										
CGC	GCG	CTG	CAG	GTG	GTG	CGC	GCC	CGC	AAG	CAG
Arg	Ala	Leu	Gln	Val	Val	Arg	Ala	Arg	Lys	Gln
56										
ATC	GTA	GCT	GGG	GTG	AAC	TAC	TTC	TTG	GAC	GTG
Ile	Val	Ala	Gly	Val	Asn	Tyr	Phe	Leu	Asp	Val
67										
GAG	CTG	GGC	CGA	ACC	ACG	TGT	ACC	AAG	ACC	CAG
Glu	Leu	Gly	Arg	Thr	Thr	Cys	Thr	Lys	Thr	Gln
78										
CCC	AAC	TTG	GAC	AAC	TGC	CCC	TTC	CAT	GAC	CAG
Pro	Asn	Leu	Asp	Asn	Cys	Pro	Phe	His	Asp	Gln
89										
CCA	CAT	CTG	AAA	AGG	AAA	GCA	TTC	TGC	TCT	TTC
Pro	His	Leu	Lys	Arg	Lys	Ala	Phe	Cys	Ser	Phe
100										
CAG	ATC	TAC	GCT	GTG	CCT	TGG	CAG	GGC	ACA	ATG
Gln	Ile	Tyr	Ala	Val	Pro	Trp	Gln	Gly	Thr	Met
111										
ACC	TTG	TCG	AAA	TCC	ACC	TGT	CAG	GAC	GCC	TAG
Thr	Leu	Ser	Lys	Ser	Thr	Cys	Gln	Asp	Ala	stop

Figure 3.3. The presence of rare *E. coli* tRNAs in the hCC gene sequence. Rosetta (Yellow), Rosetta (Yellow and red) and Codonplus (underlined) enhance the expression of the codons highlighted in the sequence.

Expression in *trxB/gor* mutants

Under physiological conditions the *E.coli* cytoplasm is maintained in a reduced state that strongly disfavours the formation of disulfide bonds. The cytoplasm contains thioredoxins and glutaredoxins that in their oxidised form catalyse the formation of disulfide bonds in peptides. However, both are maintained in the reduced state by the action of thioredoxin reductase (trxB) and glutathione respectively. Glutathione reductase (gor) is required to reduce oxidised glutathione and complete the catalytic cycle of the glutathione-glutaredoxin system. Mutants defective in the trxB and gor genes have been shown to enhance the efficiency of disulphide bond formation by maintaining thioredoxins and glutaredoxins in the oxidised state.[149] The commercially available Rosetta-gami

host strain (Novagen) have the features of the Rosetta strain but are also defective in the trxB/gor genes.

pH

The periplasm is a favourable environment for the formation of disulphide bonds, providing oxidising conditions and members of the Dsb family which are efficient catalysts of disulphide bond formation. DsbA is the most important oxidase of free sulfhydryl groups in the periplasm so choosing conditions that increase its cellular concentration may aid disulphide bond formation. Expression of DsbA occurs across the range of pH that *E. coli* grows, but the conditions that favour expression most are pH 9.0 and pH 4.1.[149-152]

Reduced temperature

The Cpx signal transduction system has been shown to activate both dsbA and degP, the major periplasmic protease in *E. coli*. DegP is activated at high temperatures and is essential for viability at temperatures above 42°C.[153, 154] Increasing the concentration of dsbA or decreasing the concentration of degP by varying growth temperature should affect the cellular concentrations of hCC by enhancing folding or minimising degradation respectively.

Selecting a highly expressing starter culture

Although optimising growth conditions made a significant improvement on the level of hCC expression there was still a considerable variation in the level of expression between cells grown under the same conditions but from different starter colonies. Rather than choosing a starter colony at random, 10 starter colonies were selected and the most highly expressing one chosen for a large scale growth.

10 x 5 ml phosphate buffered Luria Bertani media (LBK) pH 7.0 with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol were inoculated with a single colony and shaken at 250 rpm overnight at 37°C. 10 x 10 ml selective LBK pH 7.0 were inoculated with 30 μl of the corresponding overnight growth. Overnight cultures were stored at 4°C until they were required to inoculate further starter cultures. Starter cultures were shaken at 37°C until the OD_{600} reached 0.5 when they were

induced with 120 µg/ml IPTG. Following induction cultures were incubated at 37°C for a further hour.

Whole cell extracts were prepared by pelleting the starter cultures, then resuspending in 200 µl SDS-PAGE loading buffer. Samples were boiled for 10 minutes prior to being loaded onto an SDS-PAGE gel. Periplasmic extractions were prepared by pelleting starter cultures by centrifugation at 1663 g for 10 minutes. Growth media was discarded and the cell pellets resuspended in 350 µl of 20% sucrose, 100 mM EDTA, 200 mM TrisHCl pH 9.0. Cells were pelleted in a microcentrifuge at 13400 g for 10 minutes and the supernatant discarded. To release the periplasm, cells were resuspended in 350 µl cold 10 mM TrisHCl pH 9.0. Cells were centrifuged again and the supernatant containing the periplasm recovered. The presence of hCC in the periplasm was analysed by SDS-PAGE according to the method described in chapter 2.6.1.

3.2.5 Purification trials

Papain affinity chromatography

The method was adapted from Abrahamson *et al*, 1986 as a means to efficiently purify human cystatin C from the periplasmic extract.[12] A carboxymethylated papain-sepharose 4B affinity column was equilibrated with buffer A, 50 mM TrisHCl, pH 7.4, containing 0.5 M NaCl, 2 mM EDTA, 1 mM sodium azide. The periplasmic extract was adjusted so that it was the same pH and conductivity as buffer A. The column was washed with buffer A until the OD_{280} of the elute was below 0.1. Under these conditions hCC is folded so binds to inactivated papain through the hCC inhibitory site. Material bound to the carboxymethylated papain was eluted with 0.20 M trisodium phosphate, pH 12.1 containing 0.5 M NaCl, 2 mM EDTA, 1 mM sodium azide until the OD_{280} of the elute was below 0.01. At this very basic pH hCC unfolds and can no longer bind papain via its inhibitory site. The pH of the elute was neutralised with 2 M TrisHCl, pH 8.6.

Q-sepharose anion exchange chromatography

Most bacterial proteins have a pI value below 6, whereas hCC has a predicted pI value of 8.75. This difference in pI value can be utilised in Q-sepharose anion exchange chromatography to provide an effective method for purifying hCC from the majority of bacterial proteins. In the method adapted from that described

previously, hCC will not bind to the Q-sepharose at pH 9.5 whereas most of the other proteins in the periplasm will.[12] Periplasmic extract was dialysed against an approximate 200 fold greater volume of 20 mM ethanolamine pH 9.5. A Q-sepharose ion exchange column was equilibrated with equilibrated with the dialysis buffer at a rate of 2 ml/min. Dialysed periplasmic extract was loaded onto the column and the elute containing human cystatin C collected. Initially a salt gradient between 0 M and 0.5 M NaCl was included followed by a step increase to 1 M NaCl to elute any protein bound to the Q-sepharose. Human cystatin C does not bind to the Q-sepharose column under these conditions so the salt gradient step was later removed from the procedure.

SP-sepharose cation exchange chromatography

SP-sepharose cation exchange chromatography follows the same principles as Q-sepharose anion exchange chromatography to separate proteins according to their charge. Periplasmic extract was dialysed against an approximate 200 fold greater volume of 10 mM sodium phosphate buffer, pH 7.0, 2 mM sodium azide. A SP-sepharose column was equilibrated with the dialysis buffer at a flow rate of 2 ml/min. Dialysed periplasmic extract was loaded onto the SP-sepharose and the elute collected. Bound protein was eluted from the SP-sepharose using two step increases in salt concentration, initially to 200 mM then to 1 M NaCl.

Size exclusion chromatography

Size exclusion chromatography was used as a final purification step. Superdex-75 resin contained in a Amersham biosciences XK26/70 column was run at a flow rate of 3 ml/min. Once equilibrated, a maximum volume of 10 ml was loaded onto the column. Absorbance of the elute was recorded at 280 nm and 2.5 min fractions were collected. The following buffers were tested for use in SEC, 50 mM sodium phosphate buffer pH 7.0, 200 mM NaCl, 2 mM sodium azide, 50 mM sodium phosphate buffer pH 8.0, 100 mM NaCl, 2 mM sodium azide, 50 mM ammonium bicarbonate buffer pH 7.8, 100 mM NaCl, 2 mM sodium azide, 0.2 M trisodium phosphate, pH 12.1 containing 0.5 M NaCl, 2 mM EDTA, 1 mM sodium azide and PBS. PBS was found to be the most suitable buffer.

3.2.6 Nucleic acid contamination

Nuclease treatment - RNase A

A 10 mg/ml stock solution of bovine pancreatic RNase A was made up in 10 mM TrisHCl, pH 7.5, 15 mM NaCl and heated to 100°C for 15 minutes. RNase A was added to the periplasmic extract at a final concentration of 40 μg/ml and incubated at room temperature for 1 hour.

Nuclease treatment - DNase 1

The periplasmic extract was made up to 20 mM magnesium chloride and approximately 10 μg of bovine pancreatic DNase 1 was added per ml of extract. Samples were incubated at room temperature for approximately 1 hour.

Nuclease treatment - Benzonase

The periplasmic extract was made up to 20 mM magnesium chloride and 1.5 U of benzonase (Novagen) was added per ml of extract. Samples were incubated at 37°C for 15 minutes then dialysed overnight at 4°C during which time the benzonase was still active.

Polyethyleneimine precipitation

Polyethyleneimine (PEI) is positively charged at neutral and basic pH and is very efficient at precipitating nucleic acids and other acidic macromolecules. This procedure should be carried out at a pH of ~7.5 to 8, where PEI amine groups are fully cationic and nucleic acids are fully anionic. hCC has a pI of 8.75 whereas most bacterial proteins are predominantly acids with pI values below 6. Addition of PEI at pH 8.0 should therefore precipitate nucleic acid contaminants and many bacterial proteins without precipitating hCC. 3 μl/ml of 5% PEI, pH 8 solution was added to the periplasmic extract and immediately mixed by invertion. Following gentle agitation for 5 minutes the samples were centrifuged for 10 minutes at 48384 g. [155, 156]

3.2.7 Proteolytic activity

Phenylmethylsulfonylfluoride

Phenylmethylsulfonylfluoride (PMSF) irreversibly inhibits serine proteases by sulfonylation of serine residues in the protease active site. It is inactivated in aqueous solution, the rate of which increases with pH and temperature, the half-

life of a 20 µM aqueous solution of PMSF is approximately 35 minutes at pH 8.0. Due to its instability in aqueous solution a 100 mM stock solution is prepared in isopropanol and stored at -20°C. The stock solution must be warmed slightly to bring it back into solution prior to adding 100 µg/ml immediately before bursting the periplasm with osmotic shock.

Starter culture

Overgrown (15-17 hours) starter culture is likely to contain a higher level of proteolytic enzymes than a starter culture that has been grown to $OD_{600} = 0.5$ as the overgrown culture will have more dying cells in it which are releasing the enzymes as they die. 50 mls of starter culture grown to $OD_{600} = 0.5$ was used to induce 550 mls growth medium rather than 3 ml/l of overgrown culture.

Protease inhibitors

Complete Inhibitor Cocktail tablets (Roche) were added to the periplasmic extract as standard. A 25x stock solution of Complete (Roche) was made up by dissolving one tablet in 2 ml water immediately prior to use. 40 µl of stock solution was added per ml of periplasmic extract.

Protease Inhibitor Cocktail set VI (Calbiochem) was used at a 1:100 dilution as an alternative inhibitor cocktail. This inhibitor cocktail contained 200 mM AEBSF, 10 mM bestatin, 3 mM E-64 protease inhibitor, 2 mM leupeptin, 2 mM pepstatin A and 500 mM o-phenanthroline.

3.2.8   Characterising purified hCC

SEC-HPLC

Unless stated otherwise 50 mM sodium phosphate buffer pH 7.0, 2 mM sodium azide was used as the mobile phase. See materials and methods section 2.6.4 for experimental detail.

Mass spectrometry

Mass analysis was performed by Dr. A. Moir using an ABI Voyager-DE STR MALDI Mass Spectrometer operating in positive ion mode. Protein analysis was performed in reflector mode using Sinapinic acid (ABI and Fluka) as the matrix. Matrices were prepared immediately before use at a concentration of 10 mg/ml in

50% acetonitrle/0.05% trifluoroacetic acid (TFA). Typically, protein samples were diluted 3-5x with matrix & 1 μl spotted on to the MALDI plate. The spectrometer was calibrated using commercially available calibrations standards (ABI and Sigma).

^{1}H-NMR

4 mg/ml hCC in 50 mM sodium phosphate buffer pH 7.0, 2 mM sodium azide was prepared for NMR by adding 10% D_2O and transferring into a glass NMR tube. ^{1}H-NMR spectra were recorded at 298K on a Bruker DRX-500 spectrometer controlled using XWinNMR (Bruker). NMR data was processed using Felix (Accelrys).

3.3 RESULTS AND DISCUSSION

3.3.1 Analysis of the pIN-III-ompA2-hCC product

N-terminal protein sequencing of the purified product of the pIN-III-ompA2-hCC plasmid determined the first 5 amino acids to be SSPGK. This is consistent with the first 5 amino acids of mature hCC and thus confirms the correct cleavage of the ompA2 signal sequence (Figure 3.3). Mass analysis confirms the mass of full length mature hCC with no additional amino acids from the linker sequence (Figure 3.4).

Figure 3.4. MALDI-MS analysis of hCC identifies a species with a molecular mass of the predicted mass of hCC (13347 Da)
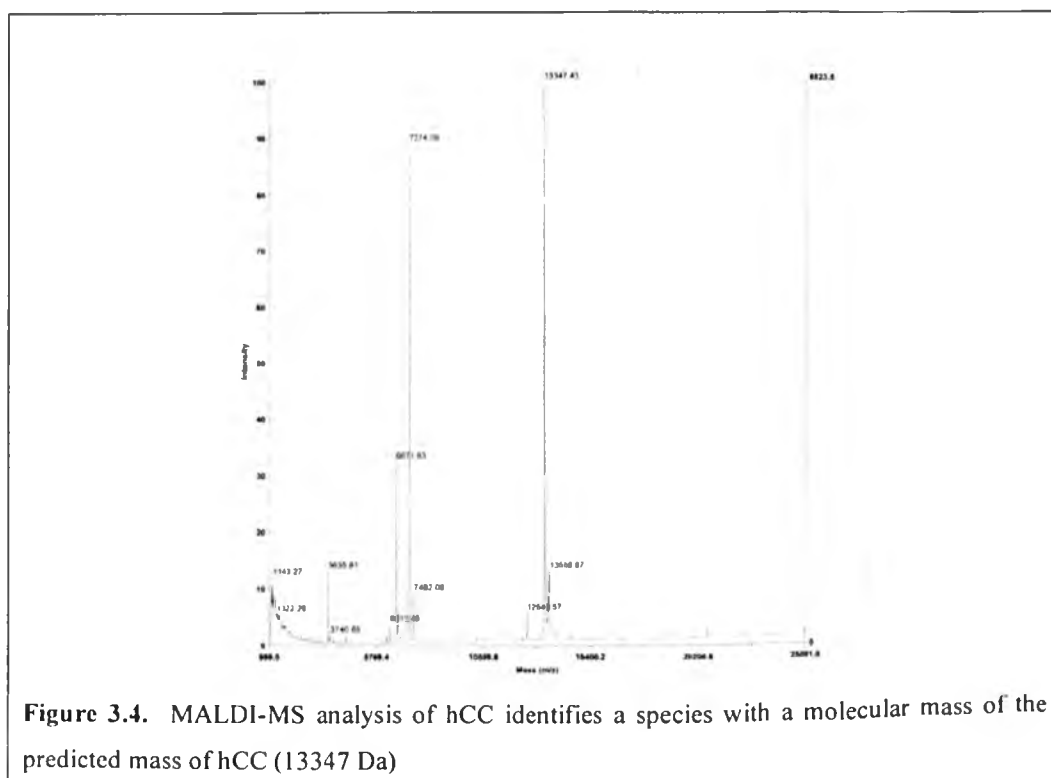
3.3.2 Western blot analysis

hCC is retarded on SDS-PAGE so appears at a higher molecular weight than expected. Therefore western blot analysis was used to confirm that the band appearing at approximately 15 kDa is hCC. In addition to identifying the hCC band on the SDS-PAGE gel, western blot analysis was used to ensure the correct functioning of the pIN-III-ompA2-hCC expression system (Figure 3.5).

Analysis of the whole cell extracts immediately before and 3 hours after induction shows that although there is some expression of hCC prior to induction, the majority of hCC that is produced occurs after induction with IPTG. hCC is not present in the growth media prior to induction, but is present at low levels 3 hours after induction. Targeting of hCC to the periplasm has been successful, although a significant amount of hCC remains in the cell pellet and is not extracted. The sonicated cell fractions show hCC in the soluble fraction and in the cell pellet. This suggests that the hCC found in the cell pellet is in the form of insoluble inclusion bodies. This analysis confirms that hCC is targeted to the periplasm and periplasmic extraction provides an efficient method of extracting the majority of the soluble hCC from the cells.

Figure 3.5. A. Duplicate SDS-PAGE gel of the gel used to transfer onto the nitrocellulose membrane. From left to right, hCC (1), LB pre-induction (2), cell pellet pre-induction (3), LB 3hrs post-induction (4), cell pellet 3hrs post-induction (5), periplasm 3hrs post-induction (6), cell pellet from periplasm extract (7), sonication supernatant (8), sonication cell pellet (9), Sigma low molecular weight markers (10). B. Image produced when the nitrocellulose membrane is treated with ECL solutions and exposed to a light sensitive film for 10 seconds.

3.3.3 Optimising the expression of hCC

In optimising the expression of hCC a high level of expression can be achieved 1 hour after induction with IPTG. Figure 3.6 shows how expression of hCC was increased significantly by enhancing the expression of rare *E. coli* tRNAs. Despite having an increased ability to form disulphide bonds, expression trials using Rosetta-gami cells showed no significant improvement in hCC expression when compared to Rosetta or TG1 host strains. In fact, Figure 3.6C shows that expression of hCC is actually decreased compared to expression in the Rosetta strain. This may be because the Rosetta-gami strain is modified to enhance

disulphide bond formation in the cytoplasm whereas the formation of the disulfide bonds in hCC should occur after hCC has been targeted to the periplasm.

Figure 3.6. SDS-PAGE of whole cell extracts prepared as described in chapter 3.2.3. 20μl of whole cell extract was loaded onto each lane of each SDS-PAGE gel. A. Expression of hCC in TG1 cells grown at 37°C, pH 7.0. Lane 2 uninduced cells, lane 3-6 1 hour, 2 hours, 3 hours and overnight after induction with IPTG. B. Expression of hCC in Rosetta cells grown at 37°C, pH7.0. Lane 2 uninduced cells, lane 3-7 1 hour, 2 hours, 3 hours, 4 hours and overnight after induction with IPTG. C. Expression of hCC in Rosetta-gami cells grown at 37°C, pH 7.0. Lane 1: uninduced cells, lane 2: 1 hour post-induction, lane 4-5: 2 hours and overnight after induction with IPTG. D. Expression of hCC in Rosetta cells grown at 30°C, pH7.0. Lane 2 uninduced cells, lane 3-7 1 hour, 2 hours, 3 hours, 4 hours and overnight after induction with IPTG.

Expression trials indicated that pH reduced the growth rate of cells whilst showing no improvement in hCC expression. Therefore, LBK media pH 7.0 was favoured for all hCC expression experiments.

As Figure 3.6 shows, reducing the growth temperature from 37°C to 30°C reduces expression of hCC. This may be partially accounted for by the decreased growth rate at 30°C, however expression is still reduced several hours after induction.

Selecting a highly expressing starter culture significantly increases the hCC yield. Initially whole cell extracts were used to select a highly expressing colony. However, this method does not select against cultures that express hCC well but do not target it to the periplasm correctly. To overcome this problem it was found that the level of hCC in the periplasm was a better indicator of a highly expressing culture. Figure 3.7 shows a typical example of the variation in expression of hCC between 10 starter colonies.

Figure 3.7. A. Variation in the presence of hCC in the whole cell extract of different starter colonies. Lane 3-14 whole cell extracts, lane 15: hCC. B. Variation in the presence of hCC in periplasmic extracts of different starter colonies. Lanes 2-10 periplasmic extracts.
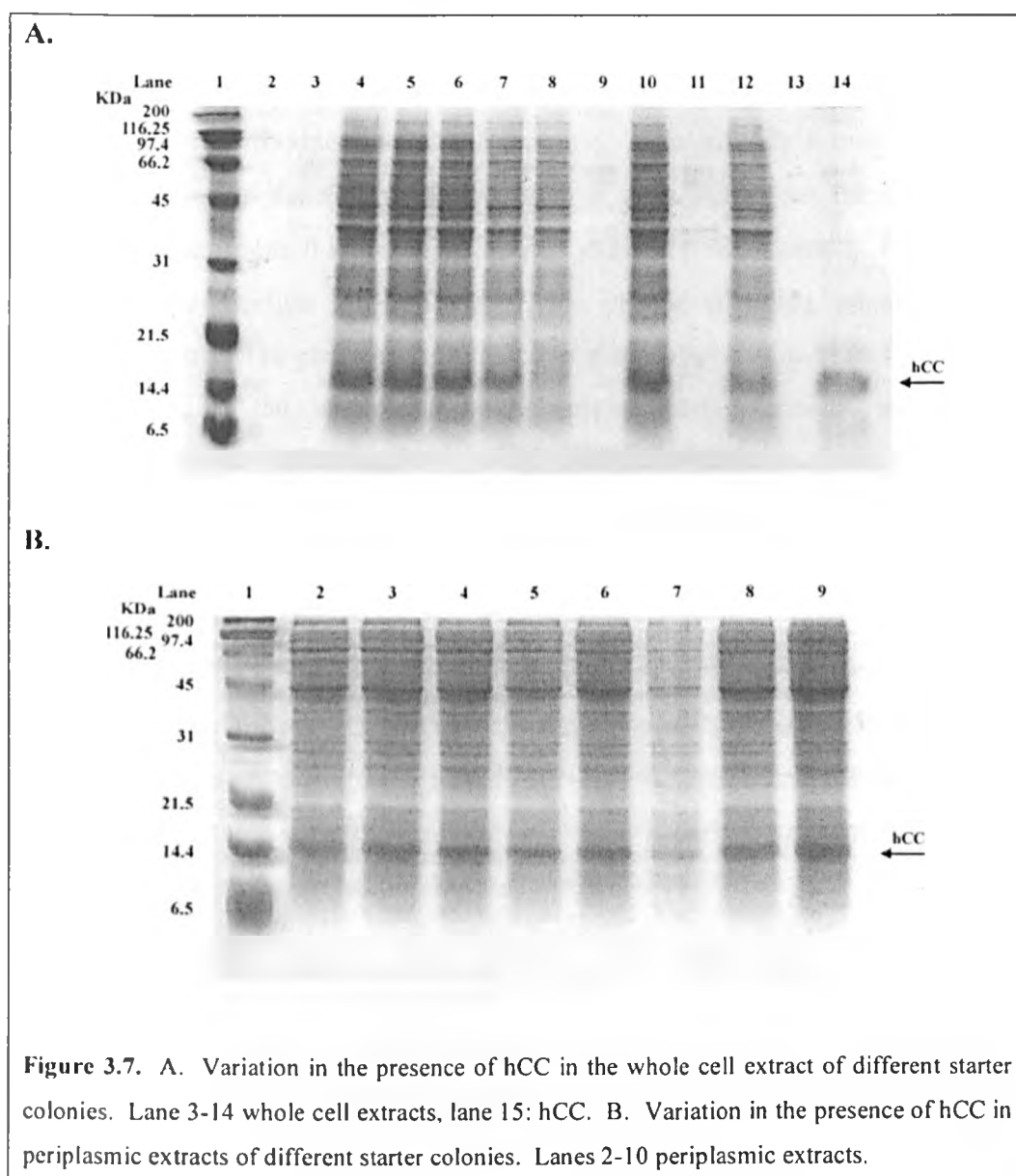
3.3.4 Purification trials

Papain affinity chromatography

Although affinity chromatography has proven to be a very efficient method of isolating chicken cystatin from a periplasmic extract it was much less effective at purifying hCC. High levels of nucleic acids were eluted along with hCC indicating non-specific binding of the nucleic acids to either hCC or the column matrix under the conditions used for the experiment. The high level of nucleic acid contamination prevented the concentration of the sample required to prepare the hCC solution for size exclusion chromatography. A second disadvantage to using this purification method is that hCC purified by this method was extremely susceptible to protein degradation. This may be due to incomplete inactivation of the papain used to produce the affinity column. Alternatively it may be because hCC is unfolded at the very basic pH used to elute hCC from the column and under these conditions it is more susceptible to protein degradation. Figure 3.8A shows the purification of hCC from the periplasm using papain affinity chromatography. The presence of heavy low molecular weight blurs in lanes 2-5 and the lack of the larger molecular weight bands normally seen in the periplasmic extract indicates the level of protein degradation found using this purification method.

Q-sepharose anion exchange chromatography

This method effectively purified hCC from most of the other proteins in the periplasm and removed the protein degradation problem associated with the papain affinity method (Figure 3.8B). However, the nucleic acid contaminant was not removed so there were still complications when concentrating the hCC sample in preparation for size exclusion chromatography.

SP-sepharose cation exchange chromatography

At pH 7.0 hCC will bind to the SP-sepharose whereas the acidic bacterial proteins will not. hCC can be eluted by increasing the ionic strength of the buffer to 200 mM. This method effectively purified hCC from the majority of the other proteins in the periplasm and separates the majority of the nucleic acid contamination from the hCC solution (Figure 3.8C).
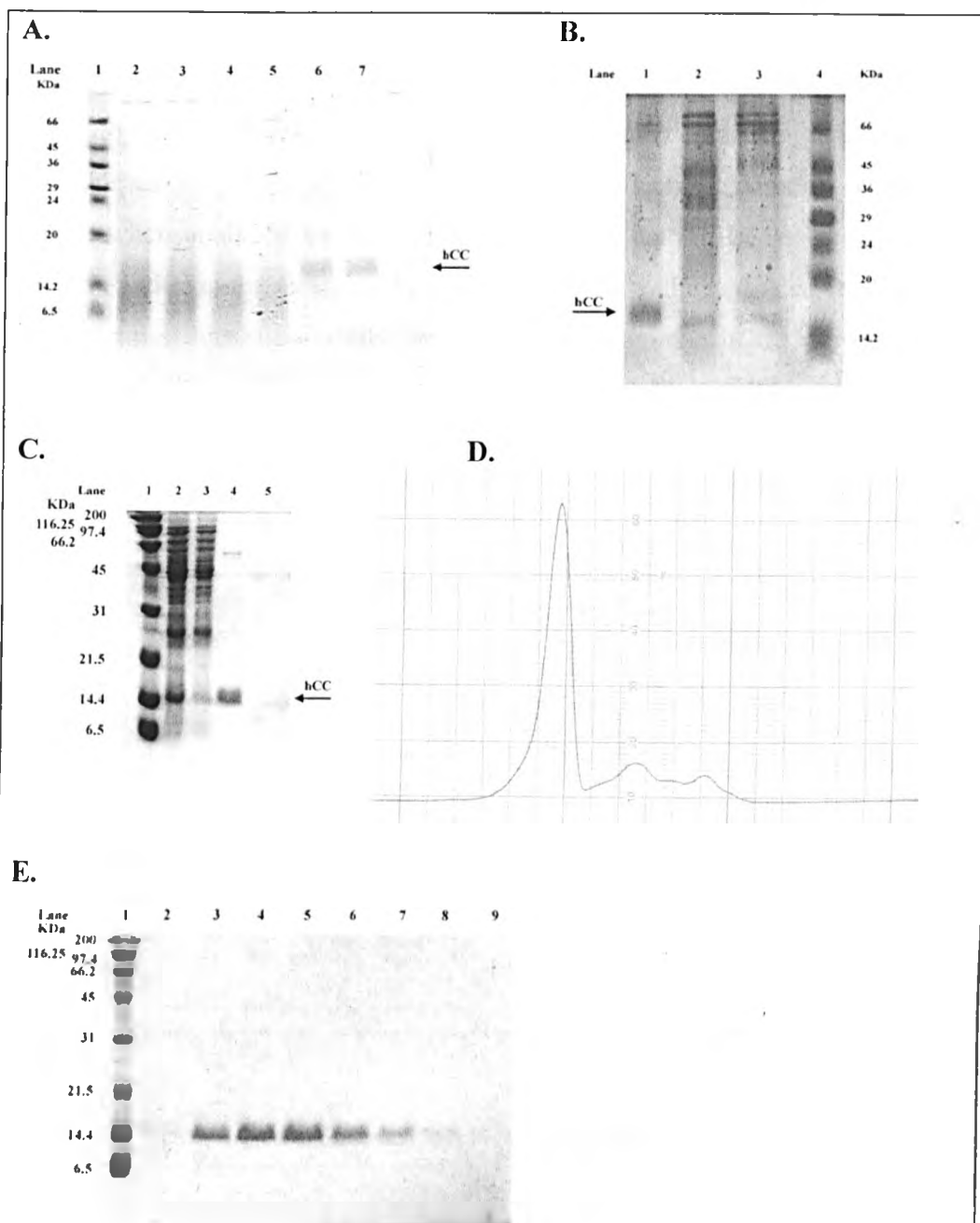
Figure 3.8. Purification trials of hCC. A. Purification using papain affinity. Lane 1: Sigma low molecular weight markers, lane 2-5: Elute in buffer A, lane 6-7: Elution of hCC in buffer B. Note the low molecular weight smearing of lanes 2-5 indicating proteolysis. B. Purification using Q-sepharose. Lane 1: elution of hCC in 20mM ethanolamine pH 9.5, lane 2: elute in 20 mM ethanolamine pH 9.5, 500 mM NaCl, lane 3: elute in 20 mM ethanolamine pH 9.5, 1 M NaCl, lane 4: Sigma low molecular weight markers. C. Purification using SP-sepharose. Lane 1: biorad broad range markers, lane 2: periplasmic extract, lane 3: elute in 50 mM sodium phosphate buffer pH 7.0, lane 4: elute in 50 mM sodium phosphate buffer pH 7.0, 200 mM NaCl, lane 5: elute in 50 mM sodium phosphate buffer pH 7.0, 1 M NaCl. D. Typical elution profile with preparative SEC. E. Typical SDS-PAGE gel of fraction collection on elution of the main peak seen in D.

Size exclusion chromatography

The concentration method used to prepare the protein solution for size exclusion chromatography separated a small amount of low molecular weight nucleic acids from the sample. hCC eluted as a single peak which was free from any other protein when analysed by SDS-PAGE and analytical HPLC as seen in Figure 3.8D-E and Figure 3.9. Absorbance spectra indicated that nucleic acid contamination has successfully been separated from the hCC. Several buffers were tested, the most suitable being PBS, under these conditions there was no evidence of interaction with the superdex matrix and hCC was not prone to degradation or precipitation.
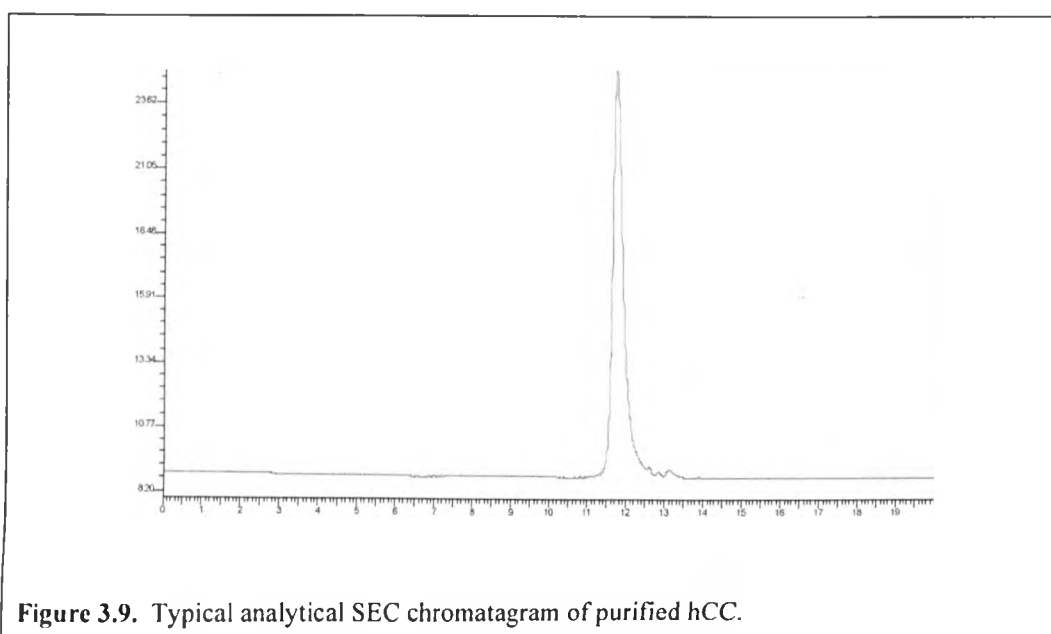
Figure 3.9. Typical analytical SEC chromatagram of purified hCC.

3.3.5 Nucleic acid contamination

Absorbance spectra of the periplasmic extract showed it to be heavily contaminated with nucleic acids. Analysis of the periplasmic extract on a 1% agarose gel stained with ethidium bromide showed the contamination to be nucleic acids approximately 200-250 bases in length. In theory both ion exchange and papain-affinity chromatography should separate hCC from the nucleic acid contaminant under the conditions used. However, in practice, the fraction containing hCC from either purification method was still heavily contaminated with nucleic acids. Size exclusion chromatography provided an effective method of purifying hCC from the contaminants, however, in order to prepare the protein sample it had to be concentrated from a volume of approximately 500 ml (for a 150 ml periplasmic extract) to a final volume of less than 10 mls which caused

heavy precipitation of the protein. The majority of the nucleic acid contamination had to be removed from the protein sample prior to preparing the sample for size exclusion chromatography. Nuclease treatment was used to degrade the contamination and then remove it by dialysis, but didn't prove to be as effective as precipitating the nucleic acids with PEI. Altering the ion exchange method used to purify was chosen as the most suitable method as it reduced the time required to purify hCC.

Nuclease treatment

Deoxyribonuclease I (DNAse 1) acts on double and single stranded DNA, preferentially at phosphodiester linkages adjacent to pyrimidines, yielding on average tetranucleotides. Incubating the periplasmic extract with DNAse 1, followed by dialysis reduced the level of nucleic acid contamination, but was not sufficient to enable preparation of the sample for SEC. DNAse 1 will only degrade DNA and it is possible that the contamination was RNA rather than DNA. Ribonuclease A (RNase A) is an endonuclease that cleaves single-stranded RNA. However, digestion with RNAse A was also insufficient at reducing the concentration of nucleic acid contamination. Incubating the periplasmic extract with both DNAse 1 and RNAse A, followed by dialysis to remove reduced the concentration of contaminants sufficiently that they no longer caused precipitation during concentration. Benzonase endonuclease is a genetically engineered endonuclease from *Serratia marcescens* that degrades both DNA and RNA whether in single stranded, double stranded, linear, circular or supercoiled form. Treatment of periplasmic extract with benzonase removed enough of the nucleic acid contamination to allow purification of hCC. In order to remove enough of the contamination using either the DNAse/RNAse combination or benzonase required many hours of dialysis which significantly extend the time required for the preparation of hCC.

Polyethyleneimine (PEI) precipitation

As shown in Figure 3.10, a final concentration of 0.25% PEI in the periplasmic extract, followed by centrifugation to remove the precipitant, is sufficient to remove a large proportion of the nucleic acid contamination and most of the bacterial proteins in the periplasm.

Figure 3.10. PEI precipitation of bacterial proteins and nucleic acids. Lane 1: Biorad broad range markers, lane 2-5: periplasmic extract and SP-sepharose purification of hCC at 0 M, 200 mM and 1 M NaCl respectively. Lane 6-10: periplasmic extract treated with PEI and SP-sepharose purification of hCC at 0 M, 200 mM and 1 M NaCl respectively.
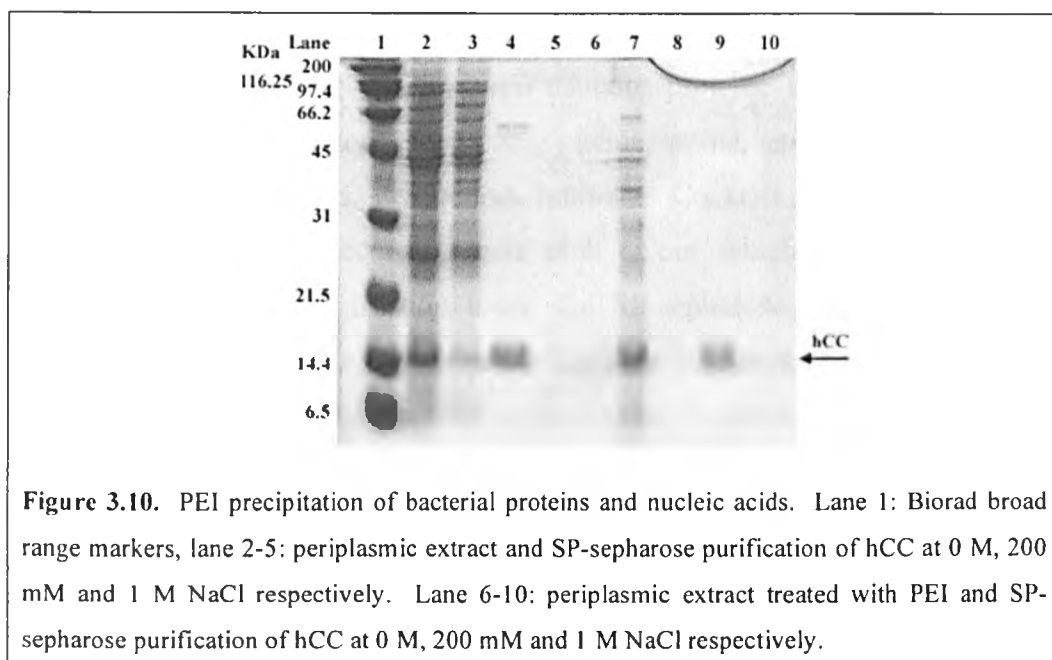
3.3.6 Proteolytic activity

Throughout the purification process hCC is very susceptible to protein degradation despite all buffers containing 2 mM sodium azide and the immediate addition of Complete protease inhibitor cocktail to the periplasmic extract. The time taken to purify hCC from the periplasmic extract was always kept to a minimum in an attempt to reduce the amount of proteolysis. When more than one purification method was available, the quickest method was often chosen in order to reduce proteolysis. Ion exchange and affinity chromatography were used at room temperature as standard, reducing the temperature to 4°C made no difference to the level of proteolysis. However, when concentration of the sample in preparation for SEC was carried out at 4°C the loss of protein between the two purification steps was significantly reduced.

