Fu, Yinan (2009) *Structure and dynamics of Pseudomonas aeruginosa ICP.* PhD thesis.

http://theses.gla.ac.uk/723/

UNIVERSITY
of
GLASGOW

Structure and dynamics of *Pseudomonas aeruginosa* ICP

Yinan Fu

University of Glasgow

Faculty of Biomedical & Life Sciences

A thesis submitted in partial fulfilment of
the requirements for the degree of Doctor
of Philosophy

April, 2009

© Yinan Fu, 2009

Abstract

Pseudomonas aeruginosa inhibitor of cysteine peptidases (PA-ICP) is a potent protein inhibitor of papain-like cysteine peptidases (CPs) identified in *Pseudomonas aeruginosa*, an opportunistic pathogenic bacteria that can cause severe infections in human. It belongs to the newly characterized natural CP inhibitors of the I42 family, designated the ICP family. The members of this family are present in some protozoa and bacterial pathogens. They can inhibit both parasite and mammalian CPs with high affinity and specificity. Whether the main biological function of the proteins in the pathogens is to regulate the hydrolytic activity of the organisms' endogenous CPs or exogenous CPs so as to facilitate the pathogens' invasion or survival is still under investigation. Although *Pseudomonas aeruginosa* contains a CP inhibitor, no CP genes are found in its genome, suggesting that the targets of PA-ICP may be exogenous. This hypothesis is supported by the presence of a putative secretion signal peptide at the N-terminus of PA-ICP which may be involved in exporting the protein to target exogenous CPs.

In order to shed light on the biological function and inhibitory specificity of PA-ICP, the structure and backbone dynamics of this protein were characterised using NMR spectroscopy. In this project, the inhibitory activity of PA-ICP to a range of mammalian model CPs was also studied. Like its previously studied homologs, PA-ICP adopts an immunoglobulin fold comprised of seven β-strands. Three highly conserved sequence motifs located in mobile loop regions form the CP binding site. The inhibitor exhibits higher affinity toward the mammalian CP cathepsin L than cathepsins H and B. Homology modelling of the PA-ICP-cathspin L interaction based on the crystal structure of the chgasin-cathpsin L complex shows that PA-ICP may inhibit the peptidases by blocking the enzyme's active site and that the interactions between chagasin and CPs may be conserved in PA-ICP-peptidase complexes. The specificity of the inhibitors may be determined by the relative flexibility of the loops bearing the binding site motifs and the electrostatic properties of certain residues near the binding sites.

Contents

List of tables

List of figures

Appendices

Acknowledgements

I hereby declare that this thesis, and the research it describes, is entirely my own work, except where explicitly stated. It has not been presented in whole, or part, for any other degree or award.

Yinan Fu
April 2009

Definitions/abbreviations

2D	two-dimensional
3D	three-dimensional
ARIA	ambiguous restraints for iterative assignment
CP	cysteine peptidase
CNS	Crystallography and NMR suite
CSI	chemical shift index
dH_2O	distilled and deionised water
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FID	free induction decay
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl-b-D-galactopyranoside
K_{cat}	turnover number
K_i	dissociation constant
K_M	Michaelis constant
MaxEnt	Maximum Entropy
NMR	nuclear magnetic resonance
NOE(SY)	nuclear Overhauser effect (spectroscopy)
PA-ICP	*Pseudomonas aeruginosa* cysteine peptidase inhibitor
PDB	protein data bank
rMD	restrained molecular dynamics
tRNA	transfer ribonucleic acid
RDC	residual dipolar coupling
RMSD	root mean square deviation
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
S/N	signal-to-noise ratio
TOCSY	total correlation spectroscopy
UV	ultra-violet radiation

CHAPTER 1

INTRODUCTION PART 1: ICP PROTEINS AND CYSTEINE PEPTIDASES

1.1 Overview

This thesis presents my work on the structure and functional studies of PA-ICP, a clan CA family C1 cysteine peptidase (CP) inhibitor of the ICP family present in the pathogenic bacteria *Pseudomonas aeruginosa.* In this chapter, I will briefly review the biological and structural characteristics of cysteine peptidases and their inhibitors of ICP family.

Pseudomonas aeruginosa is a gram-negative bacterium resident mainly in aquatic environments. It is an opportunistic pathogen infecting individuals suffering from immune deficiency (Mesaros et al., 2007). The bacteria can infect any part of the body including the respiratory tract, bones and joints, ears and eyes. The most serious infections include malignant external otitis, endophthalmitis, endocarditis, meningitis, pneumonia, and septicemia. The infections can be fatal because the bacterium resists many of today's antibiotics. New approaches to therapy for these infections are under investigation (Mesaros et al., 2007).

When infected by a pathogen, the first line of immune defense of the host is established by phagocytosis. The microorganism is taken up by the phagocytes and phagosomes are formed. The phagosomes are then converted into lysosomes which contain different types of digestive enzymes including a high concentration of peptidases that serve to degrade foreign proteins into peptides to break down the microorganism. In order to survive in this highly hydrolytic environment, pathogens have developed various skills to resist the host immune defense system. Synthesis of the peptidase inhibitors to regulate the host peptidases and so protect its own proteins from degradation is one of the approaches that are adopted by some pathogens. A good example is the protozoan parasite *Leishmania major* (Eschenlauer et al., 2009). It does not contain family S1A serine peptidases but it produces ecotin-like inhibitors to target host family S1A serine peptidases at the early stage of the infection (Eschenlauer et al., 2009). Likewise, *Pseudomonas aeruginosa* contains an ICP protein but has no clan CA family C1 cysteine peptidase present in its genome (http://merops.sanger.ac.uk/). This finding supports the hypothesis that one role of ICPs in the parasites might be to manipulate host CP activity so as to facilitate parasite invasion or survival. Given the versatility of the living environments of *Pseudomonas aeruginosa*, PA-ICP may be not an adaptation for inhibition of mammalian hosts, but of plants or other microorganisms that may be encountered by the bacteria. Additionally, comparison of the sequence of *P. aeruginosa*

ICP with other members of this family showed that it possesses an unusual N-terminal extension containing a hydrophobic region preceded by a positively charged arginine residue which has the potential to be a secretion signal peptide. Although it is unknown whether the N-terminus is involved in translocation of ICP *in vivo* and if it is, where the protein will be targeted to, it raises the possibility that the protein may be able to be exported from the bacterium to target the exogenous CPs. Therefore, understanding the binding specificity of PA-ICP to different CPs on a molecular structure basis will provide more details of the ICP-peptidase interaction and the physiological role of ICPs.

1.2 Cysteine peptidases of clan CA family C1

1.2.1 Classification of cysteine peptidases

Peptidases, also termed as proteinases or proteases, are enzymes that catalyses the proteolysis of peptide bonds. They have been found in eukaryotes, prokaryotes, fungi, and bacteria and have many physiological mechanisms. They play important roles in protein turnover, regulating protein degradation. The enzymes can be divided into two different types according to the location of the substrate cleavage site (Barrett, 1994). Endopeptidases cleave polypeptide chains at internal sites away from the termini, and exopeptidases act near the end of polypeptide chains. The enzymes can also be classified into six groups in terms of the nucleophilic residues in their catalytic center (Barrett, 1994). There are serine, cysteine, aspartic, threonine, glutamic and metallo peptidases (http://merops.sanger.ac.uk). As the name implies, cysteine peptidases utilize a cysteine residue as a nucleophile. Nine clans of CPs, designated clan CA, CD, CE, CF, CH, CL, CM, CN and CO, have been recognized so far, comprising nearly 60 families (http://merops.sanger.ac.uk). The best-investigated one is that of clan CA family C1, also termed the papain family. The members of this family can be described as "papain-like" (Rawlings and Barrett, 1994).

1.2.2 General structural features of cathepsin L, H and B and their principal catalytic mechanisms

Mammalian papain-like CPs are also known as thiol-dependant cathepsins (Lecaille et al., 2002). My project focuses on cathepsin L, H and B as they were chosen to be used as model target CPs for PA-ICP because of likelihood of PA-ICP encountering them upon infection a mammalian host (Lecaille et al., 2002). In general, the proteins are synthesized with an N-terminal signal sequence (10–20 amino acids) followed by the

zymogen which consists a pro-region (of between 38 and 250 amino acids) and the mature proteolytically active enzyme, generally 220–260 amino acids long (Lecaille et al., 2002). The mature forms of the CPs are mostly monomeric, with molecular weights in the range from 20 to 30 kDa (McGrath, 1999). The mature enzymes consist of two adjacent domains of roughly equal size (Turk et al., 2000) (fig 1.2). The N-terminal domain is dominated by three helical regions while the C-terminal domain is composed of a barrel of five to six strands. For many enzymes in this group, disulphide bonds form which stabilizes both domains. The catalytic traid of these enzymes is contributed by residues C25e, H163e and N187e (cathepsin L numbering and e for enzyme) that sit in the middle and at the bottom of the active site cleft located between the two domains of cathepsins (Turk et al., 2000). Seven possible substrate-binding sites bracketing the catalytic cysteine and histidine residues were first proposed by Schechte (Schechte.I and Berger, 1967), as illustrated in figure 1.1. The substrate orients to place its N-terminal (P) residues to interact with the S sites of the enzyme while its C-terminal (P') residues make contact with the S' subsites so that the carbonyl carbon of P1 and amide nitrogen of P1' of the substrate are readily accessible to the catalytic thiolate-imidazolium ion pair. The S1 and S1' sites bear the essential catalytic cysteine and histidine residues and interact directly with the scissile bond of the substrates while the shape and size of the sidechains of the variants in the S2 subsite of different cysteine peptidases dominates the substrate specificity in these enzymes.
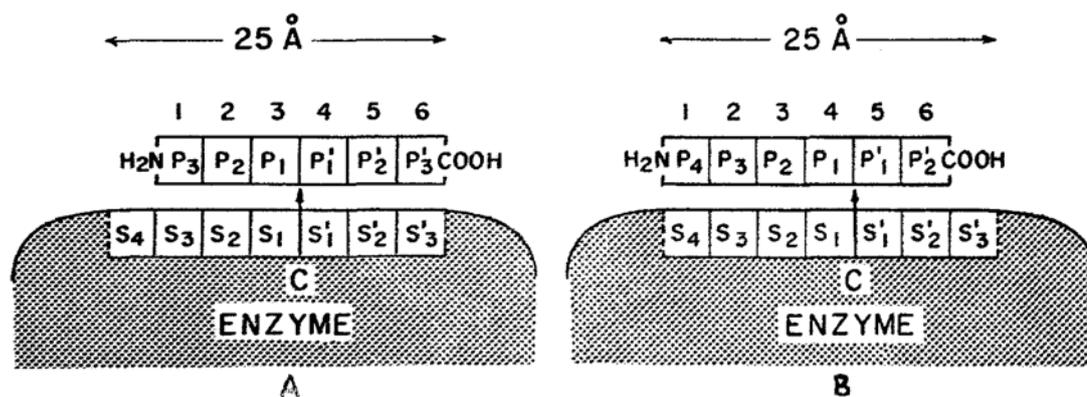
Fig 1.1   Illustration of two possible modes of interaction of papain with a hexapeptide. Seven possible substrate-binding site S and S' are located on both site of the catalytic cysteine residue forming a 25 Å-long active site. The S and S' sites interact with the N-terminal P sites and C-terminal P' sites of the peptide, respectively. (Schechte.I and Berger, 1967)

The active site clefts of these enzymes are similar and most of the enzymes are endopeptidases. Cathepsin H and B can also function as exopeptidases because the active site cleft in both peptidases is partially obstructed, restricting access to the substrate binding site so as to accommodate only the terminal residues of the substrates (Nagler et al., 1997, Vasiljeva et al., 2003). When functioning as an exopeptidase, an eight-residue mini chain occupies the S2 site of cathepsin H in a substrate-like mode held in place through a disulphide linkage between residue C205e in the mature enzyme and C80Pe in the mini chain (Guncar et al., 1998, Vasiljeva et al., 2003). Thus, only the N-terminal residues of the substrates are able to be attacked by the S1 site of the enzyme. In cathepsin B, an occluding loop covers the S2' subsite of the active site cleft and therefore allows just the C-terminal residues of the substrates to interact with the enzyme (Illy et al., 1997, Nagler et al., 1997).

The mechanism of the reaction is illustrated in figure 1.3 (Storer and Menard, 1994). The central catalytic cysteine residue is collocated in the vicinity of the active site histidine residue so as to form a thiolate-imidazolium ion pair (Cys-S$^-$/His-Im$^+$) via proton transfer at optimal pH and the nucleophile is activated. Proteolysis begins with the attack on the carbonyl carbon of the scissile peptide bond of the substrate by the nucleophilic cysteine thiolate. The cleaved C-terminal fragment of amine R'-NH$_2$ is released, resulting in the formation of a covalent intermediate: the acyl-enzyme. The subsequent steps involve the reaction of the acyl-enzyme with a water molecule to