Without using PMSF during the periplasmic extraction SDS-PAGE gels of the periplasmic extract and further purified fractions appear smeared indicating protein degradation during the extraction and purification process. The use of PMSF reduces the amount of smearing of SDS-PAGE gels and increases the yield of hCC 2-5 fold.

Despite the use of Complete protease inhibitor cocktail in the periplasmic extract proteolysis was observed during the purification process. As the composition of the Complete protease inhibitor cocktail is unavailable an alternative protease

inhibitor cocktail of known composition was used as an alternative. Protease Inhibitor Cocktail set VI (Calbiochem) contains protease inhibitors with broad specificity for the inhibition of aspartic, cysteine, serine, and metalloproteases as well as aminopeptidases. Protease Inhibitor Cocktail set VI caused the periplasmic extract to become a pale pink colour which proved difficult to separate from the hCC fraction using the SP-sepharose purification method described above. Protease Inhibitor Cocktail set VI proved not more effective than the Complete protease inhibitor cocktail so the combination of Complete and PMSF was used as the standard method of protease inhibition.

Inoculating flasks with starter culture with an OD_{600} of 0.5 rather than an overgrown starter culture improves the overall yield of hCC. Increased yield is in part due to enhanced expression of hCC and in part due to a reduction in proteolysis during purification.

3.3.7 Characterising purified hCC

The purity of prepared hCC was confirmed using analytical SEC as described in chapter 2.6.4. Both hCC supplied by Dr. M. Abrahamson and synthesised hCC elute at 11.7 minutes under these conditions, a typical chromatogram of pure hCC is shown in Figure 3.9. Mass analysis of hCC produced using the pIN-III-ompA2 expression system confirmed it to be full length protein with a mass equal to the mass of mature hCC. The ^{1}H-NMR spectra of hCC produced using the pIN-III-ompA2 system is that of a folded protein as indicated by the broad dispersion across the amide region and the upfield shifted methyl peak observed below 0.5 ppm. A comparison of the ^{1}H-NMR spectra of synthesised hCC and that of hCC provided by Dr. M. Abrahamson confirm the production of hCC (Figure 3.11).

Figure 3.11. A. ^{1}H-NMR spectra of hCC produced using the pIN-III-ompA2 system. B. ^{1}H-NMR spectra of hCC produced using the pIN-III-ompA2 system (red) overlayed with ^{1}H-NMR spectra of hCC provided by Dr. M. Abrahamson (black).
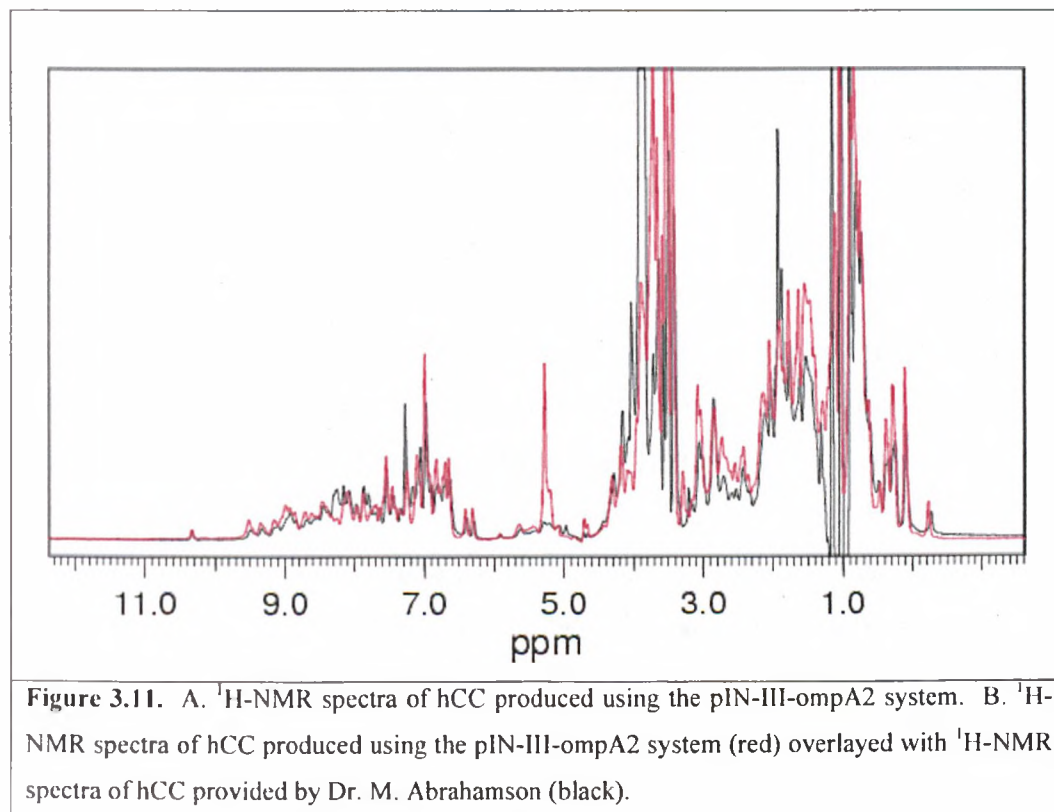
3.4 OPTIMISED PROTOCOL FOR HCC PRODUCTION USING THE PIN-III-OMPA2-HCC EXPRESSION SYSTEM

3.4.1 Unlabelled hCC

Expression

E. coli strain Rosetta cells (Novagen) were transformed with the pIN-III-ompA2-hCC plasmid according to the protocol provided with the cells. Transformants are selected on LB-agar plates containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol at 37°C overnight. 10 x 5 ml selective LBK pH 7.0 were inoculated with a single colony and shaken overnight at 37°C. 10 x 10 ml selective LBK pH 7.0 were inoculated with 30 µl of the corresponding overnight growth. Overnight cultures were stored at 4°C until they were required to inoculate further starter cultures. Starter cultures were shaken at 37°C until the OD_{600} is 0.5 when they were induced with 120 µg/ml IPTG. Cultures were incubated at 37°C for a further hour. Starter cultures were pelleted by centrifugation at 1663 g for 10 minutes. The growth media was discarded and the cell pellets resuspended in 350 µl of 20% sucrose, 100 mM EDTA, 200 mM TrisHCl pH 9.0. Cells were pelleted at 13400 g for 10 minutes and the supernatant discarded. To release the periplasm, cells were resuspended in 350 µl cold 10 mM TrisHCl pH 9.0. Cells were centrifuged again and the supernatant

containing the periplasm recovered. The presence of hCC in the periplasm was analysed by SDS-PAGE according to the method described in chapter 2.6.1. The periplasm with the highest expression of hCC was identified and the corresponding overnight growth used to inoculate 3x 5 ml LBK pH 7.0 with selective antibiotics.

1.2 mls of overnight growth of the highly expressing culture was used to inoculate 400 mls LBK in a baffled 2 l conical flask and shaken at 37°C until the OD_{600} = 0.5. Cell growth was suspended by storing the flask at 4°C overnight. 8x 50 mls of the OD_{600} = 0.5 starter culture was used to inoculate 550 mls selective LBK pH 7.0 and then shaken at 37°C until the OD_{600} = 0.5. Cells were induced with IPTG at a final concentration of 120 µg/ml and incubated at 37°C for a further hour.

Periplasmic extraction

Cultures were centrifuged at 17696 g for 40 minutes and resuspended in 35 ml/l 20% sucrose, 100 mM EDTA, 200 mM TrisHCl pH 9.0. After 10 minutes cells were pelleted by centrifuging at 48384 g for 15 minutes at 20°C. Cells were gently resuspended in 35 ml/l 10 mM TrisHCl pH 9.0 containing 100 µg/ml PMSF at 4°C, then allowed to stand for 10 minutes. Cells were removed from the periplasmic extract by centrifuging at 48384 g for 15 minutes at 4°C. The supernatant was pooled and the appropriate volume of complete inhibitor cocktail stock solution was added.

Chromatography

The periplasmic extract was diluted into 1 l 10 mM sodium phosphate buffer pH 7.0, 2 mM azide then loaded onto an SP-sepharose column equilibrated in 10 mM sodium phosphate buffer pH 7.0, 2 mM azide. Once the periplasmic extract was loaded the column was washed with 10 mM sodium phosphate buffer until the OD_{280} was below 0.05. hCC was eluted using 10 mM sodium phosphate buffer pH 7.0, 200 mM NaCl, 2 mM azide until the OD_{280} was below 0.05. The SP-sepharose column was washed with 10 mM sodium phosphate buffer pH 7.0, 1 M NaCl, 2 mM azide to elute any remaining protein bound to the column.

The fraction eluted from the SP-sepharose column with 200 mM NaCl was concentrated to between 5-10 mls with an Amicon concentrator as described in

chapter 2.6.3. The protein solution passed through a 0.2 μm syringe filter prior to being loaded onto a superdex-75 SEC preparative column (400-500 ml volume) equilibrated in PBS. As a final conformation of purity prepared hCC was analysed on SEC-HPLC to confirm the presence of a single peak at 11.7 minutes.

3.4.2 ^{15}N-labelled human cystatin C

^{15}N-labelled hCC was prepared using the ^{15}N-M9 media as the growth media (see M&M). Growth was significantly slower in M9 media so cells were left for 6 hours after induction rather than 1 hour. All other conditions were the same as for the preparation of unlabelled hCC.

3.4.3 Average yields

The yield of hCC has varied from 0.2-5 mg/l culture while optimising the expression and purification of hCC using the pIN-III-ompA2 system. Using the optimised protocol described in section 3.4 a yield of up to 4 mg/l bacterial culture can be expected with ^{15}N-labelled hCC production and 5 mg/l for unlabelled production.

Transform Rosetta cells with pINIIompA2hCC plasmid.

▼

Select 10 colonies to make starter cultures.

▼

Select a highly expressing starter culture.

▼

Grow 400 mls of starter culture to $OD_{600} = 0.5$.

▼

Induce 550 mls LBK with 50 mls starter culture. Grow to $OD_{600} = 0.5$.

▼

Induce cultures with IPTG and leave for 1 hour.

▼

Pellet the cells and resuspend in 20% sucrose solution at 20°C.

▼

Pellet the cells and resuspend in 10 mM Tris solution at 4°C.

▼

Centrifuge and retain the supernatant containing the periplasm extract.

▼

Load onto SP-sepharose column in 10 mM sodium phosphate buffer pH 7.0.

▼

Elute hCC with 10 mM sodium phosphate buffer pH 7.0, 200 mM NaCl.

▼

Concentrate hCC fraction with amicon concentrator to ≤ 10 ml.

▼

Load onto superdex-75 column equilibrated in PBS.

▼

Pool fractions containing hCC. Confirm purity with analytical HPLC-SEC and calculate yield.

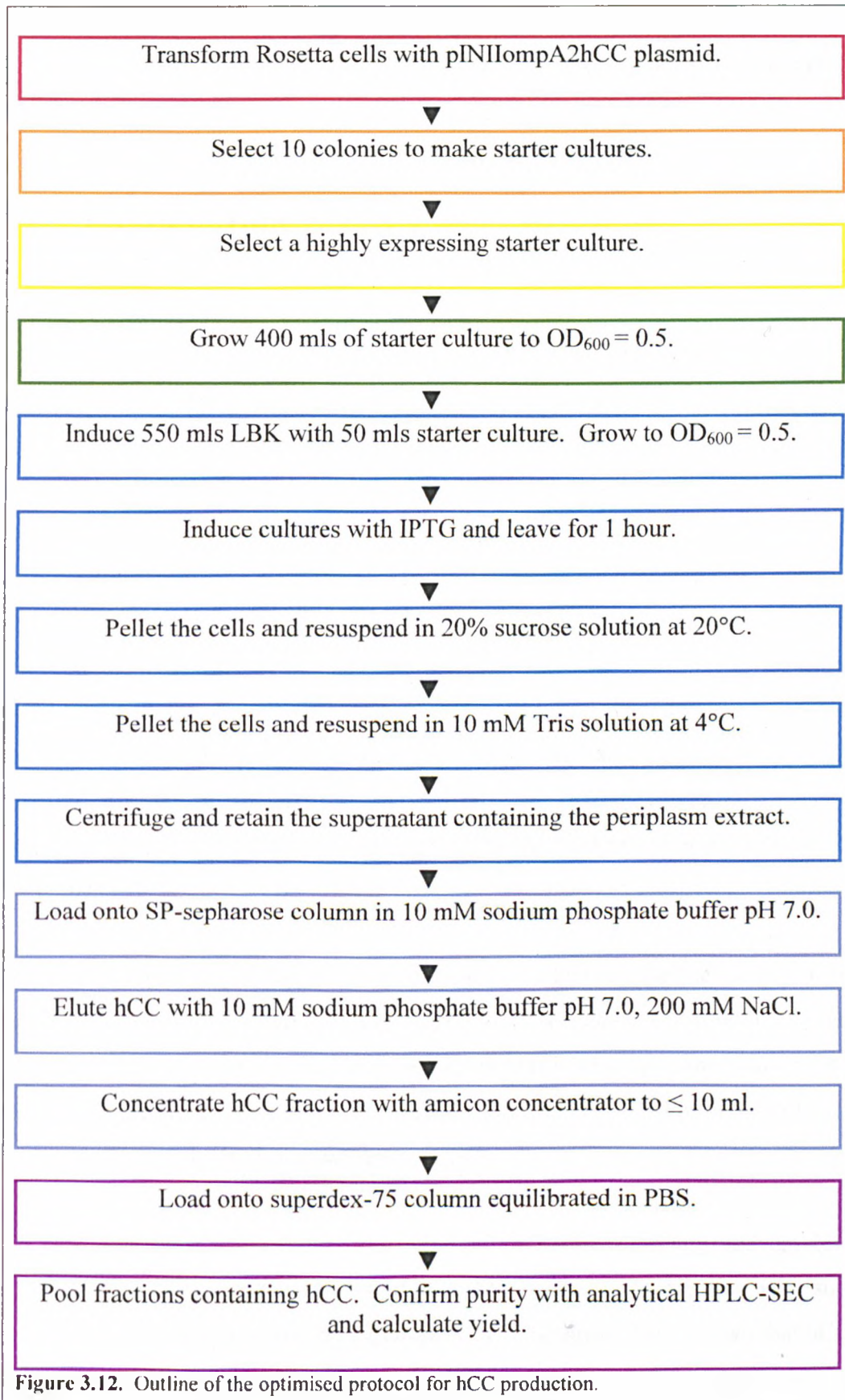**Figure 3.12.** Outline of the optimised protocol for hCC production.

CHAPTER FOUR
CHARACTERISATION OF THE FOLDING PATHWAY OF HUMAN CYSTATIN C

4.1 INTRODUCTION

The folding pathways of a number of cystatins have been studied to date, including stefin A, stefin B and chicken cystatin.[157-159] Each of these cystatins, as well as hCC, are known to form domain swapped dimers and will form amyloid at least *in vitro* if not *in vivo*. An understanding of the folding pathway of a protein can provide some insight into the relationship between folding, dimerisation and further aggregation.

In order for cystatins to form domain swapped dimers a major unfolding event is required. The α-helix must break contacts with the β-sheet and the contacts between the β-strands at the domain-swapping interface must break to allow unfolding and the formation of the dimer interface. The domain swapping reaction is slow relative to the folding and unfolding rates and requires a high level of exposure of hydrophobic side chains in the domain swapping transition state, implying that the reaction is rate limited by the association of predominantly unfolded protein.[134]

hCC has been shown to dimerise under conditions that induce the formation of amyloid. Prevention of domain swapping has also been shown to inhibit amyloid formation.[160] Therefore, the relationship between folding, dimerisation and further aggregation needs to be better characterised, beginning with an understanding of the folding pathway.

To date only limited studies on the relationship between folding and amyloid formation have been carried out for hCC. NMR spectroscopy together with size exclusion chromatograpy and gel electrophoresis techniques have shown that hCC dimerises under mildly denaturing conditions and that hCC is unfolded at temperatures above 85°C, pH lower than 2.6 and at a GdnHCl concentration above 1.6 M.[20] ^{1}H-NMR spectra show that hCC retains the dispersion signals that are characteristic of folded protein in up to 1 M GdnHCl, but are lost between

1 M and 2 M GdnHCl indicating unfolding of the protein. When pH is used to unfold hCC, there is little change in NMR spectra in the pH range 5.7-3.3, below pH 3.3 characteristic changes indicate the unfolding of hCC.[20] Features of the NMR spectra that indicate whether hCC is folded, unfolded or dimerised are discussed further in chapter five.

The preliminary experiments described above provides us with some indication of the conditions under which hCC dimerises and the conditions under which it unfolds, but provide no details of the hCC folding pathway. The experiments described in this chapter allow thermodynamic parameters to be determined from both equilibrium and kinetic data. The information these parameters provide has been used to generate a detailed description of the folding reaction of hCC.

Several optical spectroscopic techniques are widely used to monitor structural transitions such as unfolding or refolding of a protein, as the spectral properties of a protein depend on the molecular environment and mobility of its chromophores. Each technique can reveal different aspects of the protein folding reaction. This chapter describes the use of far-UV CD and tryptophan fluorescence as probes of protein folding to characterise the folding pathway of hCC.

4.2   MATERIALS AND METHODS

Unless otherwise stated all the experiments in this chapter were carried out using 10 mM sodium phosphate, pH 8.0, 2 mM sodium azide as standard buffer.

4.2.1   Circular Dichroism (CD)

Circular dichroism (CD) is used to measure the optical activity of asymmetric molecules in solution. Plane polarised light can be considered to be composed of two components that are circularly polarised, one component is polarised in the clockwise direction (right handed) and the other in the anticlockwise direction (left handed). When a chromophore is part of an asymmetric structure it absorbs left and right handed circularly polarised light differently. This difference in the extinction coefficients of the right and left circularly polarised light gives the CD signal. The polypeptide backbone is optically active in the far ultraviolet region (170-250 nm). Tryptophan and tyrosine can give CD signals in the near UV range (270-300 nm), whereas disulphide bonds give minor bands around 250 nm.

Different secondary structures produce characteristic CD spectra, α-helices show a strong characteristic CD spectrum in the far-UV region whereas signals from β-sheets are weaker and more easily obscured.

GdnHCl equilibrium unfolding of hCC

CD measurements were recorded on a J-810 spectropolarimeter (Jasco) maintained at 20°C, using a 0.1 cm pathlength cuvette. Spectra were recorded between 190-300 nm using 20 µM hCC in standard buffer containing between 0 M and 6 M GdnHCl.

pH equilibrium unfolding of hCC

CD measurements were recorded on a J-810 spectropolarimeter (Jasco) maintained at 20°C using a 0.1 cm pathlength cuvette. Specra were recorded between 190-300 nm with 10 µM hCC in 2 mM citrate buffer between pH 1 and pH 7, 2 mM sodium azide.

4.2.2 Tryptophan fluorescence

Fluorescence occurs when a photon is absorbed by a compound to give an excited state that decays by re-emmision of a photon. Fluorescence emission in proteins originates from the aromatic amino acids, the major fluorophore being tryptophan. Although tyrosine and phenylalanine fluoresce, their absorbance at the wavelength of excitation and their quantum yield of emission are considerably lower than that of tryptophan. Tryptophan is approximately 5 times more sensitive than tyrosine, whereas tyrosine is approximately 25 times more sensitive than phenylalanine. The maximum absorbance of light by tryptophan is at 280 nm and emits it with a maximum between 330-340 nm. In proteins that contain all three aromatic amino acids (such as hCC), fluorescence is dominated by the contribution of the tryptophan residue. In proteins that contain tryptophan shifts in wavelength and changes in intensity are generally observed upon unfolding. Tryptophan emission of a native protein can be greater or smaller than the emission from a free tryptophan in aqueous solution (as found in the unfolded state). Tryptophan fluoresces relatively weakly in aqueous solution, but its fluorescence may be enhanced in non-polar regions of the protein. Therefore, as a protein unfolds, fluorescence intensity tends to decrease as the tryptophan moves from a buried non-polar region to aqueous solution. However, if energy is

transferred to another group that is close to the fluorophore in the folded state then the fluorescence is said to be quenched. Fluorescence intensity will increase as the protein unfolds and is no longer close to the quenching group. In a hydrophobic environment, such as the interior of a protein tryptophan emission occurs at shorter wavelengths. As the protein unfolds, the emission maximum is usually shifted from shorter wavelengths to 350 nm, the fluorescence maximum of tryptophan in aqueous solution.

GdnHCl equilibrium unfolding of hCC

Tryptophan fluorescence measurements were recorded using either a Cary Eclipse (Varian, U.K.) or RF-5301PC (Shimadzu) fluorescence spectrophotometer. The excitation wavelength was 280 nm, emission spectra were recorded between 300 and 420 nm using 4 µM hCC in standard buffer containing between 0 M and 6M GdnHCl. Spectra were recorded approximately 10 minutes and 24 hours after exposure to denaturant.

pH equilibrium unfolding of hCC

Tryptophan fluorescence measurements were recorded using either a Cary Eclipse (Varian, U.K.) or RF-5301PC (Shimadzu) fluorescence spectrophotometer. The excitation wavelength was 280 nm, emission spectra were recorded between 300 and 420 nm. hCC samples used in the pH CD experiment described above were diluted with identical buffer to a final hCC concentration of 2.5 µM. Spectra were recorded approximately 10 minutes after exposure to denaturant.

4.2.3 Stopped-flow

The protein folding and unfolding reaction can be followed using a stopped-flow instrument. The stopped-flow instrument is essentially a rapid mixing device coupled to a fluorescence spectrometer. Whilst the dead time of manually mixing two reagents and observing the change with a spectrometer is several seconds, the rapid mixing of the stopped-flow machine reduces the dead time to approximately 5 millisecond when mixing buffers of different viscosity (as is the case with denaturing buffer containing GdnHCl and the non-denaturing buffer). By rapidly mixing denatured protein with a non-denaturing buffer the change in fluorescence can be recorded as the protein folds from the unfolded to the folded state. Likewise, by mixing folded protein with a denaturant the change in fluorescence

can be recorded as the protein unfolds from the folded to the unfolded state. The observed change in fluorescence can be fitted to determine the rates of the folding/unfolding reaction. When plotted as a function of denaturant activity and fitted to an appropriate folding mechanism, the folding rate in the absence of denaturant can be obtained by extrapolation.

Single-jump stopped-flow

All stopped-flow reactions were carried out using an Applied Photophysics SX-18MV stopped flow reaction analyser maintained at 25°C. The excitation wavelength was 280 nm and slit width 5 nm. In the single jump experiment refolding data were acquired by a 1:11 dilution of 22 µM hCC in standard buffer containing 3.3 M GdnHCl with standard buffer containing varying concentrations of GdnHCl to produce a final GdnHCl concentration between 0.3 and 1.8 M. Unfolding data were acquired by a 1:11 dilution of 22 µM hCC stock in standard buffer into standard buffer containing GdnHCl so that the final concentration of GdnHCl was between 1.3 M and 6 M.

Double-jump stopped-flow

Peptidyl-proline isomerisation can be demonstrated by double-jump stopped flow experiments. In the single jump experiments described previously protein is left under unfolding conditions for several hours. During this time the structural constraints found in the native state are not present so the conformation of the prolyl bond seen in the native state is free to interconvert between the *cis* and *trans* conformations. When the protein is refolded an equilibrium mixture of both proline isomers are present, the native isomer will rapidly refold whereas refolding of the non native isomer is decelerated to allow for the re-isomerisation to occur. In a double jump stopped-flow experiment folded protein is transferred into denaturing conditions for enough time for the protein to unfold, but restored to renaturing conditions before the prolines have had time to isomerise. As the prolines remain in their native conformation the slow transition caused by re-isomerisation is no longer observed.

The stopped-flow machine used for the single jump experiment was reconfigured for the double jump experiment. In the first push 24 µM hCC in standard buffer was mixed with an equal volume of 7 M GdnHCl in standard buffer. After 5

seconds, the second push resulted in a 1:6 dilution of the unfolded hCC with 0.02 M GdnHCl in standard buffer to produce a final GdnHCl concentration of 0.6 M. The fluorescence change was recorded over 2 and 100 seconds. All other parameters were kept the same as for the single jump experiment.

Sodium sulphate

Sodium sulphate (Na_2SO_4) is a commonly used protein stabilising agent that can be used to gain thermodynamic parameters for protein folding intermediates that are not populated in water alone. Na_2SO_4 reduces the free energy of solvation of protein hydrocarbon but does not alter the molar ability of GdnHCl to increase solvation of hydrocarbon. Therefore the effects of GdnHCl and Na_2SO_4 on solvation of hydrocarbon can be considered additive and independent. The total denaturant activity is discussed in detail later in the chapter, but the way denaturant activity relates to GdnHCl concentration and Na_2SO_4 concentration enables kinetic analysis at negative denaturant activity.[161]

All experimental detail was kept the same as for the single jump stopped flow experiment other than the presence of 0.4 M sodium sulphate in all protein stocks and buffers.

Concentration dependence

All experimental detail was kept the same as for the single jump stopped flow experiment other than the final protein concentration which was either 10 µM or 0.4 µM hCC.

4.3   DATA FITTING

4.3.1   Denaturant activity

Guanidine hydrochloride activity

The effect of increasing concentrations of denaturant on the stability of the unfolded state can be explained by the increase in solubility of the parts of the protein exposed upon denaturation. It has been shown that there is a non-linear relationship between GdnHCl concentration and the free energy of solvation of non-polar amino acids and the polypeptide backbone. The non-linearity of the free energy of solvation with GdnHCl concentration is described by the denaturant constant, $C_{0.5}$. For GdnHCl a denaturant constant of 7.5 M has been

determined and has been used to calculate the linear denaturant activity according to the equation:[162]

(Eq. 4.1.1)   Denaturant activity = $C_{0.5}$[GdnHCl] / ($C_{0.5}$+[GdnHCl])

The use of denaturant activity rather than denaturant concentration allows a more reliable extrapolation from high denaturant to conditions where no denaturant is present.[162]

Guanidine hydrochloride and sodium sulphate activity

For a given concentration of sodium sulphate denaturant activity can be calculated according to the equation:

(Eq. 4.2.1)   Denaturant activity = ($C_{0.5}$[GdnHCl] / ($C_{0.5}$+[GdnHCl])) +

$C_{0.5}\Delta G_{s,0}/(\Delta G_{s,max}-\Delta G_{s,0})$

$\Delta G_{s,0}$ represents the free energy of solvation of *N*-acetyltryptophanamide (NATA) or *N*-acetyltyrosineamide (NAYA) relative to water at a specified Na_2SO_4 concentration and in the absence of GdnHCl. $\Delta G_{s,max}$ represents the maximum difference in the free energy of solvation of NATA or NAYA at an infinite GdnHCl concentration and a specified Na_2SO_4 concentration, measured relative to water. Values for sodium sulphate activity ($C_{0.5}\Delta G_{s,0}/(\Delta G_{s,max}-\Delta G_{s,0})$) calculated for NATA and NAYA between 0-0.4 M sodium sulphate are the same within error. Sodium sulphate activity values in the presence of 0. 0.1, 0.2, 0.3 and 0.4 M sodium sulphate have been calculated as 0, -0.27, -0.51, -0.78 and -1.03 M, respectively.[161] The sodium sulphate concentration used experimentally varies between 0.4 and 0.35 M. As there is a linear relationship between Na_2SO_4 activity Na_2SO_4 concentration, values for sodium sulphate activity were calculated for each experimental sodium sulphate concentration by extrapolating the activity between 0.3 and 0.4 M.

4.3.2 Fluorescence data

Equilibrium fluorescence data

Equilibrium GdnHCl unfolding fluorescence data reports a single structural transition, therefore the simplest model that can be assumed is a two-state folding mechanism according to the equation:

(Eq. 4.3.1) $\qquad F \rightleftharpoons U$

The observed signal intensity, I, can be described according to the equation:

(Eq. 4.3.2) $\qquad I = \alpha_F I_F + \alpha_U I_U$

Where α_F and α_U are the fractional proportions of molecules in the folded and unfolded states respectively and I, I_F and I_U are the fluorescence intensities of the measured, folded and unfolded states respectively. In a two state folding mechanism it is assumed that only the folded and unfolded states are observed, therefore:

(Eq. 4.3.3) $\qquad \alpha_U = 1 - \alpha_F$

Therefore,

(Eq. 4.3.4) $\qquad I = I_U + \alpha_F(I_F - I_U)$

As α_F is also determined according to the equation:

(Eq. 4.3.5) $\qquad \alpha_F = [F] / ([F] + [U])$

Where [F] and [U] represent the concentration of the folded and unfolded state respectively. Equation Eq. 4.3.5 can be rearranged into:

(Eq. 4.3.6) $\qquad \alpha_F = K_{(F/U)} / (1 + K_{(F/U)})$

Where $K_{(F/U)}$ is the equilibrium constant at a given solvent condition. $K_{(F/U)}$ at a particular denaturant concentration is calculated according to the equation:

(Eq. 4.3.7) $\qquad K_{(F/U)} = \exp(m_{(F/U)}(D-[D]_{1/2}))$

Where $K_{(F/U)}$ is the equilibrium constant, $m_{(F/U)}$ is the m-value for the unfolded species, D is the denaturant activity and $[D]_{1/2}$ is the activity of denaturant at the midpoint of unfolding. The sloping baseline for U have been baseline corrected.

Kinetic fluorescence data

Transients of fluorescence intensity versus time were fitted to a single exponential according to the equation:

(Eq. 4.4.1) $\qquad I = I_0.\exp(-k_{obs}.t) + I_\infty$

Where I_0 is the fluorescence amplitude of the reaction, k_{obs} is the observed rate constant for the reaction and I_∞ is the endpoint fluorescence intensity. The amplitude represents the difference in signal between the initial and final equilibrium position of the system. The endpoint fluorescence is used as the offset of the fluorescence intensity.

Transient for the refolding limb were recorded over a split time course to determine the fast and slow folding rates. The fast folding rate was determined by fitting the first 2 seconds of the transient whereas the slow folding rate was determined by fitting the transient recorded over second, longer time course (100 s).

Chevron plots

The plot of $\ln k_{obs}$ versus denaturant activity for both the folding and unfolding transients produces the rate profile. In a simple two-state process the observed rate (k_{obs}) is the sum of the folding ($k_{(U-t)}$) and unfolding rates ($k_{(F-t)}$). The rate profile is V-shaped, hence it is often known as a chevron plot. The low denaturant side of the V-shape is referred to as the folding limb while the high denaturant side is referred to as the unfolding limb. In the case of hCC the chevron plot shows a distinct curvature of the folding limb at low denaturant concentrations. One explanation of this curvature is the presence of an

intermediate in the folding pathway. The formation and decay of the native state can be defined by the equation:

(Eq. 4.5.1) $k_{obs} = k_{(F-t)} + k_{(I-t)}$

Where k_{obs} is the observed rate, $k_{(I-t)}$ and $k_{(F-t)}$ are the rate constants describing the forward and reverse transitions, respectively, between the folded and intermediate states. However, because the protein must fold through a rapidly formed intermediate, I, the folding reaction is described as:

(Eq. 4.5.2) $$U \; \underset{}{\overset{K_{(I/U)}}{\rightleftharpoons}} \; I \; \underset{k_{(I-t)}}{\overset{k_{(F-t)}}{\rightleftharpoons}} \; F$$
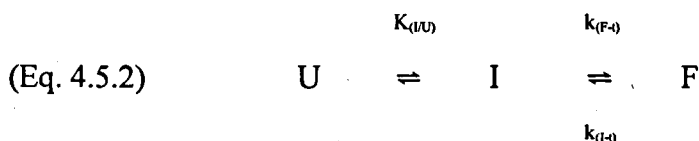
Where U, I and F represent the unfolded, intermediate and folded states, respectively, and $K_{(I/U)}$ is the equilibrium constant for the conversion between the intermediate and unfolded states. The rate of folding is determined by the rate limiting step, $k_{(I-t)}$, and by the population of the I state. Therefore:

(Eq. 4.5.3) $k_{obs} = k_{(F-t)} + k_{(I-t)} \cdot ([I] / ([U] + [I]))$

As the equilibrium ratio [I] / [U] is represented by the constant $K_{(I/U)}$ this equation can be rewritten as:

(Eq. 4.5.4) $k_{obs} = k_{(F-t)} + k_{(I-t)} / (1 + 1/K_{(I/U)})$

The denaturant dependence of $k_{(F-t)}$, $k_{(I-t)}$, and $K_{(I/U)}$ are calculated using the following equations:

(Eq. 4.5.5) $k_{(F-t)} = k_{(F-t)(w)} \cdot \exp(-m_{(t)}.D)$

(Eq. 4.5.6) $k_{(I-t)} = k_{(I-t)(w)} \cdot \exp((m_{(I)} - m_{(t)}).D)$

(Eq. 4.5.7) $K_{(I/U)} = K_{(I/U)(w)} \cdot \exp((m_{(U)} - m_{(I)}).D)$

The subscript (w) indicated the rate or equilibrium constant in water. $m_{(t)}$ represents the m values of the transition state, $m_{(I)}$ and $m_{(U)}$ represent the m-values for the intermediate and unfolded state respectively.

4.3.3 CD data

Residue ellipticity

CD data obtained from the J-810 spectropolarimeter was outputted in as millidegrees (mdeg). All CD data has been converted to residue ellipticity using the following equation:

(Eq. 4.6.1)   Residue ellipticity = $\dfrac{\theta \times 100}{[c] \times d \times N_a}$

Where θ is elipticity in degrees, [c] is molar protein concentration, d is the path length in cm, N_a is the number of amino acids per protein. The factor 100 originates from the conversion of molar concentration to $dmol/cm^3$ concentration units.

Equilibrium CD data

Equilibrium CD data reports a single structural transition, therefore, a two state folding mechanism is assumed and data have been fitted using equation Eq. 4.3.4 with equation Eq. 4.3.6 and Eq. 4.3.7 as temporary variables. The sloping baseline for U have been baseline corrected.

4.4   RESULTS

4.4.1   Equilibrium data

GdnHCl unfolding - fluorescence

GdnHCl increases the stability of the unfolded state by increasing the solubility of the parts of the protein that become exposed upon unfolding. Figure 4.2A shows the fluorescence spectra of the folded (0 M GdnHCl) and unfolded state (6 M GdnHCl) of hCC. The fluorescence of the unfolded state is red-shifted relative to the folded state and has a greater quantum yield. This indicates that the fluorescence in the folded state is quenched more than in the unfolded state. hCC contains single tryptophan residue at position W106 which is predominantly responsible for the fluorescence signal seen in Figure 4.2A. The increased fluorescence associated with the unfolded state suggests that this residue is in close proximity to quenching groups in the folded state. Tryptophan fluorescence can be quenched by arginine, aspartic acid, cysteine, glutamic acid, histidine, lysine and methionine. Exactly which of these residues is the quenching group

cannot be determined from this data. However, the close proximity (less than 5 Å) of Q107 to W106 makes it a likely candidate for the quenching group. The position of W106 and the amino acid side chains within 5 Å are shown in Figure 4.1.

Figure 4.1 The position of W106 in the loop between strands 4 and 5 of the hCC structure (coloured red). V104, P105 and Q107 (coloured green) are the only amino acid side chains within 5 Å of W106.
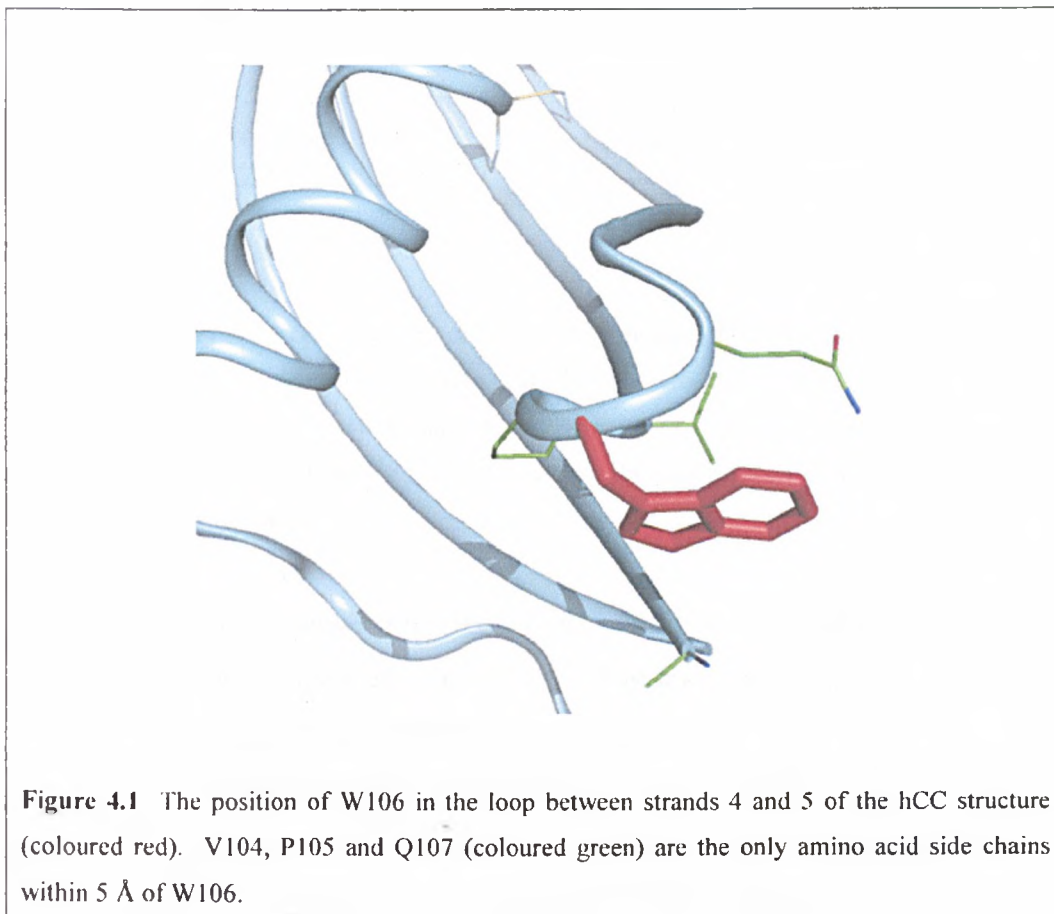
The difference in tryptophan fluorescence between the folded and unfolded state can be used as a probe to follow the equilibrium unfolding of hCC in increasing concentrations of GdnHCl. Figure 4.2B shows the change in fluorescence intensity at 375 nm with increasing denaturant activity both 10 minutes and 24 hours after addition of denaturant. The denaturation curve after 10 minutes and 24 hours overlay with each other indicating that equilibrium had been reached after 10 minutes.

Figure 4.2 **A.** Fluorescence spectra of folded hCC in standard buffer (black) and unfolded hCC in 6 M GdnHCl, standard buffer (red). **B.** Fluorescence intensity at 375 nm plotted as a function of GdnHCl activity, 10 minutes (black dots) and 24 hours (red dots) after exposure of hCC to denaturant. The black line represent the fit of the data points recorded after 10 minutes of exposure to denaturant.

The single transition between the folded and unfolded states indicates that a two-state equilibrium is observed. In order to determine the $[D]_{1/2}$, the denaturant activity at the midpoint of unfolding, and the m-value for the transition ($m_{(F/U)}$) these data have been fitted to equation Eq. 4.3.4 with equation Eq. 4.3.6 and Eq. 4.3.7 as temporary variables. The m-value for the unfolding transition is a measure of the solvent accessibility during unfolding. $[D]_{1/2}$ and the m-value are used to calculate $\ln K_{(F/U)(W)}$, according to the equation:

(Eq. 4.7.1) $\qquad [D]_{1/2} = -\ln K_{(F/U)(W)} / m_{(F/U)}$

$\ln K_{(F/U)(W)}$ is used to calculate $\Delta G_{(W)}$ according to the equation:

(Eq. 4.7.2) $\qquad \Delta G_{(W)} = -RT\ln K_{(F/U)(W)}$

The thermodynamic parameters attained from the unfolding curve of hCC using tryptophan fluorescence as a probe are summarised in the table 4.1.

Time from addition of denaturant	$\ln K_{(F/U)(W)}$	m-value (M^{-1})	$\Delta G_{(W)}$ $(Kcalmol^{-1})$	$[D]_{1/2}$ (M)
10 minutes	9.0 ± 1.8	-7.1 ± 1.2	-5.4 ± 1.1	1.3 ± 0.1
24 hours	6.8 ± 1.1	-5.7 ± 0.8	-4.1 ± 0.6	1.2 ± 0.1

Table 4.1. Summary of parameters determined from the equilibrium fluorescence unfolding curve of hCC induced by GdnHCl.
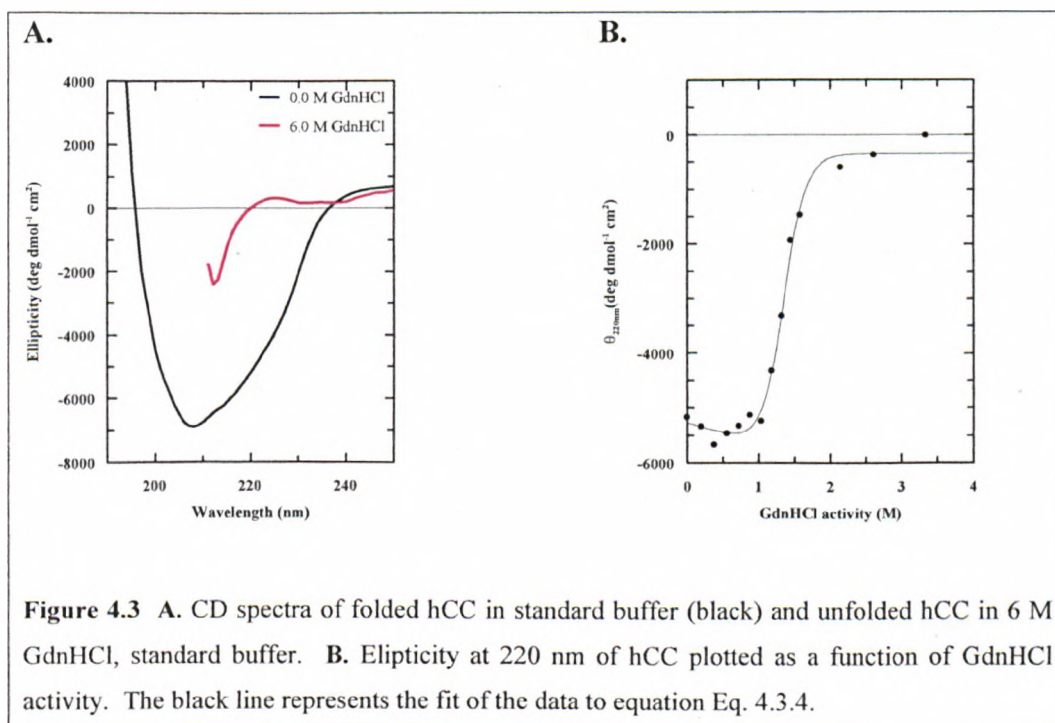
The analysis of the protein folding reactions decribed above enables the calculation of several useful parameters for the folding reaction, one of which is m-values. m-values are determined using the assumtion made in the linear free energy principle that the dependence of the free energy of the state/s involved is linear with respect to the denaturant activity. The gradient of the plot of free energy versus denaturant activity defines the m-value which has the units M^{-1}. m-values have been shown to be proportional to the change in solvent exposure on unfolding.[163] As the level of solvent exposure refects how folded a protein is m-values are used as a measure of how "folded" a state in the folding pathway is.

The subscript associated with equilibrium constants, denoted in the standard form K, indicates the product state as the numerator and the reactant state as the denominator (e.g. $K_{F/U}$). The subscript following rate constants, denoted in the standard form k, defines the reactant and the transition state (e.g. k_{F-t}). An additional subscript, (W), denotes an equilibrium or rate constant in water (e.g. $k_{F-t(W)}$). m-values are denoted by the symbol m with a subscript indicating the relevant state (e.g. m_F).

GdnHCl unfolding - CD

Figure 4.3A shows the far-UV CD spectra of the folded (0 M GdnHCl) and unfolded state (6 M GdnHCl). Far-UV CD reports the conformation of the amide backbone so is a useful probe in following the loss of secondary structure as a protein unfolds. Although hCC contains a single α-helix and a 5 stranded β-sheet the far-uv CD spectrum is dominated by CD signal from the α-helix.

As far-uv CD reports on secondary structure it can be used as a probe to follow equilibrium unfolding of hCC. Figure 4.3B shows the CD signal at 220 nm plotted against denaturant activity. As is observed with the fluorescence unfolding curve, the CD unfolding curve reports a two-state folding mechanism.

Figure 4.3 **A.** CD spectra of folded hCC in standard buffer (black) and unfolded hCC in 6 M GdnHCl, standard buffer. **B.** Elipticity at 220 nm of hCC plotted as a function of GdnHCl activity. The black line represents the fit of the data to equation Eq. 4.3.4.

The $[D]_{1/2}$ and the m-value for the transition ($m_{(F/U)}$) have been determined by fitting to equation Eq. 4.3.4 with equation Eq. 4.3.6 and Eq. 4.3.7 as temporary variables. The $[D]_{1/2}$ is used to calculate $lnK_{(W)}$ according to equation Eq. 4.7.1. $lnK_{(F/U)(W)}$ is used to calculate $\Delta G_{(W)}$ according to the equation Eq. 4.7.1. The thermodynamic parameters attained from the unfolding curve of hCC using far-uv CD as a probe are summarised in the table 4.2.

Time from addition of denaturant	$lnK_{(F/U)(W)}$	m-value (M^{-1})	$\Delta G_{(W)}$ ($Kcalmol^{-1}$)	$[D]_{1/2}$ (M)
10 minutes	8.7 ± 1.4	-6.5 ± 0.9	-5.2 ± 0.8	1.3 ± 0.1

Table 4.2. Summary of parameters determined from the equilibrium CD unfolding curve of hCC induced by GdnHCl.

The presence of a single transition in the unfolding curve of hCC in the presence of GdnHCl indicates that a simple, two-state equilibrium is observed. The far-UV CD unfolding curve is a measure of change in secondary structure where the fluorescence unfolding curve is a measure of change in tertiary structure. As shown in Figure 4.4, the unfolding curves measured by CD and fluorescence change simultaneously which is indicative of a two-state transition between the folded and unfolded states. The similarity of the unfolding transitions followed by CD and fluorescence is reflected in the parameters summarised in tables 4.1 and 4.2, which are within error of each other. Although a two-state transition has been observed at equilibrium it does not mean that no intermediate species are formed in the protein folding reaction. Rather it shows that these states are not the most populated species at equilibrium.
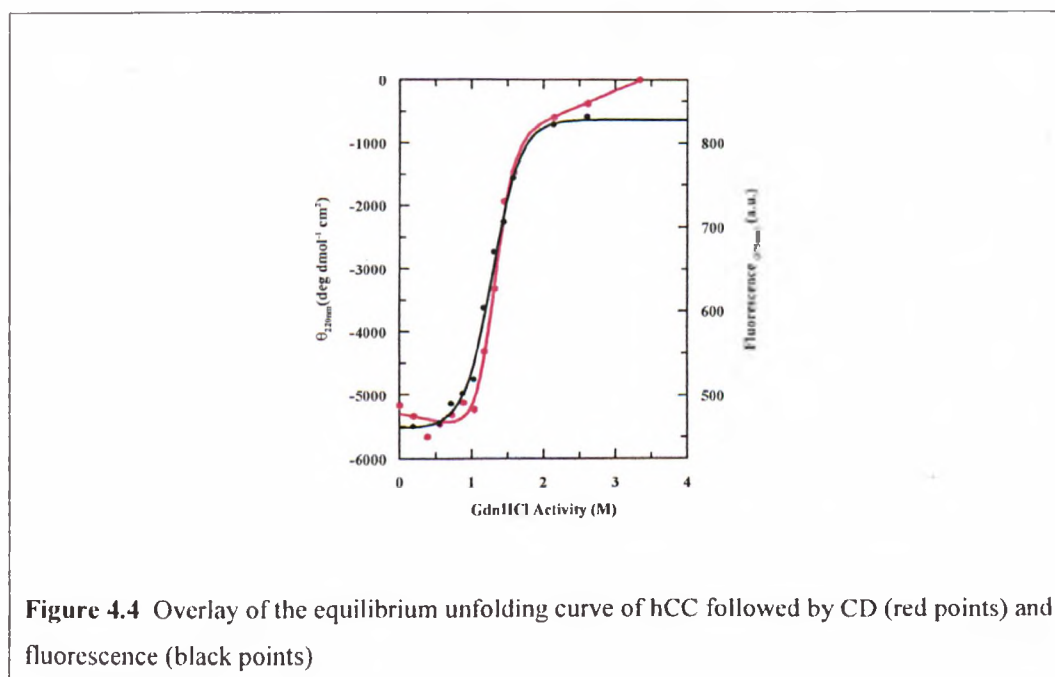

Figure 4.4 Overlay of the equilibrium unfolding curve of hCC followed by CD (red points) and fluorescence (black points)

pH unfolding – fluorescence

In addition to using a denaturant, such as GdnHCl, unfolding can be induced by lowering pH. At low pH protein becomes highly positively charged due to the protonation of aspartates and glutamates. The protein is therefore destabilised by electrostatic repulsion. In addition to this, proteins usually have buried groups that have highly perturbed pK_as due to the formation of salt bridges. At low pH the denatured state becomes protonated which causes a shift in the equilibrium towards the denatured state. Figure 4.5A shows the fluorescence spectra of the folded (pH 7) and unfolded state (pH 1). As with GdnHCl denaturation the

fluorescence of the unfolded state is red-shifted relative to the folded state and has a greater quantum yield. However, the increase in quantum yield seen with acid unfolding of hCC is significantly less than that seen with GdnHCl unfolding of hCC. This suggests that there may not be a complete loss of structure during the acid unfolding of hCC.
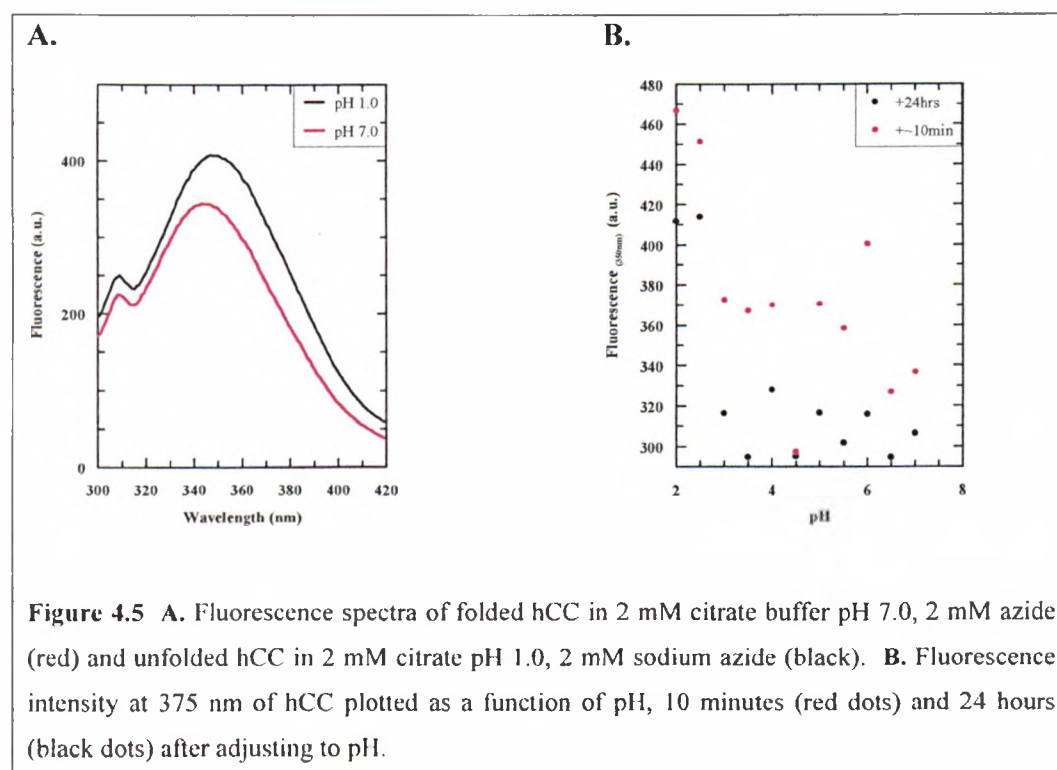

Figure 4.5 **A.** Fluorescence spectra of folded hCC in 2 mM citrate buffer pH 7.0, 2 mM azide (red) and unfolded hCC in 2 mM citrate pH 1.0, 2 mM sodium azide (black). **B.** Fluorescence intensity at 375 nm of hCC plotted as a function of pH, 10 minutes (red dots) and 24 hours (black dots) after adjusting to pH.

The difference in tryptophan fluorescence between the folded and unfolded state has been used as a probe to follow the equilibrium unfolding of hCC with pH. Figure 4.5B shows the change in fluorescence intensity at 350 nm with pH, 10 minutes and 24 hours after adjusting to the desired pH. The two data sets do not directly overlay although they both pass through a similar transition. It is possible that the reduction in fluorescence intensity is caused by fluctuations in the lamp of the fluorescence spectrophotometer. The reduced fluorescence intensity may also be caused by aggregation of hCC. It is know that partially denaturing conditions destabilise hCC enough for dimerisation and further aggregation to be induced. The prolonged period over which hCC is incubated at lowered pH may provide sufficiently denaturing conditions to induce dimerisation and further aggregation of hCC. As the dimeric form is very similar in structure to the monomeric form it is likely that the fluorescence intensity would be quenched relative to the unfolded

form. Therefore, any dimerisation would result in a reduction in the fluorescence intensity.

It is also clear from Figure 4.5 that lowering the pH to 1.0 is not sufficient to completely unfold hCC. In Figure 4.2B a clear sigmoidal curve is seen that represents the change from folded protein, through the transition region to the unfolded protein. In the case of Figure 4.5B, the folded protein and the majority of the transition region is observed, however, there is no levelling out of the transition region to represent the unfolded protein.

pH unfolding – CD

Far-UV CD can be used to study acid unfolding of hCC. Figure 4.6A shows the far-UV CD spectra of the folded (pH 7) and unfolded state (pH 1). As hCC unfolds due to acid unfolding there is a loss of CD signal due to a loss of secondary structure. Figure 4.6B shows the CD signal at 220 nm plotted against pH.

Figure 4.6 A. CD spectra of folded hCC in 2 mM citrate buffer pH 7.0, 2 mM azide (black) and unfolded hCC in 2 mM citrate buffer pH 1.0, 2 mM sodium azide (red). **B.** Ellipticity at 220 nm of hCC plotted as a function of pH.
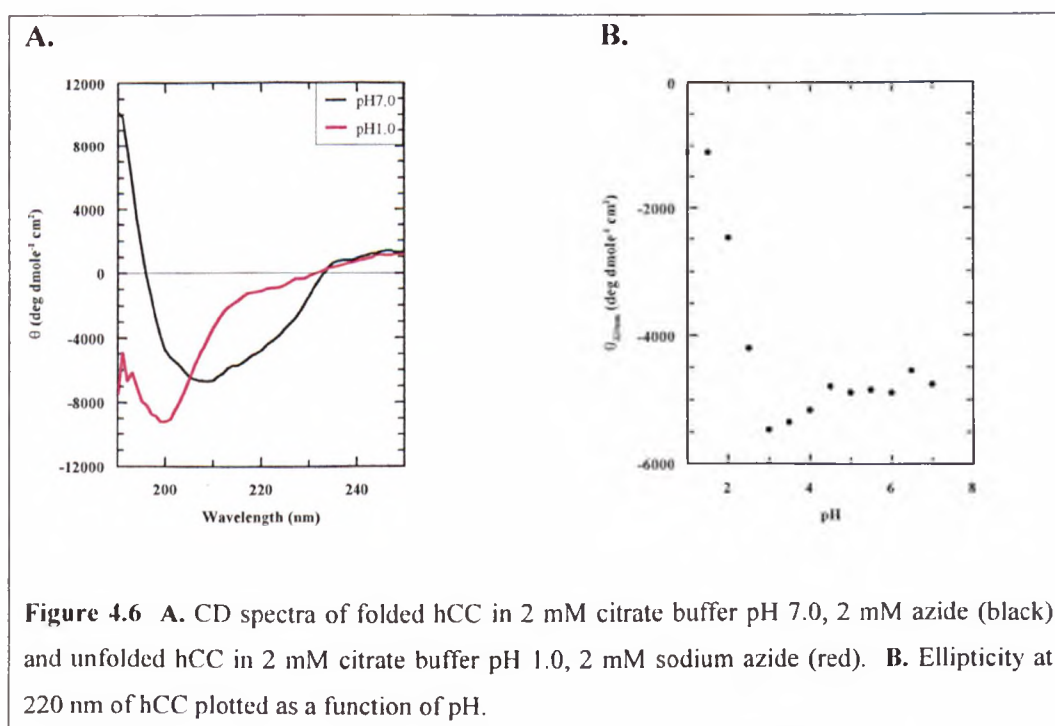
Figure 4.3A shows that when GdnHCl is used as a denaturant there is a complete loss of CD signal at 220 nm indicating that there is a complete unfolding of hCC. In contrast, Figure 4.6A shows that there is still residual CD signal at pH 1.0,

indicating that there is not a complete loss of secondary structure at pH 1.0 during the acid unfolding of hCC.

Together the pH unfolding data presented here indicates that lowered pH is not sufficient to completely unfold hCC. Although the level of tertiary and secondary structure is significantly reduced in comparison to the folded protein, there is still some residual structure at pH 1.0. GdnHCl is considered to be a very strong denaturant and, as the data presented above shows, is sufficient to completely unfold hCC as shown by a complete loss of secondary and tertiary structure.

4.4.2 Kinetic data

Stopped flow

Fluorescence has been used to take kinetic measurements of the folding and unfolding of hCC. Folded hCC was rapidly mixed with denaturant using a stopped flow machine and the change in fluorescence recorded over time. Likewise, unfolded hCC was rapidly transferred into folding conditions and the fluorescence change recorded over time. Figure 4.7 shows a typical example of the folding and unfolding transitions recorded with stopped flow. Each unfolding transient was fitted to a single exponential according to equation Eq. 4.4.1 to calculate the observed rate and the fluorescence amplitude. Refolding transients were recorded over a split timescale as two rates were observed, one seen in the first 2 seconds of the refolding reaction and a second much slower rate recorded over 100 seconds. The refolding transients were treated as two transients and each fitted to a single exponential according to equation Eq. 4.4.1.
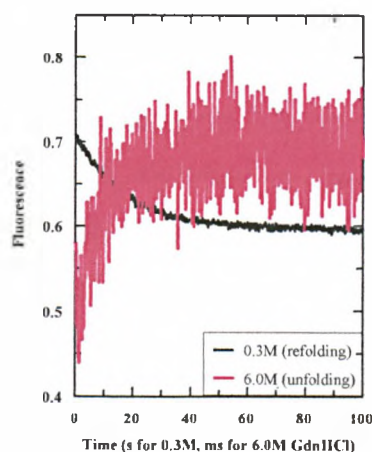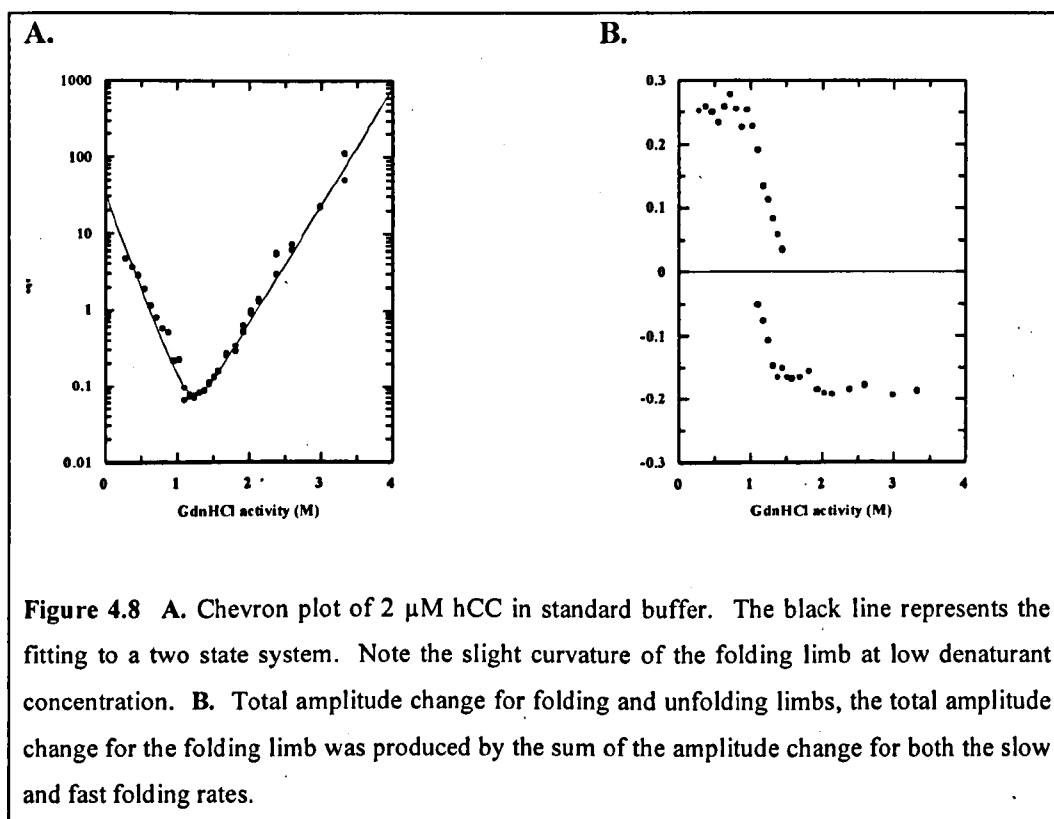
Figure 4.7 Typical folding (red) and unfolding (black) transients observed by stopped-flow fluorescence.

Chevron plot

The observed rates produced from fitting the fluorescence transients have been plotted against denaturant activity to produce the chevron plot seen in Figure 4.8. The characteristic V-shape is seen in the chevron plot indicating a 2-state folding reaction. If this was the case equation Eq. 4.5.1 would apply and the curve can be fitted using $k_{(F-t)} = k_{(F-t)(W)}.exp(-m_t.D)$ and $k_{(U-t)} = k_{(U-t)(W)}.exp((m_{(u)}-m_{(t)}).D)$. However, the non-linearity of the folding limb (left-hand side of the V-shape) of the chevron plot at low denaturant concentration suggests the presence of an intermediate. The chevron plot is poorly defined at low levels of GdnHCl activity, but the GdnHCl concentration can not be lowered further to better define the chevron plot due to the limiting dilution factor of the asymmetric mix of the stopped-flow machine. In order to determine folding rates at lower denaturant activity the folding and unfolding rates have been determined in the presence of the stabilizing agent sodium sulphate.

The amplitude change associated with each of the k_{obs} shown on the chevron plot provide us with some useful information about the folding reaction. Once both the fast and slow folding rates were identified the total amplitude change for the folding limb could be calculated by adding the amplitude associated with both rates. There is no amplitude difference between the unfolded protein and the start of the folding transients indicating that no burst phase amplitude is observed on folding. This indicates that there is no significant amplitude loss during the dead
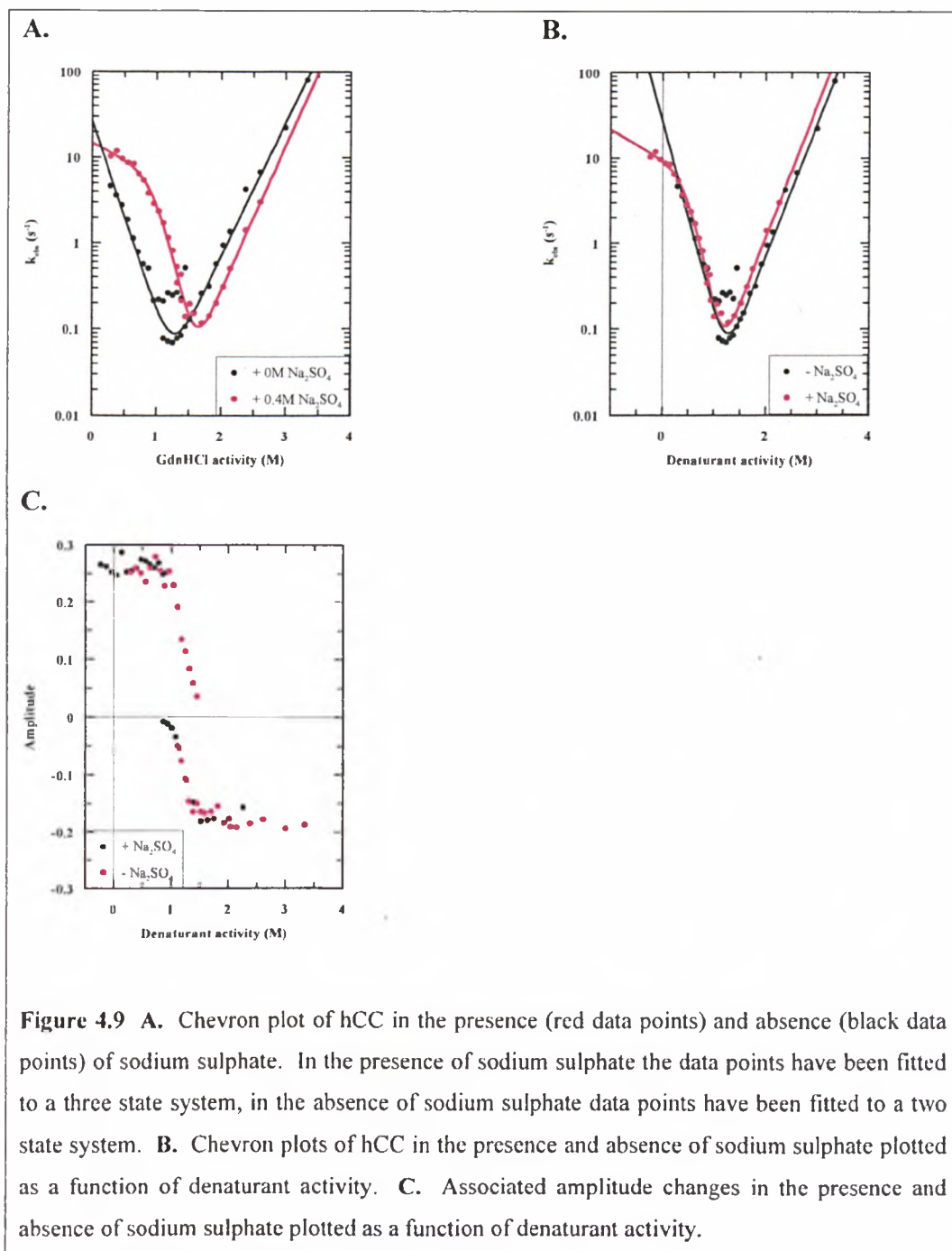
time of the stopped-flow machine. When plotted as a function of GdnHCl activity the amplitude change seen upon folding and unfolding is equal. Although this information supports the presence of a two state system it does not rule out the possibility of a three state system. In fact, as seen later in this chapter, the curvature of the folding limb observed in the presence of sodium sulphate indicates the presence of at least one intermediate in the folding pathway. The fact that the amplitude of the folding reaction is the same under conditions where the intermediate is populated and where it is not indicates that the fluorescence of the intermediate and the unfolded state are the same.

Figure 4.8 **A.** Chevron plot of 2 μM hCC in standard buffer. The black line represents the fitting to a two state system. Note the slight curvature of the folding limb at low denaturant concentration. **B.** Total amplitude change for folding and unfolding limbs, the total amplitude change for the folding limb was produced by the sum of the amplitude change for both the slow and fast folding rates.

Sodium sulphate

Sodium sulphate can be considered to act in an analagous, but opposite way to GdnHCl. As the effects of GdnHCl and sodium sulphate are additive and independent it is possible to produce conditions where the denaturant activity is very low and even negative. The stabilizing effect of sodium sulphate can be seen in Figure 4.9A where it effectively shifts the whole chevron plot to the right. When k_{obs} are plotted against denaturant activity determined by equation Eq. 4.2.1, the chevron plots in the presence and absence of sodium sulphate do not superimpose. The values used to calculate the denaturant activity of sodium

sulphate are an average value derived from the GdnHCl dependence of the free energy of change of solvation of N-acetyltryptophanamide and N-acetyltyrosineamide. The relationship between the solvation energy and the denaturant concentration will vary for each amino acid side chain so the actual sodium sulphate activity may vary depending on the composition of a protein. The calculation of the denaturant activity of sodium sulphate is also subject to an average error of 20%. The chevron plots in the presence and absence of sodium sulphate will superimpose if the stabilising activity of sodium sulphate is reduced by half as seen in Figure 4.9B. Figure 4.9C shows the plot of amplitude against denaturant activity in the presence and absence of sodium sulphate. The associated amplitudes overlay indicating that there is no difference in the amplitude change at each denaturant activity. Together this demonstrates that the addition of sodium sulphate does not significantly alter the conformation of the different states on the folding pathway.

Figure 4.9 A. Chevron plot of hCC in the presence (red data points) and absence (black data points) of sodium sulphate. In the presence of sodium sulphate the data points have been fitted to a three state system, in the absence of sodium sulphate data points have been fitted to a two state system. **B.** Chevron plots of hCC in the presence and absence of sodium sulphate plotted as a function of denaturant activity. **C.** Associated amplitude changes in the presence and absence of sodium sulphate plotted as a function of denaturant activity.

The rollover of the folding limb of the chevron plot produced in the presence of sodium sulphate is much better defined indicating the presence of at least one intermediate in the folding pathway. The chevron plot can be fitted to equation Eq. 4.5.4 with Eq. 4.5.5, Eq. 4.5.6 and Eq. 4.5.7 as temporary variables. The fit can be seen by the solid black line in Figure 4.9B and 4.10.
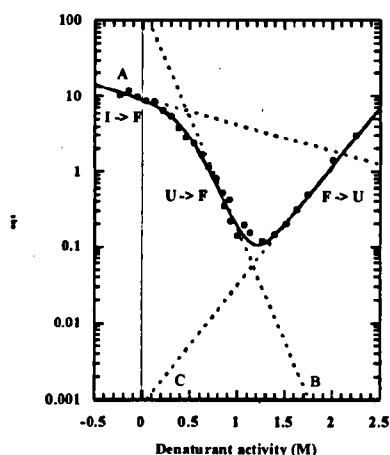
Figure 4.10 Chevron plot of hCC. Dashed lines represent lines A-C, the solid black line represents the fitting of the data points to equation Eq. 4.5.4.

The curve can be split into 3 lines, A, B and C that represent the transition I-F, U-F and F-U respectively which are marked as dashed lines in Figure 4.10. For each line the slope is the difference between the m-value of the ground state and the m-value of the transition state and the intercept is the $\ln k_{(W)}$. For line A, the slope is defined as $m_{(I)} - m_{(t)}$ and its intercept defines $k_{(I-t)(W)}$. The gradient of B is defined as $m_{(U)} - m_{(t)}$ with intercept $k_{(I-t)(W)}.K_{(I/U)(W)}$. The slope of C is defined as $-m_{(t)}$ with the intercept defined as $k_{(F-t)}$. In the case of line C the ground state is $m_{(F)}$ which, as the other m-values are measured relative to $m_{(F)}$, is defined as 0. The thermodynamic parameters ascertained from the fitting are summarised in table 4.3

$k_{(F-t)}$	$9.4 \pm 0.9 \text{ s}^{-1}$	$m_{(t)}$	$-3.6 \pm 0.2 \text{ M}^{-1}$	$\Delta G_{(I-F)}$	$-5.6 \text{ kcalmol}^{-1}$
$K_{(I-t)}$	$8.5 \times 10^{-4} \pm 3 \times 10^{-4} \text{ s}^{-1}$	$m_{(U)}$	$-10.5 \pm 0.7 \text{ M}^{-1}$	$\Delta G_{(U-I)}$	$-1.7 \text{ kcalmol}^{-1}$
$K_{(I/U)}$	16.7 ± 12	$m_{(I)}$	$-4.4 \pm 0.5 \text{ M}^{-1}$	$\Delta G_{(I-t)}$	8.3 kcalmol^{-1}
				$\Delta G_{(F-t)}$	$13.9 \text{ kcalmol}^{-1}$

Table 4.3 Kinetic folding parameters for human cystatin C

Proline isomerisation

The equilibrium constant for the normal peptide bond in proteins favours the *trans* conformation by a factor of 10^3-10^4. The peptidyl-prolyl bond is an exception with in the region of 2-20% in the *cis* conformation. The interconversion of *cis* to *trans* in solution at room temperature and neutral pH is quite slow, with a half life of 10-100 s. In the native protein each prolyl peptide has a unique conformation, either *cis* or *trans*. During unfolding the conformational restraints found in the native state are no longer present and the prolyl bonds are free to isomerise leading to a mixture of unfolded states with either the correct or incorrect prolyl isomer. Refolding of the unfolded state with the non-native isomer is slowed because of the coupling of refolding with the isomerisation of the incorrect prolyl bond.