release the N-terminal fragment and recovery of the free enzyme to continue a new catalytic cycle.
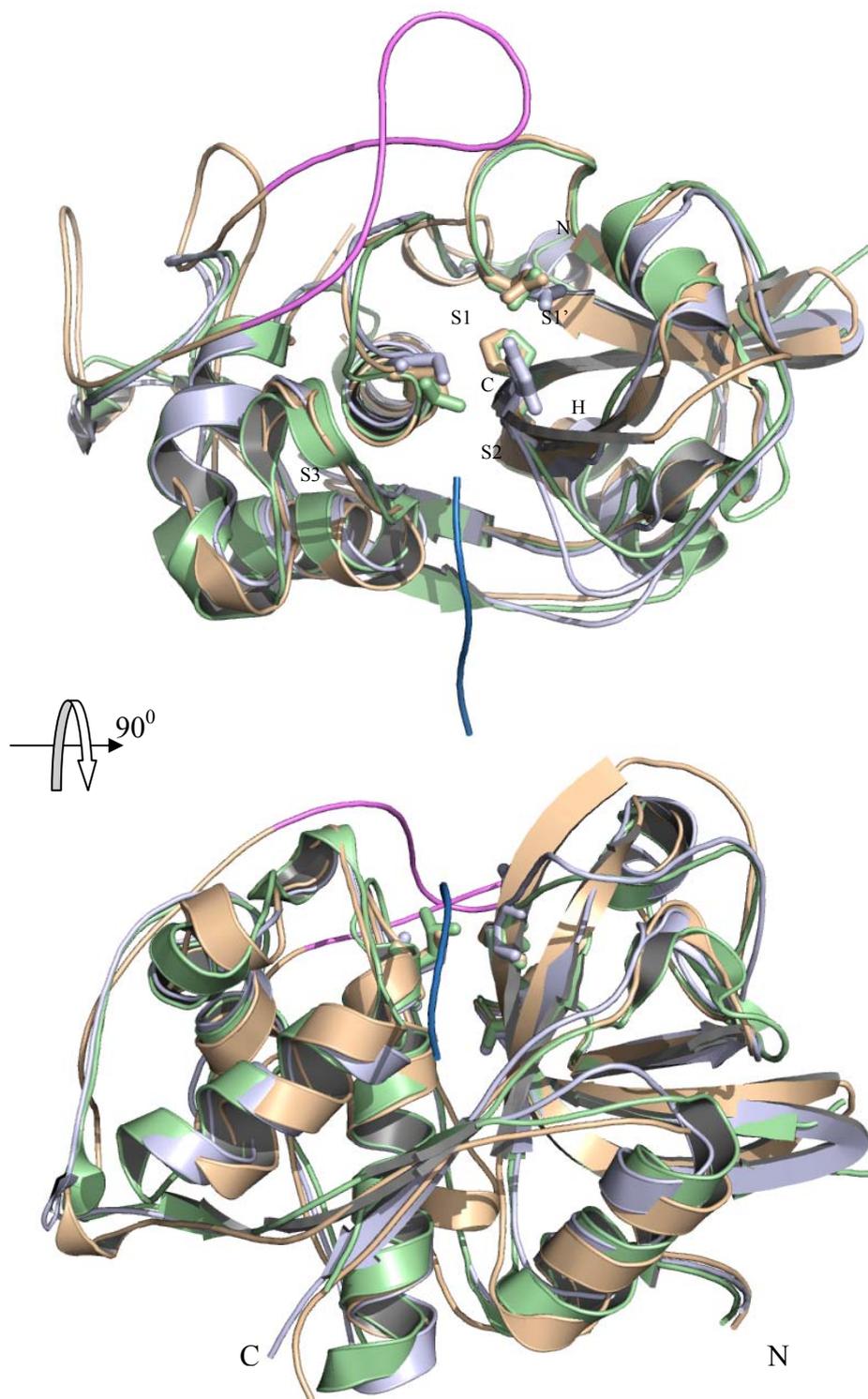
Fig 1.2   A superimposition of the cartoon representations of cathepsin L (green) (PDB 3BC3 (Chowdhury et al., 2008)), H (gray) (PDB 8PCH (Guncar et al., 1998)) and B (yellow) (PDB 2IPP). The active site C, H and N residues are highlighted with sticks. The mini chain in cathepsin H and the occluding loop in cathepsin B are coloured in blue and red respectively. The substrate binding sites S1', S1, S2 and S3 are also indicated. Pictures were produced using PYMOL.
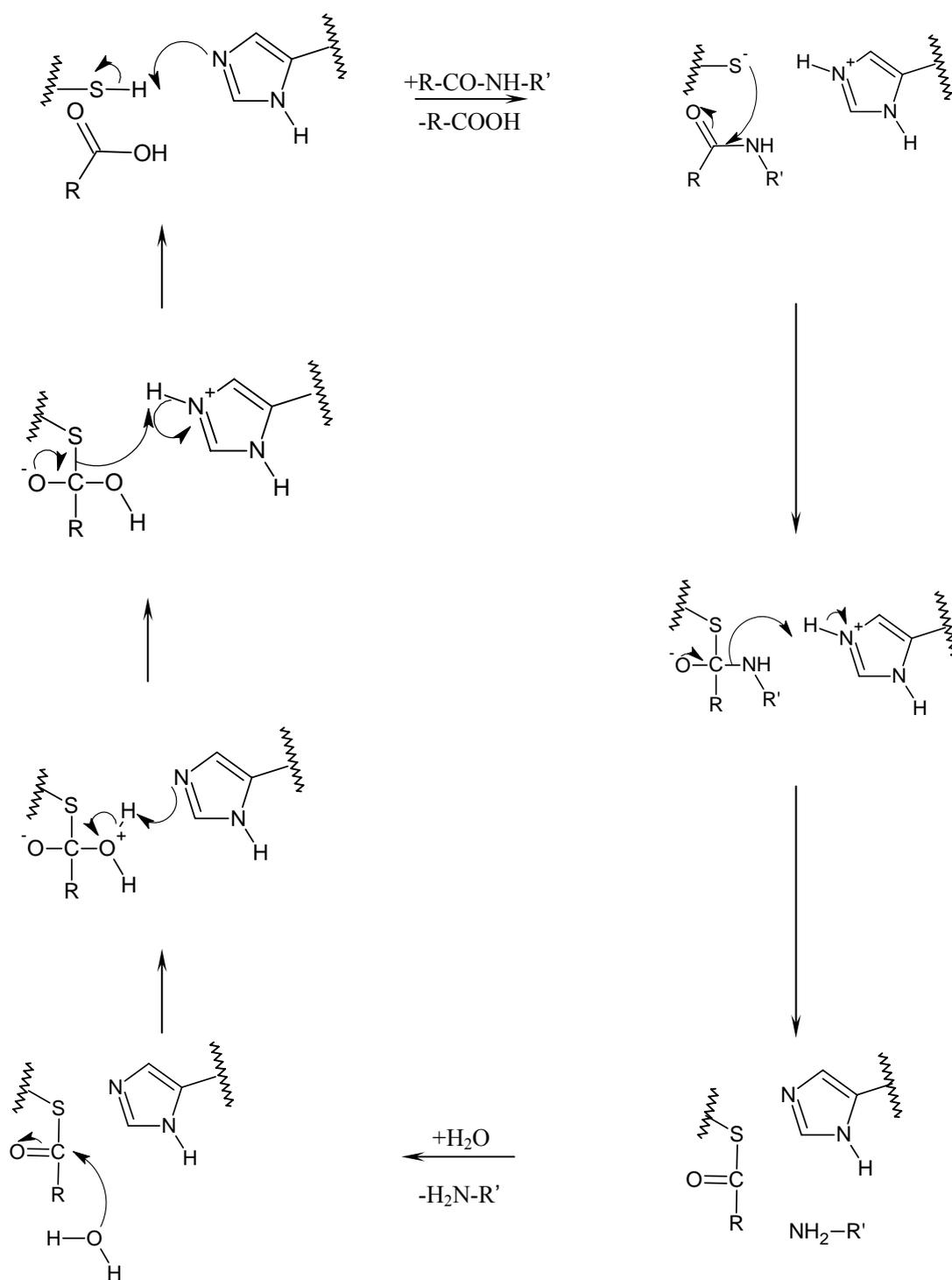
Fig 1.3   The mechanism of the cleavage of substrate by cysteine peptidase.

1.3 ICP proteins

1.3.1 ICP family

Natural cysteine peptidase inhibitors of the I42 family, designated the ICP or chagasin family, have been recently identified (Lo Conte et al., 2000, Rawlings et al., 2004). The members of this family are found in parasitic bacteria, achaea and protozoa (http://merops.sanger.ac.uk/). The proteins bind tightly to papain-like CPs with K_i values in the picomolar to nanomolar range (Sanderson et al., 2003a). The inhibition is competitive and reversible, specific to clan CA family C1 CPs, with a higher affinity to cathepsin-L like CPs than cathepsin-B like (Sanderson et al., 2003a). They have no significant sequence similarity to any other previously characterized cysteine peptidase inhibitors, nor do they share common fold (Rigden et al., 2002). Previous structural studies of these proteins nevertheless suggested convergent evolution of inhibitor function from an immunoglobulin fold (Smith et al., 2006, Salmon et al., 2006).

Although it has been shown that their targets *in vivo* are clan CA family C1 CPs, the question whether the inhibitors' major function is to regulate the activity of the endogenous or the exogenous CPs remains open. Chagasin, the first discovered ICP protein that occurs in *Trypanosome cruzi*, has been characterized as functioning as an endogenous cysteine peptidase inhibitor regulating the activity of the parasite's lysozomal cysteine peptidase, cruzipain (Santos et al., 2005). However, it has also been suggested that the ICP protein in *Leishmania mexicana* may inhibit the CPs of the host (Besteiro et al., 2004).

1.3.2 The overall fold and dynamics of chagasin and *L. mexicana* ICP

The structures of two homologs of PA-ICP, chagasin and *Leishmania mexicana* (*L. mexicana)* ICP, have been extensively studied (Smith et al., 2006, da Silva et al., 2007, Salmon et al., 2006). Both chagasin and *L. mexicana* ICP adopt an immunoglobulin fold consisting seven or eight β-strands (fig 1.4). The strands assemble into two β-sheets to form a "Greek-key" β sandwich. The protein sequence alignment (appendix E) of the ICP family reveals a number of conserved hydrophobic residues and three highly conserved sequence motifs. The conserved hydrophobic residues appear alternately along the β strands contributing the majority of the hydrophobic core of the proteins. The three highly conserved motifs are located in three loop regions on the same end of the proteins, namely L2, L4 and L6 in chagasin and the BC, DE, FG loops in *L.*

mexicana ICP. The NMR solution structures of chagasin and *L. mexicana* ICP show that the L4 loop in chagasin and DE loop in *L. mexicana* ICP are poorly defined in both proteins. Chagasin and *L. mexicana* ICP fold in a similar way, although in the solution structure of chagasin an additional N-terminal β-strand is identified while it is not resolved in the *L. mexicana* ICP NMR structure. This may be due to the N-terminal truncation of *L. mexicana* ICP at residue S6 during protein sample preparation (Smith et al., 2006) . The BC loop of *L. mexicana* ICP contains a 3_{10} helix which is not defined in the NMR structure of chagasin but is identified in its crystal structure (da Silva et al., 2007).

Backbone dynamics studies of the ^{15}N relaxation rates of chagasin and *L. mexicana* ICP indicate that the majority of the residues in chagasin and *L. mexicana* ICP display dynamic properties as expected in folded globular proteins except for the L4 and DE loop which undergo internal motion in both proteins reflecting backbone flexibility in these regions (Salmon et al., 2006, Smith et al., 2006).
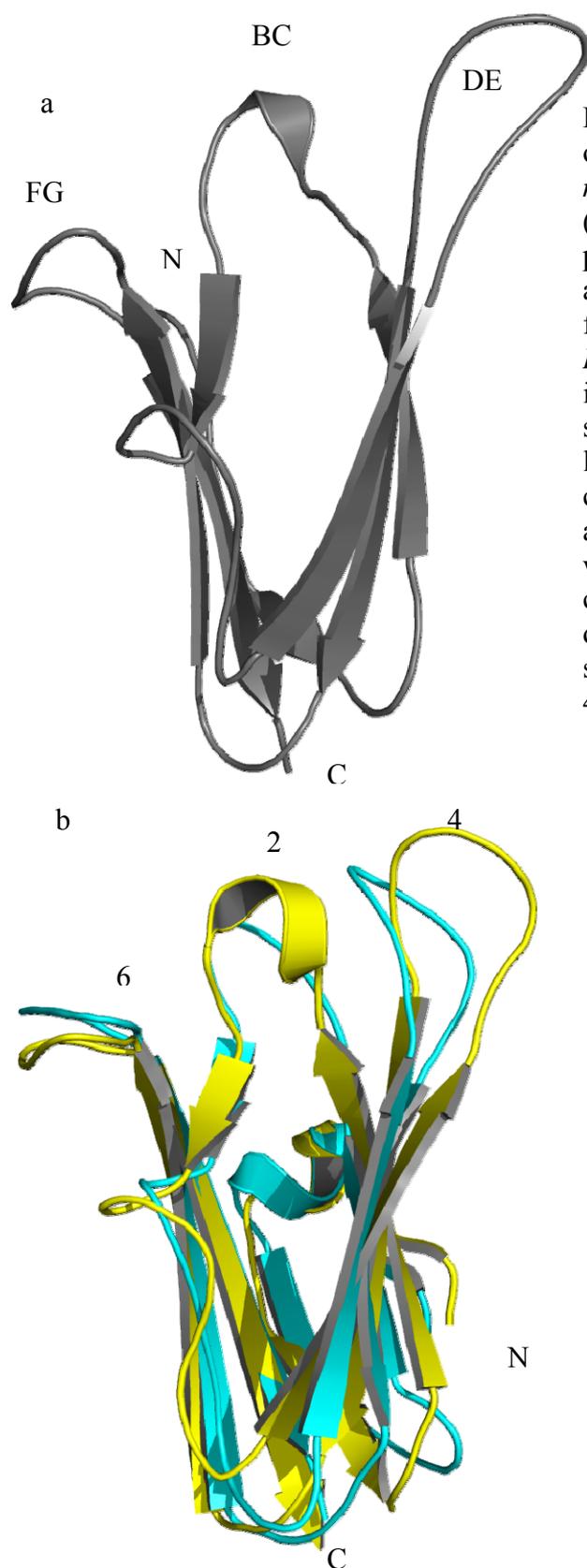
Fig 1.4 Cartoon representation of the tertiary structures of *L. mexicana* ICP (grey, PDB 2C34) (a) and chagasin (b). The proteins are mainly β stranded and adopt an immunoglublin fold. The DE, BC, FG loops in *L. mexicana* ICP and loop 2, 4, 6 in chagasin form the inhibitory site. A 3_{10} helix is observed for loop 2 in the crystal structure of chagasin (yellow, PDB 2H7W) and BC loop in *L. mexicana* ICP while in the solution structure of chagasin (blue, PDB 2FO8), it does not adopt a regular 3_{10} structure. The DE loop and loop 4 are highly mobile.