In addition to the fast folding rates that are observed for hCC a second slow folding rate, shown in Figure 4.11, is also observed. During the single jump stopped-flow experiment hCC is unfolded for a prolonged period of time. In this time it is possible for a post-unfolding modification, such as proline isomerisation, to occur. In order for the protein to fold correctly, molecules with the proline in the *cis* conformation need to isomerise to the *trans* conformation. As the conversion from *cis* to *trans* takes around 10-100 s, proline isomerisation will be the rate limiting step in the folding reaction. Therefore, proline isomerisation is proposed as the cause of the slow folding rate of hCC. This hypothesis has been tested using a double jump stopped-flow experiment. In this experiment folded hCC is transferred into unfolding conditions for sufficient time for it to unfold, but insufficient time for proline isomerisation to occur, it is then transferred back to folding conditions and the fluorescence transient recorded. In transients recorded when hCC is refolded to 1M GdnHCl and 0.6 M GdnHCl a similar fast rate is observed but there is no evidence of the slow folding rate, as seen in Figure 4.11. This confirms that a post-unfolding modification occurs to hCC when it is unfolded for a prolonged period of time.
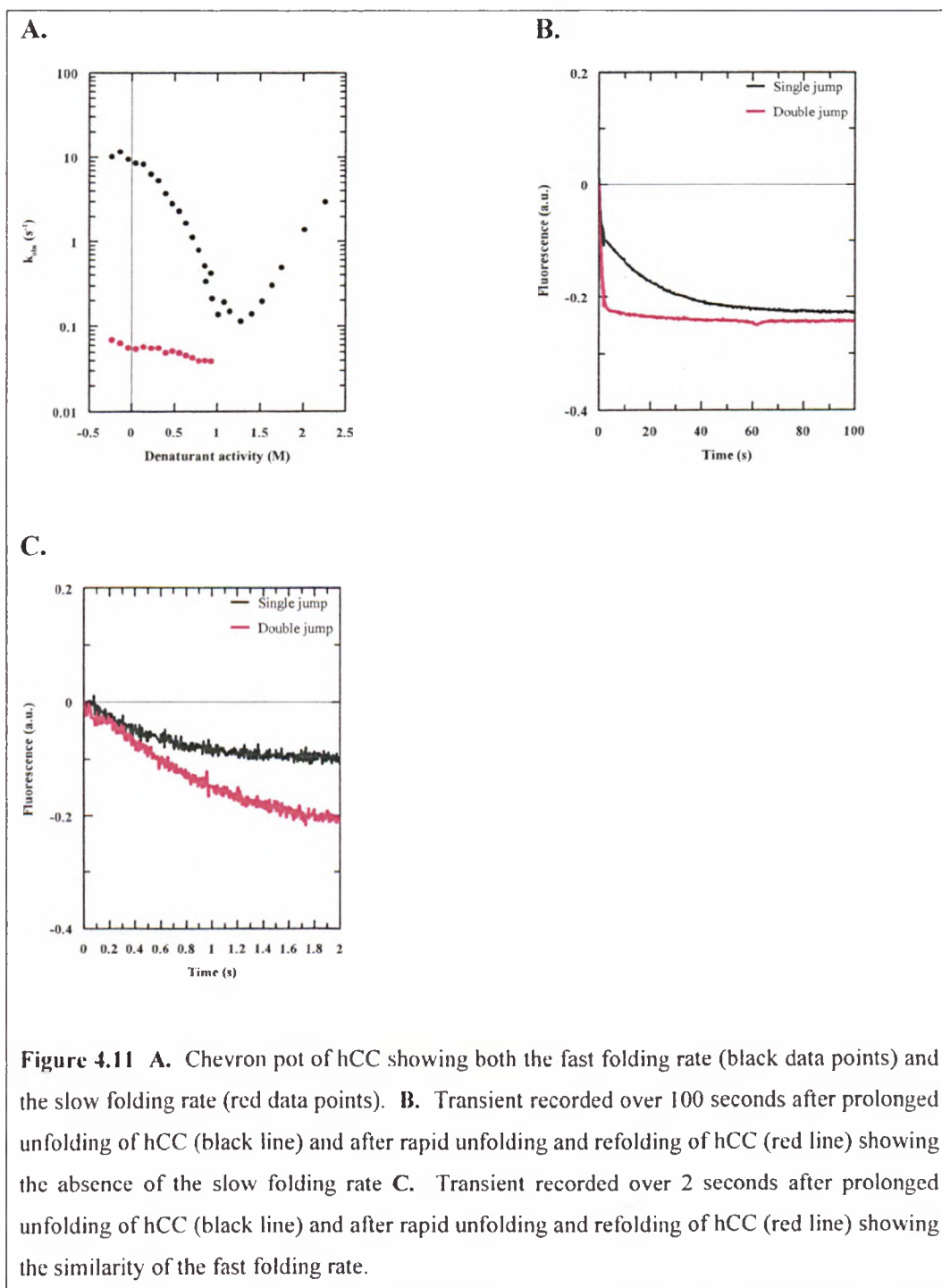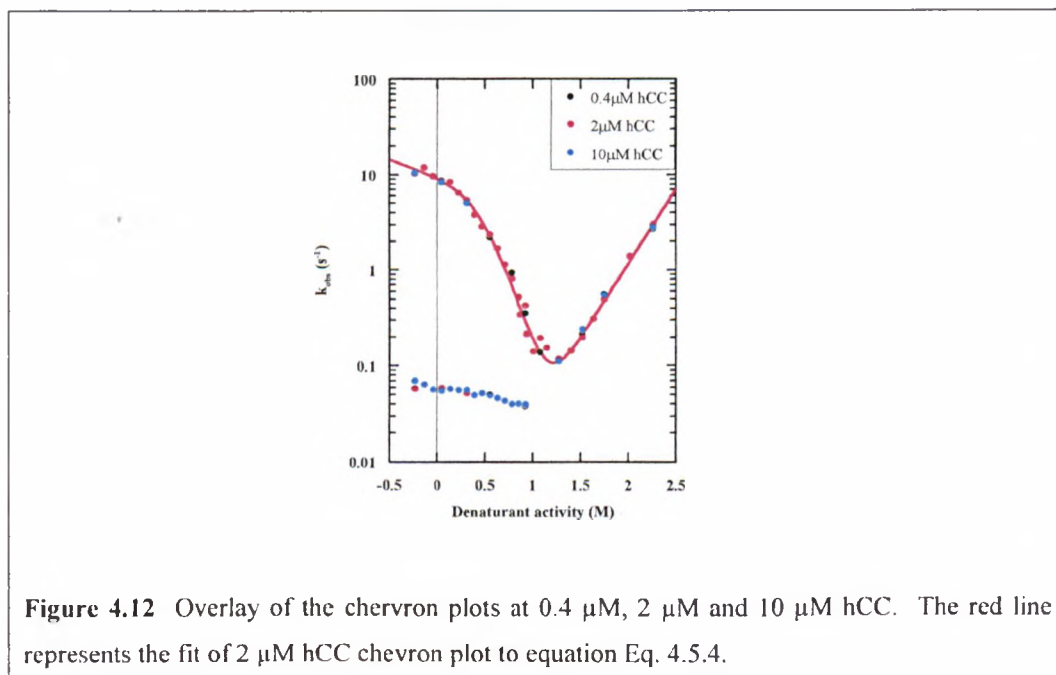
Figure 4.11 A. Chevron pot of hCC showing both the fast folding rate (black data points) and the slow folding rate (red data points). **B.** Transient recorded over 100 seconds after prolonged unfolding of hCC (black line) and after rapid unfolding and refolding of hCC (red line) showing the absence of the slow folding rate **C.** Transient recorded over 2 seconds after prolonged unfolding of hCC (black line) and after rapid unfolding and refolding of hCC (red line) showing the similarity of the fast folding rate.

Concentration dependence

Folding intermediates and denatured states are prone to aggregation because of exposed hydrophobic surfaces. Intermediate states may be wrongly identified because of artefacts in the kinetics caused by aggregation. Therefore kinetic measurements should be carried out over a range of protein concentrations to check for aggregation. As two or more protein molecules need to meet in order for aggregation to occur, aggregation is concentration dependant. Protein folding is a unimolecular process so folding is not concentration dependant. In order to

show that there is no significant aggregation during the timescale of the kinetic experiments, kinetic measurements have been recorded at 0.4 µM, 2 µM and 10 µM hCC. Figure 4.12 shows an overlay of chevron plots of hCC at 0.4 µM, 2 µM and 10 µM. There is no significant variation in the rates of folding and unfolding across this concentration range indicating that the rollover seen on the refolding limb is not an aggregation artefact.

Figure 4.12 Overlay of the chervron plots at 0.4 µM, 2 µM and 10 µM hCC. The red line represents the fit of 2 µM hCC chevron plot to equation Eq. 4.5.4.

4.5 DISCUSSION

The kinetic data indicates the presence of an intermediate and has been modelled to the folding pathway:

(Eq. 4.8.1) $\qquad U \; \rightleftharpoons \; I \; \rightleftharpoons \; F$

However, these studies can not determine whether this intermediate is on-pathway or off-pathway. In the on-pathway model shown above (4.8.1), I is a productive intermediate that serves as the ground state from which the rate-limiting transition state is acquired. The same kinetic profile would be observed if the intermediate was off pathway:

(Eq. 4.8.2) $\qquad I \; \rightleftharpoons \; U \; \rightleftharpoons \; F$

In this model, I is a misfolded species and the rate-limiting step results from a combination of creating a more unfolded ground state by the breaking of non-native interactions and the slow conversion to the folded state. If I needs to unfold before hCC can fold then the rate of conversion from U to F can be calculated by multiplying the observed folding rate with the folding equilibrium constant. In the case of hCC, if the intermediate is off-pathway then the rate of conversion between U and F would be approximately 160 s^{-1}. Whilst this is a feasible rate for the folding of humancystatin, in the case of cC this rate is 4 orders of magnitude faster making the folding rate so large that it unlikely to be applicable to the formation of a complex α/β structure such as cystatin.[157, 164] Given the similarity of the structures and folding pathways of hCC and cC it is unlikely that an intermediate would be on-pathway in the case of cC and off-pathway in the case of hCC. Therefore, it is most likely that the folding intermediate is an obligatory, on pathway intermediate in the folding of hCC.

The folding pathway is further complicated by proline isomerisation that occurs when hCC is in the unfolded state:

(Eq. 4.8.3) $$U_{iso} \rightleftharpoons U \rightleftharpoons I \rightleftharpoons F$$

Where U_{iso} represent the unfolded state with the incorrect proline isomer to allow folding and U, I and F represent the unfolded, intermediate and folded with the correct proline isomer. The incorrect proline isomer could be permitted in the intermediate state so that:

$$U_{iso} \rightleftharpoons I_{iso}$$
$$\updownarrow \qquad \updownarrow$$

(Eq. 4.8.4) $$U \rightleftharpoons I \rightleftharpoons F$$
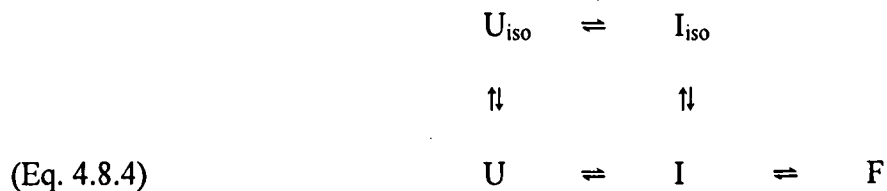
Analysis of the m-values of hCC indicates that the rapidly formed intermediate is relatively compact, with approximately half of the solvent excluded from the protein interior. Whether the relatively compact intermediate can form with the incorrect proline isomer present can not be determined from this data.

Comparison of the folding pathways of hCC and cC

The folding pathway deduced for hCC shows many similarities with that of cC. Table 4.4 shows a summary of the thermodynamic parameters for both cystatins at equilibrium.

Cystatin	m-value (M^{-1})	$\Delta G_{(w)}$ (Kcalmol^{-1})	[D]$_{1/2}$ (M)
hCC	-6.6 ± 1.7	-5.0 ± 1.5	1.3 ± 0.1
cC	-8.6 ± 0.3*	-15.6 ± 1.0*	3.1 ± 0.3*

Table 4.4 Comparison of the thermodynamic parameters determined from the equilibrium unfolding of hCC and cC. * data taken from [157]

The similarity of the m-values indicates that the solvent exposure during unfolding is similar for both proteins, which is consistent with the similarity of the structures of hCC and cC. The values of $\Delta G_{(w)}$ and [D]$_{1/2}$ are significantly larger for cC compared to hCC which reflects the increased stability of cC relative to hCC.
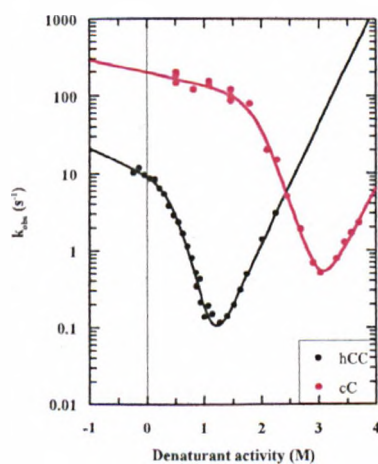

Figure 4.13 Comparison of the chevron plots of hCC (black data points) fitted to a three state system (black line) and cC (red data points) fitted to a three state system (red line).

Figure 4.13 shows an overlay of the chevron plots of hCC and cC. The comparison of the equilibrium unfolding of hCC and cC has already demonstrated

that hCC is significantly less stable than cC. Therefore, hCC unfolds at lower denaturant concentrations effectively shifting the chevron plot to the left of the cC plot. However, the similarity in the shape of the chevron plot indicates that hCC and cC fold through very similar pathways. Table 4.5 summarises the parameters ascertained from the kinetic analysis of each protein.

	hCC	cC
$K_{(F-t)}$ (s^{-1})	9.4 ± 0.9	192 ± 30
$K_{(I-t)}$ (s^{-1})	$8.5 \times 10^{-4} \pm 3 \times 10^{-4}$	$1.4 \times 10^{-5} \pm 4 \times 10^{-6}$
$K_{(I/U)}$	16.7 ± 12	8755 ± 8200
$M_{(t)}$ (M^{-1})	-3.6 ± 0.2	-3.26
$M_{(F)}$ (M^{-1})	-10.5 ± 0.7	-8.43
$M_{(I)}$ (M^{-1})	-4.4 ± 0.5	-3.65
$\Delta G_{(U)}$ $(kcalmol^{-1})$	0	0
$\Delta G_{(I)}$ $(kcalmol^{-1})$	-1.69	-5.28
$\Delta G_{(t)}$ $(kcalmol^{-1})$	6.64	1.31
$\Delta G_{(F)}$ $(kcalmol^{-1})$	-7.28	-10.25

Table 4.5 Comparison of the parameters determined from the kinetic analysis of hCC and cC. cC data is taken from [157]. ΔG values are calculated using the equation $\Delta G = -RT\ln K$. $\Delta G_{(I)}$ is calculated from $K_{(I/U)}$. In order to determine $\Delta G_{(F)}$, $K_{(F/I)}$ is calculated by from $k_{(F)}/k_{(I)}$. $\Delta G_{(I-t)}$ and $\Delta G_{(F-t)}$ were calculated from the equation $k = (kT/h)e^{(-\Delta G/RT)}$ where k is the Boltzmann constant, h is the Planks constant. An empirical estimate of 10^7 is assumed for (kT/h).[165]

The ΔG and m-values have been used to produce the free energy profile seen in Figure 4.14. The profile shows the difference in stability between cC and hCC, but also highlights the similarities between the two folding pathways. Both fold via a relatively compact intermediate before passing through a transition state barrier to the folded state.

It has been proposed that the major contributor to the folding barrier of cC is the removal of water. Two processes are required for the I to F transition to occur, the removal of water from the protein core and the immobilisation of sidechains. Reduced cC has a highly compact, molten globule structure in which side chains are mobile across the whole molecule, but the secondary structure and
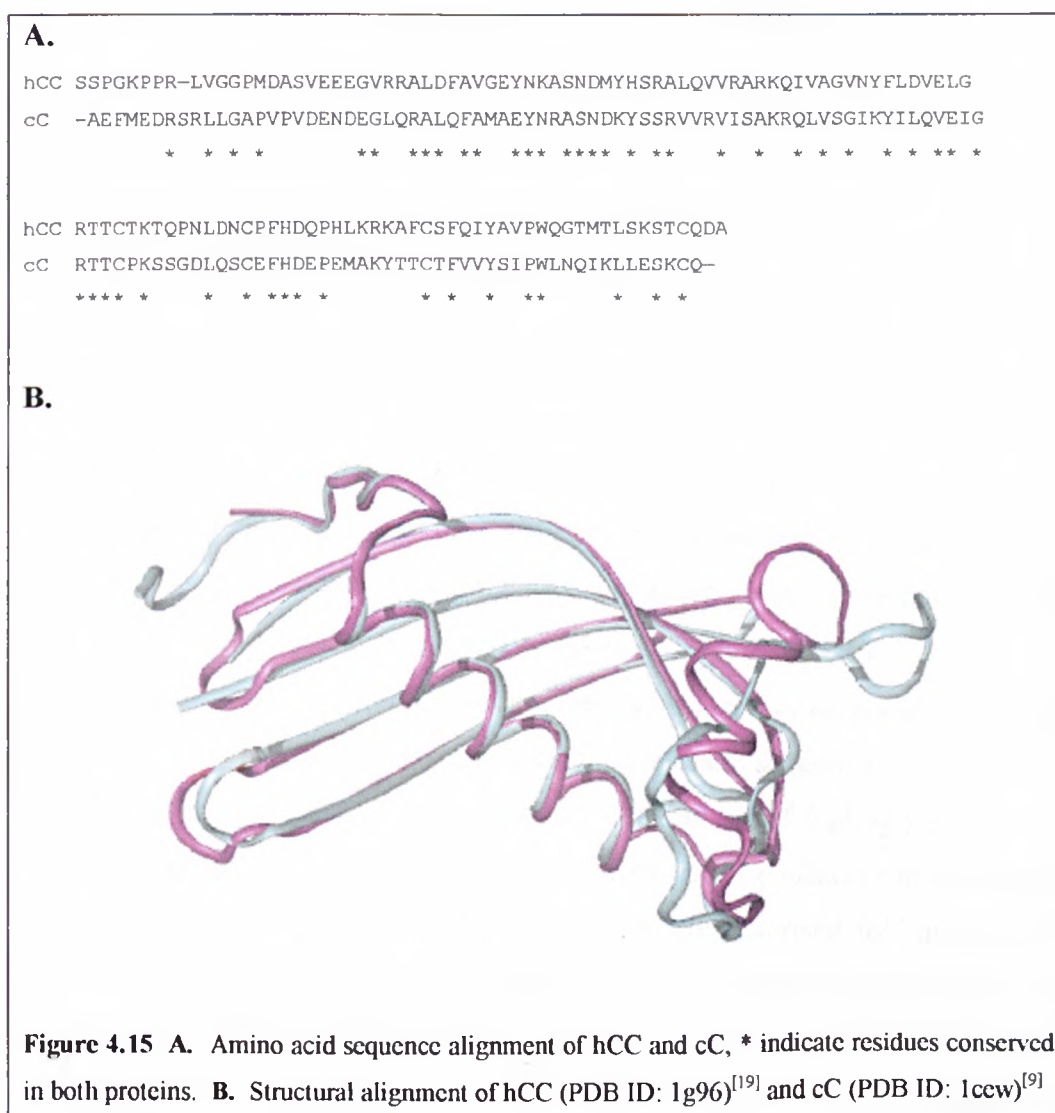
Page missing in original

analogous pathways to closely related structures, it is interesting to consider what factors influence the folding rates the two proteins.

One obvious difference between hCC and cC is the difference in their amino acid sequences. Figure 4.15 shows the sequence alignment of hCC and cC, highlighting that there is only 42% sequence identity between the two proteins. However, it has been reported that large changes in amino acid sequence that don't alter the overall topology of a protein don't greatly effect the folding rate.[166] Rather, theoretical and experimental studies of protein folding indicate that the topology of the native state, not its amino acid sequence, is the most important factor in determining the folding pathway of a protein.[167, 168]

A.

```
hCC  SSPGKPPR-LVGGPMDASVEEEGVRRALDFAVGEYNKASNDMYHSRALQVVRARKQIVAGVNYFLDVELG
cC   -AEFMEDRSRLLGAPVPVDENDEGLQRALQFAMAEYNRASNDKYSSRVVRVISAKRQLVSGIKYILQVEIG
         *   *  *  *      **  *** **  *** **** *  **   *   *   *  *  *   *  * **  *

hCC  RTTCTKTQPNLDNCPFHDQPHLKRKAFCSFQIYAVPWQGTMTLSKSTCQDA
cC   RTTCPKSSGDLQSCEFHDEPEMAKYTTCTFVVYSIPWLNQIKLLESKCQ-
     **** *     *  *  ***  *        *  *  *  **     *   *  *
```

B.

Figure 4.15 A. Amino acid sequence alignment of hCC and cC, * indicate residues conserved in both proteins. **B.** Structural alignment of hCC (PDB ID: 1g96)[19] and cC (PDB ID: 1ccw)[9]

Relative contact order can be used as a measure of topological complexity as it reflects the relative importance of local and non-local contacts to a proteins native structure. It is calculated by averaging sequence distance between all pairs of contacting residues normalised by the total length of the protein. Proteins with predominantly long range interactions have high contact order whilst predominantly local interactions are of a low contact order. For small, two-state proteins the natural logarithm of the folding rate is well correlated with relative contact order. This correlation holds for small proteins that fold according to a two-state model, but additional factors may complicate the folding of larger non-two-state proteins.[166-168] While topology is a critical determinant of folding kinetics, mutations that alter the folding rate but don't significantly change the contact order show that other factors clearly contribute. [168]

Although it is possible to have two proteins with similar topologies, but quite different contact order values, this is not the case for hCC and cC. hCC and cC not only share a common topology they also share a common pattern of contacts between amino acids in the polypeptide chain. Clearly in the case of these two cystatins other factors than contact order strongly influence the folding rate of the two proteins.

It has been proposed that global stability can play a key role in dictating the folding rates among members of a structurally similar family. The correlation between folding rate and stability suggests that stability plays an important role in determining the folding rates of α/β proteins or β-sheet proteins. Stability has also been used to predict relative folding rates across sets of homologous proteins and topologically similar proteins lacking significant sequence identity.[168-170] Stability may reflect an important secondary determinant of folding kinetics as, at least among structurally similar proteins, stability-specific effects can account for up to 2 orders of magnitude in the range of the characterised folding rates.[168] The difference in folding rates between hCC and cC may be attributed to the difference in stability between the two cystatins. This suggests that it is the relative strength of native-like contacts rather than the distribution of contacts that dictates the folding rate of these two proteins.

CHAPTER FIVE

EARLY AGGREGATES IN THE FORMATION OF CYSTATIN AMYLOID

5.1   INTRODUCTION

Amyloid deposits are a hallmark of several common, incureable diseases including Alzheimers' disease, Parkinsons' disease and type II diabetes. Each clinical syndrome is associated with a specific protein that, although normally soluble, in the disease state undergoes a conformational change that causes it to aggregate into insoluble amyloid fibres. Identifying the factors that influence the initiation and progression of amyloid formation remains a central issue in the development of treatments against amyloid disease. Emerging evidence indicates that pre-fibrillar oligomeric species rather than amyloid fibres are the toxic species in amyloid disease.[82, 96] This highlights the need for a detailed understanding of the molecular mechanism of amyloidogenesis prior to the development of targeted therapies against amyloid disease.

In studying the mechanism of amyloidogenesis it is important to gain an understanding of the early stages of protein assembly. The cystatins have proven to be a good model system for characterising the initial stable species in the amyloid pathway. In the case of cC, monomers associate to form dimers, tetramers and larger molecular weight species under conditions where the native state is destabilised.[136] This indicates that dimers and tetramers are intermediates, whether on- or off- pathway, that preceed the formation of amyloid. Dimeric cC has been proposed as the assembly competent species as cC multimers are formed from dimers in a similar manner to tetramers.[68] Tetramers are thought to be an off-pathway species that is formed from an early termination or cyclisation event.[68]

hCC has been shown to form domain swapped dimers under mildly denaturing conditions but no tetrameric species have been isolated.[20] Mutants of hCC stabilised against domain swapping by the insertion of disulphide bridges across the domain swapping interface have a drastically reduced ability to produce dimers and amyloid fibres under conditions known to induce fibrillisation of wild-

type hCC.[160] Prevention of amyloid formation by blocking domain swapping indicates that dimer is an obligatory intermediate in the formation of amyloid. hCC will also form doughnut shaped oligomers that are fibril-assembly intermediates. Mutants of hCC that are stabilised against domain swapping do not produce oligomers, indicating that domain swapping is involved in their formation.[171]

In this chapter size exclusion chromatography-high pressure liquid chromatography (SEC-HPLC) has been used to characterise the early stages of amyloidogenesis of both wild-type hCC and P103A cC. Whilst it has the advantage of being able to resolve the size of early cystatin aggregates, it is limited in the detection of larger aggregates. For this reason, when necessary, EM has been used for the detection of larger aggregates and amyloid.

5.2   MATERIALS AND METHODS

5.2.1   Formation of domain swapped hCC dimers

A 200 µM hCC sample in 50 mM sodium phosphate pH 6.0, 2 mM sodium azide was prepared for NMR by the addition of 10% D_2O. A ^{1}H-NMR spectrum was recorded of the monomeric sample prior to the induction of dimerisation by the addition of aliquots of a 7 M GdnHCl stock solution in 50mM sodium phosphate pH 6.0 to a final concentration of 1 M GdnHCl. ^{1}H-NMR spectra were recorded to determine whether the dimerisation of hCC was complete in 1 M GdnHCl, and following the removal of GdnHCl by buffer exchange. A second 50 µM hCC sample was prepared and the protein was allowed to dimerise in the same conditions. All spectra were recorded at 298°K on a Bruker DRX-500 spectrometer controlled using XWinNMR software (Bruker). Spectra were processed using Felix (Accelerys). Both samples were buffer exchanged into 50 mM sodium phosphate pH 7.0 to remove GdnHCl and stored at 25°C.

5.2.2   Analytical size exclusion chromatography (SEC)

Analytical SEC is a chromatographic method that can be used to separate proteins based on their molecular size, which for globular proteins is generally approximately proportional to their molecular weight. During SEC there is a continual partitioning of the sample, and therefore the resolution is dependant on the column length. In order to reduce the run time the column is packed with an

efficient matrix that is able to withstand the high operating pressures used in SEC-HPLC. This enables the analytical SEC-HPLC column to be relatively short and operate at a higher flow rate than a low pressure SEC column, while producing equivalent resolution. This reduces the run time of the column to around 20 minutes whereas an equivalent resolution using a low pressure SEC column would take 1-2 hours. SEC-HPLC is therefore a quick, efficient method of analysing the oligomeric state of proteins during the early stages of aggregation.

Testing hCC dimer stability

A portion of the dimeric 50 µM hCC sample described in section 5.2.1 was diluted to prepare a 5 µM sample of dimeric hCC. This dimeric sample, and the two described in section 5.2.1 were analysed using SEC-HPLC according to the method described in section 2.6.4 to confirm that dimerisation remained complete. Aliquots of each dimeric sample were removed at regular time intervals and analysed immediately by SEC-HPLC for up to 4 days after dimerisation.

Guanidine hydrochloride dependence of hCC dimerisation

60 µM hCC in 50 mM sodium phosphate pH 7.0 was incubated at 25°C with either 0.5 M, 0.75 M, 1 M, 1.25 M, 1.5 M or 1.75 M GdnHCl. Aliquots were removed and immediately analysed using SEC-HPLC for up to 43 days after induction of dimerisation. Samples were gently agitated prior to the removal of aliquots to ensure the removal of a representative distribution of species.

P103A cC aggregation

100 µM P103A cC in 10 mM potassium phosphate buffer, pH 7.0 was incubated at 85°C to induce aggregation. 25 µl aliquots were removed at regular time intervals and transferred into 25 µl ice-cold 50 mM potassium phosphate buffer pH 7.0 to quench the reaction. Samples were stored at 4°C for up to 24 hours until they were analysed by SEC-HPLC. To prevent evaporation, the sample was incubated in a parafilm sealed Eppendorf tube. To maintain a uniform sample for analysis the sample tube was gently agitated prior to aliquot removal. In order to minimise any dilution error, the same pipette, was used throughout.

5.2.3   Transmission Electron Microscopy

In transmission electron microscopy (TEM) an electron beam is transmitted through a specimen that is semitransparent to electrons. The electrons that pass through the specimen can be used to provide structural information about the specimen. Electrons that remain in the beam are detected on a fluorescent screen where dark areas represent areas where the electrons have been scattered. Contrast can be enhanced by the use of heavy metal compounds such as uranyl formate. Such stains enhance structural detail because their dense nuclei scatter the electrons out of the optical path.

Wt hCC grids

2 µl aliquots of each dimerisation reaction were used to prepare EM grids according to the method described in section 2.7.4. Grids were analysed and photographed with the help of Dr. M. Conroy.

P103A cC grids

Dimerisation an oligomerisation of P103A cC was induced according to the method described above. 2 µl of the quenched sample 81 hours after induction of dimerisation was used to prepare an EM grid according to the method described in section 2.7.4. Grids were analysed and photographed by Dr. C. Jelinska.

5.3   DATA FITTING

Unless otherwise stated data manipulation and fitting was carried out using Grafit software (Erithacus software Ltd).

HPLC data – rates

In the dimerisation reaction two monomers (M) become a dimer (D):

(Eq. 5.1.1) $2M \rightarrow D$

For a bimolecular reaction the rate at which the reactants meet is proportional to their concentration so that:

(Eq. 5.1.2) $d[M]/dt = -k[M]^2$

The change in monomer concentration can be integrated between the concentration at the start of the reaction, $[M]_0$ (represented in equation Eq. 5.1.3 as x) and the concentration at time t, $[M]_t$ (represented in equation Eq. 5.1.3 as y). The change in rate can be integrated between time at the start of the reaction, i.e. 0, and time t.

(Eq. 5.1.3) $\quad \int_x^y d[M] / [M]^2 = \int_0^t k\, dt \quad$ or $\quad (1/[M]_t) - (1/[M]_0) = kt$

This rearranges to

(Eq. 5.1.4) $\quad [M]_t = [M]_0 / (1 + kt[M]_0)$

According to equation Eq. 5.1.1 the rate of dimer formation should be equivalent to the rate of monomer loss. In which case

(Eq. 5.1.5) $\quad [D]_t = [M]_t - [M]_0$

By combining equations Eq. 5.1.4 and Eq. 5.1.5

(Eq. 5.1.6) $\quad [D]_t = ([M]_0 / (1 + kt[M]_0)) - [M]_0$

Which can be rearranged to

(Eq. 5.1.7) $\quad [D]_t = ([M]_0 - [M]_0.(1 + kt[M]_0)) / (1 + kt[M]_0)$

or

(Eq. 5.1.8) $\quad [D]_t = -kt[M]_0^2 / (1 + kt[M]_0)$

Equation Eq. 5.1.8 has been used to fit the appearance of dimer. In order to fit to the dissapearance of monomer, equation Eq. 5.1.5 and Eq. 5.16 are combined and rearranged to:

(Eq. 5.1.9) $\quad [M]_t = [M]_0 / (1 + kt[M]_0)$

GdnHCl dependence of rates

Bimolecular rate constants for dimerisation vary logarithmically with GdnHCl concentration according to the equation:
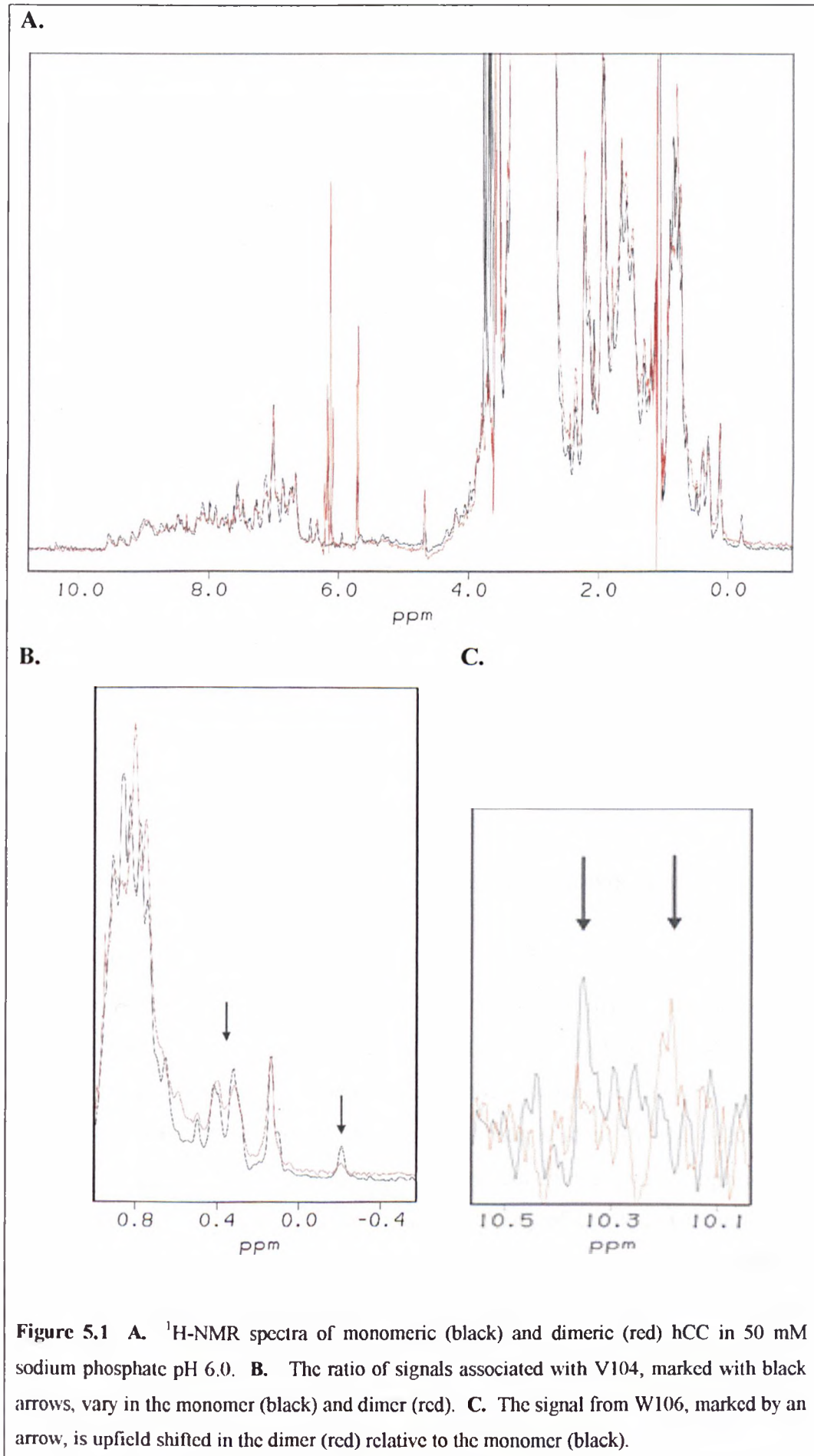
(Eq. 5.2.1) $\ln k_{obs} = \ln k_{(w)} + m.[GdnHCl]$

Where k_{obs}, k_W, m, $[GdnHCl]$ represents the observed rate constant, the rate constant in the absence of denaturant, the m-value and the GdnHCl concentration respectively. The bimolecular rate constants have been plotted against denaturant activity and fitted to equation Eq. 5.2.1 to extrapolate to a rate constant in the absence of water ($k_{(w)}$). GdnHCl activity rather than GdnHCl concentration has been used to take into account the nonlinear dependence of the free energy of folding with GdnHCl (see equation Eq. 4.1.1)[162]. The m-value for dimerisation provides information on the solvent exposure of hydrophobic groups in the transition state relative to the folded monomer.

5.4   RESULTS

5.4.1   Dimerisation of hCC

hCC has been shown to self associate to dimers through three dimensional domain swapping under mild denaturing conditions, such as the presence of denaturant, low pH or high temperature.[19, 20] ^{1}H-NMR spectra confirm the formation of a domain swapped dimer of hCC in the sample used here upon induction of dimerisation with 1 M GdnHCl (within 1 hour for a 200 μM sample). Figure 5.1A shows an overlay of the ^{1}H-NMR spectra of monomeric and dimeric hCC. As would be expected of a domain swapped dimer, the overall features of the two spectra look remarkably similar. However, certain characteristic signal shifts are associated with the formation of dimer that can be used to identify the domain swapped dimer.[20] The clearest signal shifts in the ^{1}H-NMR spectra are those of Val[104] and Trp[106]. Figure 5.1B and 5.1C show the change upon dimerisation in the peaks associated with Val[104] and Trp[106] respectively.

Figure 5.1 **A.** ¹H-NMR spectra of monomeric (black) and dimeric (red) hCC in 50 mM sodium phosphate pH 6.0. **B.** The ratio of signals associated with V104, marked with black arrows, vary in the monomer (black) and dimer (red). **C.** The signal from W106, marked by an arrow, is upfield shifted in the dimer (red) relative to the monomer (black).

The elution profile of monomeric and dimeric hCC, shown in figure 5.2, shows that monomer and dimer can clearly be resolved using SEC-HPLC at the chosen flow-rate of 1 ml/min. Monomeric hCC elutes in 11.7 mls whilst dimeric hCC elutes in 11 mls. For globular proteins the elution volume is approximately proportional to the logarithm of its molecular weight. When compared against a calibration curve the elution volumes of both monomeric and dimeric hCC are retarded relative to their molecular weights. The predicted molecular weight of dimeric hCC from the calibration curve, 20 kDa, shows that it is double the predicted molecular weight of monomeric hCC, 10 kDa. The molecular weight of monomeric hCC has been confirmed by MS (section 3.2.7).
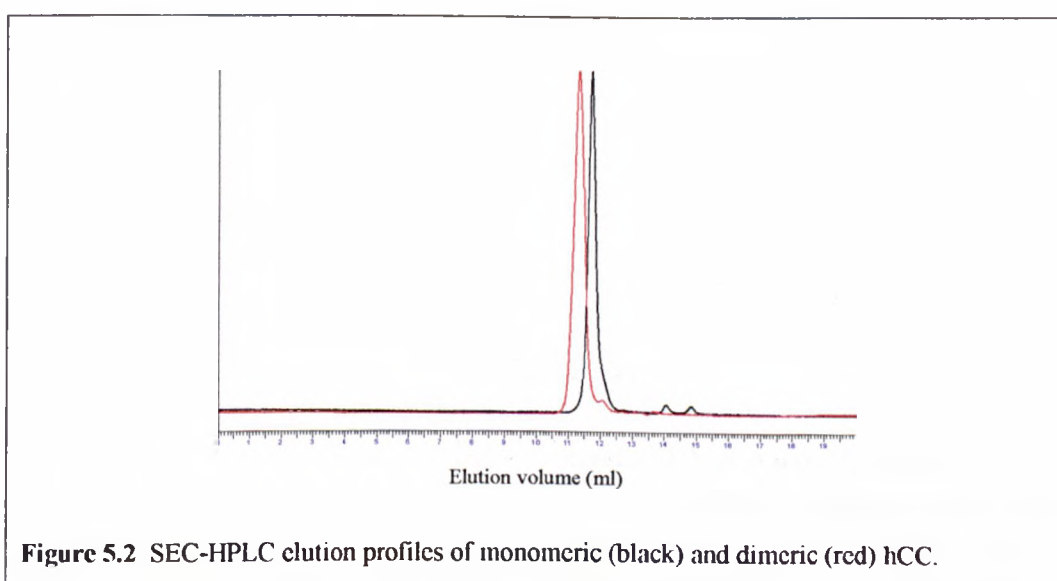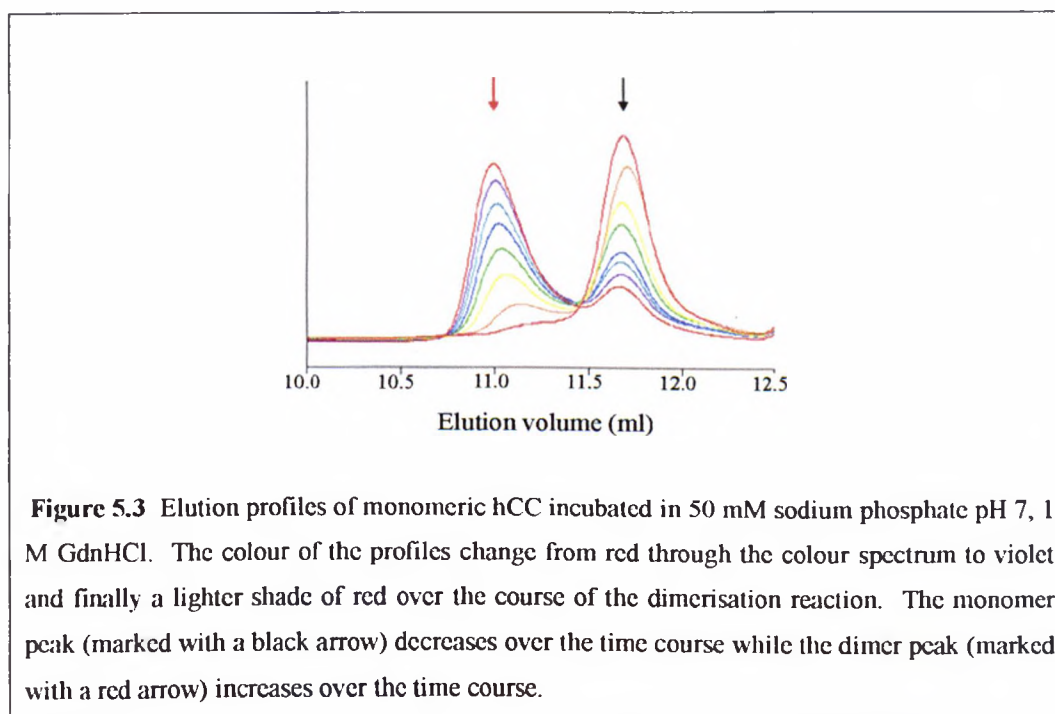

Figure 5.2 SEC-HPLC elution profiles of monomeric (black) and dimeric (red) hCC.

Dimer stability

The stability of hCC dimers in 50 mM sodium phosphate pH 7.0 was determined by analysing aliquots of dimerised hCC by SEC-HPLC. In four days SEC-HPLC chromatograms showed no evidence of monomeric species or larger oligomeric species over the hCC concentration range that was studied (5 µM – 200 µM). Although dimerisation may occur under these conditions over a significantly longer time scale, this experiment proves the stability of isolated hCC dimer over the time scale required for the experiments described below. The high stability of the domain swapped dimer is consistent with what is observed for cC where the formation of dimer is irreversible under the experimentally accessible conditions. This contrasts with the general case where the free energy difference between monomers and domain-swapped dimers is small.[134]
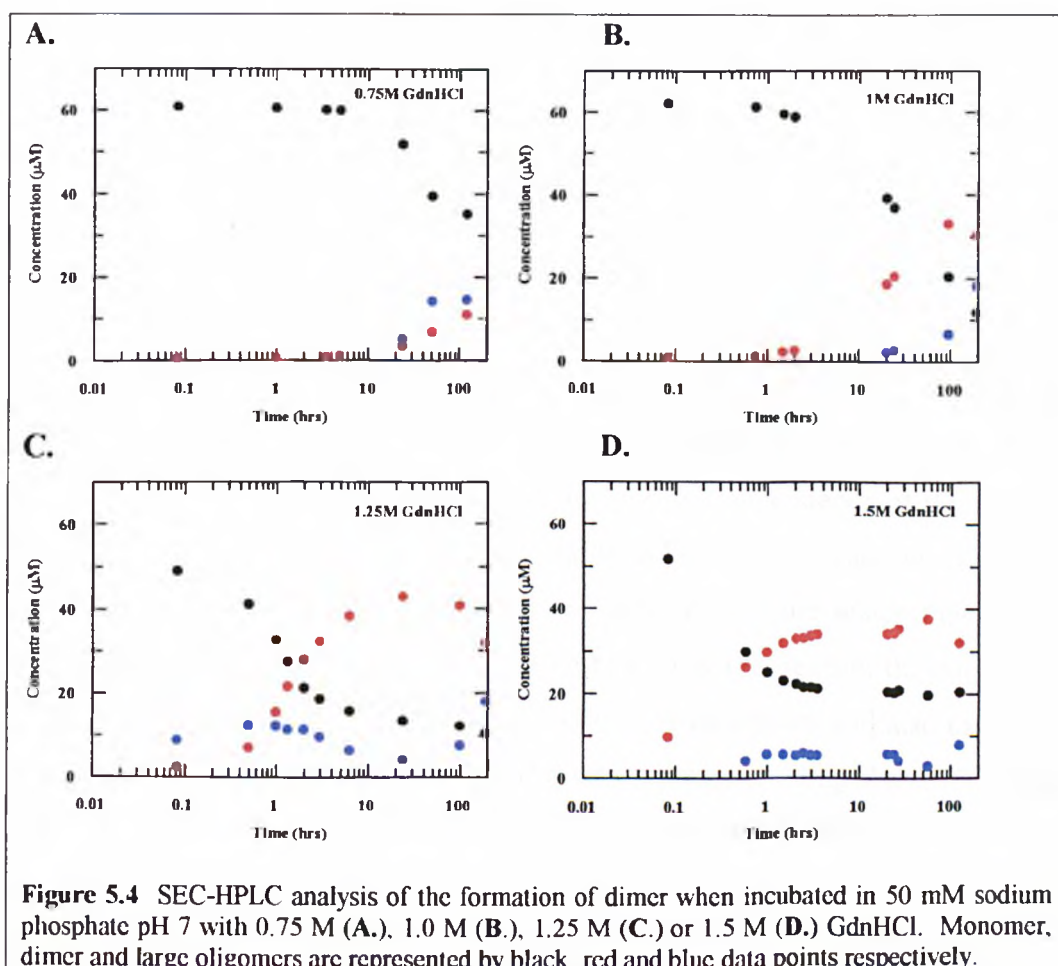
Guanidine hydrochloride induced dimerisation of hCC

The ability to clearly resolve between monomer and dimer using analytical SEC-HPLC makes it a useful method for characterising the dimerisation reaction of hCC. Following the induction of dimerisation by the addition of GdnHCl, the dimerisation reaction was followed by regularly removing aliquots from the reaction and analysing by SEC-HPLC. Figure 5.3 shows the gradual appearance of the dimer paired with the gradual decrease of the monomer over time. No other species are observed by SEC-HPLC despite prolonged incubation under dimer inducing conditions.

Figure 5.3 Elution profiles of monomeric hCC incubated in 50 mM sodium phosphate pH 7, 1 M GdnHCl. The colour of the profiles change from red through the colour spectrum to violet and finally a lighter shade of red over the course of the dimerisation reaction. The monomer peak (marked with a black arrow) decreases over the time course while the dimer peak (marked with a red arrow) increases over the time course.

The concentration of monomer and dimer at each time interval has been determined from the peak height scaled to the peak height of 60 μM hCC in the absence of denaturant. A series of GdnHCl concentrations were used to induce dimerisation over a range that destabilises the monomer, but does not cause complete unfolding. For each reaction at a given GdnHCl concentration, the concentration of monomer and dimer has been plotted as a function of time (figure 5.4). Monomer and dimer are the only species that are observed by SEC-HPLC. However, the sum of the monomer and dimer concentrations does not account for the original protein concentration at the start of the reaction. In the case of cC this "missing" protein concentration is accounted for by large oligomeric species that are unable to penetrate the matrix of the SEC column.

Therefore, the "missing" protein concentration representing multimers has been plotted alongside the concentration of monomer and dimer in figure 5.4. No significant dimerisation was observed in the dimerisation reaction in 0.5 M GdnHCl. Extrapolating the GdnHCl dependance of the dimerisation rate (shown later in the chapter) indicate that at 0.5 M GdnHCl dimerisation would not be observed during the time course of the experiment. No population of dimer is seen in the dimerisation reaction in 1.75 M GdnHCl. As shown in chapter 4, the equilibrium between folded and unfolded protein is significantly shifted towards unfolded protein at this concentration of denaturant. The monomeric protein that is observed by SEC-HPLC is likely to be unfolded protein that has refolded when the denaturant is removed during the SEC. In each time course shown in figure 5.4 there remains a residual amount of monomeric protein. As the level of monomer is greater the higher the concentration of GdnHCl it is likely that this reflects the proportion of unfolded protein present in the sample.

Figure 5.4 SEC-HPLC analysis of the formation of dimer when incubated in 50 mM sodium phosphate pH 7 with 0.75 M (**A.**), 1.0 M (**B.**), 1.25 M (**C.**) or 1.5 M (**D.**) GdnHCl. Monomer, dimer and large oligomers are represented by black, red and blue data points respectively.

Guanidine hydrochloride dependence of the bimolecular rate constants

For each dimerisation reaction shown in figure 5.4 the bimolecular rate constant has been determined by fitting the appearance of dimer to equation Eq. 5.1.8 and the depletion of monomer to equation Eq. 5.1.9. The rates determined from fitting this data are summarised in table 5.1.

[GdnHCl]	lnk monomer decrease	lnk dimer formation
1.50 M	3.52 ± 0.20	3.78 ± 0.49
1.25 M	2.16 ± 0.12	1.08 ± 0.64
1.00 M	-2.07 ± 0.12	-1.13 ± 0.91
0.75 M	-2.40 ± 0.19	-5.77 ± 0.59

Table 5.1 Summary of the bimolecular rate constants (k) determined from fitting the dimerisation reactions in figure 5.4 to equation Eq. 5.1.8 and Eq. 5.1.9. The units of k are $M^{-1}s^{-1}$. Errors are taken from the standard deviation of the fit calculated by Grafit software.

The rate constant for the transition from monomer to dimer is dependant on several factors including temperature, pressure and denaturant concentration. The dependance of the reaction rate on GdnHCl activity reports the degree of hydrophobic group exposure of the transition state for the reaction. This is often taken as a measure of how folded the transition state is relative to the folded state, the greater the effect on the observed rate the more unfolded is the transition state. The rate constants in table 5.2 have been plotted against GdnHCl activity and fitted to equation Eq. 5.2.1 to determine the GdnHCl dependence of the rate and the rate in the absence of denaturant (figure 5.5). In the reaction in 0.75 M GdnHCl very few data points define the appearance of dimer which may lead to an inaccurate calulation of the dimerisation rate. For this reason, the data points in figure 5.5 have been fitted including the 0.75 M data point and also excluding the 0.75 M data point. The fit of all the data points excluding the 0.75 M data provides an m-value of 11.9 ±1 .5 M^{-1} and a $k_{(w)}$ of 1.1×10^{-5} $M^{-1}s^{-1}$.
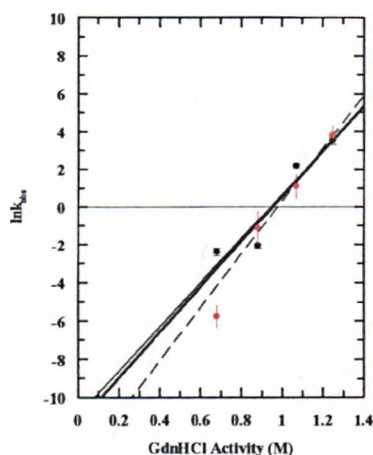
Figure 5.5 GdnHCl dependance of dimerisation rates. The black data points shown the dimerisation rate calculated from the decrease in monomer concentration, the thin black line represents the fit of these points to equation Eq. 5.2.1. Red data point show the dimerisation rates calculated from the increase in dimer concentration, the dashed line represent the fit of these points to equation Eq. 5.2.1. The heavy black line represents the fit of both sets of data points excluding the rate determined from the appearance of dimer at 0.75 M GdnHCl.

The m-values shown in table 5.2 are a measure of the level of folding in the transition state. If the transition state of the bimolecular reaction was as solvent exposed as the unfolded monomer then the m-value for the dimerisation reaction would be twice the m-value of the unfolding equilibrium constant, $2m_{(F/U)} = 15.4 \pm 1.4 \ M^{-1}$. Therefore, structure of the transition state of the bimolecular reaction is very close to the structure of the unfolded state and is more unfolded than the kinetic intermediate identified in the hCC folding pathway. This data reflects the high level of disruption of the monomeric structure that is required to form the domain swapped dimer.
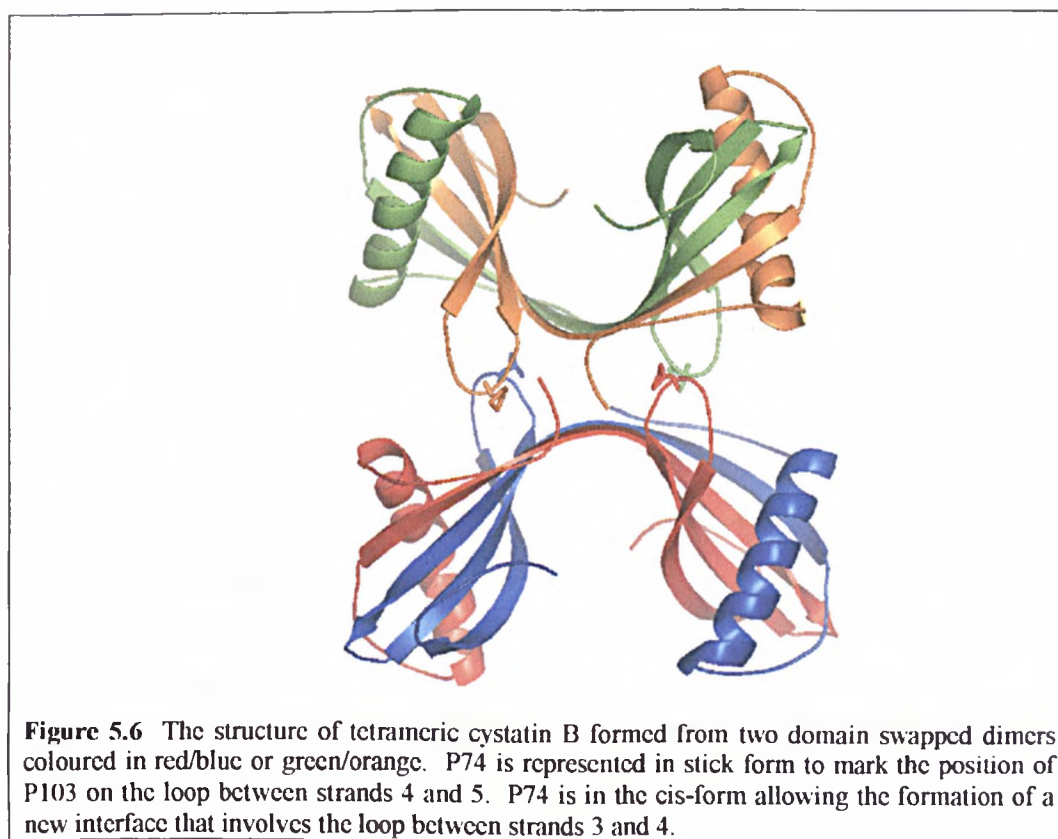
		Monomer decrease	Dimer formation
m-value	(M^{-1})	11.6 ± 2.7	16.4 ± 1.7
$Lnk_{(w)}$		-10.9 ± 2.7	-16.4 ± 1.7
$K_{(w)}$	$(M^{-1}s^{-1})$	1.8×10^{-5}	7.5×10^{-8}

Table 5.2 Summary of the parameters determined from the GdnHCl dependence of the bimolecular rate constants.

5.4.2 Proline isomerisation

Introduction to P103A cC

A recent crystal structure of tetrameric cystatin B (PDB ID: 2oct)[172] shows that it is assembled from two domain swapped dimers with P74 (equivalent to P103 in cC or P105 in hCC) in the *cis* conformation (figure 5.6). With P74 in the *cis* conformation a new interface is formed between the two dimers with interactions involving loop II and the N-terminus.[172] As shown in figure 5.10, this proline is widely conserved throughout the cystatins where it is found in the *trans* conformation in monomeric structures.

Figure 5.6 The structure of tetrameric cystatin B formed from two domain swapped dimers coloured in red/blue or green/orange. P74 is represented in stick form to mark the position of P103 on the loop between strands 4 and 5. P74 is in the cis-form allowing the formation of a new interface that involves the loop between strands 3 and 4.

m-value analysis of cC oligomerisation indicates that the transition state of tetramerisation is significantly more folded than the transition state of dimerisation. Under the conditions described above, tetramerisation of cC is a unimolecular process so is rate limited by a slow conformational rearrangement of the dimer rather than the association of assembly-competent dimers. It was originally proposed that the conversion of dimer to tetramer was rate limited by partial unfolding of the dimer that would allow the propagation of domain swapping.[136] However, in light of the recent tetramer structure, it is possible that
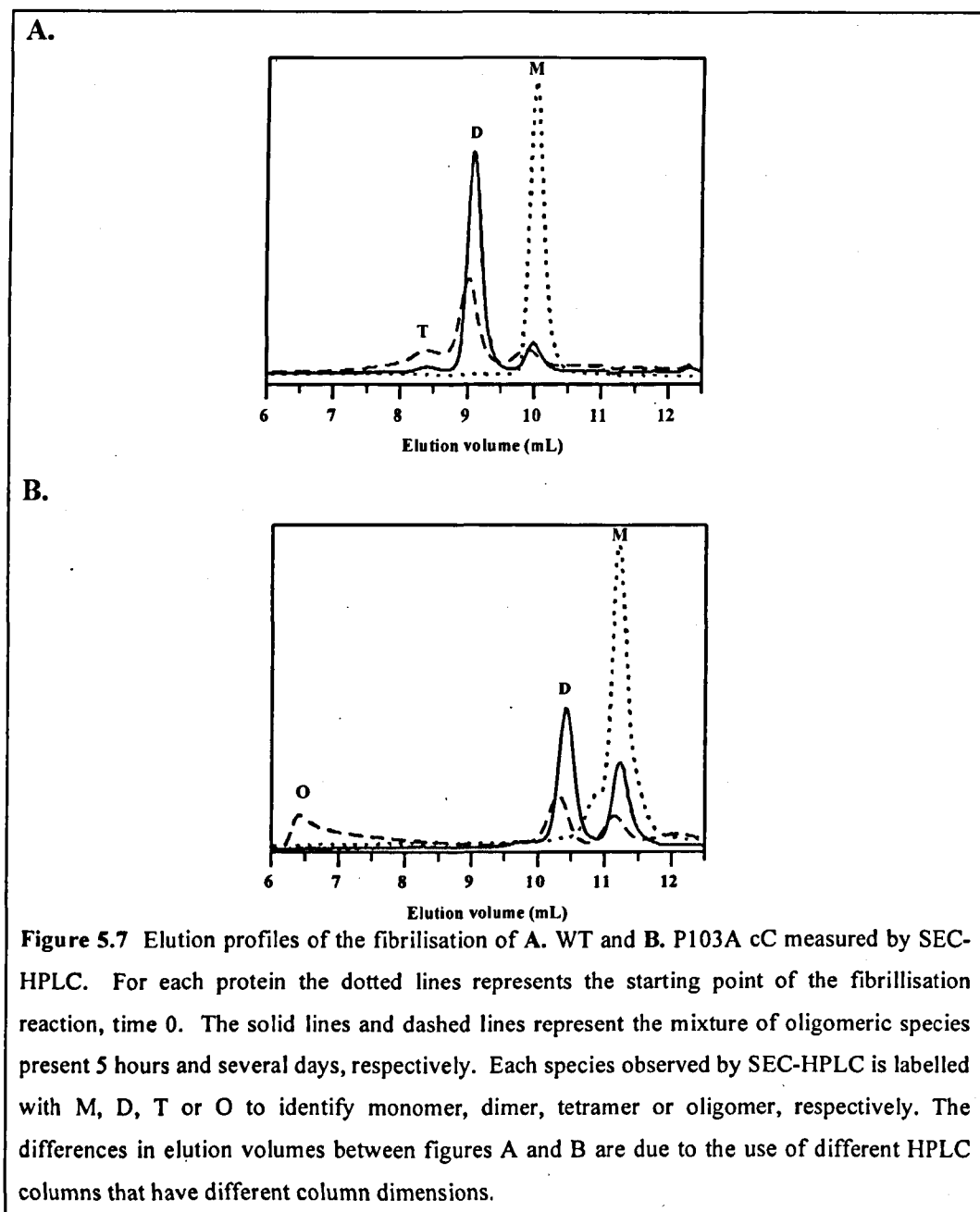
the rate limiting step in tetramerisation is the isomerisation of the conserved proline from the native *trans-* conformer to the *cis-* conformer.

SEC-HPLC analysis of the P103A cC fibrilisation reaction

In contrast to previously described experiments, analysis of the effect of mutating the conserved proline in the cystatins has been carried out using cC rather than hCC. Given the difficulties in producing recombinant hCC described in detail in Chapter 3, P103A cC was chosen as an alternative to P105A hCC. The initial stages of amyloid formation have been previously well characterised.[68] As is discussed later in this chapter there are many similarities in the initial stages of aggregation of cC and hCC.

A mutant of cC with the conserved proline, P103, mutated to an alanine has been studied in order to determine whether isomerisation of P103 is necessary for the formation of tetramer and amyloid. Although the *cis* orientation of the peptide bond is not exclusive to proline, for other residues it is so rarely populated that if necessary for tetramer or amyloid formation the rate of formation would be drastically reduced.
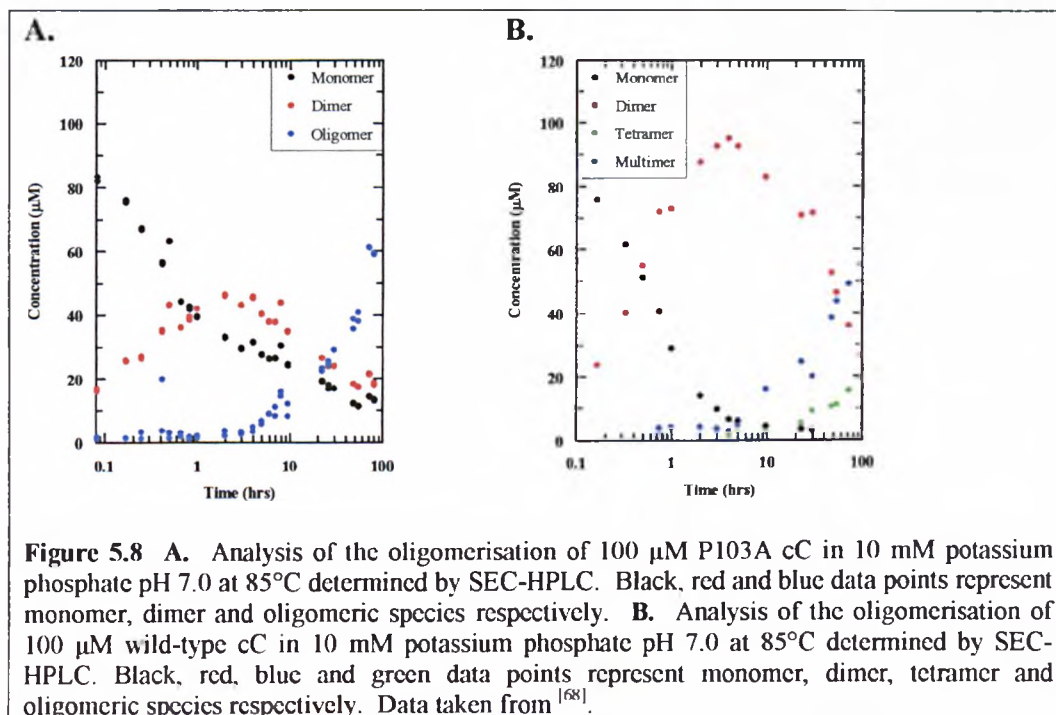
Differences in the fibrillisation reaction of wt and P103A cC are instantly seen by a comparison of the elution profiles of wt and P103A cC (figure 5.7). In contrast to wt cC, P103A cC does not form tetramers but forms a significant amount of resolvable aggregates. The size of these aggregates range from approximately 200 kDa to 1000 kDa, the upper limit of the column.

Figure 5.7 Elution profiles of the fibrilisation of A. WT and B. P103A cC measured by SEC-HPLC. For each protein the dotted lines represents the starting point of the fibrillisation reaction, time 0. The solid lines and dashed lines represent the mixture of oligomeric species present 5 hours and several days, respectively. Each species observed by SEC-HPLC is labelled with M, D, T or O to identify monomer, dimer, tetramer or oligomer, respectively. The differences in elution volumes between figures A and B are due to the use of different HPLC columns that have different column dimensions.

At each time interval the concentration of monomer, dimer and oligomer P103A cC has been determined from the peak area scaled to the peak area of 100 μM monomeric P103A cC at 25°C. In figure 5.8, the concentration of monomer, dimer and oligomer has been plotted as a function of time. For comparison the time course of wild-type cC under the same conditions is also shown in figure 5.8.

Once formed, the dimer population assembles to higher order species making fitting of the dimerisation time curves difficult. As dimers are formed from the aggregation of monomers the dimerisation rate can be determined by fitting the depletion of monomer rather than the appearance of dimer. Although

significantly less dimer is populated in the P103A reaction, the bimolecular rate constant for P103A at 85°C is 7.7 ± 0.7 $M^{-1}s^{-1}$. This is within error of the biomolecular rate constant determined for wt cC under these conditions (6.6 ± 0.7 $M^{-1}s^{-1}$).

Figure 5.8 **A.** Analysis of the oligomerisation of 100 μM P103A cC in 10 mM potassium phosphate pH 7.0 at 85°C determined by SEC-HPLC. Black, red and blue data points represent monomer, dimer and oligomeric species respectively. **B.** Analysis of the oligomerisation of 100 μM wild-type cC in 10 mM potassium phosphate pH 7.0 at 85°C determined by SEC-HPLC. Black, red, blue and green data points represent monomer, dimer, tetramer and oligomeric species respectively. Data taken from [68].

Electron microscopy analysis of the P103A fibrilisation reaction

At the end of the wild-type cC time course shown in figure 5.8B, the most predominant species are large aggregates that are unable to penetrate the matrix of the SEC column. These large aggregates have been visualised by EM and found to have the expected dimensions of amyloid (figure 5.9A). The morphology of the oligomeric species populated during the P103A cC time course is strikingly different to WT cC (figure 5.9B). There is no evidence of amyloid fibres, rather the grids are evenly distributed with short curly filaments with an average diameter of 7.85 ± 0.8 nm and no longer than 100 nm in length.
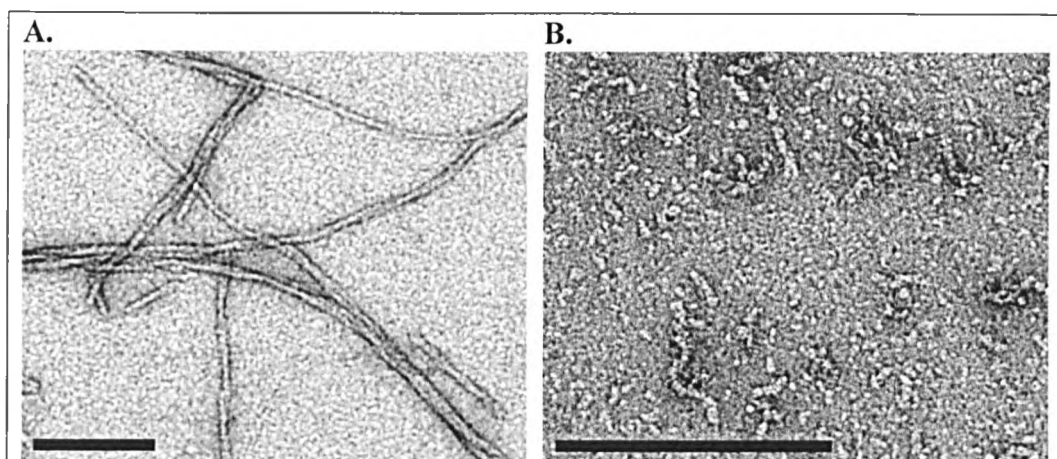
Figure 5.9 A. Amyloid fibres of WT cC present in the end point of the time course shown in figure 5.8B. Individual fibrils have average diameters of 11nm and measure several microns in length. **B.** In contrast to WT cC, only short, curly structures are observed in the end point of the time course shown in figure 5.8A. These species average 8 nm in diameter with a length no longer than 100 nm. Scale bars in both images are 200 nm.

5.5   DISCUSSION

5.5.1   Analysis of the early stages of amyloidogenesis of hCC

The initial process in the aggregation of hCC into amyloid is the formation of a domain-swapped dimer. The high GdnHCl dependence of dimerisation indicates that dimerisation occurs via a predominantly unfolded transition state. The dimerisation reaction can be considered to occur in two phases. The first phase is unimolecular as it involves the unfolding of monomeric hCC (M) into an aggregation competent species (M*), because it is a unimolecular process the rate at which M* is formed is independent of the protein concentration. The second phase is bimolecular as it requires the successful collision of two aggregation competent species resulting in the formation of dimer (D). The rate at which this process occurs is dependent on protein concentration as the aggregation competent species are more likely to collide at high protein concentration. ^{1}H-NMR spectra used to follow dimerisation at various concentrations of hCC (as described in section 5.2.1) suggest that dimerisation of hCC is concentration dependant at the concentration used for the experiment described in section 5.2.2. Therefore, the dimerisation of hCC described in section 5.2.2 is rate limited by the bimolecular phase of the reaction. The dimerisation process is described by the following scheme:

(Eq. 5.3.1)   2M $\rightleftharpoons$ 2M* $\rightleftharpoons$ 2D

If the rate at which this bimolecular process occurs becomes faster than the rate at which monomer unfolds to aggregation competent species then the dimerisation reaction becomes limited by a unimolecular transition. To estimate the monomer concentration at which this transition occurs the unfolding rate is used as an estimate of the rate the aggregation competent species is formed. In the case of hCC the concentration required for the bimolecular rate to be faster than the unimolecular rate is beyond the upper limits of the accessable protein concentration range.

hCC further aggregates into large oligomeric species that are too large to penetrate the matrix of the SEC column indicating that they have a molecular weight above 1000 kDa. Attempts to view the large oligomeric species by EM proved unsuccessful. The grids generally had very little material on them. What little was present was in the form of amorphous aggregate. Previously, cC amyloid fibres have also proved difficult to analyse by EM, with grids also having very little material or amorphous aggregate on them. Additional experiments, such as further attempts at EM analysis or staining with congo red or thioflavin T, are required to determine whether the large oligomeric species are amyloid.

5.5.2 Comparison between hCC and wild type cC aggregation

SEC-HPLC has been used to characterise the initial stages of cC amyloid formation, where not only monomer and dimer, but also tetramer can easily be resolved.[134] The dimerisation reation of wt cC shows many similarities to that of hCC, validating the use of cC as a model for hCC amyloidogenesis. The dimerisation of cC has been shown to be limited by a bimolecular transition over the range of experimentally accessible protein concentration. hCC has been estimated to be limited by a bimolecular transition of the experimentally accessible protein concentrations, but further study is required to confirm this.
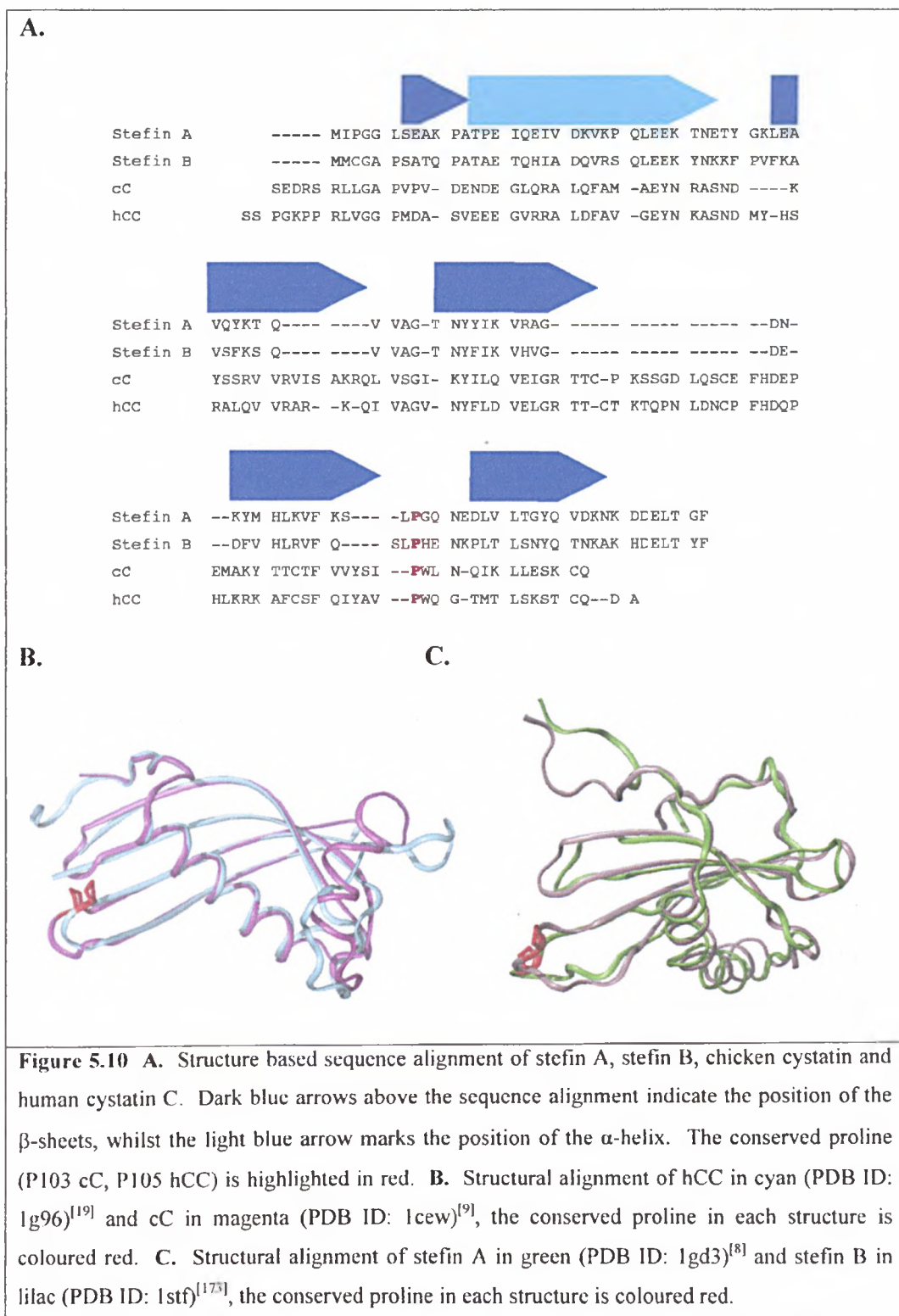
The m-value of the dimerisation reaction reflects the level of solvent exposure of hydrophobic groups in the transition state relative to the folded monomer. The m-value for the hCC and wt cC dimerisation reactions are within experimental error of each other (12 ± 1.5 M^{-1} and 10 ± 2 M^{-1}[134], respectively). The similarity of the m-value indicates that both proteins have the same level of solvent exposure in the transition state of the dimerisation reaction. For both hCC and cC the structure of

the transition state of dimerisation is closer to the unfolded state than the folded state, being less folded than the kinetic intermediate populated in the monomeric folding pathway.

Whilst there are many similarities in the dimerisation reaction of hCC and wt cC, SEC-HPLC analysis shows that wt cC populates a tetramer structure in addition to monomer and dimer. No tetramer structure is observed by SEC-HPLC for hCC. Kinetic data suggests that tetramer formation is possibly off-pathway to amyloid formation.[136] As no tetrameric species is observed in the early aggregation of hCC and the cystatins are likely to aggregate through a similar pathway, the hCC data supports a model of amyloidogenesis where tetramer is a non-obligatory intermediate.

5.5.2 Proline isomerisation

The tetramerisation and multimerisation of wt cC has been shown to be a first order process so is rate-limited by a unimolecular event. M-value analysis indicates that the transition state of both the tetramerisation and multimerisation reaction of wt cC is considerably more folded than the transition state of the dimerisation reaction. Whilst this clearly shows that a major unfolding event such as domain swapping is not involved in the tetramerisation and multimerisation processes, at least a partial unfolding event is required.[68] It is proposed that the first order event that limits tetramerisaton and multimerisation of cystatin is a partially unfolding event that allows isomeration of a proline that is conserved throughout the cystatins (see figure 5.10).

Figure 5.10 A. Structure based sequence alignment of stefin A, stefin B, chicken cystatin and human cystatin C. Dark blue arrows above the sequence alignment indicate the position of the β-sheets, whilst the light blue arrow marks the position of the α-helix. The conserved proline (P103 cC, P105 hCC) is highlighted in red. **B.** Structural alignment of hCC in cyan (PDB ID: 1g96)[19] and cC in magenta (PDB ID: 1cew)[9], the conserved proline in each structure is coloured red. **C.** Structural alignment of stefin A in green (PDB ID: 1gd3)[8] and stefin B in lilac (PDB ID: 1stf)[173], the conserved proline in each structure is coloured red.

Characterisation of the multimerisation pathway of P103A cC along with several pieces of data from the study of other cystatins has led to the development of a model of cystatin amyloid formation in which the rate limiting process is proline isomerisation.

No tetrameric P103A cC is observed by SEC-HPLC, which suggests that proline isomerisation is essential for the formation of tetramers. This is consistent with the published crystal structure of tetrameric stefin B in which P103 is in the *cis-* conformation rather than the *trans-* conformation that is observed in the monomer. For proline isomerization to occur a partial unfolding is required, this is consistent with the m-value analysis of the tetramerisation reaction where a relatively small unfolding event is required in the tetramerisation process.

In the developing cystatin amyloid model, the tetramer is an off-pathway intermediate in the formation of amyloid. This conclusion is supported data from several other cystatin experiments. Firstly, kinetic analysis of wt cC indicated that multimerisation is dependent on the concentration of dimer rather than tetramer.[136] This suggests that cystatin amyloid is formed through further aggregation of dimers, not the aggregation of tetramers. Secondly, no tetramer has been observed in the multimerisation of hCC although data suggests a common mechanism of cystatin amyloid formation. In addition to this, the β-strands between the dimers do not align to form a continuous β-sheet in the tetrameric structure. This prevents the tetramers stacking together into the cross-β structure that is observed in the amyloid fibre.