1.3.3 ICP-peptidase interactions

Site-directed mutagenesis studies of chagasin and *L. mexicana* ICP have revealed that the three highly conserved motifs NPTTGY/F, GXGG and RPW/F in loops L2, L4 and L6 respectively are responsible for peptidase inhibition (dos Reis et al., 2008, Smith et al., 2006). The crystal structures of chagasin and *Trypanosoma brucei* ICP with presumed target or model cysteine peptidases, from both host and parasites, have also been reported and provided detailed information on the inhibitor-peptidase interactions (Alphey and Hunter, 2006, Ljunggren et al., 2007, Redzynia et al., 2008, Wang et al., 2007, Redzynia et al., 2009). The overall architecture of the complex and mode of the interactions are similar in all cases. The highly conserved motifs protrude from one end of the protein and together comprise the binding site that interacts with the enzymes (fig 1.5). The short central loop L2 interacts directly with the catalytic triad of the peptidases while the other two loops fill the active site cleft to prevent substrate binding with loop 4 filling the S subsites and loop 6 occupying the S' subsites. The interaction of loop 2 with the peptidases is achieved by direct insertion of the highly conserved NPTTGY/F motif into the active site cleft, mainly through residue T31, whose main chain carbonyl is hydrogen bonded to the peptidase active site cysteine via a mediating water molecule (dos Reis et al., 2008). The backbone conformation of this residue protects the carbonyl carbon from being attacked by the active site cysteine. This may explain why the inhibitor is not cleaved when in complex with the active enzyme. The conformation of the loop is defined by the proline through hydrophobic interaction with Y57 of chagasin. The importance of this loop in peptidase binding is confirmed by a synthesized heptapeptide GNPTTGF which is able to inhibit the activity of papain with a K_i value of 1.5 μM (Riekenberg et al., 2005). The interaction of the RPW/F motif in loop 6 plays an important role in the recognition of crucial residues in the catalytic cleft. The aromatic residue packs on top of a hydrophobic cluster. Together with the aromatic residue of the NPTTGY/F in loop 2, it serves to anchor the side chain of the arginine residue in an extended conformation. This allows the guanidium group of the arginine to be hydrogen bonded to the side chain carbonyl group of the enzyme's N18e of papain, cathepsin L, cathepsin B, falcipain and cruzipain (Alphey and Hunter, 2006, dos Reis et al., 2008, Ljunggren et al., 2007, Redzynia et al., 2009). This component of the interaction appears to be stronger for parasites' CPs and cathepsin B, as the asparagine residue in cathepsin L and papain is replaced by an aspartate which forms a salt bridge with the arginine. The proline helps maintain the optimal conformation of the loop.
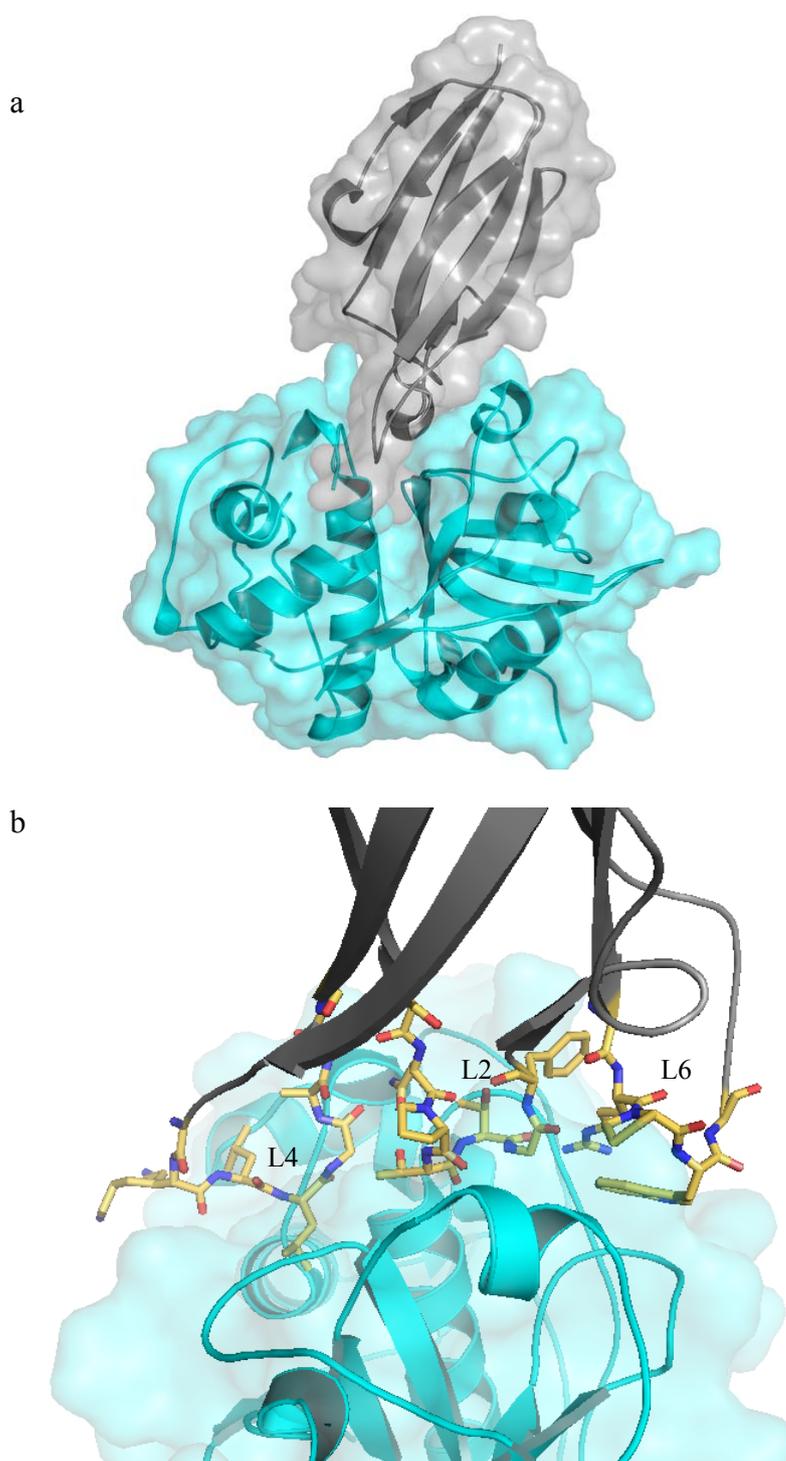
Fig 1.5   a: Cartoon representation of the crystal structure of chagasin in complex with cathepsin L (PDB 2NQD), revealing that the binding of the inhibitor to the peptidase is performed by a tripartite binding motif blocking the enzyme's active site. b: Loops 2, 4 and 6 (yellow) form the chagasin inhibitory wedge. The intermolecular interactions are dominated by hydrogen bonding and hydrophobic interactions. The interactions are likely conserved in other ICP-CP interactions

The interaction of loop 4 of chagasin with peptidases is based on crucial hydrophobic contacts with the enzyme's S2 subsite. The importance of this interaction is highlighted by the finding that the replacement of the hydrophobic residues with hydrophilic ones diminished the inhibitory activity of *L. Mexicana* ICP toward its endogenous clan CA family C1 cysteine peptidase CPB.

In addition to the abovementioned interactions, additional intermolecular contacts are present in the chagasin-cathepsin B complex. The conformation of region H190e-G198e of the enzyme is shaped by a few direct or water mediated hydrogen bonds with residues N55-Y57 in loop 4 of chagasin. A strong hydrogen bonding interaction is seen between residue Y57 of chagasin and E194e of cathepsin B. These contacts are not observed in the chagasin-cathepsin L, chagasin-papain and chagasin-falcipain 2 complexes. This may be because the disulphide bonds C156e-C209e in cathepsin L, C153e-C200e in papain and C151e-C212e in falcipain serve to hold the equivalent loops in distinct conformations.

The occluding loop of cathepsin B is pushed out of the catalytic cleft by loop 6 of chagasin. This process would require additional energy and may explain why chagasin binds better to cathepsin L-like than cathepsin B-like enzymes.

CHAPTER 2

INTRODUCTION PART 2: PROTEIN NMR TECHNIQUE

2.1 Overview

Some nuclei possess the property of aligning themselves with respect to an applied magnetic field in such a way that they can have several states, each with a different energy level. The nuclei that have this property are termed spins. Excitation of these spins with a radio frequency matching the difference between these energy levels causes absorption of energy and can give rise to NMR signals. This is the fundamental principal of NMR spectroscopy. NMR techniques have long been considered useful tools for structural characterisation of biomacromolecules complementary to X-ray crystallography with the ability to probe the dynamics of the molecules (Kay, 2005, Wuthrich, 1990). This chapter reviews the basic theoretical background of the methods that were applied to investigation of the PA-ICP protein structure and dynamics using NMR.

2.2 Multi-dimensional triple resonance experiments for chemical shift assignment

Conventional NMR experiments carried out on protein samples from natural sources are restricted to utilizing magnetization transfer based solely on protons because hydrogen is the only nucleus that has a naturally abundant isotope with a nuclear spin of ½ (Wüthrich, 1986). With more advanced isotopic labelling techniques, additional magnetization transfer pathways can be generated by replacing the naturally abundant ^{12}C and ^{14}N with the spin ½ nuclei of ^{13}C and ^{15}N and spin 1 nucleus of ^2H in biological macromolecules (Gardner and Kay, 1998). This allows the application of multidimensional triple resonance NMR experiments to protein structure and dynamics studies (Sattler et al., 1999). Triple resonance experiments rely exclusively on heteronuclear $^1J/^2J$ coupling to achieve magnetization coherence transfer. The magnetization of a spin can be transferred through bonds to another spin and then transferred back the same way (out-and-back), or the magnetization is transferred to another spin where it stays for acquisition (out-and-stay). Typically, the name of each experiment is composed of the nuclei involved in the magnetization transfer according to its transfer pathway with the spins whose chemical shifts are not evolved are put in parentheses.

The chemical shift of each observed spin has to be assigned to a specific atom in order to interpret the NMR data for structure calculations and dynamics studies. Assignment strategies have been established to fulfill this purpose with the aid of multiple

dimensional triple resonance experiments. The experiments are able to correlate ^{1}H, ^{15}N and ^{13}C spin frequencies sequentially along the polypeptide chain through backbone amide linkages as well as within specific spin systems. Resonance assignment commonly starts from the protein backbone by sequentially connecting the backbone atoms to their related chemical shifts. The assignments are then extended from the backbone atoms to associated side chain carbons and protons. This technique has two advantages. One is that multi-dimensional triple resonance experiments utilize relatively large heteronuclear ^{1}J/^{2}J coupling constants (fig 2.1) to perform coherence transfers faster than experiments that rely on relatively small homonuclear ^{1}H ^{3}J couplings to achieve magnetization transfer. It minimizes the loss of magnetization due to relaxation during the pulse sequences (Sattler et al., 1999). Another is that, each crosspeak has higher information content, which reduces signal overlap and simplifies data interpretation (fig 2.2).

2.3 Structure studies

2.3.1 NMR restraints

NMR restraints used for protein structure determination have been extensively reviewed by Guntert (Guntert, 1998). Despite the increasing availability of other types of restraints, distance restraints, which provide distance information between two spins close in space, remain the most useful and abundant type of restraint that can be incorporated in the structure calculation of a protein using NMR. Distance restraints are usually derived from NOE experiments and as proton are the most enriched spins in proteins, the interproton contacts are most easily obtained. The majority of NOE crosspeaks arise from spins in the same residue or between near neighbours in the primary sequence, while the most important ones for structure determination are the NOE restraints obtained from spins that are far apart in the primary structure but close to each other in the tertiary structure.

In addition to restraints based on NOE enhancements, hydrogen bond restraints can also supply structural information between hydrogen bonded atoms. Residual dipolar coupling restraints measured in weakly aligned media have emerged as a useful tool to define long-range order of biomolecules (Bax and Grishaev, 2005). Torsion angle restraints provide conformational information of the protein backbone or side chains. They can be used for structure refinement and validation.
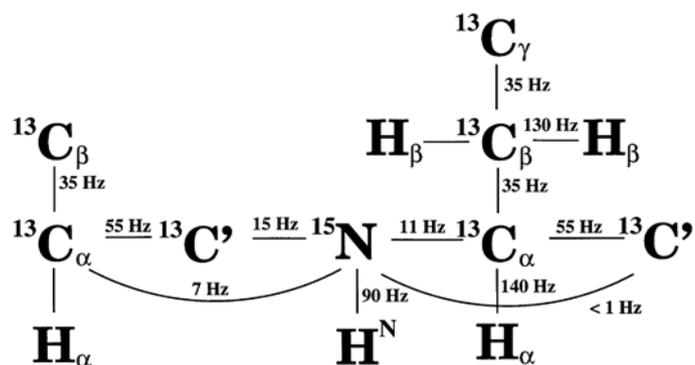
Fig 2.1　　　　The ^1J and ^2J coupling constants that are used for magnetization transfer in ^{13}C-, ^{15}N-labelled proteins.
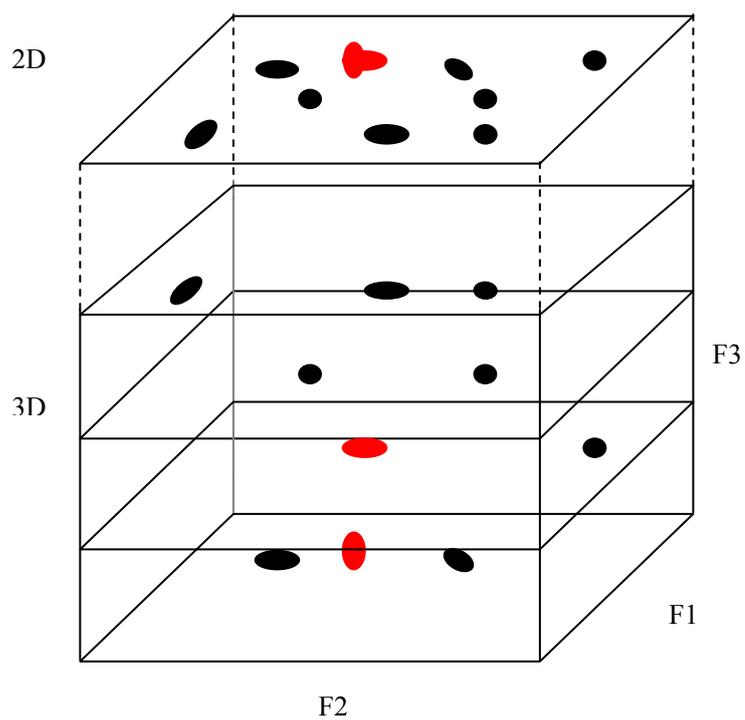
Fig 2.2　　　　A representation of a 3D experiment extended from a 2D experiment.

2.3.1.1 Distance from NOEs

The intensity of nuclear Overhauser effect (I_{ij}) due to dipolar interactions between neighboring protons i and j is dependent on the inverse sixth power of the internuclear distance (r_{ij}^{-6}) (NEUHAUS and WILLIAMSON, 2000) and a correlation function

$$I_{ij} \propto (r_{ij})^{-6} \times f(\tau_c) \qquad (2.1)$$

where τ_c refers to the correlation time. Under the simple circumstances where the system under investigation is rigid and tumbles isotropically, τ_c remains constant for all interproton pairs, the relative intensity is solely proportional to r_{ij}^{-6}. Therefore, if a reference distance r_{ref} can be set with known NOE intensity I_{ref}, the unknown distance r_{ij} between two protons giving rise to an NOE can be determined as follows

$$r_{ij} = r_{ref} \times (\frac{I_{ref}}{I_{ij}})^{\frac{1}{6}} \qquad (2.2)$$

2.3.1.2 Residual dipolar coupling

Nuclear spins possessing a magnetic moment can interact through space via dipole-dipole interactions. A spin ½ nucleus A can be considered a magnetic dipole generating a local magnetic field that affects the magnetic field of the neighbouring spin B, resulting in a frequency shift of spin B dependent on the distance of the two nuclei and the angle between the internuclear vector and the applied magnetic field B_0 (Bax et al., 2001). The dipolar coupling is given by

$$D^{AB} = -\frac{\mu_0 h \gamma_A \gamma_B (3\cos 2\theta - 1)}{16\pi^3 r_{AB}^{\ 3}} \qquad (2.3)$$

where r_{AB} is the internuclear distance and θ the angle between the internuclear vector and B_0, μ_0, the magnetic permeability of vacuum, h, Planck's constant, γ_x, gyromagnetic ratio of nucleus x. For a given r_{AB}, the maximum dipolar coupling D^{AB}_{max} is obtained when $\theta = 0^0$ while the dipolar coupling becomes zero when $\theta = 54.7^0$.

As the dipolar coupling is dependent on the orientation of the internuclear vector with respect to the applied magnetic field B_0, the dipole-dipole interactions in spin ½ enriched biomacromolecule contain valuable orientation information. In solution NMR the relative positions of the two spins in space alter due to fast isotropic molecular tumbling and consequently, the angle θ alters. If the molecule tumbles faster than the dipolar coupling (typically $0-10^{-5}$ Hz), the dipolar coupling averages and a narrow signal is observed. The useful orientational information contained by dipolar interaction is lost. To extract this information for structure calculation, the motion of the molecule needs to be restricted so as to align the molecule with respect to the applied magnetic field. However, if the molecular motion is restricted entirely, the dipolar interactions occurring over large networks of spins together with chemical shift anisotropy will make data interpretation difficult. Also, solid state is not physiological for soluble proteins. Practically, the molecule is partially aligned to a small degree leading to an incomplete averaging of the spatially anisotropic dipolar interaction without sacrificing the spectral resolution too much and the residual dipolar coupling between two neighbour nuclei can be measured (Bax, 2003). The measured RDC refers to the orientation of the internuclear vector in an axis system defined by the global molecular frame, which is in turn partially oriented with respect to the static magnetic field. The alignment can be defined by an alignment tensor A, a real valued, symmetric, traceless 3X3 matrix with 5 degrees of freedom. The principal components A_{xx}, A_{yy} and A_{zz} of A reflect the tendency of the x, y and z axes to be parallel to B_0 with $|A_{zz}| \geq |A_{yy}| \geq |A_{xx}|$ and $A_{xx}+A_{yy}+A_{zz}=0$. The dipolar coupling between nuclei A and B is determined as

$$D^{AB}\left(\theta,\ \varphi\right) = D_a^{\ AB}\left[\left(3\cos^2\theta - 1\right) + \frac{3}{2}R\sin^2\theta\ \cos 2\varphi\right] \quad (2.4)$$

where $D_a^{\ AB} = -\dfrac{\mu_0 h S \gamma_A \gamma_B}{16\pi^3 r_{AB}^{\ 3}}$ is referred to the magnitude of the dipolar coupling

tensor. $R = \dfrac{D_r^{\ AB}}{D_a^{\ AB}}$ is the rhombicity and $D_a^{\ AB} = \dfrac{1}{3}[D_{zz}^{\ AB} - \dfrac{D_{xx}^{\ AB} + D_{yy}^{\ AB})}{2}]$,

$D_r^{\ AB} = \dfrac{1}{3}[D_{xx}^{\ AB} - D_{yy}^{\ AB}]$. θ and φ are the spherical coordinates describing the orientation

of the *AB* vector in the principal axis system of the alignment tensor. S is the general order parameter describing the effect of the internal motion and $\langle r_{AB}\rangle$ is the vibrational averaged internuclear distance (Ottiger and Bax, 1998).

2.3.2 Restrained molecular dynamics, a method for protein NMR structure calculation

The most popular method used to calculate protein structures from NMR data is restrained molecular dynamics (rMD) (NEUHAUS and WILLIAMSON, 2000). Restrained molecular dynamics calculates the molecular structure by monitoring the changes in the potential energy of the molecule under the influence of terms describing our knowledge of its covalent structure and the experimental data during a molecular dynamics simulation. A force field representing the potential energy of the molecular system is applied to each atom. In general, the following force field is used to calculate the potential energies imposed on an atom,

$$E_{total} = E_{covalent} + E_{non\text{-}covalent} + E_{experimental} \quad (2.5)$$

The covalent and non-covalent potential energies can be defined by

$$E_{covalent} = E_{bond} + E_{angle} + E_{dihedral} + E_{improper} \quad (2.6)$$

$$\&$$

$$E_{noncovalent} = E_{van\,der\,Waals} \quad (2.7)$$

with $E_{improper}$ representing terms which enforce the chirality of chiral centers and planarity of aromatic rings and peptide groups. The covalent and non-covalent terms model energy penalties for deviations of the structure from the ideal values of the molecular geometry. The force field parallhdg 5.3 designed by Linge *et al*. (Linge et al., 2003b) was used for structure calculation of PA-ICP. The details of this force field have been extensively described in the literature and will not be further discussed here. The experimental term is to restrain the simulated molecular motion such that the structures satisfy the experimental observations. The simulation process gives some kinetic energy to each atom in the molecule with its initial position determined by a chosen starting molecular conformation. The kinetic energy is defined by an initial velocity of the atom in a randomly chosen direction, which is in turn calculated from a given temperature for all atoms in the molecule as a whole according to a Maxwellian distribution. As the atom moves, the velocity and the molecular geometry change (i.e., its kinetic energy and potential energy change) according to the Newton's laws of motion. The movement of the atoms can be simulated over time by repeatedly calculating their positions, kinetic and potential energies at time intervals Δt shorter than the highest frequency of motion in

the system. The kinetic energy is properly adjusted after each Δt step to ensure the given temperature is kept constant during simulation. The purpose of this simulation is to find the sets of coordinates that satisfy the experimental data as well as possible so that the calculated structures have the lowest potential energy.

2.3.3 Simulated annealing as a method for target function optimization

Simulated annealing (Kirkpatrick et al., 1983) is a powerful optimization tool for locating the minima of complex functions. Annealing is simulated by a high-temperature search phase followed by a slow cooling phase. An initial velocity is assigned to each atom corresponding to the initial simulated annealing temperature. At this stage, the restraining forces are set to be negligible so that the kinetic energy will dominate the molecular system to allow the atoms to move about freely to be able to pass any folding barriers. The system is then cooled down slowly, reducing the kinetic energy, while the weight on the potential energy terms is increased gradually until their functions are fully restored. To avoid the atoms being trapped in local energy minima, the cooling should in theory be infinitely slow, although this cannot be achieved in practice. Therefore, the system is kept in pseudo-equilibrium throughout the cooling phase. The increases in the forces can be considered to have the same effect as decreasing the temperature as they both decrease the freedom of the atomic movements. Varying the forces has other advantages in that different forces can be introduced at different stages of the calculation and that the weight of the different forces in the calculation can be tuned by allocating user-defined force constants. The size of the atoms and the repulsion between them are varied at different calculation stages to determine how easily they can pass by each other. In addition, for non-bonded interactions, a simplified molecular system is used at the early stages of the calculation to improve efficiency with the consideration of the interaction of all atoms re-established at the later calculation stages. As the atoms move, the forces applied to them at each step are calculated based on the gradient of the target function. In theory, the atoms settle down gradually to positions with lowest potential energy and the global potential energy minimum of the structure will be achieved if there is a unique solution. To find the energy minimum, a set of structures are calculated from molecular systems with randomly chosen starting positions for each atom. During simulated annealing, the randomized initial coordinates are driven by the experimental restraints to minimize the potential energy of the molecular system. With appropriately chosen simulated annealing parameters, a good fraction of calculated structures will end up with similar molecular conformations and minimal potential energy at the end of the

simulation. Because the experimental data only cover a small proportion of the geometry information for a protein molecule and they are restricted in certain ranges instead of being absolute values, the calculated structures represent the possible conformations of the molecule compatible with the experimental restraints and theoretical values. Therefore, it is appropriate to characterize the protein structure calculated from NMR data by an ensemble of the calculated structures rather than by a single one.

2.3.4 Potential energy of the distance restraints E_{noe}

The principal experimental terms of the force used in structure calculation come from NOE distance restraints. The potential energy E_{noe} of a single distance restraint I is defined by a flat-bottem-hamonic-wall (FBHW) potential (Nilges et al., 1988) stated as follows:

$$E_{noe}= \begin{cases} k_{noe}\sum_i (L_i - \overline{D}_i)^2 & \text{if } \overline{D}_i < L_i \\ \\ 0 & \text{if } L_i \le \overline{D}_i \le U_i \\ \\ k_{noe}\sum_i (\overline{D}_i - U_i)^2 & \text{if } U_i < \overline{D}_i < n \end{cases} \qquad (2.8)$$

where k_{noe} refers to the force constant, while L_i and U_i are the lower and upper bounds of NOE-derived distance, and $\overline{D}_i$ the calculated distance or summed r^{-6} distance in current structure. When $\overline{D}_i > U_i + n$, the potential switches from harmonic to asymptotic by

$$A + B(\overline{D}_i - n)^{-1} + C(\overline{D}_i - n) \quad \text{if } \overline{D}i > n \qquad (2.9)$$

where A determines the asymptotic gradient, B and C are the coefficients to make the potential continuous and differentiable.

It is obvious that the larger the deviation of the calculated value from the allowed experimental distance range, the larger the potential energy E_{noe} that will be seen. Therefore, the degree of the inconsistence in the calculated structures caused by violated distance restraints can be easily spotted from the calculated distance potential energy term.

2.4 Dynamics analysis

2.4.1 From relaxation to dynamics

In NMR, after the application of a radio frequency pulse, the disturbed spins have the tendency to come back to their thermal equilibrium. This phenomenon is called relaxation and can be characterized by the longitudinal relaxation time T_1 and transverse relaxation time T_2 (Hore, 1995). The relaxation mainly results from two types of interactions for spin ½ nuclei caused by randomly fluctuating internal magnetic fields due to molecular motion, leading to spreading of frequency. The dipolar interaction, the effect of one nuclear magnetic field on another, is the dominant relaxation mechanism between the magnetic moments of the spins and typically lies in the range $0{\sim}10^5$ Hz (Derome, 1987). Chemical shift anisotropy is another source of relaxation which is caused by different orientations of the molecule giving rise to different electron shielding. This effect also typically lies in the range $0{\sim}10^5$ Hz. The rates of relaxation are related to overall rotational correlation time and internal motions.

For a spherical protein, the overall molecular tumbling would be isotropic and the correlation times similar throughout the protein backbone. A rod-like molecule, however, gives rise to anisotropy of the overall rotational motion leading to different rates of reorientations about the various molecular axes according to the anisotropy of the diffusion tensor. The rotation about the longest axis will be more rapid than the rotation about other axes.

The relaxation rates of individual spins in a molecule are by no means the same. The relaxation properties of each spin are also affected by the internal motions of the spin relative to the overall rotational molecular motion. The internal motions of protein and peptide cover a wide range of time scales from picoseconds to milliseconds and longer. Dynamics occurring in the picosecond and nanosecond time range are considered fast internal motion. It arises from local motion of individual bonds or small groups of atoms faster than the overall molecular tumbling and can be detected by measuring T_1, T_2 relaxation times and heteronuclear NOE. Internal motions involving large groups of atoms such as domain motions usually happen on the nanosecond to millisecond range, whereas protein folding and unfolding process and enzyme catalysis take microseconds to milliseconds and even longer. Slow conformational or chemical exchange processes occur on the millisecond to microsecond time scale. The transverse relaxation rate is

very sensitive to this type of motion, leading to a decrease in T_2 relaxation time.

2.4.2 Lipari-Szabo model free analysis of relaxation data

Dynamics studies of protein backbones are carried out by measuring backbone amide ^{15}N T_1, T_2 relaxation times and heteronuclear {^{1}H, ^{15}N} steady NOE. The relaxation constants R_1 (=1/T_1), R_2 (=1/T_2) and heteronuclear NOE are determined by the spectral density functions, J(ω) as follows (Osborne et al., 2001),

$$R_1 = \frac{d^2}{4}[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)] + \frac{c^2}{3}J(\omega_N) \qquad (2.10)$$

$$R_2 = \frac{d^2}{8}[4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] +$$
$$\frac{c^2}{18}[4J(0) + 3J(\omega_N)] + R_{ex} \qquad (2.11)$$

$$NOE = 1 + \frac{d^2}{4R_1}\frac{\gamma_H}{\gamma_N}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \qquad (2.12)$$

where $d = \dfrac{\mu_0 \hbar \gamma_H \gamma_N <r^{-3}{}_{NH}>}{8\pi^2}$, $c = \Delta\sigma\omega_N$, μ_0 is the permeability of free space; $\hbar$ is Planck's constant; γ_H and γ_N are the gyromagnetic ratio of ^{1}H and ^{15}N respectively, r_{NH} is the N-H bond length; $\Delta\sigma$ is the axially symmetric CSA tensor of ^{15}N with the symmetry axis collinear with the H-N bond vector; ω_H and ω_N are the Larmor frequencies of ^{1}H and ^{15}N. Because they are related to the external magnetic field used for measurement, the relaxation experiments are primarily sensitive to motions on picosecond to nanosecond range as the typical magnetic field available at present are from 50 MHz to 900 MHz for ^{1}H and ^{15}N (Lipari and Szabo, 1982a). R_{ex} is used to account for chemical exchange processes that contribute to the decay of transverse magnetization, implying motions on the microsecond to millisecond time scale.

A comprehensive knowledge of the internal motions of proteins helps the understanding of their biological function. The relaxation data obtained above are influenced by both overall molecular tumbling and internal motions. Therefore, it is important to distinguish the influence of the overall rotational motion in order to better study the internal motions. In addition, a good estimation of the rate and the degree of the internal motion together with the identification of the chemical and/or conformational exchange processes in the molecule will be very informative. This can be achieved by analysis of the relaxation

data using a Lipari-Szabo model-free formalism (Lipari and Szabo, 1982a, Lipari and Szabo, 1982b). For molecules tumbling isotropically in solution, the amplitude and rate of the motions can be deduced from the spectral density function $J(\omega)$

$$J(\omega) = \frac{2}{5}[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2}] \qquad (2.13)$$

where $\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$, τ_m is the isotropic rotational correlation time of the molecule, τ_e is the effective correlation time for internal motions, S^2 is the square of the general order parameter characterizing the amplitude of the internal motions, defined by value between 1.0 (absolute restriction) and 0.0 (free rotation). If the molecule experiences a high degree of rotational anisotropy, its impact on the spectral density function needs to be further considered (Bruschweiler et al., 1995).

CHAPTER 3

SAMPLE PREPARATION AND BIOLOGICAL CHARACTERISATION OF PA-ICP

3.1 Overview

Isotopic labelling is now routinely used for NMR study of proteins in excess of 10 kDa. Due to the cost of the isotopic labelling, obtaining high yields of soluble protein from minimal supply of the labelling medium is desired. In addition, because the acquisition of the NMR experiments for structure and dynamics studies can take weeks at typical temperature of 15-40^{0}C, the stability of the protein sample has to be taken into account. Thus, to meet the requirements for NMR sample production, the PA-ICP protein had to be overexpressed to good yield and purified to homogeneity. The first part of this chapter describes the optimization of the sample preparation of PA-ICP for NMR study.

The second part of this chapter investigates the translocation and inhibitory activity of PA-ICP. This information, together with the structural knowledge of PA-ICP, helps shed light on its biological functions.