Although P103A cC forms oligomers, striking differences are observed in the oligomeric species populated by the P103A cC mutant and wt cC. SEC-HPLC resolves a population of broad range mid-molecular weight oligomers that are not observed in wt cC. EM analysis indicates that mid-molecular weight P103A cC oligomers are not intermediates in the formation of amyloid as no P103A cC amyloid is observed under conditions in which wt cC does form amyloid. These data suggest that proline isomerisation plays a role in the aggregation of dimers into multimers.

Hydrogen exchange measured on cystatin B shows that loop I is protected in the fibril. Therefore, it is likely that the domain swapped interface is maintained within the fibril.[174] This observation is consistent with the rate of fibrilisation being dependent on dimer concentration. However, the formation of cystatin amyloid by the aggregation of dimer is prevented by the inherent twist in the β-sheet of the dimer. Considerable flattening of the twist in the β-sheet is required

to form amyloid. The helix, which is protected in the monomer and dimer structure, shows no measurable protection in the fibril.[174] Figure 5.11A. shows a model of a dimer-like unit in which the domain-swapping interface is maintained and a β-sheet is formed from strands 2-5. If these dimer-like units are stacked together the result is a steric clash between the loop IIs of adjacent dimers, as is shown in figure 5.11B. This steric clash is alleviated if the P103 in loop II adopts the *cis*- conformation, as is shown in figure 5.11C..
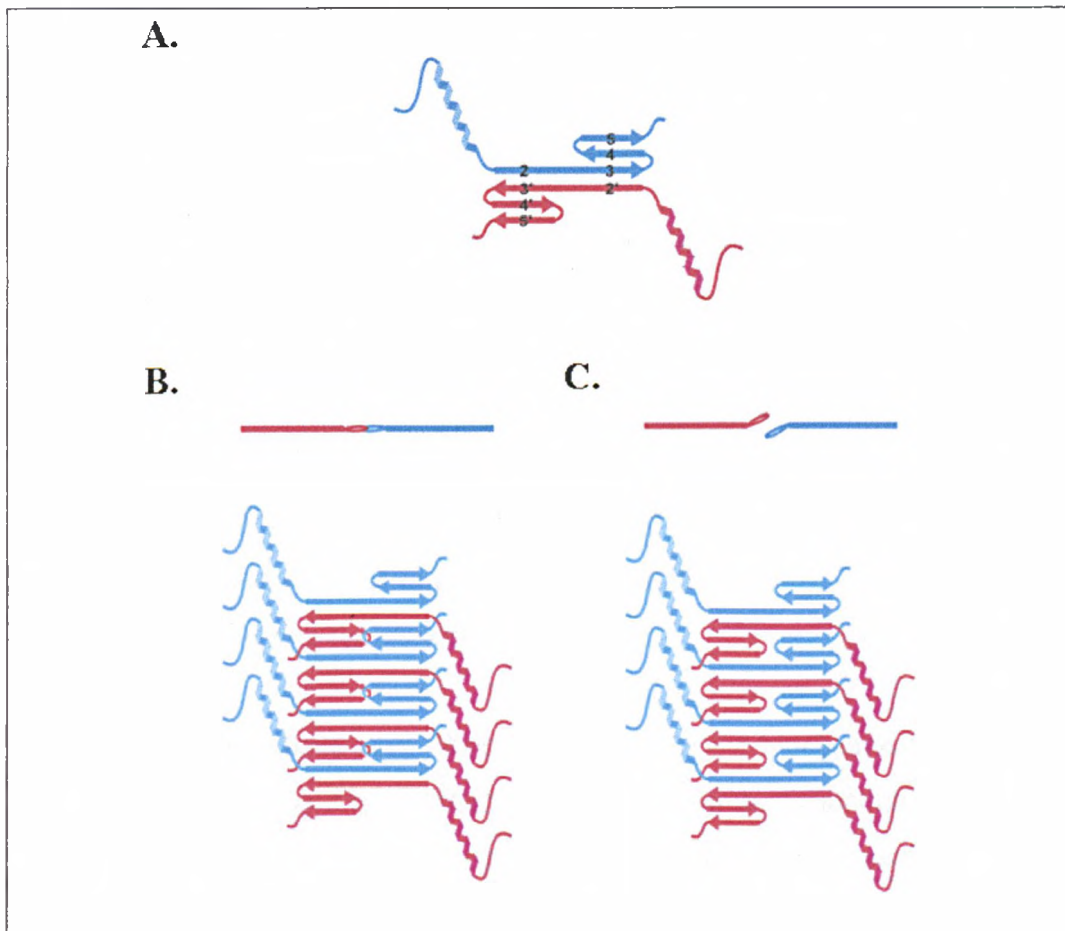
Figure 5.11 A. A model of a dimer-like stacking unit based on hydrogen exchange data.[174] The helicies are shown for clarity but their exact position has not been determined. Each cystatin is shown in either red or blue. The β-sheet is formed from strand 2 from one cystatin and strand 3, 4 and 5 of the second cystatin. **B.** Two perpendicular representations of the steric clash in between adjacent loop 2s if the stacking unit has the conserved proline in the *trans*-conformation. **C.** The steric clash represented in B is alleviated if the conserved proline is in the *cis*- conformation as this cause the position of loop 2 to move.

CHAPTER SIX

ANALYSIS OF THE INTERACTION BETWEEN HUMAN CYSTATIN C AND $A\beta_{1-40}$

6.1 INTRODUCTION

Several immunological observations and a genetic link between hCC and AD has led to the proposal that hCC may play a role in the pathology of AD and several other $A\beta$ amyloid diseases. These observations and the subsequent experiments to define the interaction between hCC and $A\beta$ are discussed briefly here, further details can be found in chapter 1.2.5.

Immunostaining has identified the co-localisation of $A\beta$ and hCC in several $A\beta$ amyloid diseases.[56-58] However, where hCC is the main amyloidogenic protein, no co-immunostaining could be detected.[55, 58] Although they co-localise, hCC is not an intrinsic part of the $A\beta$ amyloid fibril.[98] Rather, a distinct layering of hCC and $A\beta$ has been observed where an immunonegative core surrounded by a layer of hCC is covered with a further layer of $A\beta$.[112] Further evidence of a role of hCC in $A\beta$ amyloid diseases is that polymorphisms in the hCC gene have been identified as risk-factors for late-onset AD.[110, 111]

Several experiments have been devised to determine the biochemical interaction between hCC and $A\beta$ and what effect this may have on the formation of amyloid. Immunoprecipitation assays indicate that hCC binds to full length APP and secreted APP in the $A\beta$ region of APP. Analysis of the association between hCC and $A\beta$ using ELISA suggests a specific, saturable high affinity 1:1 binding between hCC and $A\beta$, with a dissociation constant in the nM range. A binding competition assay has identified that the amino-terminus of $A\beta$ binds to hCC.[113]

The association of hCC and $A\beta$ results in a concentration dependent inhibition of $A\beta$ fibril formation. Co-incubation of hCC with monomeric $A\beta$ significantly reduces formation of $A\beta$ oligomers and protofibrils. These observations suggest that hCC is playing a protective role against the formation of $A\beta$ amyloid.[113, 114] As hCC inhibits the formation of $A\beta$ amyloid, the interaction site is potentially a useful target in the development of drug therapies against $A\beta$ diseases.

Despite recent interest in this area of research, there is little structural information available regarding the formation of the complex between hCC and Aβ. NMR enables residue-specific information about protein structure, stability and interactions to be easily obtained. Therefore, it is an excellent tool for studying the interaction between hCC and Aβ.

6.2 MATERIALS AND METHODS

^{15}N-labelled hCC was expressed and purified according to the method described in section 3.4.2. The purity and monomeric state of samples was confirmed by SEC-HPLC (described in chapter 2.6.4) prior to initiating NMR experiments. All NMR spectra described in this chapter were recorded on a Bruker DRX spectrometer operating at 500 MHz, controlled using XWinNMR (Bruker). Spectra were processed using Felix 2004 (Accelrys). All heteronuclear single quantum coherence (HSQC) experiments were acquired using 1024 increments in the proton dimension and 512 increments in the nitrogen dimension. The spectral widths of the proton and nitrogen dimensions were 7507.5 Hz and 2128.4 Hz, respectively.

6.2.1 ^{1}H-^{15}N HSQC spectrum of hCC

A HSQC experiment measures the chemical shifts of the proton and nitrogen nuclei of every bonded ^{15}N-H pair. This is achieved by modulating each proton signal with the signal of the attached nitrogen. Processing software is used to deconvolute the two frequencies and generate a two dimensional plot with a peak for every amide at the intersection of the proton and nitrogen chemical shifts. The chemical shift of a nucleus is directly related to its chemical environment. Therefore, factors that alter the chemical environment of the amide can be detected by changes in the HSQC spectrum. Peaks may shift position, but can also broaden, change intensity or disappear completely. In order to make sense of any changes in the spectrum it is necessary to determine which amide is responsible for which peak in the spectrum in a process known as resonance assignment. An assignment has been published for 200 μM hCC in 50 mM sodium phosphate pH 6.0, recorded at 303°K.[25] A 500 μl sample of 200 μM hCC in 50 mM sodium phosphate pH 6.0 was prepared for NMR by the addition of 10% D$_2$O. A 1D ^{1}H-spectrum was recorded prior to recording the HSQC spectrum at 303°K.

6.2.2 Assignment of hCC

The backbone assignment of hCC under the desired experimental conditions, 15 mM tris-TFA, pH 7.5, 278°K, was determined by tracking changes in the HSQC spectra as conditions were gradually changed from the previously assigned conditions. The new assignment was determined in two-phases, first the pH was gradually increased, and then the temperature was gradually lowered, as decribed below.

pH titration

The sample prepared for section 6.2.1 was used in the pH titration. The sample was removed from the NMR tube so that the pH could be adjusted by the addition of aliquots of 0.5 M Na_2HPO_4. Spectra were recorded at pH 6.0, pH 6.5, pH 7.0 and pH 7.5. At each pH interval a 1D ^{1}H-spectrum was recorded prior to recording the HSQC spectra at 303°K.

Temperature titration

A 500 µl sample of 50 µM hCC in 15 mM tris-TFA, pH 7.5 was prepared for NMR by the addition of 10% D_2O. The initial spectrum was recorded once the sample was equilibrated to 303°K. The assignment of 200 µM hCC, 50 mM sodium phosphate pH 7.5, 303°K transferred well onto this spectrum. Following this, spectra were recorded as the temperature was lowered in 5°K intervals to 278°K. At each new temperature the sample was allowed to equilibrate to the lower temperature before spectra were recorded. At each temperature interval a 1D ^{1}H-spectrum was recorded prior to recording the HSCQ spectra.

6.2.3 Preparation of Aβ$_{1-40}$

Aβ$_{1-40}$ was purchased as a TFA salt (rPeptide) and prepared using an adaptation of the Teplow procedure.[175] Aβ$_{1-40}$ undergoes time and concentration dependent aggregation in the acetonitrile-water used in its preparation.[176] Therefore, it is important to dissagregate Aβ$_{1-40}$ and generate a monomeric, random coil structure when preparing Aβ samples. The Teplow procedure is designed to prevent the pH falling between pH 4 and pH 7, the pH range where Aβ$_{1-40}$ aggregation is maximised. This method was not suitable for preparing Aβ$_{1-40}$ in the desired

buffer condition for NMR so was adapted maintaining the principles of the Teplow method.

1 mg $A\beta_{1-40}$ was dissolved in HFIP, divided into 0.5 mg samples and sonicated for 15 minutes. Samples were freeze-dried with the help of Dr. A. Moir. $A\beta_{1-40}$ was resuspended in 375 µl 20 mM tris and the pH adjusted to pH 7.5 using TFA. Finally, the volume was made up to 500 µl with water to produce a 15 mM tris-TFA buffer, pH 7.5. The concentration of $A\beta_{1-40}$ was determined using the method described in chapter 2.6.2 and where necessary adjusted to 50 µM $A\beta_{1-40}$ by the addition of 15 mM tris-TFA, pH 7.5. The $A\beta_{1-40}$ sample was prepared for NMR by the addition of 10% D_2O. A 1D ^1H spectra was recorded at 278°K.

6.2.4 Titration of unlabelled $A\beta_{1-40}$ into ^{15}N-hCC

The sensitivity of the NMR spectra to changes in the chemical environment of a nuclei make it a useful method for mapping regions of proteins that are involved in binding to a ligand. The regions of hCC that interact with $A\beta_{1-40}$ have been determined by titrating $A\beta_{1-40}$ into a ^{15}N-labelled hCC sample. The hCC sample used in the temperature titration described 6.2.2 was subsequently used for the $A\beta_{1-40}$ titration described here. All spectra were recorded at 278°K. At each stage of the titration, 50 µl of 50 µM $A\beta_{1-40}$ prepared according to the method described in 6.2.3 was added to the NMR tube containing the hCC sample, and gently inverted to produce a homogenous sample. After reaching a 1:1 stoichiometry, an additional 100 µl of $A\beta_{1-40}$ was added to the hCC sample which equates to a 1:1.2 stoichiometry. At each stage a 1D ^1H spectrum and 9 HSQC spectra were recorded. Once processed, the 9 HSQC matrix files were combined linearly into a single matrix file. Spectra were aquired in this manner rather than as a single HSQC spectrum to ensure that any changes in the HSQC spectrum resulting from the addition of $A\beta_{1-40}$ were complete prior to a further addition of $A\beta_{1-40}$.
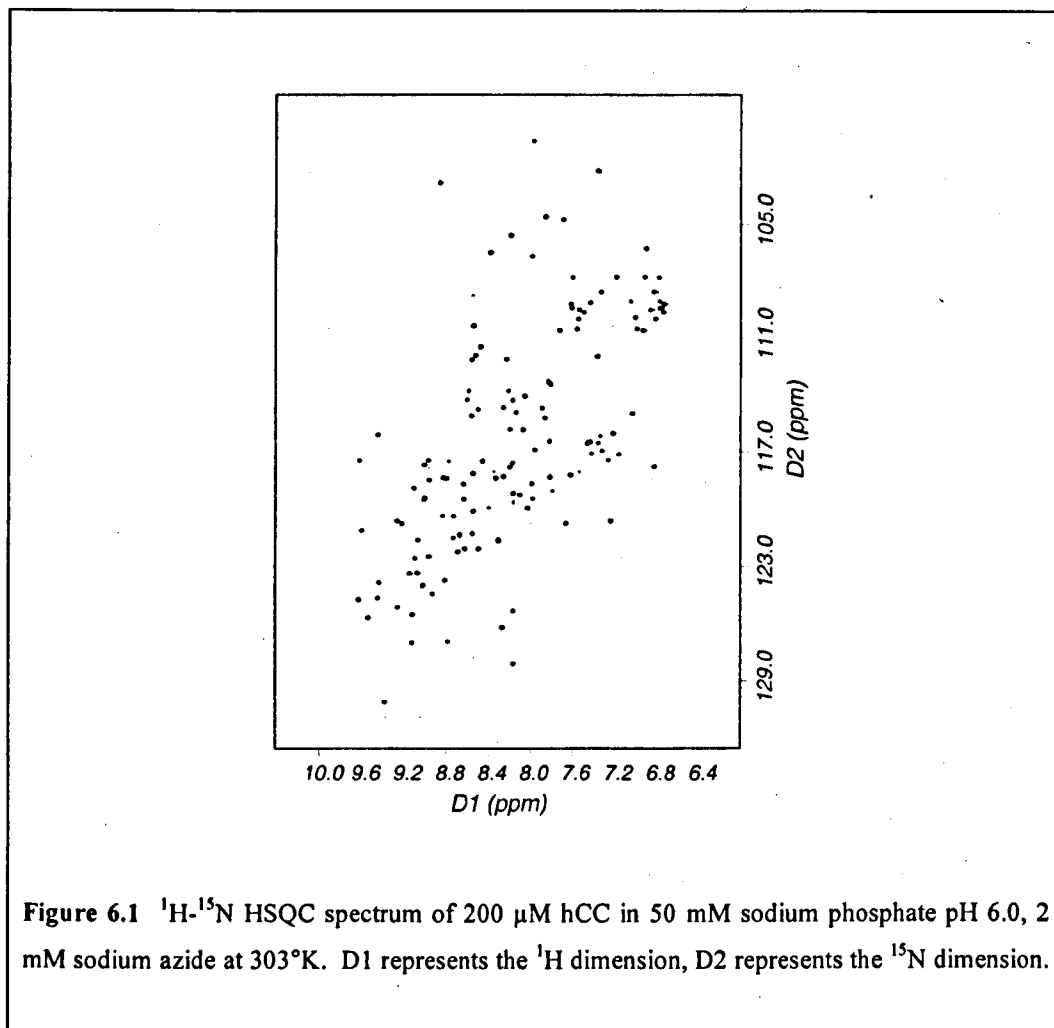
As a control, the experiment was repeated replacing the addition of $A\beta_{1-40}$ with 15 mM tris-TFA pH 7.5. Less data points were recorded as buffer was added in 100 µl aliquots rather than 50 µl aliquots. All other experimental details remained the same.

6.3 RESULTS

6.3.1 ^{1}H-^{15}N HSQC spectrum of hCC

The initial ^{1}H spectrum of 200 µM hCC in 50 mM sodium phosphate pH 6.0 at 303°K shows a wide dispersion of amide proton resonances (6-10 ppm) and upfield (below 0 ppm) aliphatic proton peaks indicating that hCC is folded. In addition to this, the wide dispersion of amide chemical shifts observed in HSQC spectra also indicate that hCC is folded under these conditions (see figure 6.1). The distribution of amide peaks in the HSQC spectrum aligns well with the published assignment of hCC under these conditions so that each amide in the published assignment could be assigned to a peak in the HSQC spectrum.

The reproducibility of the published data is a useful quality control measure in the preparation, recording and processing of the hCC NMR sample. Even relatively minor changes in the chemical environment of the amides are reflected in the HSQC spectra. Therefore, the similarity of the two spectra indicates that the two hCC samples are in an almost identical chemical environment despite being prepared in separate laboratories using alternative methods. It is an additional indication of the quality of hCC prepared using the new method described in chapter 3. This also gives confidence in the published assignment, which is an essential element to nearly all the conclusions presented in this chapter.

Figure 6.1 ^{1}H-^{15}N HSQC spectrum of 200 µM hCC in 50 mM sodium phosphate pH 6.0, 2 mM sodium azide at 303°K. D1 represents the ^{1}H dimension, D2 represents the ^{15}N dimension.

6.3.2   Assignment of 50 µM hCC in 15 mM tris-TFA pH 7.5, 278°K

Aβ_{1-40} has been described as "the peptide from hell" because of the difficulty in preparing and maintaining a homogenous, un-aggregated sample.[177] Buffer type, ionic strength, pH, temperature and peptide concentration are all factors that will influence the ability of Aβ to aggregate. Therefore, they must be considered when determining the conditions of an experiment in which Aβ is involved.[177] Unfortunately, the experimental conditions in which hCC was assigned previously[25] are conditions in which Aβ is highly prone to aggregate. A two-phased approach was used to assign hCC in conditions in which Aβ is stable, 15 mM tris-TFA pH 7.5, 303°K.

pH

In the first phase of the assignment, changes in the HSQC spectrum were successfully tracked as the pH was gradually adjusted from pH 6.0 to pH 7.5. Figure 6.2 shows the change in chemical shift of each amide in the HSQC

spectrum. The majority of peaks can be tracked from their position at pH 6 to their position at pH 7.5. However, a few peaks fade or completely disappear during the course of the titration. This is seen more clearly in figure 6.3, which shows the change in intensity of each assigned amide. The majority of residues change very little in intensity, the gentle decrease that is observed can be attributed to the minor dilution factor caused by adjusting the pH of the sample. However, several residues show a marked decrease in intensity as the pH is lower. For example, residues N39, R93 and C117 completely disappear whereas residues in the N-terminal region show a significant reduction in intensity, but are still present in the HSQC spectrum. The peaks of residues H43 and V60 and K75 and N82 merge so they are indistinguishable in the HSQC spectrum.

Figure 6.2 The associated chemical shift changes observed in the ^{1}H-^{15}N HSQC spectra of hCC as the pH is adjusted from pH 6.0 by half pH unit steps (blue positions) to pH 7.5 (black positions). The spectra were recorded at 303°K using 200 μM hCC in 50 mM sodium phosphate, 2 mM sodium azide. D1 represents the ^{1}H dimension, D2 represents the ^{15}N dimension.
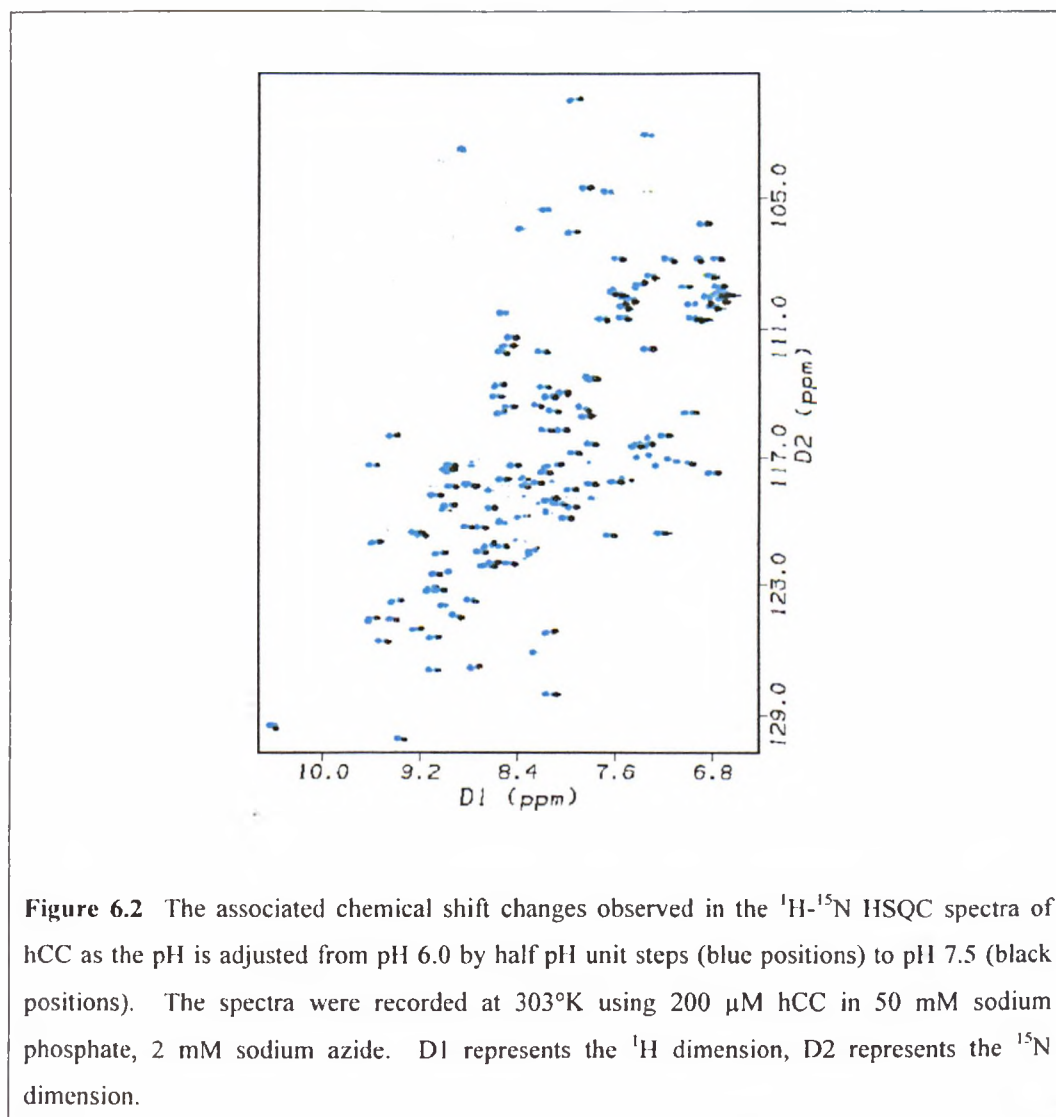
The amide hydrogen in each amino acid residue can exchange with hydrogen atoms in the solvent by a base catalysed mechanism. The rate of exchange

depends on several factors including temperature and the composition and pH of the solvent. Amides in a protein are often found to exchange at different rates due to the protection caused by secondary and tertiary structure. Hydrogen bonds within the protein significantly reduce the exchange rate and steric effects reduce solvent access to the core of the protein.

The amide cross peaks that show a significant decrease in intensity during the pH titration can be mapped to regions of hCC that are solvent exposed. Residues that are hydrogen bonded, and therefore protected against solvent exchange, show very little intensity decrease (coloured blue in Figure 6.3). This evidence strongly suggests that the observed changes during the pH titration are due to solvent exchange.
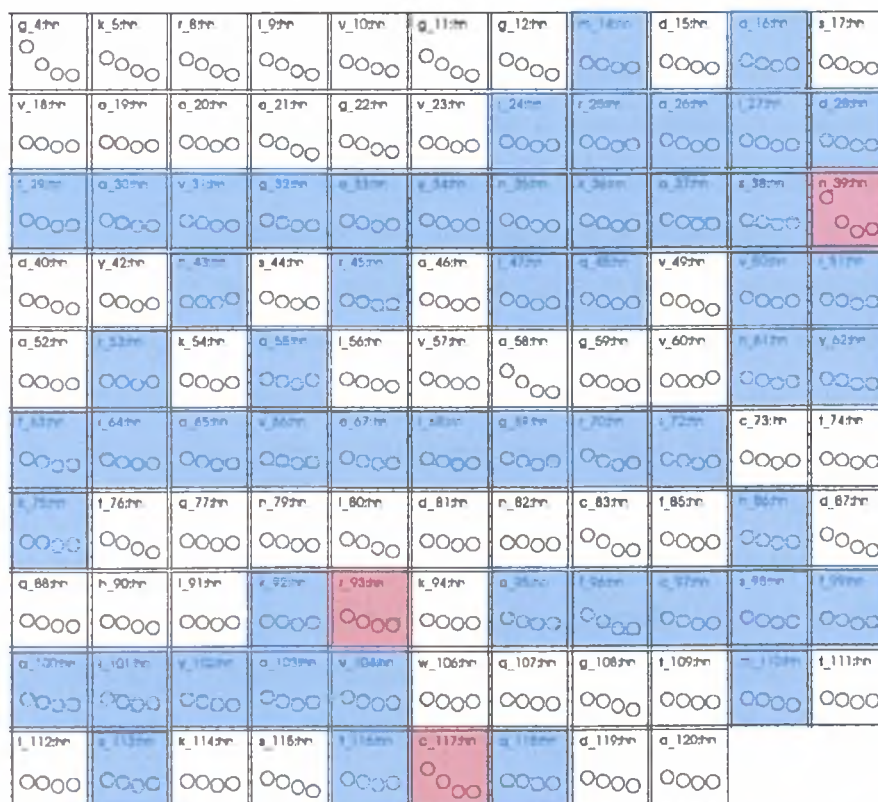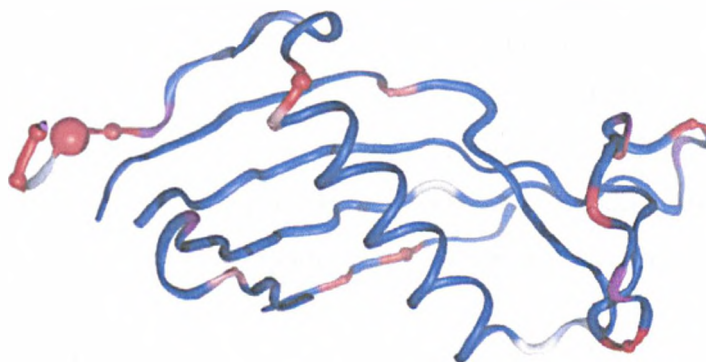
Figure 6.3 A. The intensity change observed in the ^{1}H-^{15}N HSQC experiment of hCC as the pH is adjusted from pH 6.0 by half pH unit steps to pH 7.5.   The upper and lower limit of the x-axis is pH 8.0 and pH 5.5, respectively. The upper and lower limit of the y-axis is the relative intensity of 5 and -1 respectively. The squares coloured in blue indicate hydrogen bonded amides observed in the dimeric hCC structure (PDB ID: 1g96)[19]. The squares coloured red indicate residues where signal intensity is lost during the titration.   B.   A structural representation of hCC (based on the domain swapped dimer of hCC (PDB ID: 1g96)[19]) showing the residues that change intensity as pH is adjusted to pH 7.5. A58, which is present in the loop between strands 2 and 3, is not shown in this representation because of its position of the domain swapped dimer. Blue colouring represents areas of little intensity change, through white to red, which shows areas of strong intensity change. The spheres represent the relative size of the intensity change. Purple residues represent proline residues.

Proteins in solution do not have a fixed, rigid structure, rather they are dynamic molecules that can adopt a number of different conformations. The probability of a particular structural state being populated is dependant on the stability of that conformation. In NMR, data is recorded on a vast number of molecules, and because of this, the spectra can give an indication of the state of this population. If two states are in fast exchange then a single peak is observed at a position that reflects an average of the two conformations. This is because the exchange is so rapid a group flipping between two states will experience both conformations many times in the spectral time scale (i.e. the time when the FID is being recorded). In contrast, when the exchange rate is slow two discrete peaks are observed. Two peaks are observed because in the time scale of the spectra the group in exchange does not have time to move between states. However, if two states exchange at an intermediate rate peaks generally broaden or disappear. Therefore, if the pH titration caused a shift in the conformational exchange rate between different states it could produce a similar result to that seen in Figure 6.3.

Either solvent exchange or conformational exchange, or a combination of both, could cause the changes observed in the HSQC spectra. However, the behaviour of the four disulphide bonded cysteine residues suggests that solvent exchange is the major cause of the observed changes. The two cysteines in each disulphide bond, C73-C83 and C97-C117, behave quite differently to each other. Given that there is a physical link between the two residues it seems unlikely that a conformational change would occur that affects each cysteine substantially differently.

Temperature

The assigned peaks in the 200 μM hCC, 50 mM sodium phosphate pH 7.5, 303°K spectrum were all identified in the 50 μM hCC in 15 mM tris-TFA pH 7.5, 303°K spectrum. Only minor chemical shift changes were observed between the two spectra, and these can be attributed to the difference in solvent and protein concentration.

The second phase of the assignment titration was to follow chemical shift changes as the temperature was lowered from 303°K to 278°K. Figure 6.4 shows the temperature induced change in chemical shift of each amide in the HSQC

experiments. Clearly, the assignment of the majority of peaks can be determined at low temperature, although some residues are lost during the temperature titration. For example, the peaks of residues S44 and T111 completely dissapear from the spectrum whereas the peaks of residues K5 and V10 and N35 and D81 merge so they are indistinguishable from each other.

The rate at which a protein tumbles in solution is proportional to the temperature of the sample. As temperature is reduces the tumbling rate slows due to the reduced energy available in the system. The rate at which the NMR signal decays reflects the rate at which the protein tumbles, the slower the tumbling the faster the decay rate of the NMR signal. The result is general broadening of the signal as the temperature is reduced.

Figure 6.4 The chemical shift changes observed in the ^{1}H-^{15}N HSQC spectra of hCC as the temperature is lowered from 303°K (initial blue position) to 278°K (black position) in 5°K steps. The spectra were recorded using 50 μM hCC in 15 mM tris-TFA pH 7.5, 2 mM sodium azide. D1 represents the ^{1}H dimension, D2 represents the ^{15}N dimension.
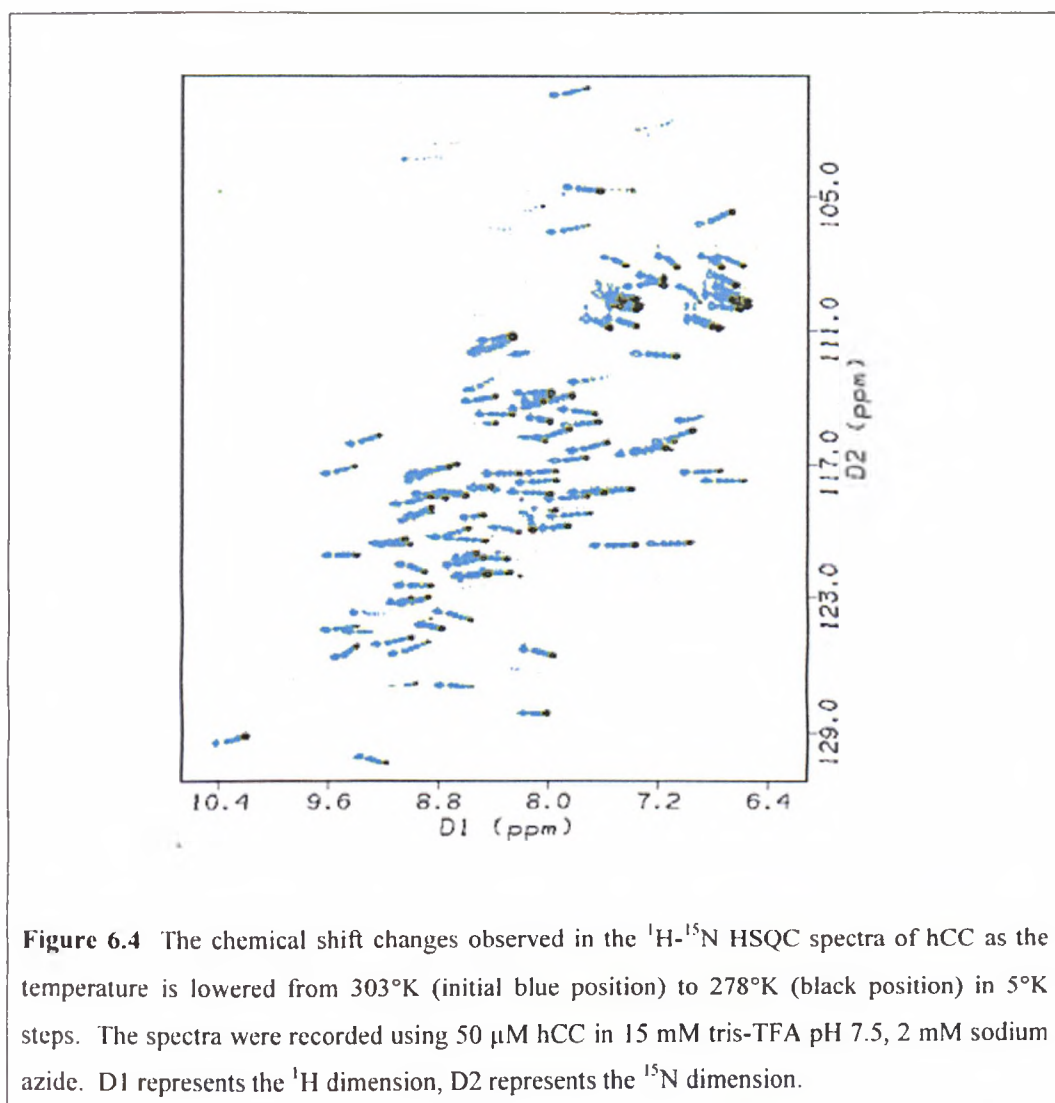
Figure 6.5 shows the temperature induced changes in intensity for each assigned cross peak. The intensity of some residues appears relatively unaffected by the change in temperature. For example, many residues in the N-terminal region, S17, G22, I56 and L80 show very little change in intensity despite changing the temperature. This is somewhat unexpected as lowering temperature should result in a decrease in intensity. The regions that show very little intensity change can again be mapped to areas of hCC that are solvent exposed. Lowering temperature reduces the rate of solvent exchange, which should be reflected in an increase in intensity. As two factors are present, one that is lowering intensity and one that is increasing intensity, the two factors counteract each other and the result is only a very small change in intensity. Mobile residues may also be affected only a little by cooling.

A.
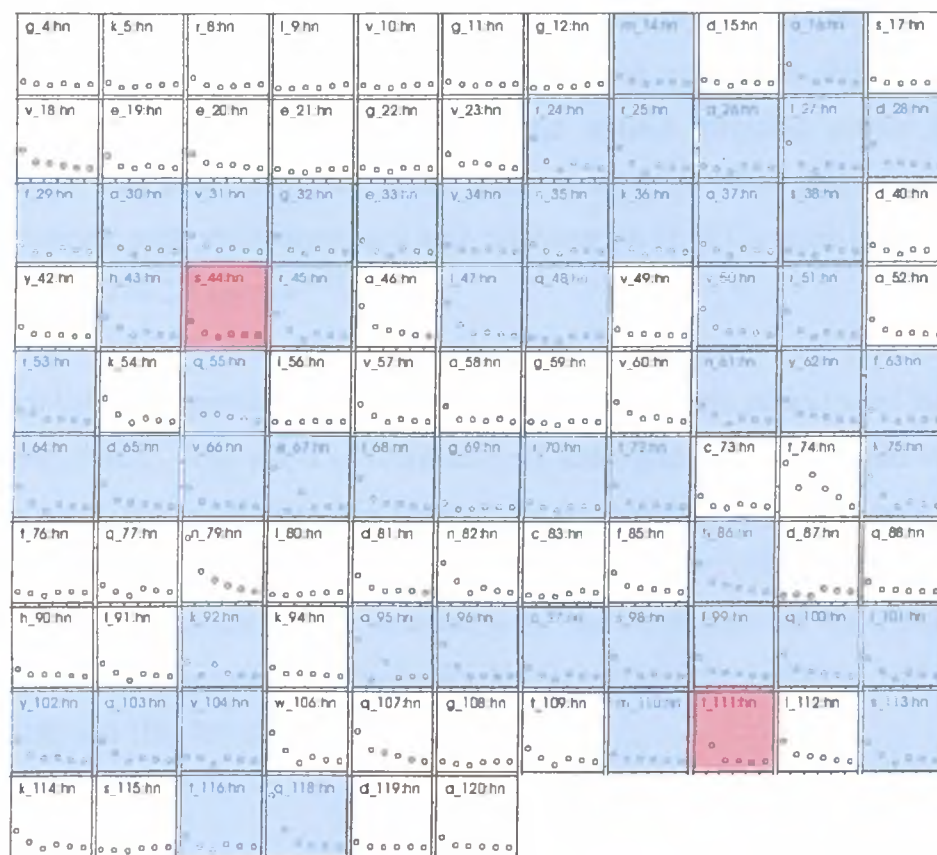
B.
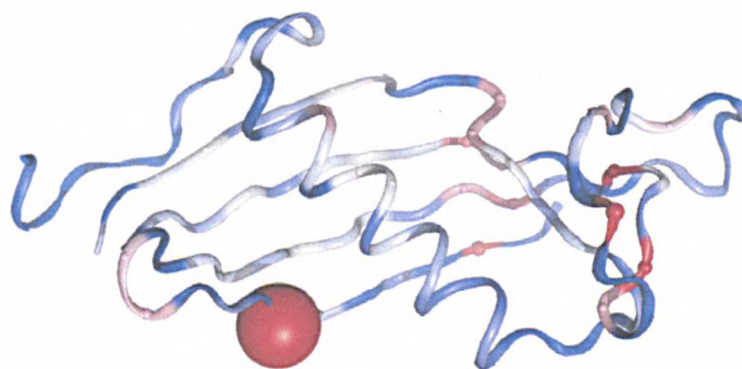
Figure 6.5 A. The intensity change observed in the ^{1}H-^{15}N HSQC spectra of hCC as the temperature is lowered in 5°K steps from 303°K to 278K. The upper and lower limit of the x-axis is 308°K and 273°K, respectively. The upper and lower limit of the y-axis is the relative intensity of 10 and -1 respectively. The squares coloured in blue indicate hydrogen bonded amides observed in the dimeric hCC structure (PDB ID: 1g96)[19]. The red squares represent residues where signal intensity is lost during the titration. B. A structural representation of hCC (based on the domain swapped dimer of hCC (PDB ID: 1g96)[19]) showing the residues that change intensity as temperature is reduced. A58, which is present in the loop between strands 2 and 3, is not shown in this representation because of its position of the domain swapped dimer. Blue colouring represents areas of little intensity change, through white to red, which shows areas of strong intensity change. The spheres represent the relative size of the intensity change.