3.2 Protein sample preparation

3.2.1 Optimization of protein expression

E.coli. BL21 (DE3) strain was transformed with a pET28a-derived plasmid (pBP 109) containing the PA-ICP gene encoding the full-length of the protein with a molecular weight of approximately 17 kDa including an N-terminal His tag. The amino acid sequence of the protein is characterized in appendix F. The recombinant PA-ICP protein was expressed as described in section 8.2.1. The expression level was insufficient and a large proportion of the target protein remained insoluble (fig 3.1 a). Many efforts were pursued to increase the amount of soluble protein and to improve the expression level. The insolubility may be due to formation of inclusion bodies, which happens when a heterologous protein expressed in *E.coli* using a strong promoter fails to fold to its native conformation (Rudolph and Lilie, 1996). To increase the solubility of the protein, the expression was tested under low temperature and reduced IPTG concentration for induction to decelerate translation. Nevertheless, reduction in temperature to 20 (fig 3.1 b) and 18 ^{0}C following the IPTG induction (data not shown) made no significant improvement in solubilization of PA-ICP protein and adjustment of IPTG concentration to 0.05 mM did not help either (data not shown). Examining the DNA sequence of PA-ICP reveals that 3 prolines (CCC) in PA-ICP use a codon that is infrequently used by *E.coli* and codon bias may cause low expression level. Thus, the protein was also expressed in BL21 codon plus (DE3)-RIPL strain that supplies additional copies of

specific tRNA genes that are rare in *E. coli*. The expression was tested in LB medium with induction at 20 ^{0}C and assayed using SDS-PAGE. As can be seen in figure 3.2 a, the expression level was improved. However, after cell lysis, the majority of the protein remained insoluble (fig. 3.2 b).
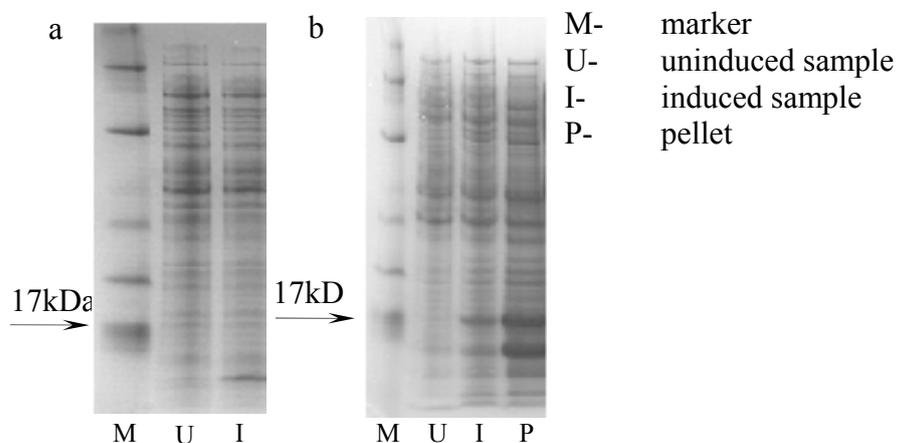
Fig 3.1   SDS-PAGE assay of the expression of PA-ICP in BL 21 (DE3) with 37 ^{0}C (a) and 20 ^{0}C (b) induction. The expression level increased with lower temperature induction. However the majority of the protein remained insoluble after lysis.
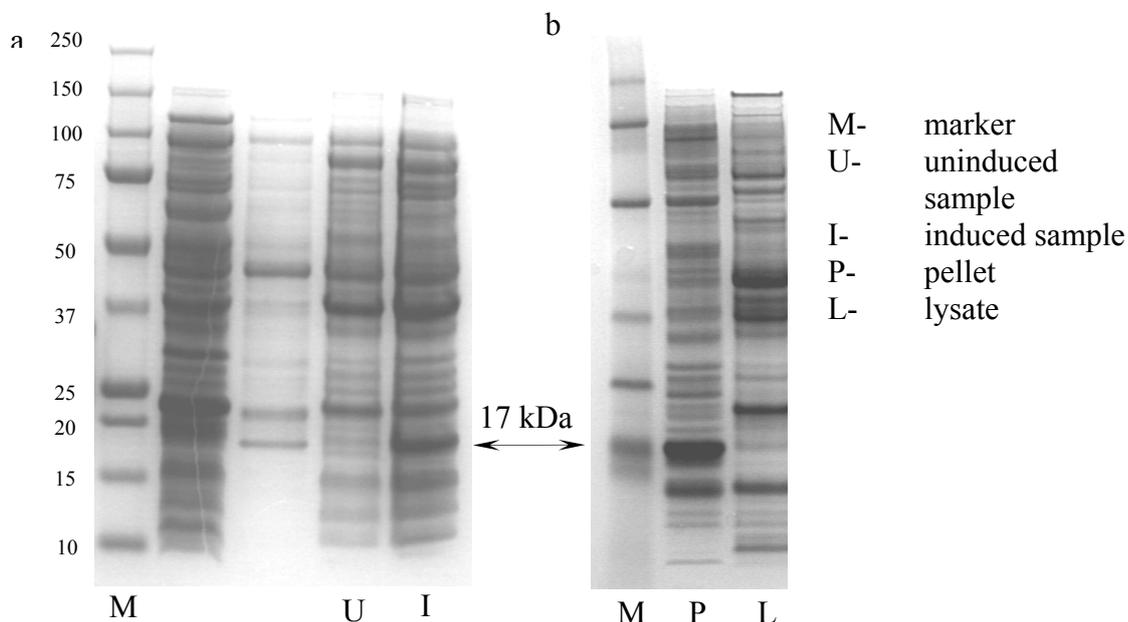
Fig 3.2   a: SDS-PAGE assay of the expression of PA-ICP in BL21 codon plus (DE3)-RIPL strains. b: The majority of the protein remained insoluble after lysis using bugbuster.

Amino acid sequence alignment of the ICP family of proteins revealed that PA-ICP has

a hydrophobic N-terminal extension composed of about 20 residues (fig 3.3). It has been recognized as a potential signal peptide with two possible cleavage sites by the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen et al., 1997). The hydrophibicity of the putative secretion signal peptide may result in misfolding and aggregation if not recognized by *E.coli* secretion mechanism (Baneyx, 1999). On the basis of the sequence alignment, deletion of the N-terminal extension is unlikely to disrupt the tertiary structure of PA-ICP and its function. Hence, a new construct was engineered to produce a PA-ICP truncated at residue Q23, resulting a His-tagged protein with a molecular weight of about 14.5 kDa. The new plasmid was expressed in BL 21 (DE3) cells and the cells were lysed. SDS-PAGE showed increases in expression level and solubility (fig 3.4).

```
                         1                    ↓                          50
P. aeruginosa ICP  (1)  MSFSPSRLLLPLSIVVLAGCAGQQKPVVTLDDADDCSPLKLTQGQELVLT
   L.mexicana ICP  (1)  -----------------------MIAPLSVKDNDKWVDTHVGKTTEIH
    T. brucei ICP  (1)  -----------------------MSHNLFTEEDNNKTIRMVIGETFTIE
     T. cruzi ICP  (1)  -----------------------MSHKVTKAHNGATLTVAVGELVEIQ


                        51                                              100
P. aeruginosa ICP (51)  LPSNPTTGFRWELRNPAASVLKR--LGPEVYSNSEEDSG--------LVG
   L.mexicana ICP (26)  LKGNPTTGYMWTRVGFVGKDVLSDEILEVVCKYTPTPSS------TPMVG
    T. brucei ICP (27)  LKSNPTTGYTWLRSGLAGTELSDCTFAIQSKFNNRAPHDNHKNHRRLLVG
     T. cruzi ICP (26)  LPSNPTTGFAWYFEGGTKESPNESMFTVENKYFPPDSKL---------LG


                        101                                            145
P. aeruginosa ICP (91)  SGGESTWRFRVAASGDDRLELVYRRPWEKDAEPAESFSCAIQVR-
   L.mexicana ICP (70)  VGGIYVVLVKPRKRGHHTLELVYTRPFEGIKPENERYTLHLNVK-
    T. brucei ICP (77)  AGGTMVLEVKALKXGKHTLSLAYGRPWVGFNAAAKRYNIHVEATA
     T. cruzi ICP (67)  AGGTEHFHVTVKAAGTHAVNLTYMRPWTGPSHDSERFTVYLKAN-
```

Fig 3.3   Amino acid sequence alignment of the ICP family. The N-terminal hydrophobic extension is coloured in red and the potential signal peptidase cleavage sites in blue. The arrow indicates to the N-terminal residue of the new construct.

M-   protein molecular weight
      markers
U-   uninduced cells
I-   induced cells
P-   pellet after lysis
L-   soluble lysate after lysis
FL-   flow through after Ni- column
W-   wash
E-   elution
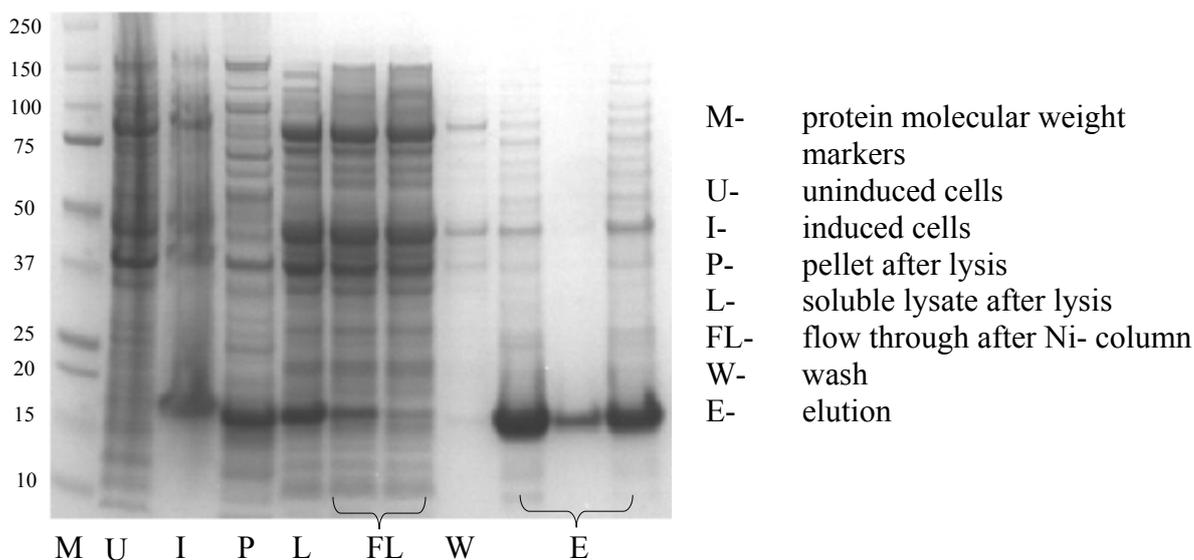
Fig 3.4   SDS-PAGE assay of the expression and Ni^{2+} affinity chromatography of
               the new construct in BL21 (DE3). The expression and solubility of PA-
               ICP were improved significantly.

3.2.2 Optimization of protein purification

The protein was first purified using the strategy as follows. After expression, the cells were harvested and lysed as described in section 8.2. With the aid of the His tag, the protein was first purified using immobilized metal ion affinity chromatography (fig 3.4), after which, the His tag was selectively cleaved using thrombin (fig 3.5), resulting in a protein of approximately 12.4 kDa. Thrombin and the cleaved His tag were then removed using ion-exchange chromatography (fig 3.6). Under reducing conditions, the protein appeared to be pure in SDS-PAGE and the quantity was sufficient for NMR study. However, the sample degraded after a few days, which did not meet the requirement of stability for NMR study. A 2D ^{15}N HSQC spectrum acquired on a ^{15}N uniformly-labelled sample at 308 K showed that many crosspeaks were nicely dispersed in the ^{1}H dimension, covering the frequency range from 6 to 11 ppm. it indicates these crosspeaks arise from amide groups in properly folded molecules. However, a cluster of broad signals were also observed in the region between 8 to 8.6 ppm in the ^{1}H dimension which is typical for the random coil chemical shifts, suggesting that a proportion of the protein molecules in the sample are in a denatured state (fig 3.8 a). The broadening of the signals may result from slow conformational change or/and an unexpected increase in the overall correlation time due to oligomerization. A more sophisticated purification strategy had to be investigated. Since the presence of the His tag does not interfere with the inhibition of CPs to PA-ICP (Sanderson et al., 2003b), the removal of the His tag

was skipped to avoid unexpected hydrolysis by thrombin. Also, ion exchange chromatography was replaced by size exclusion chromatography after Ni^{2+} affinity chromatography.

M-   protein molecular weight markers
E-   eluate from Ni^{2+} column
TC-   after thrombin treatment

M   E   TC
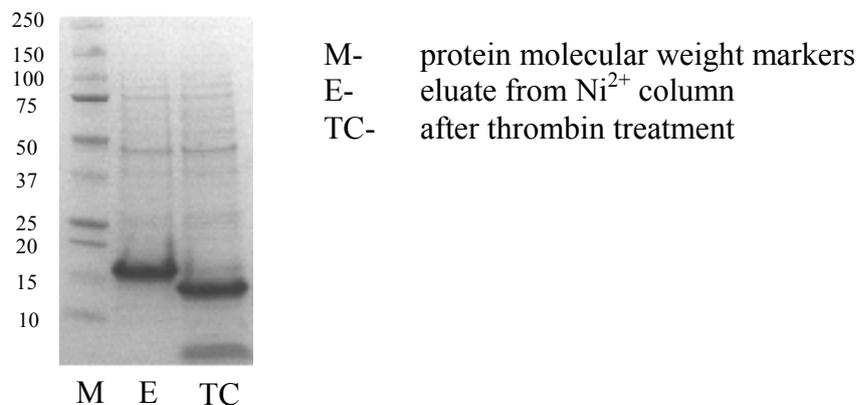
Fig 3.5   SDS-PAGE of His tag cleavage of PA-ICP by thrombin. After immobilized metal ion affinity chromatography, the protein sample were subject to thrombin treatment to remove the His tag.

The elution profile of gel filtration was monitored by absorbance at 280 nm and showed two peaks (fig 3.7). The elution volumes of the two peaks allowed the estimation of the protein molecular weight to be about 28 and 14 kDa respectively, which were close to the dimeric and monomeric size of the PA-ICP protein. Protein sequence revealed PA-ICP has two cysteines, C13 and C106. Examining the protein samples using SDS-PAGE revealed that the main protein band showing up at 28 kDa under non-reducing conditions was shifted to 14 kDa when the sample was reduced, indicating that the 28 kDa protein peak was an intermolecular disulphide bonded dimer of PA-ICP (fig 3.7). Also, there are two protein bands that show up in non-reducing SDS-PAGE both with molecular weights close to 14 kDa and the lower band merged into the upper one under reducing conditions. This implied that the two protein bands represented two redox states of PA-ICP monomers, an intra molecular disulphide bonded monomer migrating as the lower band and a reduced monomer as the upper band. The eluent containing the dimeric PA-ICP was examined using ^{15}N HSQC spectra (fig 3.8 b). The very broad peaks in the ^{15}N HSQC spectrum indicate that the dimer undergoes slow conformational exchange. The signals are clustered in the typical random coil chemical shift region, implying the dimer is unstructured, presumably due to the formation of the intermolecular disulphide bond.