Assignment of 50 µM hCC, 15 mM tris-TFA pH 7.5, 278°K

The HSQC spectrum of 50 µM hCC, 15 mM tris-TFA pH 7.5, 278°K is shown in figure 6.6. The ^{1}H spectrum recorded under these conditions shows a wide dispersion of amide proton resonances and upfield aliphatic proton peaks indicating that hCC is folded under these conditions. No evidence of the characteristic peak shifts associated with dimerisation of hCC can be identified in either the 1D or 2D spectra.[25]

The published assignment of hCC identifies 110 amide cross peaks out of the 120 residues of hCC. The effect of increasing pH and decreasing temperature on the amide cross peaks of hCC has resulted in disappearance of cross peaks for residues G4, G11, N39, M41, A58, C83, R93, T111 and C117. Further to this, several peaks that are clearly resolvable in the conditions of the published assignment are now overlapped including K5/V10, N35/D81, H43/V60, E67/A95, K75/N82 and H90/M110.

Figure 6.6 ^{1}H-^{15}N HSQC spectrum of 50 μM hCC in 15 mM tris-TFA pH 7.5, 2 mM sodium azide at 278°K. D1 represents the ^{1}H dimension, D2 represents the ^{15}N dimension.
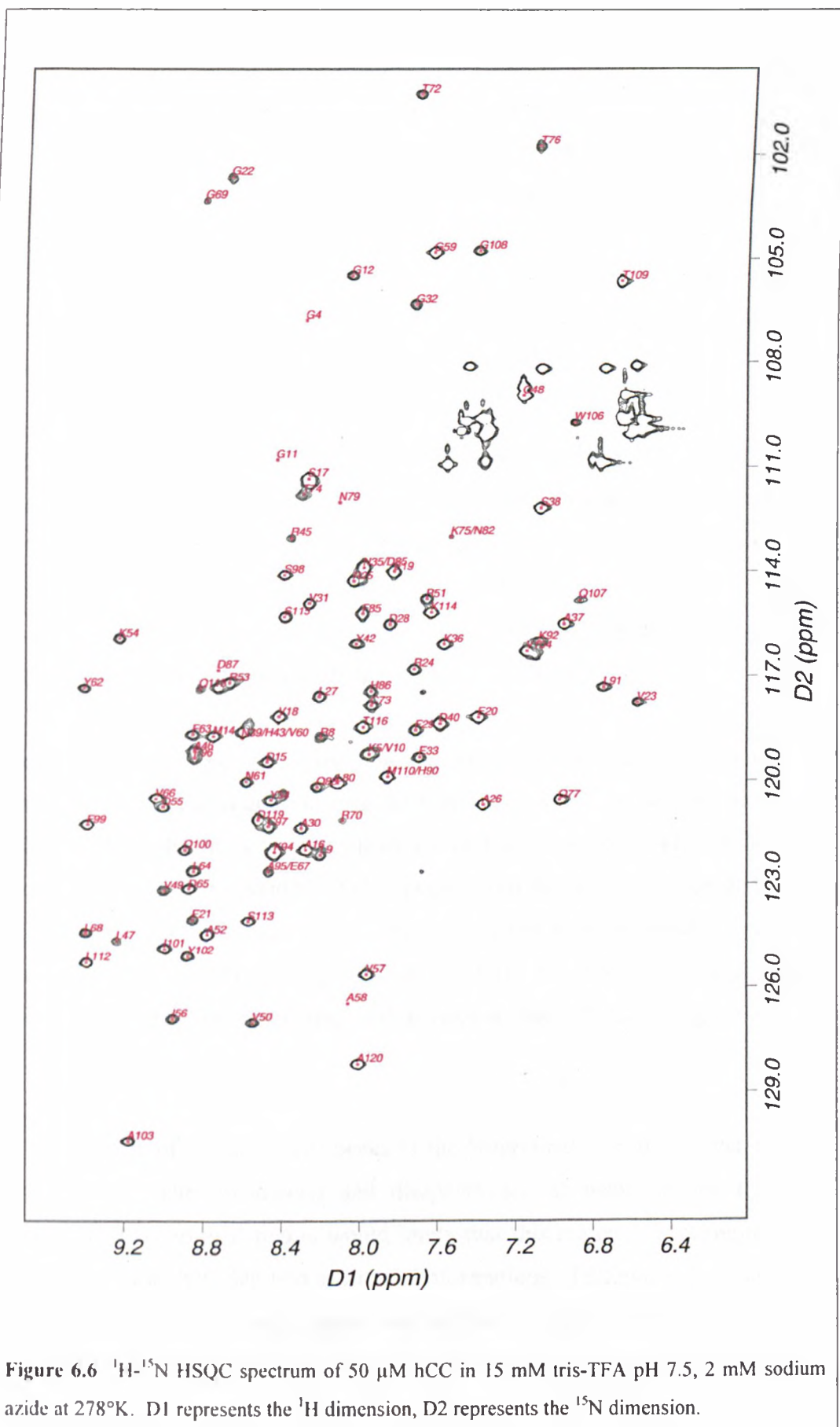
6.3.3 Titration with Aβ$_{1-40}$

Figure 6.7A shows, somewhat surprisingly, there are only very minor chemical shift changes observed in the titration of Aβ$_{1-40}$ into hCC as described in section 6.2.4. It is remarkable that despite the addition of beyond a 1:1 equivalence, there appears to be virtually no difference in the chemical environment of each amide of hCC. The few changes that are observed are due to the emergence of peaks that occurs as Aβ$_{1-40}$ is titrated into the sample. Figure 6.7B shows the chemical shift changes of the control experiment in which buffer is titrated into the hCC sample.

Figure 6.8 shows the intensity change of each amide in the Aβ titration and in the control experiment. The majority of the hCC amides show no change in intensity with the exception of the N-terminal region and A58. There is some indication of peaks emerging that representing the amides of M41 and C83 although is difficult to distinguish from the background noise of the experiment. The control experiment shows no intensity changes in any of the amides.

Figure 6.9 shows the individual amide peak for each residue that shows an increased intensity change following the addition of Aβ$_{1-40}$. In addition the amide peak of E20 is shown as an example of a peak that shows no change in intensity. For each residue the expanded HSQC spectra from the Aβ$_{1-40}$ titration and buffer titration are shown. Amide signals from G4, G11 and A58 are not observed in the absence of Aβ$_{1-40}$. Amide signals from K5/V10, R8, L9, and G12 appear as relatively broad peaks that sharpen and increase in intensity following the addition of Aβ$_{1-40}$.

In the absence of Aβ$_{1-40}$, amide peaks of the N-terminal region are broadened or not present. The broadening and disappearance of peaks in the relatively unstructured N-terminal region would imply that this region is exchanging at an intermediate rate between two or more conformations. Following the addition of Aβ$_{1-40}$, these peaks gradually appear and sharpen. This indicates that Aβ$_{1-40}$ has influenced the ability of hCC to change between these conformations.
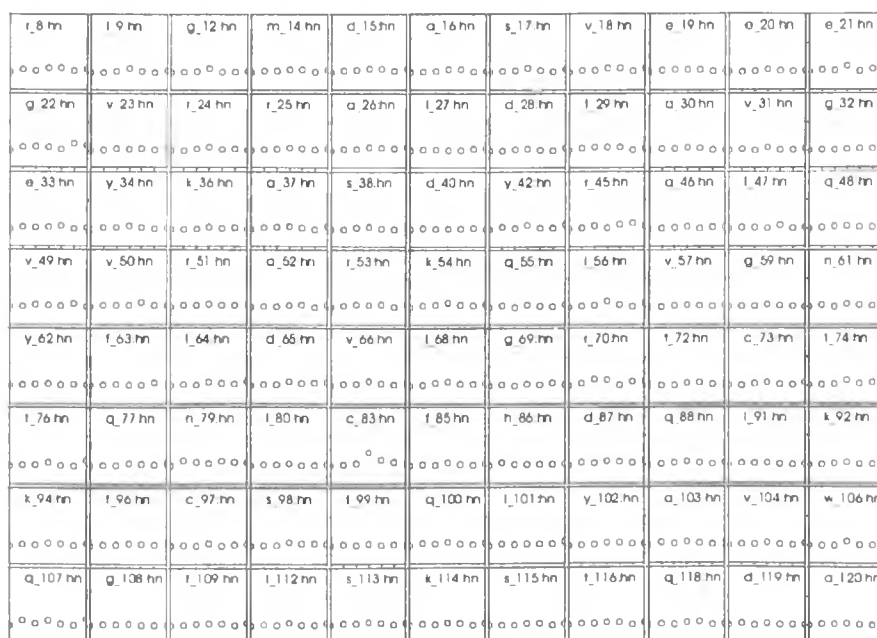
A.
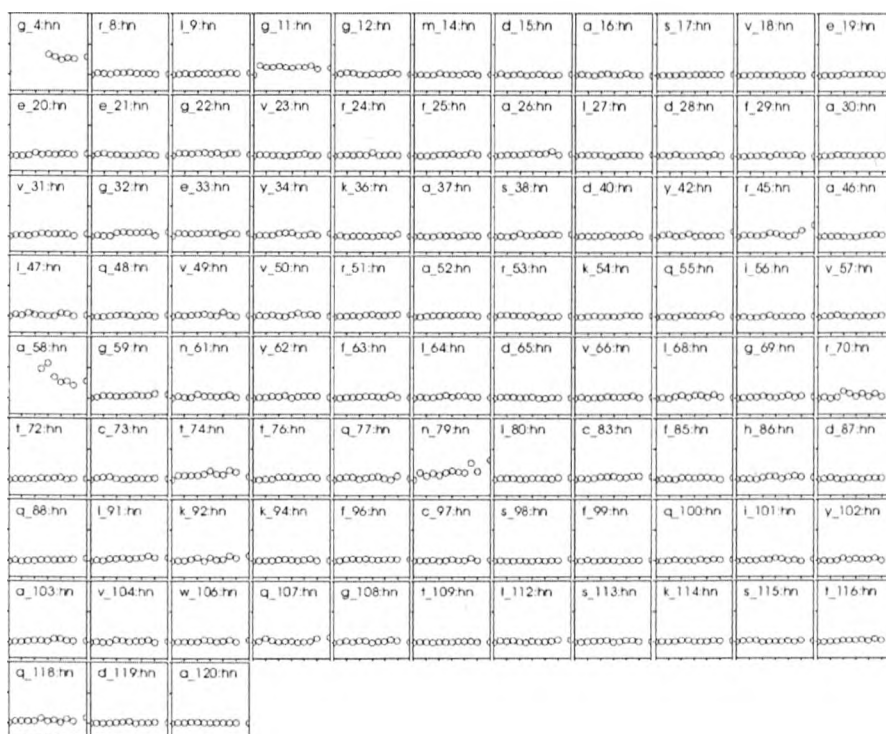
B.

Figure 6.7 The changes in chemical shift observed in the ^{1}H-^{15}N HSQC spectra of 50 µM hCC in 15 mM tris-TFA pH 7.5, 2 mM sodium azide at 278°K following the addition of Aβ$_{1-40}$ (**A.**) or buffer (**B.**). The x-axis shows the volume of Aβ$_{1-40}$/buffer added ranging from 0 to 600 µl. The y axis shows the chemical shift change ranging from -0.01 to 0.1. To account for the dilution factor in the titration and necessary changes in receiver gain during the titration, data have been normalised against residue A120.
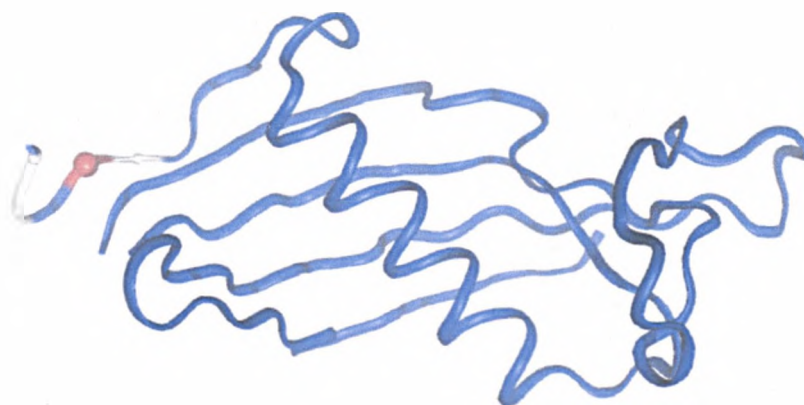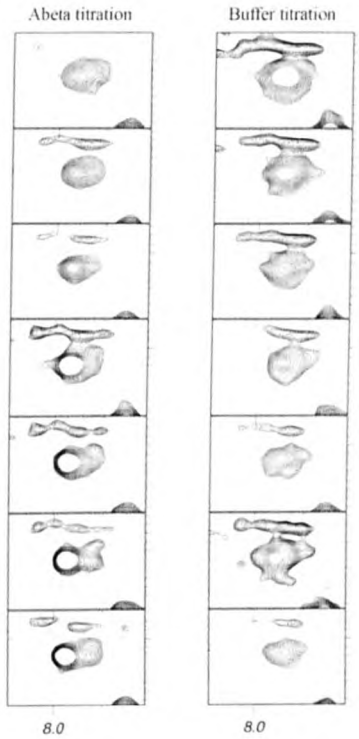
Figure 6.8 The changes in intensity observed for each amide in the ^{1}H-^{15}N HSQC spectra of 50 µM hCC in 15 mM tris-TFA pH 7.5, 2 mM sodium azide at 278°K following the addition of Aβ$_{1-40}$ (**A.**) or buffer (**B.**). The x-axis shows the volume of Aβ$_{1-40}$/buffer added ranging from 0 to 600 µl. The y axis shows the relative intensity ranging from 0 to 10. To account for the dilution factor in the titration and necessary changes in receiver gain during the titration, data have been normalised against residue A120. **C.** A structural representation of hCC (based on the domain swapped dimer of hCC (PDB ID: 1g96)[19]) showing the residues that change intensity following the addition of Aβ$_{1-40}$. A58, which is present in the loop between strands 2 and 3, is not shown in this representation because of its position of the domain swapped dimer. Blue colouring represents areas of little intensity change, through white to red, which shows areas of strong intensity change. The spheres represent the relative size of the intensity change. Purple residues represent proline residues.

Figure 6.9 Figures **A-G** track the changes of residues that show a difference between the $A\beta_{1-40}$ titration and the control buffer titration. Figure **H.** is included as an example of a residue that shows no significant difference between the $A\beta_{1-40}$ titration and the control titration. At the top of each column is the HSQC of 50 μM hCC in 15 mM tris-TFA pH 7.5, 2 mM sodium azide at 278°K, descending down the column each HSCQ is recorded following the addition of a further 100 μl of $A\beta_{1-40}$/buffer. **A.-F.** represent G4, K5/V10, R8, L9, G11, G12, A58 and E20, respectively.
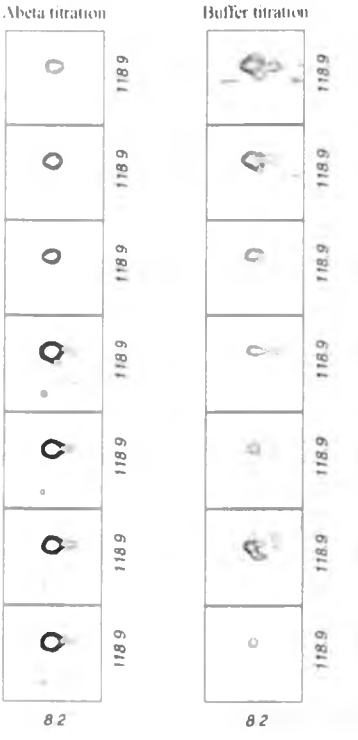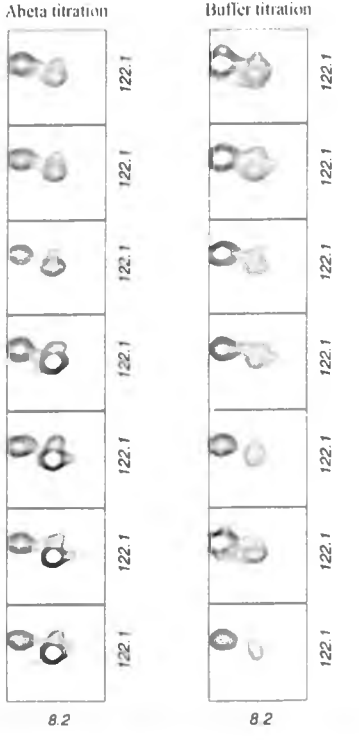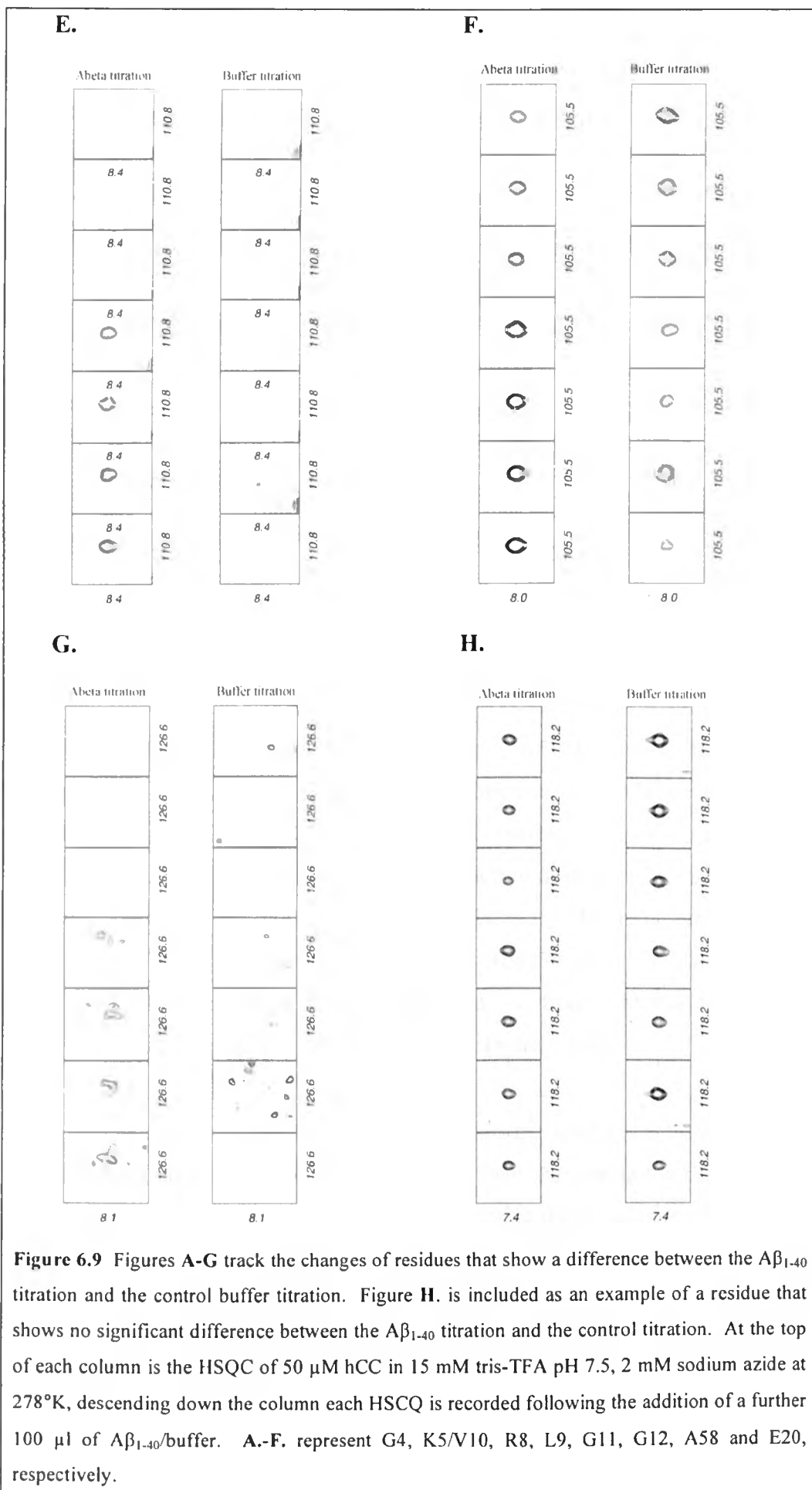
6.4 DISCUSSION

The data obtained from the $A\beta_{1-40}$ titration has been used to build a model describing the interaction between hCC and $A\beta_{1-40}$. The broadened or missing peaks of amides in the N-terminal region and residue A58 suggests that in the absence of $A\beta_{1-40}$ these residues populate at least two conformations. The simplest scenario is represented in Figure 6.10A. Here, two conformations can be populated, one in which the N-terminal region interacts with A58 and the other in which it does not interact. If the two conformations are in intermediate exchange then the amide peaks for each residue will broaden or disappear.

In this model, as $A\beta_{1-40}$ is titrated into the hCC sample it prevents the interaction between the N-terminal region of hCC and A58, thus favouring the alternative conformation. As the N-terminus can no longer interact with A58, the effects caused by intermediate exchange are removed. The result of this would be a sharpening of amide peaks as is observed in the titration data.
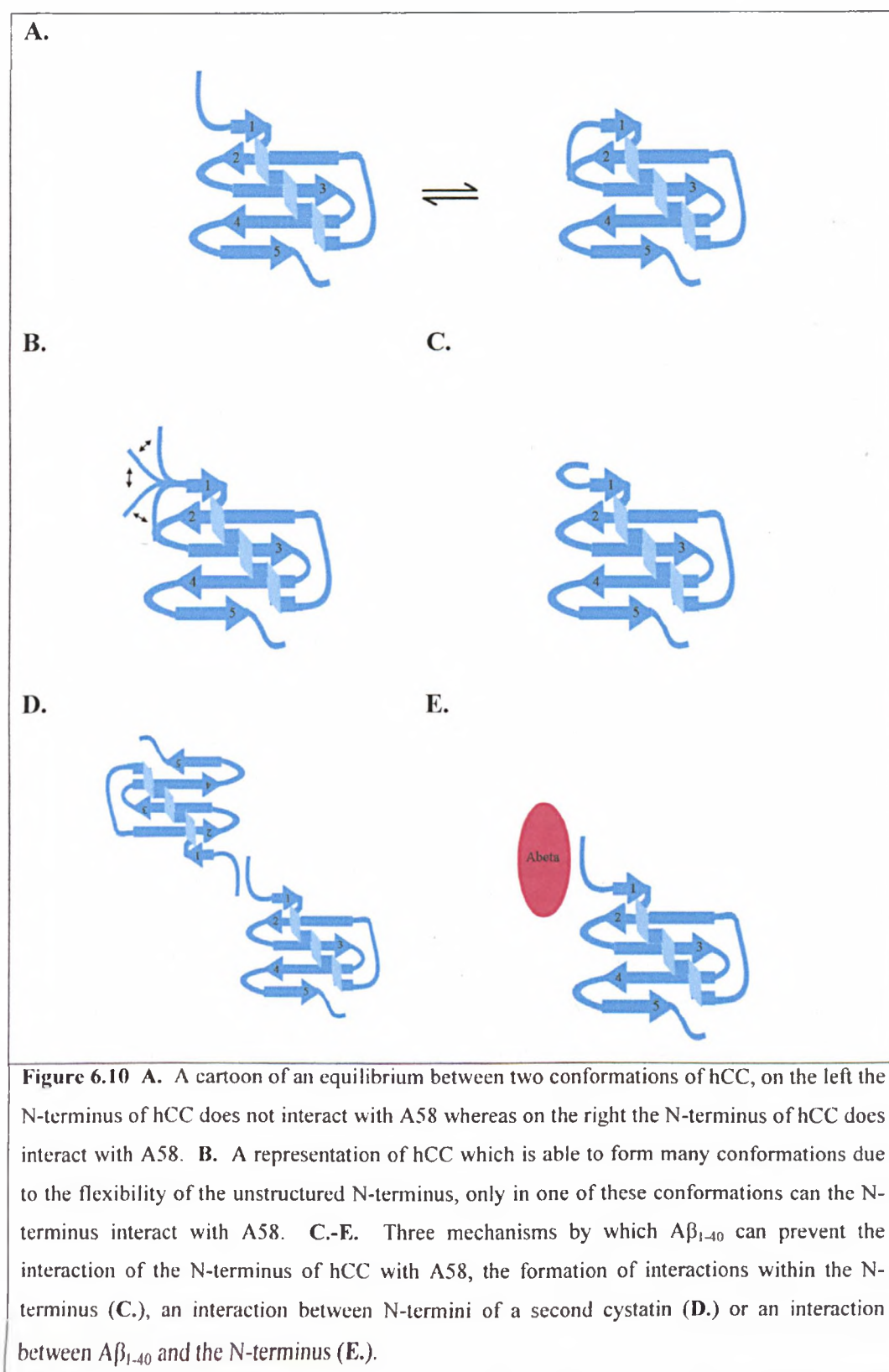
The chemical shift data shown in figure 6.7 does not support such a simple model. Unless the chemical shift differences between the two conformations were unresolvable, then a change in chemical shift should be observed as the conformation where the N-terminus does not interact with A58 becomes the predominant species. It seems unlikely that approximately 10 residues would have no resolvable chemical shift changes between the two conformations. Therefore, the model has been adapted to account for the chemical shift data. Rather than two conformation being populated, multiple conformations can be populated. This is represented by the cartoon in figure 6.10B.

In the adapted model, the idea that $A\beta_{1-40}$ prevents the interaction between the N-terminus and A58 still holds. However, rather than preventing the population of one out of two possible conformation, it is preventing the population of one out of many conformations. The observed chemical shift represents an average of all the chemical environments that an amide populates. Therefore, preventing the exposure to one chemical environment out of many will have a minimal effect on the observed chemical shift.

In order for the model to account for the change in intensity presented in Figure 6.8, the conformation that interacts with A58 must exchange with other states at an intermediate rate. However, the rest of the conformations need to be in fast exchange with each other. This is observed in the HSQC spectrum as a single sharp peak that is broadened only when the conformation that interacts with A58 can be populated.

The interaction that prevents the N-terminus interacting with A58 must occur within the N-terminal region. The three mechanisms by which this can occur are illustrated in Figure 6.10C-E. Either $A\beta_{1-40}$ causes the N-terminus to interact with itself (6.10C), with another N-terminal region (6.10D) or with $A\beta_{1-40}$ itself (6.10E).

The data presented in this chapter conflicts with previously published data that suggests that hCC binds tightly to the N-terminus of $A\beta_{1-40}$.[113] $A\beta_{1-40}$ is highly prone to aggregation, readily forming both oligomeric species and amyloid. The model discussed above only describes the interaction between monomeric hCC and monomeric $A\beta_{1-40}$. A specific interaction may occur between hCC and oligomeric species of $A\beta_{1-40}$ which would not be observed in the experiments described in this chapter. Further experimentation is required to determine whether hCC interacts specifically with oligomeric $A\beta_{1-40}$ species and $A\beta_{1-40}$ amyloid.

Figure 6.10 **A.** A cartoon of an equilibrium between two conformations of hCC, on the left the N-terminus of hCC does not interact with A58 whereas on the right the N-terminus of hCC does interact with A58. **B.** A representation of hCC which is able to form many conformations due to the flexibility of the unstructured N-terminus, only in one of these conformations can the N-terminus interact with A58. **C.-E.** Three mechanisms by which $A\beta_{1-40}$ can prevent the interaction of the N-terminus of hCC with A58, the formation of interactions within the N-terminus (**C.**), an interaction between N-termini of a second cystatin (**D.**) or an interaction between $A\beta_{1-40}$ and the N-terminus (**E.**).

CHAPTER SEVEN
FINAL CONCLUSIONS AND FUTURE WORK

The main aim of this project was to compare the amyloidogenicity of hCC with that of other members of the cystatin superfamily. Identifying common behaviour across the cystatins helps to validate the use of other, better characterised cystatins in the development of a model of cystatin amyloidogenesis. Given that at least partial unfolding of the folded state is required to produce amyloid precursor states that trigger the formation of amyloid, a detailed understanding of the folding pathway is a first step to the identification of these states. For this thesis, the folding pathway of hCC was characterised and compared with that of cC (chapter 4). Following on from that, the initial stages of amyloidogenesis of hCC, including dimerisation by domain-swapping, were observed in order to identify behaviour common to all cystatins and factors that may increase the general propensity of proteins to form amyloid.

Prior to conducting the experiments described in chapters 4, 5 and 6 it was necessary to produce sufficient quantities of recombinant hCC. The expression and purification protocol described in chapter 3 was adapted from a cC protocol and optimised to produce a yield of approximately 5mg/L growth media. Despite the optimisation of the protocol there remains a significant variation in the yield of hCC. The hCC gene contains several codons that are rarely used in *E. coli*, within this group of rare codons one proline codon is particularly favoured. In order to compensate for this bias *E. coli* expressing rare tRNAs were used in the production of recombinant hCC. The expression of hCC was significantly improved, but there were a number of inconsistencies with the behaviour of this cell line. The growth rate of the cells varied considerably, which could significantly increase the time taken to produce recombinant hCC. Time is probably the most important factor when considering the role of contaminants in the degradation of hCC. Producing the rare tRNAs and the antibiotic resistance needed to maintain their production adds additional stress to the cells. A high concentration of nucleic acids were found in the periplasmic extract which caused difficulties in the purification process. The need to use this cell line would be removed if the rare codons were replaced with codons that are used more frequently in *E. coli*. Alternatively the whole hCC gene could be replaced with a

synthetic gene that favours the codon bias of *E. coli*. In conclucsion, improved expression levels and the removal of nucleic acids from the periplasmic extract should improve the overall yield of hCC.

The folding pathway of cC has been well characterised, but there was previously very little information available on the folding pathway of hCC. Analysis of the data described in chapter 4 shows that the folding pathway of hCC is in fact very similar to that of cC. hCC and cC both fold via a relatively compact intermediate before passing through a transition state barrier to the folded state. The stability of the folding intermediate of hCC is very poor (-1.7 kcal mole^{-1}), predictably so considering the difference in stability of the folded proteins ($\Delta\Delta G_{F/U}$ = 10.6 kcal mole^{-1}, for hCC vs. cC). Other cystatins whose folding behaviour has been characterized are less stable than hCC, with a $\Delta G_{F/U}$ of 4.3 kcal mole^{-1} for cystatins B.[178] This suggests that although other cystatins may well be undergoing similar structural changes during folding, for a number of these proteins, 2-state folding is observed as I_{kin} becomes too unstable to be populated.

Amyloid precursor states are the link between the folding and misfolding pathways. For disordered states to become involved in bimolecular reactions, they need to exist for sufficient time for these to interact. Short of stabilising disordered states of proteins, another way of encouraging these states is to make it kinetically difficult for the protein to get out of this state, in other words, a "kinetic trap". An example of this is seen for β-microglobulin where the amyloid precursor state is believed to be trapped as a partially folded state by the isomerisation of a proline residue.[137, 138] In the case of cystatins, there are several different hypotheses. Native state hydrogen exchange (Staniforth *et al.*, unpublished data) has been used to investigate any population of species that are in exchange with the native state of cC. The GdnHCl dependence of the exchange rates have been determined and this has led to the suggestion that the reduced state may be a precursor to dimerisation and amyloid formation (unpublished data).[134] A reduced state has been identified as a amyloid precursor state of human prion protein. Human prion protein switches from the native α conformation to a monomeric form rich in β structure following the reduction of the disulphide bond. The soluble β form of human prion protein is a direct precursor amyloid fibres.[179] *In vivo* it is certainly plausible that the necessary

reducing conditions are encountered for the reduced state to be a realistic amyloid precursor state. Whether or not the reduced state is a feasible amyloid precursor in hCC remains to be determined. Other members of the cystatin superfamily, specifically cystatins A and B have been studied using hydrogen exchange.[134, 180] The overall picture that is emerging may, in fact, be that these proteins exchange their hydrogens through local structural fluctuations, rather than by any cooperative unfolding event. Examining the behaviour of a cystatin with a stable intermediate between that of cystatin B and cC, i.e. hCC may help resolve these conflicting hypotheses.

As well as the reduced state of hCC or cC, a proline isomer populated during the refolding of this protein may provide an ideal candidate for an amyloid forming precursor. In chapter 4, I show that proline isomerisation is responsible for the slow folding (5.5×10^{-2} s^{-1}) of a population of the unfolded hCC. A recent study of cystatin B amyloidogenesis suggests that the isomerisation of proline 74 is essential for the formation of a tetrameric form of this protein. The structural change in the protein, caused by this isomerisation has been proposed to be essential for amyloid formation by cystatins.[172] In this thesis, I contribute to resolving this issue by mutating the proline that is conserved across the cystatins, P103 in cC. I show that this significantly alters the propensity for fibrillisation whilst having no effect on the ability to dimerise. In the future, when problems of availability of hCC, are resolved, analysis of the equivalent mutant in hCC, P105A, will determine whether this behaviour is common to all cystatins, and more importantly, determine whether the slow folding rate that is observed in the wt hCC folding pathway is due to the isomerisation of this conserved proline (P105). If isomerisation of P105 is required for hCC amyloid formation, the population of a partially folded state with the correct P105 isomer for aggregation may be an important factor in the propensity for fibrillisation.

Prior to the experiments described in chapter 5, it was known that the domain-swapped dimers of hCC and cC are structurally very similar.[19, 134] hCC and cC both form domain swapped dimers by incorporating loop 1 of the active site into the β-sheet thus forming a continuous β-strand from strands 2 and 3. Although the end product of the dimerisation reaction, the domain-swapped dimers, are structurally similar no information was available on the mechanisms by which

they are formed. Over the range of conditions studied in this thesis, the dimerisation of hCC was shown to be rate limited by a bimolecular process. Furthermore, m-value analysis indicated that the structure of the dimerisation transition state is very close to the structure of the unfolded state and is more unfolded than the kinetic intermediate identified in the hCC folding pathway. The analysis of the hCC dimerisation reaction shows it to be very similar to that of cC. The difference in the observed rates of dimerisation between hCC and cC are directly related to the difference in unfolding rates and are thus predictable from the observed difference in the stability of the folded proteins. Evidence of a common behaviour of dimerisation in cystatins lends support to the use of several cystatins in the development of a model of amyloidogenesis in cystatins. It also provides additional routes to the characterisation of the domain-swapping phenomenon in proteins as very few domain-swapping proteins have experimentally accessible dimerisation reactions. Dimerisation of the cell cycle regulatory protein p13suc1 is one of the few domain-swapping reaction that has been well characterised, here the monomer-dimer equilibrium is controlled by two conserved prolines in the hinge loop that connects the exchanging domains.[131]

cC has been shown to form dimers, tetramers and oligomers under fibrillisation conditions prior to the formation of amyloid fibres. Under dimerisation conditions, there was no evidence of a tetrameric hCC species. Further work is required to determine conditions under which hCC can be induced to form amyloid fibres. Analysis of the initial stages of amyloidogenesis of hCC is required to determine which oligomeric species are populated under amyloidogenic conditions. A comparison of the amyloidogenesis of hCC and cC will indicate what behaviour is common to the cystatin superfamily and help identify what factors influence the amyloidogenicity of the different cystatins.

The work described in chapter 6 is a series of preliminary experiments investigating the nature of the interaction between monomeric hCC and monomeric $A\beta_{1-40}$. By assigning the ^{1}H-^{15}N HSQC spectrum of hCC any changes due to the interaction with $A\beta$ can be mapped onto the structure of hCC. Analysis of this data shows quite clearly that there is no interaction between monomeric hCC and monomeric $A\beta_{1-40}$ under these conditions. Given that hCC has been shown to inhibit $A\beta$ amyloid formation, hCC must interact with one of the

oligomeric species of Aβ that is populated during amyloidogenesis. Further experimentation is required to determine the exact nature of the interaction between hCC and Aβ. The bound state of hCC to either oligomeric Aβ or Aβ amyloid would exceed the upper size limit of NMR experiment described in chapter 6 and cause a complete loss of NMR signal. However, although the exact residues involved the interaction could not be identified the loss of the NMR signal in itself could be used as evidence of the interaction between hCC and Aβ. Alternatively, analytical ultracentrifugation (AUC) in combination with diffusion experiments or light scattering, including quasi-elastic light scattering (QLS)[181], may shed light on this problem.