The monomeric sample was chosen for structural study as there is no evidence showing that the ICP proteins have to be dimerized to function and the intermolecular disulphide bonded dimer is considered to be a side-effect of over expression. The majority of the crosspeaks in the ^{15}N HSQC spectra acquired on the intramolecular disulphide bonded and reduced monomers showed no significant changes in peak positions, indicating the redox states of the two cysteines do not have significant impact on the protein structure (fig 3.9). For structural studies, the protein sample is required to be homogenous. However, the disulphide bonded monomer did not convert to the reduced form at low pH at an appreciable rate even with 10 mM DTT present. Moreover, the degradation still occurred and could only be decelerated by the addition of 1 mM EDTA which, on the other hand, prevented the formation of the disulphide bonds. Hence, to make a stable homogenous sample for structure determination, the purified monomeric protein was left for over a week at 4 ^{0}C to allow sufficient oxidation of the protein and EDTA was added at this point to prevent degradation. The line width and dispersion of the crosspeaks in the ^{15}N HSQC spectrum collected on a ^{15}N uniformly labelled sample suggested that the optimized expression and purification protocols were sufficient to produce a homogenous PA-ICP sample stable for weeks for NMR study (fig 3.8 c).
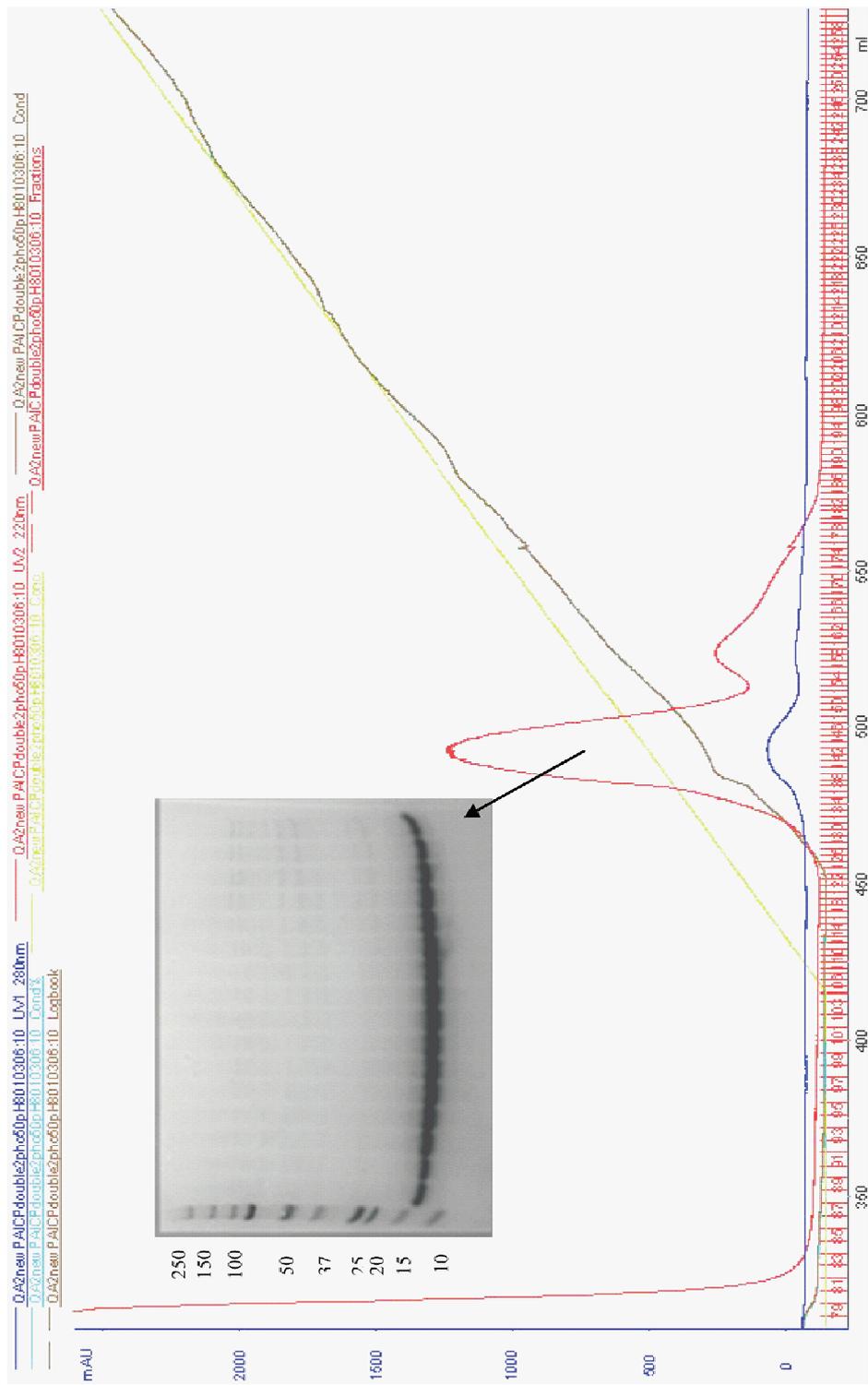
Fig 3.6   The elution profile of ion exchange chromatography of PA-ICP using anion exchange (-$CH_2N^+(CH_3)_3$) Q sepharose with a linear 0 to 1M salt gradient (50 mM NaH_2PO_4, pH 8.0).
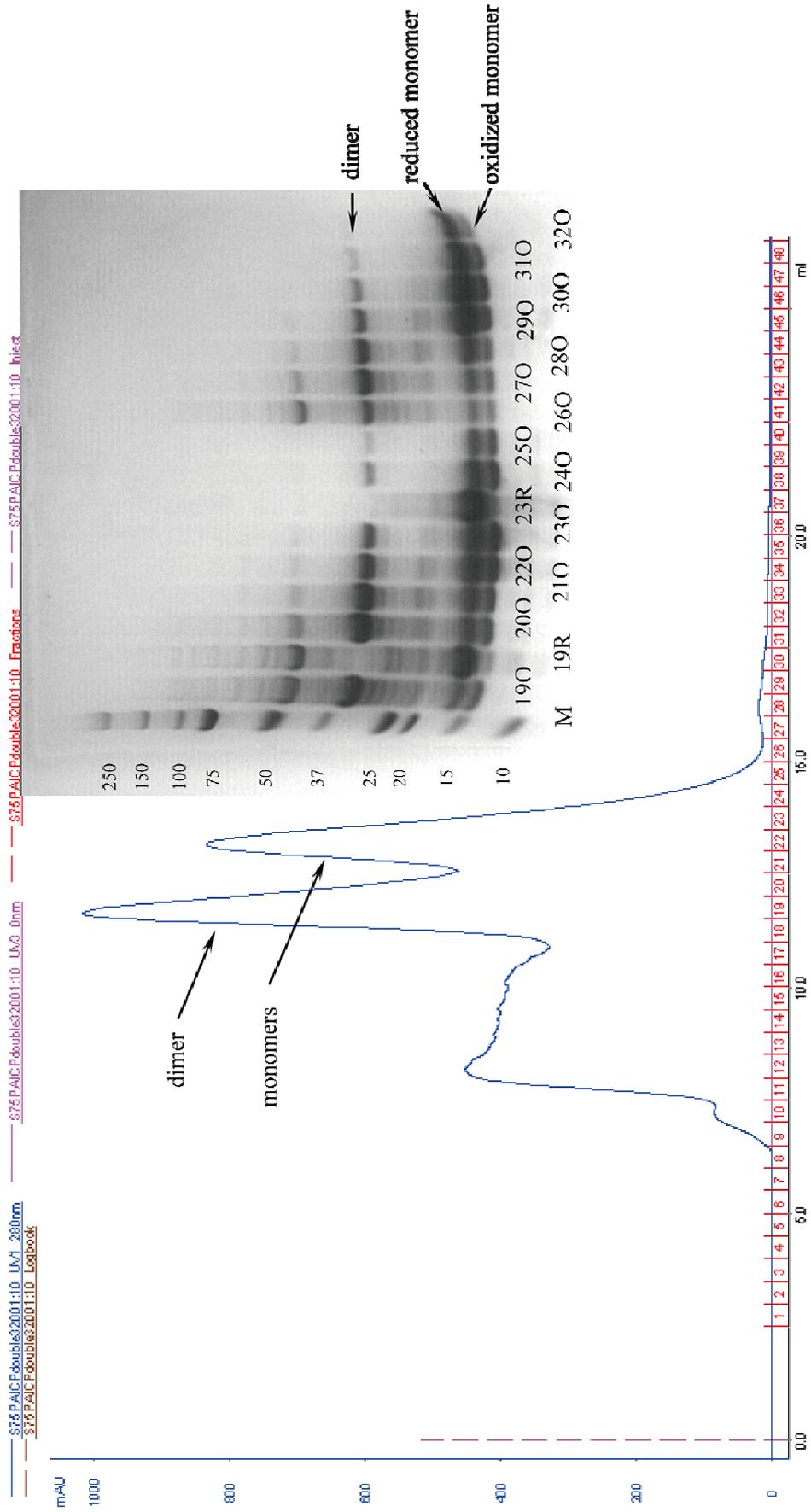
Fig 3.7   The elution profile of gel filtration of PA-ICP using a superdex 75 GL 10/300 column (25mM Na_2HPO_4, 75mM NaCl, pH 7.2). The intermolecular disulphide bonded dimer protein was eluted before the monomer mixture of intra molecular disulbhide bonded and reduced PA-ICP.
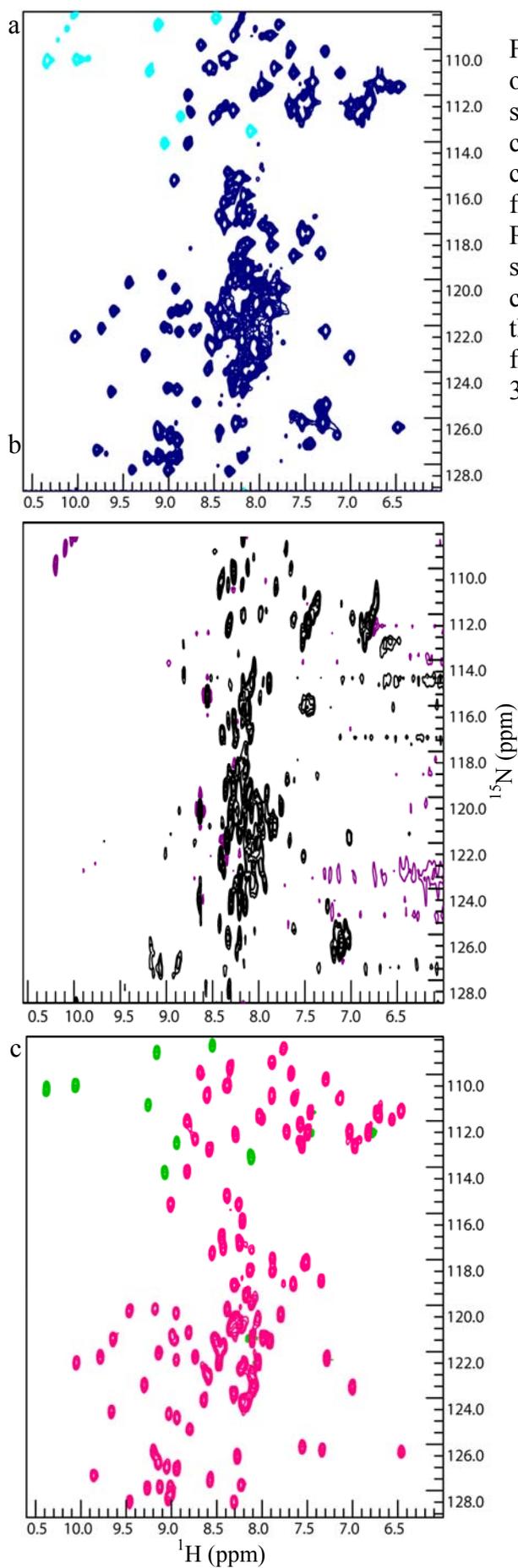
Fig 3.8 The ^{15}N HSQC spectra acquired on ^{15}N uniformly labelled PA-ICP protein samples purified using ion exchange chromatography (a) and size exclusive chromatography (b and c). The gel filtration eluent containing the dimeric PA-ICP gave rise to spectrum b while spectrum c was recorded on the sample concentrated from the eluent containing the monomeric protein a week after gel filtration. All spectra were collected at 308 K, pH 6.0
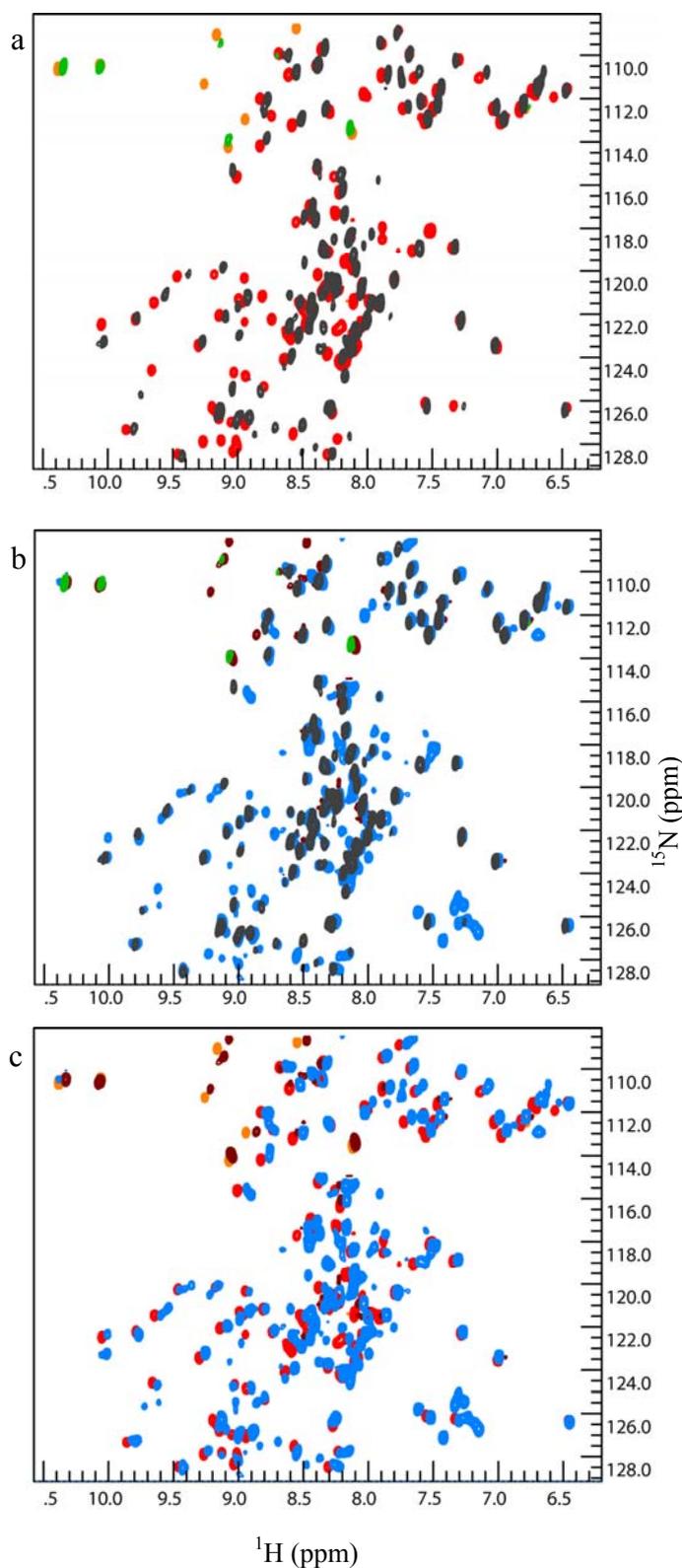
Fig 3.9 A comparison of the ^{15}N HSQC spectra of reduced (black and green), oxidized (red and orange) monomeric PA-ICP at pH 6.8 and 7.2 respectively and a mixture of both (blue and brown) at pH 5.0. The samples were purified using the optimized protocol. Despite the pH difference, the reduced and disulphide bonded monomers show no significant differences in the ^{15}N HSQC spectra for the majority of the crosspeaks. a: The ^{15}N-HSQC spectra of reduced in comparison with that of oxidised PA-ICP. b and c: the ^{15}N HSQC spectra of reduced (b) and oxidised PA-ICP (c) in comparison with that of the mixture of both.

^{1}H (ppm)

3.3 Redox studies of PA-ICP

As stated above, it was difficult to reduce the protein with 10 mM DTT at pH 7.2. Nevertheless, the protein could be reduced at pH 8.0. This is not surprising because DTT works more effectively under alkaline conditions. This is because the pKa of the thiol groups is typically 8.3 and only the negatively charged thiolate form -S$^-$ is reactive in DTT. The formation of disulphide bonds *in vitro* can be extremely slow even in the presence of oxygen or other strong oxidants due to the lack of enzymatic systems essential for formation of the disulphide bonds *in vivo* (Bardwell, 1994). Different methods were applied to make a stable, homogenously disulphide bonded monomer. An attempt to rearrange the disulphide bonds using the glutathione disulphide shuffling method (Clark, 1998) was not successful because the protein degraded after being oxidized (fig 3.10 a). The experiment also failed with the presence of 1 mM EDTA to prevent the degradation because the redox state remained unchanged (fig 3.10 b). This may be due to the chelation of the metal ion (copper in this case) needed for metal catalyzed disulphide bond formation by EDTA.

a

M-   protein markers
R-   PA-ICP samples run
      under   reducing
      conditions after
      glutathione shuffling
N-   PA-ICP sample run
      under non-reducing
      conditions after
      glutathione shuffling

b

M-   protein marker
N&R- PA-ICP protein sample
      run under non-reducing
      and reducing conditions
      respectively
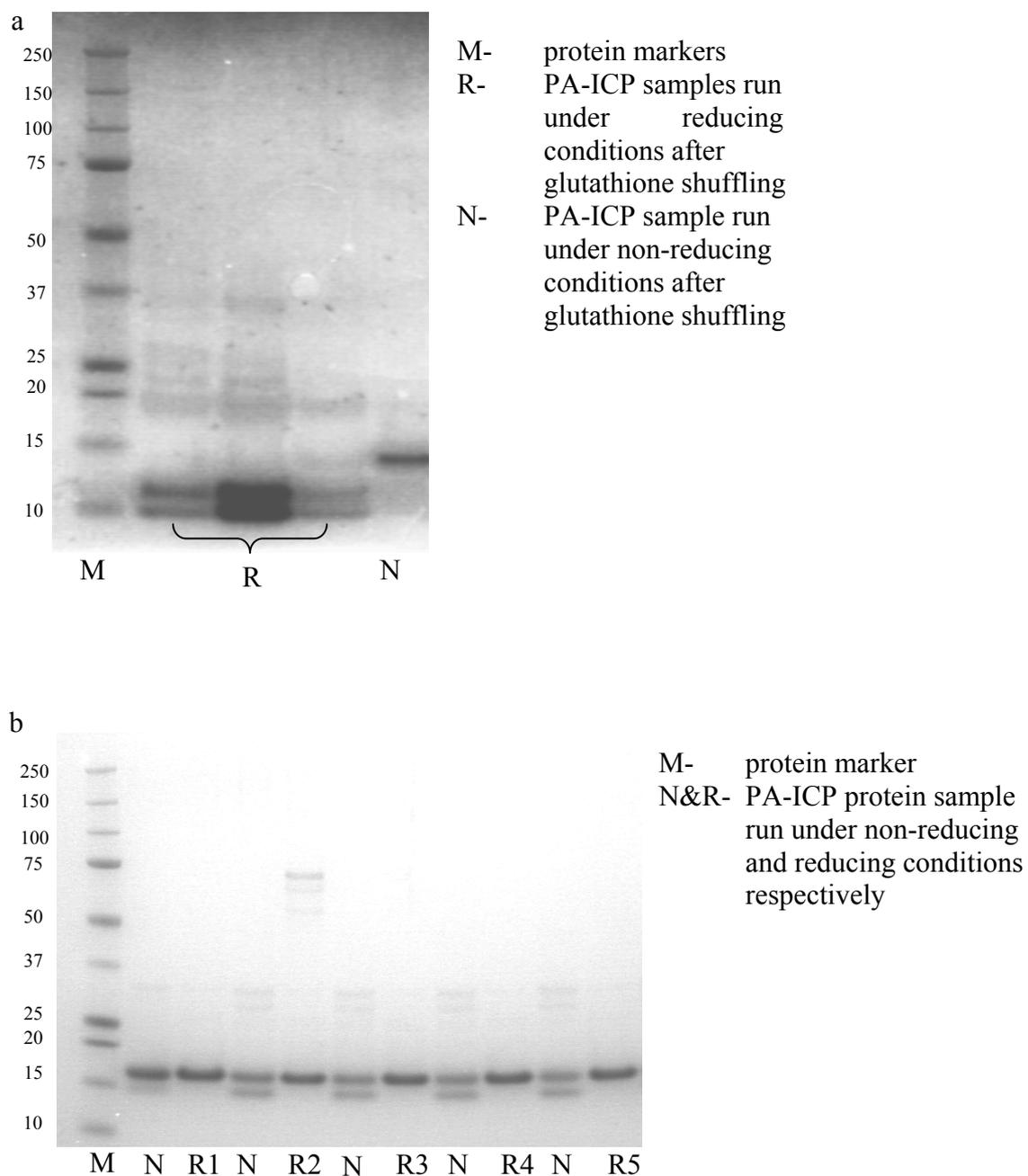
Fig 3.10   SDS-PAGE assays of a: Glutathione disulphide shuffling of the
            disulphide bond in PA-ICP. b: Glutathione disulphide shuffling of the
            disulphide bond in PA-ICP with addition of 1mM EDTA. 1-5: PA-ICP
            was treated with 5 different ratios of reduced and oxidized glutathione.
            Ratios of 10:1, 8:1, 6:1, 4:1, 2:1 were used with the concentration of
            oxidized glutathione kept to 1 mM.

3.4 Secretion of PA-ICP from *Pseudomonas aeruginosa* PAO1

The N-terminal hydrophobic region on PA-ICP was predicted to be a potential secretion signal peptide. The protein is thus likely to be exported by a general secretion system of *Pseudomonas aeruginosa* by which the protein is first targeted across the inner membrane into the periplasmic space before crossing the outer membrane (Kerr, 2000, Filloux et al., 1998). Therefore, the protein is expected to be detectable in the periplasm or the culture supernatant. The release of periplasmic material from *Pseudomonas aeruginosa* PAO1 was carried out using osmotic shock (Jensch and Fricke, 1997, Nossal and Heppel, 1966). *Pseudomonas aeruginosa* bacteria were first subject to a high concentration of sucrose, followed by a rapid reduction in the osmotic strength. The sudden change in osmotic stress of the cells could lead to the liberation of the periplasmic material without penetrating the cells. Assessment of the osmotic shock products and the overnight culture supernatant using western blot revealed that the majority of PA-ICP protein remained in the cells after the osmotic shock treatment and no detectable PA-ICP was found in the growth culture (fig 3.11). The efficiency of the release of periplasmic material by osmotic shock was assessed using acid phosphatase as a periplasmic marker protein. The enzyme assay showed acid phosphate exhibited higher activity towards the chromogenic phosphatase substrate in lysed cells after osmotic shock than released material, indicating that the release of the periplasmic material from *Pseudomonas aeruginosa* using osmotic shock was inefficient (fig 3.12).

M-	protein molecular weight markers
SC-	soluble lysate of the bacterial lysed after osmotic shock
PC-	pellet of the bacterial lysed after osmotic shock
R-	released periplasmic material from osmotic shock
OM-	overnight growth medium
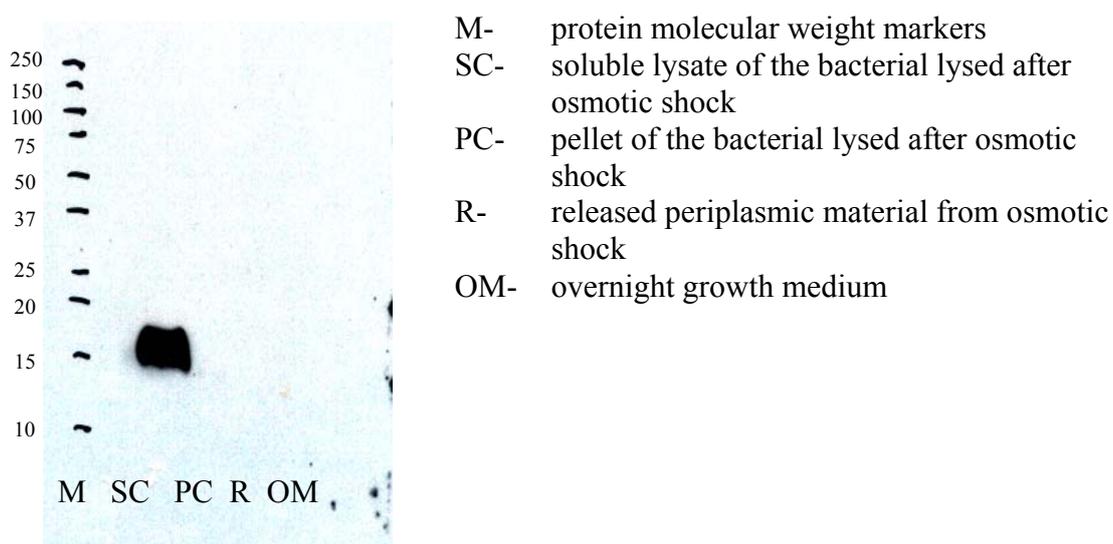
Fig 3.11 Detection of PA-ICP in osmotic shock products and growth medium using western blot.
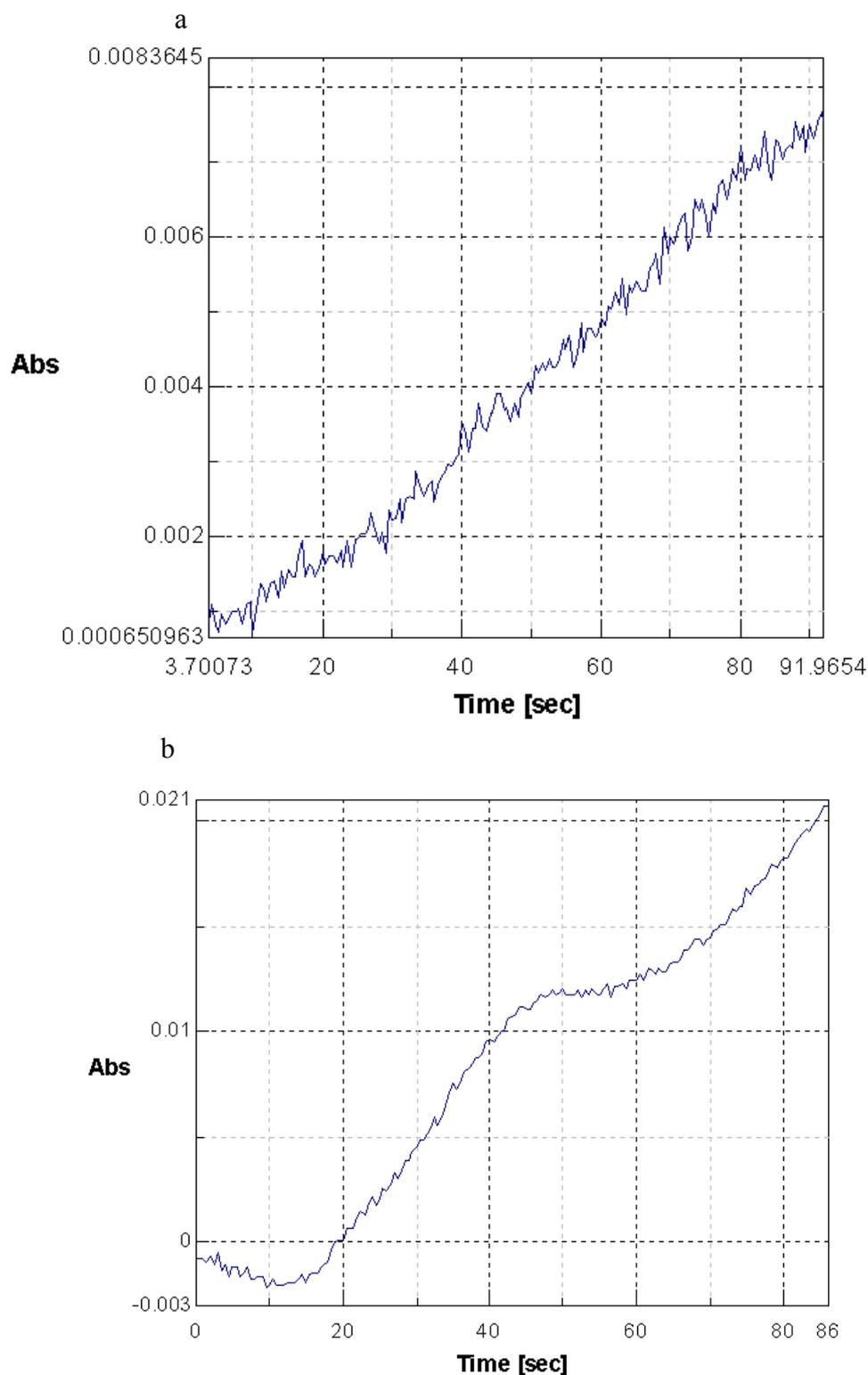
Fig 3.12   a: The measurement of the activity of acid phosphatase towards *npdh* in released material. An activity of 0.00510± 2.175e-04 units was observed. b: The activity of acid phosphatase towards *npdh* in lysed cells after osmotic shock was also measured with the activity reading of 0.02594±4.451e-04 units. The lysate was diluted to the same volume as the released material before measurement.