The main limitation to the work described in this thesis has been the ability to produce sufficient quantities of recombinant hCC. Despite the optimisation of the protocol for hCC expression and purification, there is still significant variation in the yield of hCC. Future modification of the pINIII-ompA2-hCC plasmid should significantly improve the yield and reduce the variability in the yield, providing a reliable, high yield source of hCC (consistently ≥ 5mg/L culture). Given sufficient recombinant hCC, there are a number of useful experiments that would add support to the data presented in this thesis. Despite the limited supply of hCC, the analysis of the folding and dimerisation of hCC presented in this thesis highlights common features in the folding and aggregation of the cystatin superfamily.

REFERENCES

1. Turk, V. and Bode, W. (1991) The Cystatins - Protein Inhibitors of Cysteine Proteinases. *FEBS Lett.*, **285** (2) 213-219.

2. Grzonka, Z., *et al.* (2001) Structural Studies of Cysteine Proteases and their Inhibitors. *Acta Biochim. Pol.*, **48** (1) 1-20.

3. Barrett, A. J. (1987) The Cystatins: a New Class of Peptidase Inhibitors. *TIBS*, **12** 193-196.

4. Bobek, L. A. and Levine, M. J. (1992) Cystatins - Inhibitors of Cysteine Proteinases. *Crit. Rev. Oral Biol. Med.*, **3** (4) 307-332.

5. Mussap, M. and Plebani, M. (2004) Biochemistry and Clinical Role of Human Cystatin C. *Crit. Rev. Clin. Lab. Sci.*, **41** 467-550.

6. Otto, H. H. and Schirmeister, T. (1997) Cysteine Proteases and their Inhibitors. *Chem. Rev.*, **97** (1) 133-171.

7. Brown, W. M. and Dziegielewska, K. M. (1997) Friends and Relations of the Cystatin Superfamily - New Members and Their Evolution. *Protein Sci.*, **6** 5-12.

8. Shimba, N., *et al.* (2000) Structural Comparison Between Wild-type and P25S Human Cystatin A by NMR Spectroscopy. Does this Mutation Affect the Alpha-helix Conformation? *J. Struct. Funct. Genom.*, **1** (1) 26-42.

9. Bode, W., *et al.* (1988) The 2.0 A X-Ray Crystal Structure of Chicken Egg White Cystatin and Its Possible Mode of Interaction with Cysteine Proteinases. *Embo J.*, **7** (8) 2593-2599.

10. Abrahamson, M. (1988) Human Cysteine Proteinase Inhibitors - Isolation, Physiological Importance, Inhibitory Mechanism, Gene Structure and

Relation to Hereditary Cerebral Hemorrhage. *Scand. J. Clin. Lab. Invest.,* **48** 21-31.

11.   Grubb, A. and Lofberg, H. (1985) Human Gamma-Trace - Structure, Function and Clinical Use of Concentration Measurements. *Scand. J. Clin. Lab. Invest.,* **45** 7-13.

12.   Abrahamson, M., *et al.* (1986) Isolation of 6 Cysteine Proteinase Inhibitors from Human Urine - Their Physicochemical and Enzyme Kinetic Properties and Concentrations in Biological Fluids. *J. Biol. Chem.,* **261** (24) 1282-1289.

13.   Abrahamson, M., *et al.* (1988) Efficient Production of Native, Biologically Active Human Cystatin C by *Escherichia Coli. FEBS Lett.,* **236** (1) 14-18.

14.   Abrahamson, M. and Grubb, A. (1994) Increased Body Temperature Accelerates Aggregation of the Leu-68-Gln Mutant Cystatin C, the Amyloid Forming Protein in Hereditary Cystatin C Amyloid Angiopathy. *Proc. Natl. Acad. Sci. U. S. A.,* **91** (4) 1416-1420.

15.   Abrahamson, M., *et al.* (1987) Molecular Cloning and Sequence Analysis of cDNA Coding for the Precursor of the Human Cysteine Proteinase Inhibitor Cystatin C. *FEBS Lett.,* **216** (2) 229-233.

16.   Dalboge, H., *et al.* (1989) High-Level Expression of Active Human Cystatin C in *Escherichia Coli. Gene,* **79** (2) 325-332.

17.   Auerswald, E. A., *et al.* (1991) Purification and Characterization of a Chicken Egg White Cystatin Variant Expressed in an *Escherichia Coli* pIN-III-OmpA System. *Eur. J. Biochem.,* **200** (1) 131-138.

18.   Ghrayeb, J., *et al.* (1984) Secretion Cloning Vectors in *Escherichia Coli. Embo J.,* **3** (10) 2437-2442.

19. Janowski, R., *et al.* (2001) Human cystatin C, an Amyloidogenic Protein, Dimerizes Through Three-dimensional Domain Swapping. *Nat. Struct. Biol.*, **8** (4) 316-320.

20. Ekiel, I. and Abrahamson, M. (1996) Folding Related Dimerization of Human Cystatin C. *J. Biol. Chem.*, **271** (3) 1314-1321.

21. Jaskolski, M. (2001) 3D Domain Swapping, Protein Oligomerization, and Amyloid Formation. *Acta Biochim Pol*, **48** (4) 807-827.

22. Levy, E., *et al.* (1989) Stroke in Icelandic Patients with Hereditary Amyloid Angiopathy Is Related to a Mutation in the Cystatin C Gene, an Inhibitor of Cysteine Proteases. *J. Exp. Med.*, **169** (5) 1771-1778.

23. Gerhartz, B., *et al.* (1998) Two Stable Unfolding Intermediates of the Disease Causing L68Q Variant of Human Cystatin C. *Biochemistry*, **37** (49) 17309-17317.

24. Calero, M., *et al.* (2001) Distinct Properties of Wild Type and the Amyloidogenic Human Cystatin C Variant of Hereditary Cerebral Hemorrhage with Amyloidosis, Icelandic Type. *J. Neurochem.*, **77** (2) 628-637.

25. Ekiel, I., *et al.* (1997) NMR Structural Studies of Human Cystatin C Dimers and Monomers. *J. Mol. Biol.*, **271** (2) 266-277.

26. Barrett, A. J., *et al.* (1984) The Place of Human Gamma-Trace (Cystatin-C) Amongst the Cysteine Proteinase Inhibitors. *Biochem. Biophys. Res. Commun.*, **120** (2) 631-636.

27. Sipe, J. D. and Cohen, A. S. (2000) Review: History of the Amyloid Fibril. *J. Struct. Biol.*, **130** (2-3) 88-98.

28. Jenko, S., *et al.* (2003) Crystal Structure of Stefin A in Complex with Cathepsin H: N-terminal Residues of Inhibitors can Adapt to the Active Sites of Endo- and Exopeptidases. *J. Mol. Biol.,* **326** (3) 875-885.

29. Abrahamson, M., *et al.* (1987) Identification of the Probable Inhibitory Reactive Sites of the Cysteine Proteinase Inhibitors Human Cystatin C and Chicken Cystatin. *J. Biol. Chem.,* **262** (20) 9688-9694.

30. Hall, A., *et al.* (1995) Structural Basis for the Biological Specificity of Cystatin C - Identification of Leucine-9 in the N-Terminal Binding Region as a Selectivity Conferring Residue in the Inhibition of Mammalian Cysteine Peptidases. *J. Biol. Chem.,* **270** (10) 5115-5121.

31. Alvarez-Fernandez, M., *et al.* (1999) Inhibition of Mammalian Legumain by Some Cystatins is Due to a Novel Second Reactive Site. *J. Biol. Chem.,* **274** (27) 19195-19203.

32. Dando, P. M., *et al.* (1999) Pig Kidney Legumain: an Asparaginyl Endopeptidase with Restricted Specificity. *Biochem. J.,* **339** 743-749.

33. Taupin, P., *et al.* (2000) FGF-2-responsive Neural Stem Cell Proliferation Requires CCg, a Novel Autocrine/paracrine Cofactor. *Neuron,* **28** (2) 385-397.

34. Warfel, A. H., *et al.* (1987) Constitutive Secretion of Cystatin C (Gamma-trace) by Monocytes and Macrophages and its Downregulation After Stimulation. *J. Exp. Med.,* **166** (6) 1912-1917.

35. Leung-Tack, J., *et al.* (1990) Modulation of Phagocytosis Associated Respiratory Burst by Human Cystatin C: Role of the N-terminal Tetrapeptide Lys-Pro-Pro-Arg. *Exp Cell Res,* **188** (1) 16-22.

36. Leung-Tack, J., *et al.* (1990) Neutrophil Chemotactic Activity is Modulated by Human Cystatin C, an Inhibitor of Cysteine Proteases. *Inflammation,* **14** (3) 427-458.

37. Bjorck, L., *et al.* (1989) Bacterial Growth Blocked by a Synthetic Peptide Based on the Structure of a Human Proteinase Inhibitor. *Nature,* **337** (6205) 385-386.

38. Bjorck, L., *et al.* (1990) Cystatin C, a Human Proteinase Inhibitor, Blocks Replication of *Herpes simplex* Virus. *J. Virol.,* **64** (2) 941-943.

39. Korant, B. D., *et al.* (1985) Cystatin, a Protein Inhibitor of Cysteine Proteases Alters Viral Protein Cleavages in Infected Human Cells. *Biochem. Biophys. Res. Commun.,* **127** (3) 1072-1076.

40. Keppler, D., *et al.* (1996) Tumor Progression and Angiogenesis: Cathepsin B & Co. *Biochem. Cell Biol.,* **74** 799-810.

41. Fehrenbacher, N., *et al.* (2004) Sensitization to the Lysosomal Cell Death Pathway upon Immortalization and Transformation. *Cancer Res.,* **64** (15) 5301-5310.

42. Liu, C., *et al.* (2003) Overexpression of Legumain in Tumors Is Significant for Invasion/Metastasis and a Candidate Enzymatic Target for Prodrug Therapy. *Cancer Res.,* **63** (11) 2957-2964.

43. Joyce, J. A., *et al.* (2004) Cathepsin Cysteine Proteases are Effectors of Invasive Growth and Angiogenesis During Multistage Tumorigenesis. *Cancer Cell,* **5** (5) 443-453.

44. Sloane B.F., *et al.* (1981) Lysosomal cathepsin B: correlation with metastatic potential. *Science,* **212** 1151-1153.

45. Zheng, X., *et al.* (2004) Senescence Initiated Reversal of Drug Resistance: Specific Role of Cathepsin L. *Cancer Res.,* **64** (5) 1773-1780.

46. Keppler, D. (2006) Towards novel anti-cancer strategies based on cystatin function. *Cancer Letters,* **235** (2) 159-176.

47.   Nishikawa, H., *et al.* (2004) The Role of Cathepsin B and Cystatin C in the Mechanisms of Invasion by Ovarian Cancer. *Gynecol. Oncol.,* **92** (3) 881-886.

48.   Zore, I., *et al.* (2001) Cathepsin B/cystatin C Complex Levels in Sera from Patients with Lung and Colorectal Cancer. *Biol. Chem.,* **382** 805-810.

49.   Domej, W., *et al.* (2002) Cystatin C of Pleural Effusion as a Novel Diagnostic Aid in Pleural Diseases of Different Aetiologies *Clin. Sci.,* **102** 373-380.

50.   Chauhan, S. S., *et al.* (1991) Expression of Cathepsin L in Human Tumors. *Cancer Res.,* **51** (5) 1478-1481.

51.   Sokol, J. P., *et al.* (2005) The Use of Cystatin C to Inhibit Epithelial Mesenchymal Transition and Morphological Transformation Stimulated by Transforming Growth Factor-beta. *Breast Cancer Res.,* **7** (5) 844-853.

52.   Gudmundsson, G., *et al.* (1972) Hereditary Cerebral Haemorrhage with Amyloidosis. *Brain,* **95** (2) 387-404.

53.   Ghiso, J., *et al.* (1986) Hereditary Cerebral Amyloid Angiopathy - the Amyloid Fibrils Contain a Protein Which Is a Variant of Cystatin C, an Inhibitor of Lysosomal Cysteine Proteases. *Biochem. Biophys. Res. Commun.,* **136** (2) 548-554.

54.   Revesz, T., *et al.* (2003) Cerebral Amyloid Angiopathies: A Pathologic, Biochemical and Genetic View. *J. Neuropathol. Exp. Neurol.,* **62** (9) 885-898.

55.   Haan, J. and Roos, R. A. C. (1992) Comparison between the Icelandic and Dutch Forms of Hereditary Cerebral Amyloid Angiopathy. *Clin. Neurol. Neurosurg.,* **94** S82-S83.

56. Levy, E., *et al.* (2001) Codeposition of cystatin C with amyloid-beta protein in the brain of Alzheimer disease patients. *J. Neuropathol. Exp. Neurol.,* **60** (1) 94-104.

57. Deng, A., *et al.* (2001) Elevation of Cystatin C in Susceptible Neurons in Alzheimers' Disease. *Am. J. Pathol.,* **159** (3) 1061-1068.

58. Vinters, H. V., *et al.* (1990) Immunoreactive-A4 and Gamma-Trace Peptide Colocalization in Amyloidotic Arteriolar Lesions in Brains of Patients with Alzheimers Disease. *Am. J. Pathol.,* **137** (2) 233-240.

59. Vattemi, G., *et al.* (2003) Cystatin C Colocalizes with Amyloid-beta and Coimmunoprecipitates with Amyloid Beta Precursor Protein in Sporadic Inclusion Body Myositis Muscles. *J. Neurochem.,* **85** (6) 1539-1546.

60. Dobson, C. M. (2004) Principles of Protein Folding, Misfolding and Aggregation. *Semin. Cell Dev. Biol.,* **15** (1) 3-16.

61. Rochet, J. C. and Lansbury, P. T. (2000) Amyloid Fibrillogenesis: Themes and Variations. *Curr. Opin. Struct. Biol.,* **10** (1) 60-68.

62. Makin, O. S. and Serpell, L. C. (2005) Structures for Amyloid Fibrils. *FEBS J.,* **272** (23) 5950-5961.

63. Sunde, M., *et al.* (1997) Common Core Structure of Amyloid Fibrils by Synchrotron X-ray Diffraction. *J. Mol. Biol.,* **273** (3) 729-739.

64. Jimenez, J. L., *et al.* (2002) The Protofilament Structure of Insulin Amyloid Fibrils. *Proc. Natl. Acad. Sci. U. S. A.,* **99** (14) 9196-9201.

65. Makin, O. S. and Serpell, L. C. (2002) Examining the Structure of the Mature Amyloid Fibril. *Biochem. Soc. Trans.,* **30** 521-525.

66. LeVine 3rd, H. (1993) Thioflavine T Interaction with Synthetic Alzheimer's Disease Beta-Amyloid Peptides: Detection of Amyloid Aggregation in Solution. *Prot. Sci.,* **2** (3) 404-410.

67. Elghetany, M., *et al.* (1989) The Congo Red Stain Revisited. *Annal. Clin. Lab. Sci.,* **19** (3) 190-195.

68. Sanders, A. (2004) The tetramerisation, multimerisation and amyloidogenesis of cystatins. Doctor of Philosophy, University of Sheffield

69. Khurana, R., *et al.* (2005) Mechanism of Thioflavin T Binding to Amyloid Fibrils. *J. Struct. Biol.,* **151** (3) 229-238.

70. Sachse, C., *et al.* (2006) Quaternary Structure of a Mature Amyloid Fibril from Alzheimer's Aβ(1-40) Peptide. *J. Mol. Biol.,* **362** 347-354.

71. Jimenez, J. L., *et al.* (1999) Cryo-Electron Microscopy Structure of an Sh3 Amyloid Fibril and Model of the Molecular Packing. *Embo J.,* **18** (4) 815-821.

72. Serpell, L. C. and Smith, J. M. (2000) Direct Visualisation of the Beta-sheet Structure of Synthetic Alzheimer's Amyloid. *J. Mol. Biol.,* **299** (1) 225-231.

73. Makin, O. S. and Serpell, L. C. (2002) Amyloidogenic Proteins Involved in Neurodegeneration and Therapeutic Implications. *Biochem. Soc. T.,* **30** (4) 521-525.

74. Tycko, R. (2006) Characterization of Amyloid Structures at the Molecular Level by Solid State Nuclear Magnetic Resonance Spectroscopy. *Method Enzymol.,* **413** 103-122.

75.   Petkova, A. T., *et al.* (2002) A Structural Model for Alzheimer's Beta - amyloid Fibrils Based on Experimental Constraints from Solid State NMR. *Proc. Natl. Acad. Sci. U. S. A.,* **99** (26) 16742-16747.

76.   Petkova, A. T., *et al.* (2006) Experimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry,* **45** (2) 498-512.

77.   Correia, B. E., *et al.* (2006) A Structural Model of an Amyloid Protofilament of Transthyretin. *Protein Sci.,* **15** (1) 28-32.

78.   Iwata, K., *et al.* (2006) 3D Structure of Amyloid Protofilaments of Beta2-Microglobulin Fragment Probed by Solid-State NMR. *Proc. Natl. Acad. Sci. U. S. A.,* **103** (48) 18119-18124.

79.   Luhrs, T., *et al.* (2005) 3D Structure of Alzheimers' Amyloid-{beta}(1-42) Fibrils. *Proc. Natl. Acad. Sci. U. S. A.,* **102** (48) 17342-17347.

80.   Poulsen, S. A., *et al.* (2000) Solution Structures in Aqueous SDS Micelles of Two Amyloid Beta Peptides of A beta(1-28) Mutated at the Alpha-secretase Cleavage Site (K16E, K16F). *J. Struct. Biol.,* **130** (2-3) 142-152.

81.   Dawbarn, D. and Allen, S. J. (2001) Alzheimers' Disease: Past, Present and Future Themes. Oxford University Press

82.   Caughey, B. and Lansbury, P. T. (2003) Protofibrils, Pores, Fibrils and Neurodegeneration: Separating the Responsible Protein Aggregates from the Innocent Bystanders. *Annu. Rev. Neurosci.,* **26** 267-298.

83.   Zhang, S., *et al.* (2000) The Alzheimer's Peptide A Beta Adopts a Collapsed Coil Structure in Water. *J. Struct. Biol.,* **130** (2-3) 130-141.

84.   Coles, M., *et al.* (1998) Solution Structure of Amyloid beta Peptide(1-40) in a Water Micelle Environment. Is the Membrane Spanning Domain Where we Think it is? *Biochemistry,* **37** (31) 11064-11077.

85.   Serpell, L. C. (1999) Alzheimer's Amyloid Fibrils: Structure and Assembly. *Biochim. Biophys. Acta-Mol. Basis Dis.*, **1502** (1) 16-30.

86.   Thirumalai, D., *et al.* (2003) Emerging Ideas on the Molecular Basis of Protein and Peptide Aggregation. *Curr. Opin. Struct. Biol.*, **13** (2) 146-159.

87.   Kelly, J. W. (1998) The Alternative Conformations of Amyloidogenic Proteins and their Multi-step Assembly Pathways. *Curr. Opin. Struct. Biol.*, **8** (1) 101-106.

88.   Stromer, T. and Serpell, L. C. (2005) Structure and Morphology of the Alzheimer's Amyloid Fibril. *Microsc. Res. Tech.*, **67** 210-217.

89.   Lashuel, H. A., *et al.* (2003) Mixtures of wild-type and a pathogenic (E22G) form of Abeta 40 *in vitro* accumulate protofibrils, including amyloid pores. *J. Mol. Biol.*, **332** 795-808.

90.   Kheterpal, I., *et al.* (J. Mol. Biol.) Structural Differences in Abeta Amyloid Protofibrils and Fibrils Mapped by Hydrogen Exchange - Mass Spectroscopy with On-line Proteolytic Fragmentation. *J. Mol. Biol.*, **361** 785-795.

91.   Malinchik, S. B., *et al.* (1998) Structural analysis of Alzheimer's beta(1-40) amyloid: protofilament assembly of tubular fibrils. **74** (1) 537-545.

92.   Wetzel, R., *et al.* (2007) Plasticity of Amyloid Fibrils. *Biochemistry*, **46** (1) 1-10.

93.   Williams, A. D., *et al.* (2004) Mapping A[beta] Amyloid Fibril Secondary Structure Using Scanning Proline Mutagenesis. *J. Mol. Biol.*, **335** (3) 833-842.

94. Shivaprasad, S. and Wetzel, R. (2006) Scanning Cysteine Mutagenesis Analysis of Abeta-(1-40) Amyloid Fibrils. *J. Biol. Chem.*, **281** (2) 993-1000.

95. Balbach, J. J., *et al.* (2002) Supramolecular Structure in Full Length Alzheimers' Beta-Amyloid Fibrils: Evidence for a Parallel beta-Sheet Organization from Solid State Nuclear Magnetic Resonance. *Biophys. J.*, **83** (2) 1205-1216.

96. Kayed, R., *et al.* (2003) Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis. *Science,* **300** (5618) 486-489.

97. Bucciantini, M., *et al.* (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature,* **416** (6880) 507-511.

98. Yamada, M. (2000) Cerebral Amyloid Angiopathy: An Overview. *Neuropathology,* **20** (1) 8-22.

99. Yamada, M., *et al.* (1997) Association of Presenilin-1 Polymorphism with Cerebral Amyloid Angiopathy in the Elderly. *Stroke,* **28** (11) 2219-2221.

100. Bugiani, O. (2004) A-beta Related Cerebral Amyloid Angiopathy. *Neurol Sci,* **25** Suppl 1. S1-2.

101. McLaurin, J. and Chakrabartty, A. (1996) Membrane Disruption by Alzheimer Beta Amyloid Peptides Mediated Through Specific Finding to either Phospholipids or Gangliosides - Implications for Neurotoxicity. *J. Biol. Chem.,* **271** (43) 26482-26489.

102. Kakio, A., *et al.* (2002) Interactions of Amyloid Beta Protein with Various Gangliosides in Raft Like Membranes: Importance of GM1 Ganglioside Bound Form as an Endogenous Seed for Alzheimer Amyloid. *Biochemistry,* **41** (23) 7385-7390.

103. Subasinghe, S., *et al.* (2003) Cholesterol is Necessary both for the Toxic Effect of A Beta Peptides on Vascular Smooth Muscle Cells and for A Beta Binding to Vascular Smooth Muscle Cell Membranes. *J. Neurochem.*, **84** (3) 471-479.

104. Gan, L., *et al.* (2004) Identification of Cathepsin B as a Mediator of Neuronal Death Induced by A-beta Activated Microglial Cells using a Functional Genomics Approach. *J. Biol. Chem.*, **279** (7) 5565-5572.

105. Maruyama, K., *et al.* (1990) Immunohistochemical Characterization of Cerebrovascular Amyloid in 46 Autopsied Cases Using Antibodies to Beta Protein and Cystatin C. *Stroke,* **21** (3) 397-403.

106. Gilbert, J. J. and Vinters, H. V. (1983) Cerebral Amyloid Angiopathy - Incidence and Complications in the Aging Brain. 1. Cerebral Hemorrhage. *Stroke,* **14** (6) 915-923.

107. Ishii, N., *et al.* (1984) Amyloid Angiopathy and Lobar Cerebral Hemorrhage. *J. Neurol. Neurosurg. Psychiatry,* **47** (11) 1203-1210.

108. Ellis, R. J., *et al.* (1996) Cerebral Amyloid Angiopathy in the Brains of Patients with Alzheimers' Disease: The CERAD experience .15. *Neurology,* **46** (6) 1592-1596.

109. Jensson, O., *et al.* (1987) Hereditary Cystatin C (Gamma-Trace) Amyloid Angiopathy of the CNS Causing Cerebral Hemorrhage. *Acta Neurol. Scand.,* **76** (2) 102-114.

110. Crawford, F. C., *et al.* (2000) A Polymorphism in the Cystatin C Gene is a Novel Risk Factor for Late-onset Alzheimers' Disease. *Neurology,* **55** (6) 763-768.

111. Finckh, U., *et al.* (2000) Genetic Association of a Cystatin C Gene Polymorphism with Late-onset Alzheimers' Disease. *Arch. Neurol.,* **57** (11) 1579-1583.

112. Steinhoff, T., *et al.* (2001) Increased Cystatin C in Astrocytes of Transgenic Mice Expressing the K670N-M671L Mutation of the Amyloid Precursor Protein and Deposition in Brain Amyloid Plaques. *Neurobiol. Dis.,* **8** (4) 647-654.

113. Sastre, M., *et al.* (2004) Binding of Cystatin C to Alzheimer's amyloid [beta] Inhibits *in vitro* Amyloid Fibril Formation. *Neurobiol. Aging,* **25** (8) 1033-1043.

114. Selenica, M. L., *et al.* (2007) Cystatin C Reduces the *in vitro* Formation of Soluble Abeta(1-42) Oligomers and Protofibrils. *Scand. J. Clin. Lab. Invest.,* **67** (2) 179 - 190.

115. Dobson, C. M. (2003) Protein Folding and Misfolding. *Nature,* **426** (6968) 884-890.

116. Wei, L. H., *et al.* (1998) Instability of the Amyloidogenic Cystatin C Variant of Hereditary Cerebral Hemorrhage with Amyloidosis, Icelandic Type. *J. Biol. Chem.,* **273** (19) 11806-11814.

117. Wetzel, R. (1996) For Protein Misassembly, It's the "I" Decade. *Cell,* **86** (5) 699-702.

118. Nelson, R. and Eisenberg, D. (2006) Structural Models of Amyloid-Like Fibrils. *Advances in Protein Chemistry,* **73** 235-282.

119. Fandrich, M., *et al.* (2001) Amyloid Fibrils from Muscle Myoglobin. *Nature,* **410** (6825) 165-166.

120. Perutz, M. F., *et al.* (2002) Amyloid Fibers are Water-filled Nanotubes. *Proc. Natl. Acad. Sci. U. S. A.,* **99** (8) 5591-5595.

121. Sikorski, P. and Atkins, E. (2005) New Model for Crystalline Polyglutamine Assemblies and Their Connection with Amyloid Fibrils. *Biomacromol.*, **6** (1) 425-432.

122. Perutz, M. F., *et al.* (2002) Aggregation of Proteins with Expanded Glutamine and Alanine Repeats of the Glutamine Rich and Asparagine Rich Domains of Sup35 and of the Amyloid Beta-peptide of Amyloid Plaques. *Proc. Natl. Acad. Sci. U. S. A.*, **99** (8) 5596-5600.

123. Serag, A. A., *et al.* (2002) Arrangement of Subunits and Ordering of [beta]-Strands in an Amyloid Sheet. *Nat Struct Mol Biol*, **9** (10) 734-739.

124. Serag, A. A., *et al.* (2001) Identification of a Subunit Interface in Transthyretin Amyloid Fibrils: Evidence for Self-Assembly from Oligomeric Building Blocks. *Biochemistry*, **40** (31) 9089-9096.

125. Nelson, R., *et al.* (2005) Structure of the Cross-[beta] Spine of Amyloid-like Fibrils. *Nature*, **435** (7043) 773-778.

126. Ivanova, M. I., *et al.* (2004) An Amyloid-forming Segment of {beta}2-microglobulin Suggests a Molecular Model for the Fibril. *Proc. Natl. Acad. Sci. U. S. A.*, **101** (29) 10584-10589.

127. Benyamini, H., *et al.* (2003) [beta]2-Microglobulin Amyloidosis: Insights from Conservation Analysis and Fibril Modelling by Protein Docking Techniques. *J. Mol. Biol.*, **330** (1) 159-174.

128. Bennett, M. J., *et al.* (2006) Deposition Diseases and 3D Domain Swapping. *Structure*, **14** (5) 811-824.

129. Liu, Y. and Eisenberg, D. (2002) 3D domain swapping: As domains continue to swap. *Protein Sci.*, **11** (6) 1285-1299.

130. Sambashivan, S., *et al.* (2005) Amyloid-like Fibrils of Ribonuclease A with Three Dimensional Domain Swapped and Native like Structure. *Nature*, **437** (7056) 266-269.

131. Rousseau, F., *et al.* (2003) The Unfolding Story of Three-Dimensional Domain Swapping. *Structure*, **11** (3) 243-251.

132. Zegers, I., *et al.* (1999) Trimeric Domain-swapped Barnase. *Proc. Natl. Acad. Sci. U. S. A.*, **96** (3) 818-822.

133. Rousseau, F., *et al.* (2001) Three-dimensional Domain Swapping in p13suc1 Occurs in the Unfolded State and is Controlled by Conserved Proline Residues. *Proc. Natl. Acad. Sci. U. S. A.*, **98** (10) 5596-5601.

134. Staniforth, R. A., *et al.* (2001) Three-dimensional Domain Swapping in the Folded and Molten Globule States of Cystatins, an Amyloid Forming Structural Superfamily. *Embo J.*, **20** (17) 4774-4781.

135. Serpell, L. C., *et al.* (2000) The Protofilament Substructure of Amyloid Fibrils. *J. Mol. Biol.*, **300** 1033-1039.

136. Sanders, A., *et al.* (2004) Cystatin Forms a Tetramer Through Structural Rearrangement of Domain-swapped Dimers Prior to Amyloidogenesis. *J. Mol. Biol.*, **336** (1) 165-178.

137. Jahn, T., *et al.* (2006) Amyloid Formation under Physiological Conditions Proceeds Via a Native-Like Folding Intermediate. *Nat. Struct. Mol. Biol.*, **13** (3) 195-201.

138. Eakin, C., *et al.* (2006) A Native to Amyloidogenic Transition Regulated by a Backbone Trigger. *Nat. Struct. Mol. Biol.*, **13** (3) 202-208.

139. MacRaild, C. A., *et al.* (2004) Non-fibrillar Components of Amyloid Deposits Mediate the Self-association and Tangling of Amyloid Fibrils. *J Biol Chem*, **279** (20) 21038-21045.

140. Abraham, C. R., *et al.* (1988) Immunochemical Identification of the Serine Protease Inhibitor Alpha-1-Antichymotrypsin in the Brain Amyloid Deposits of Alzheimers' Disease. *Cell,* **52** (4) 487-501.

141. Bales, K. R., *et al.* (2002) Apolipoprotein E, Amyloid and Alzheimer disease. *Mol. Intervent.,* **2** (6) 363-375.

142. Namba, Y., *et al.* (1991) Apolipoprotein-E Immunoreactivity in Cerebral Amyloid Deposits and Neurofibrillary Tangles in Alzheimers Disease and Kuru Plaque Amyloid in Creutzfeldt-Jakob Disease. *Brain Res.,* **541** (1) 163-166.

143. McCarron, M. O., *et al.* (1999) The Apolipoprotein E Epsilon 2 Allele and the Pathological Features in Cerebral Amyloid Angiopathy Related Hemorrhage. *J. Neuropathol. Exp. Neurol.,* **58** (7) 711-718.

144. Nicoll, J. A. R., *et al.* (1997) High Frequency of Apolipoprotein E Epsilon 2 Allele in Hemorrhage due to Cerebral Amyloid Angiopathy. *Annal. Neurol.,* **41** (6) 716-721.

145. Itoh, Y., *et al.* (1996) Influence of Apolipoprotein E Genotype on Cerebral Amyloid Angiopathy in the Elderly. *Stroke,* **27** (2) 216-218.

146. Greenberg, S. M., *et al.* (1995) Apolipoprotein-E Type-4 Allele and Risk of Intracerebral Hemorrhage Associated with Cerebral Amyloid Angiopathy. *Annal. Neurol.,* **38** (2) 285-286.

147. Sambrook, J., *et al.* (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press

148. Anastasi, A., *et al.* (1983) Cystatin, a Protein Inhibitor of Cysteine Proteinases - Improved Purification from Egg White, Characterization and Detection in Chicken Serum. *Biochem. J.,* **211** (1) 129-138.

149. Bessette, P. H., *et al.* (1999) Efficient Folding of Proteins with Multiple Disulfide Bonds in the *Escherichia coli* Cytoplasm. *Proc. Natl. Acad. Sci. U. S. A.*, **96** (24) 13703-13708.

150. Stancik, L. M., *et al.* (2002) pH-dependent Expression of Periplasmic Proteins and Amino Acid Catabolism in *Escherichia coli. J. Bacteriol.*, **184** (15) 4246-4258.

151. Winter, J., *et al.* (2000) Increased Production of Human Proinsulin in the Periplasmic Space of *Escherichia coli* by Fusion to DsbA. *J. Biotechnol.*, **84** (2) 175-185.

152. Raina, S. and Missiakas, D. (1997) Making and Breaking Disulphide Bonds. *Ann. Rev. Microbiol.*, **51** 179-202.

153. Langen, G. R., *et al.* (2001) Absence of the Outer Membrane Phospholipase A Suppresses the Temperature Sensitive Phenotype of *Escherichia coli* degP Mutants and Induces the Cpx and {sigma}E Extracytoplasmic Stress Responses. *J. Bacteriol.*, **183** (18) 5230-5238.

154. Pogliano, J., *et al.* (1997) Regulation of *Escherichia coli* Cell Envelope Proteins Involved in Protein Folding and Degradation by the Cpx Two-component System. *Genes Dev.*, **11** (9) 1169-1182.

155. Jerala, R., *et al.* (1994) Improved Expression and Evaluation of Polyethyleneimine Precipitation in Isolation of Recombinant Cysteine Proteinase Inhibitor Stefin B. *Protein Expr. Purif.*, **5** 65-69.

156. Lovrien, R. E. and Matulis, D. (1997) Selective Precipitation of Proteins. John Wiley and Sons, Inc.

157. Staniforth, R. A., *et al.* (2000) The Major Transition State in Folding Need not Involve the Immobilization of Side Chains. *Proc. Natl. Acad. Sci. U. S. A.*, **97** (10) 5790-5795.

158. Zerovnik E, *et al.* (1998) On the mechanism of human stefin B folding: II. Folding from GuHCl unfolded, TFE denatured, acid denatured, and acid intermediate states. *Proteins, 32* (3) 304-313.

159. Zerovnik E, *et al.* (1998) On the mechanism of human stefin B folding: I. Comparison to homologous stefin A. Influence of pH and trifluoroethanol on the fast and slow folding phases. *Proteins, 32* (3) 296-303.

160. Nilsson, M., *et al.* (2004) Prevention of Domain Swapping Inhibits Dimerization and Amyloid Fibril Formation of Cystatin C. Use of Engineered Disulfide Bridges, Antibodies and Carboxymethylpapain to Stabilize the Monomeric Form of Cystatin C. *J Biol Chem, 279* (23) 24236-24245.

161. Parker, M. J., *et al.* (1997) Acquisition of Native beta-Strand Topology During the Rapid Collapse Phase of Protein Folding. *Biochemistry, 36* (43) 13396-13405.

162. Parker, M. J., *et al.* (1995) An Integrated Kinetic Analysis of Intermediates and Transition States in Protein Folding Reactions. *J. Mol. Biol., 253* (5) 771-786.

163. Pace, C. N. (1975) The Stability of Globular Proteins. *CRC Crit. Rev. Biochem., 3* 1-43.

164. Clarke, A. R. and Waltho, J. P. (1997) Protein Folding Pathways and Intermediates. *Curr. Opin. Biotech., 8* (4) 400-410.

165. Parker, M. J., *et al.* (1998) Thermodynamic properties of transient intermediates and transition states in the folding of two contrasting protein structures. *Biochemistry, 37* 2538-2545.

166. Baker, D. (2000) A Surprising Simplicity to Protein Folding. *Nature, 405* 39-42.

167. Plaxco, K. W., *et al.* (1998) Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol.,* **277** (4) 985-994.

168. Plaxco, K. W., *et al.* (2000) Topology, Stability, Sequence, and Length: Defining the Determinants of Two-State Protein Folding Kinetics. *Biochemistry,* **39** (37) 11177-11183.

169. Dinner, A. R. and Karplus, M. (2001) The Roles of Stability and Contact Order in Determining Protein Folding Rates. *Nat. Struct. Biol.,* **8** (1) 21-22.

170. Sato, S., *et al.* (2001) On the Relationship Between Protein Stability and Folding Kinetics: a Comparative Study of the N-terminal Domains of RNase HI, *E. coli* and *Bacillus stearothermophilus* L9. *J. Mol. Biol.,* **312** 569-577.

171. Wahlbom, M., *et al.* (2007) Fibrillogenic Oligomers of Human Cystatin C Are Formed by Propagated Domain Swapping. *J. Biol. Chem.,* **282** (25) 18318-18326.

172. Jenko-Kokalj, S., *et al.* (2007) Essential Role of Proline Isomerization in Stefin B Tetramer Formation. *J. Mol. Biol.,* **366** (5) 1569-1579.

173. Stubbs, M. T., *et al.* (1990) The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *Embo J.,* **9** 1939-1947.

174. Morgan, G. (2006) Structural studies on the cystatin amyloid fibril. Doctor of Philosophy, University of Sheffield

175. Fezoui, Y., *et al.* (2000) An Improved Method of Preparing the Amyloid β Protein for Fibrillogenesis and Neurotoxicity Experiments. *Amyloid: Int. J. Exp. Clin. Invest.,* **7** 166-178.

176. Hou, L., *et al.* (2004) Solution NMR Studies of the A-beta(1-40) and A-beta(1-42) Peptides Establish that the Met35 Oxidation State Affects the Mechanism of Amyloid Formation. *J Am Chem Soc,* **126** (7) 1992-2005.

177. Zagorski, M. G., *et al.* (1999) Methodological and Chemical Factors Affecting Amyloid Beta Peptide Amyloidogenicity. *Method Enzymol.,* **309** 189-204.

178. Kenig, M., *et al.* (2006) Folding and amyloid-fibril formation for a series of human stefins' chimeras: Any correlation? *Proteins,* **62** (4) 918-927.

179. Jackson, G. S., *et al.* (1999) Reversible conversion of monomeric human prion protein between native and fibrilogenic conformations. *Science,* **283** (5409) 1935-1937.

180. Morgan, G. J., *et al.* (2008) Exclusion of the native alpha-helix from the amyloid fibrils of a mixed alpha/beta protein. *J. Mol. Biol,* **375** (2) 487-498.

181. Harper, J. D. and Lansbury, P. T. (1997) Models of Amyloid Seeding in Alzheimer's Disease and Scrapie: Mechanistic Truths and Physiological Consequences of Time-Dependent Solubility of Amyloid Proteins. *Annual Review of Biochemistry,* **66** 385-407.