3.5 Inhibitory activity of PA-ICP against target CPs

Pseudomonas aeruginosa can infect the pulmonary tract, urinary tract, burns, wounds, and also causes other blood infections (Lyczak et al., 2000). At least 7 clan CA family C1 cysteine peptidases are identified in the lysosomes of the human lung (Buhling et al., 2004) which might be encountered by PA-ICP during *Pseudomonas aeruginosa* infection. They are cathepsins L, B, H, S, W and X, among which cathepsin H, L, B and S are commercially available. To assess the inhibitory specificity of PA-ICP against these cysteine peptidases, a series of enzyme assays were carried out.

A chromogenic substrate was used for the determination of the initial reaction velocities. A substrate peptide is synthetically attached to a chemical group which, when released after the enzyme cleavage, gives rise to colour. This gives excellent readings on the spectrophotometer for the measurement of the residual activity of the enzymes. However, in measurement of the dissociation constants which requires the detection of the functional activity of an enzyme at very low concentration, a fluorogenic substrate should be used. When interacting with the peptidase, the fluorescent AMC group of the substrate is cleaved and the fluorescence emission detected by a spectrofluorometer is typically several orders of magnitude more sensitive than the chromogenic substrate.

3.5.1 Determination of the functional concentration of CPs and PA-ICP

The PA-ICP protein was expressed as described above and then purified using immobilized metal ion affinity chromatography, followed by buffer exchange into lysis buffer using a PD10 column for storage.

The irreversible inhibitor E64 was used as a reference in the measurement of the functional concentrations of cathepsins and papain. The low molecular weight inhibitor E64 inhibits papain-like cysteine peptidases such as papain, cathepsin L, H and B, specifically and irreversibly at 1:1 ratio (Barrett et al., 1982). The crystal structure of E64 in complex with papain identified a covalent bond formed between the C_2 carbon atom of the E64-c oxirane ring and the S_γ atom of the papain active site's C25 (fig 3.13), occupying the S subsites of the enzyme (Varughese et al., 1989). Other low molecular weight thiol compounds or thiol dependant enzymes have little or no effect on the inhibition activity of E64, making it a good titrant for active site titration of papain-like cysteine peptidases (Barrett et al., 1982).
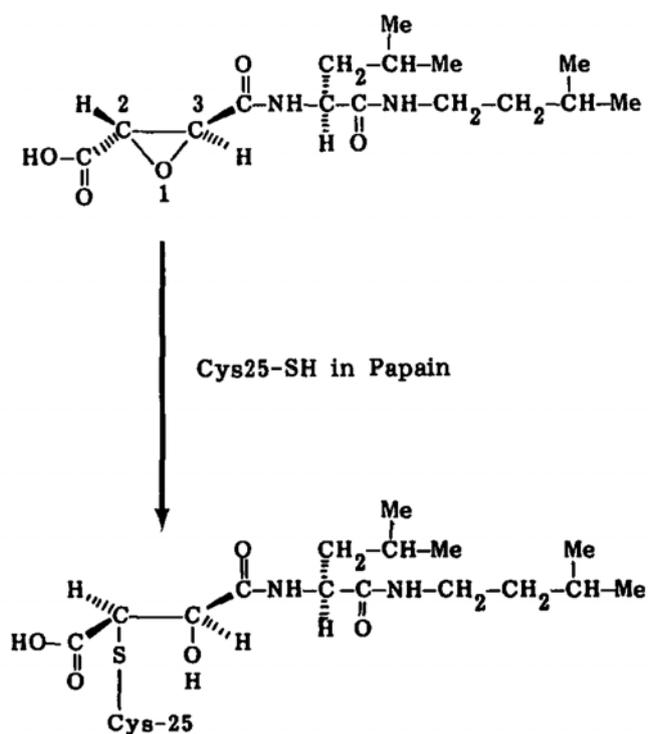
Fig 3.13   S-alkylation of the active site C25 of papain with opening of the epoxide ring of E64 (Matsumoto et al., 1989)

The functional concentrations of cathepsin L, B and papain could be determined using the chromogenic substrate. On the other hand, because of the small turnover number of cathepsin S for the substrate, the initial velocity of the reaction was not detectable with the chromogenic substrate unless the enzyme was used at very high concentration. Instead, the fluorogenic substrate z-FR-AMC was used.

Due to the cost of the reagents, assay points were limited to 4 or 5 to roughly estimate the active concentration of each enzyme. For irreversible inhibition, the activity should decline linearly with the increasing amount of inhibitor. For cathepsin L, B and S, it was easy to fit a straight line through the experimental points (fig 3.14). However, as reported in the literature (Barrett et al., 1982), a linear relationship between the activity and the residual molarity of cathepsin H was not observed due to the incompleteness of the reaction of the enzyme with E64 at low concentration of both reactants. Therefore, the active concentration of cathepsin H was not measured and was estimated according to the protein concentration provided by the manufacturer.

The functional concentration of papain was also measured by active site titration with E64 (fig 3.15 a). The enzyme was completely inhibited with 0.1 μM E64, indicating that its functional molarity was no higher than 0.1 μM. Having determined the functional molarity of papain, the concentration of active PA-ICP was determined using papain at known concentration in a pseudo-irreversible manner, assuming that the ratio of the reaction between PA-ICP and target CPs is 1:1 (Sanderson et al., 2003b) (fig 3.15 b). The concentrations of the enzyme and PA-ICP in the reaction are sufficiently high that the dissociation of the peptidase-inhibitor complex is negligible and the reaction can be considered practically irreversible. As 0.1 μM papain was completely inhibited by 1.07 μM PA-ICP (concentration measured by Bradford assay), there was 0.1 μM active PA-ICP in 1.07 μM protein sample. This could be due to the presence of the unfolded PA-ICP in the sample used for the assay contributing to the measurement of the total concentration of the protein because the protein sample was purified using only immobilized metal affinity chromatography.
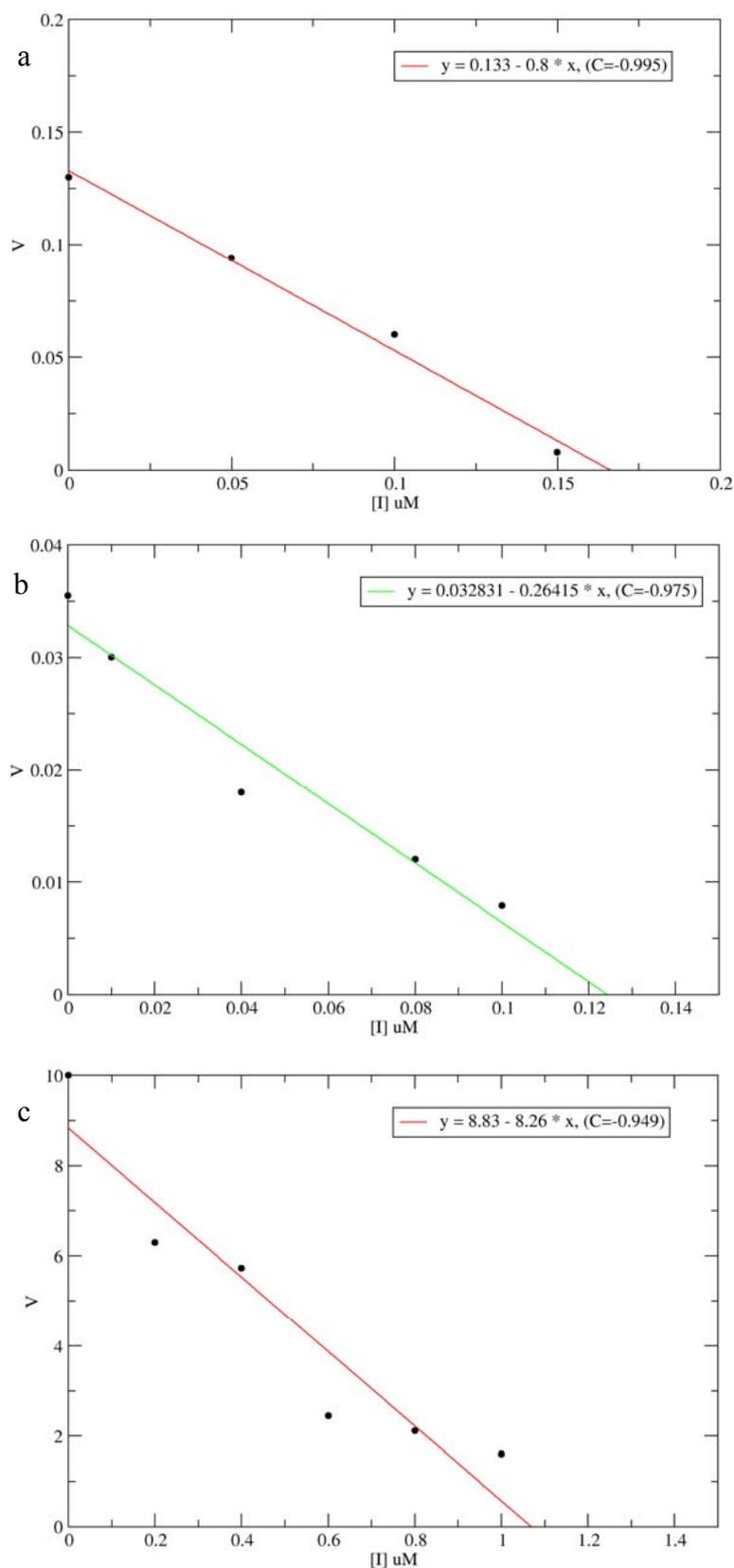
Fig 3.14   Active site titrations of cathepsin B (a), L (b) and S (c) with E64. The fitting functions are listed in the legend boxes with the fitting correlation coefficients C.
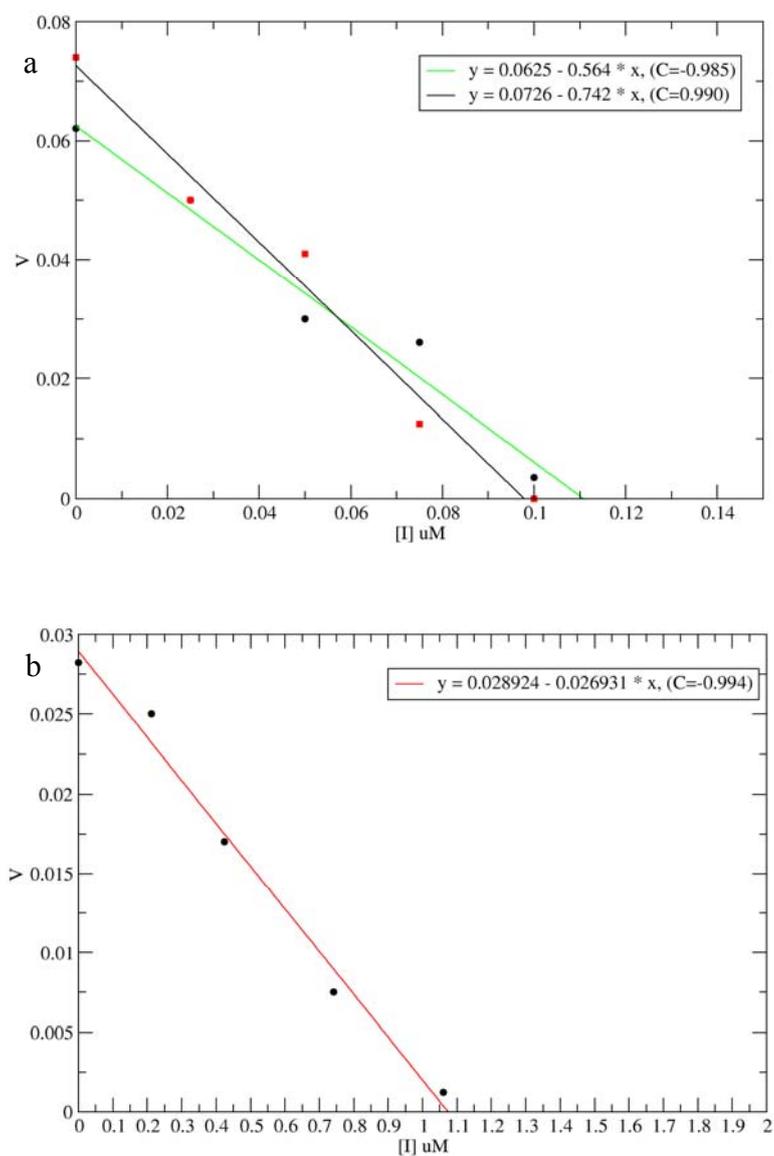
Fig 3.15   a: Active site titration of papain with E64. Two sets of independent sets of measurements are shown. b: Measurement of active concentration of PA-ICP with known concentration of papain. The fitting functions are listed in the legend boxes with the fitting correlation coefficients C.

3.5.2 Determination of dissociation constants K_i

With known active concentrations of PA-ICP and cathepsins, the dissociation constant of inhibition of PA-ICP to each cathepsin, K_i, can be determined as described in section 8.3.5.2 (Sanderson et al., 2003b). During pre-incubation of PA-ICP with CPs, the inhibitory activity of PA-ICP declined rapidly due to the enzymes' high proteolytic activities. Therefore, long time incubation of cathepsins with PA-ICP was avoided to minimize the hydrolysis of the inhibitor. Each experiment was performed in duplicate to allow error estimation except for cathepsin B (fig 3.16). The K_i values were calculated by fitting the experimental points to the following equation:

$$V = \frac{V_{max}}{1 + \frac{K_M}{[S]}(1 + \frac{[I]}{K_i})} \qquad (3.1)$$

The V_{max} of each reaction was calculated when [I] =0, thus,

$$V_{max} = V_0(1 + \frac{K_M}{[S]}) \quad (3.2)$$

where V_0 refers to the initial velocity of the reaction with no inhibitor present. The K_M values of the enzymes to the specific substrates were taken from literature and [S] was kept constant for each experiment (table 3.1).

cathepsin	B	L	H	S
substrate	z-FR-AMC	z-FR-AMC	H-R-AMC	z-FR-AMC
K_M µM	51	2.4	150	18.2
K_{cat} s^{-1}	364	17	2.5	1.6

Table 3.1   The K_M and K_{cat} values of cathepsin B, L, H and S toward fluorogenic substrates z-FR-AMC and H-R-AMC. (Buttle et al., 1988)
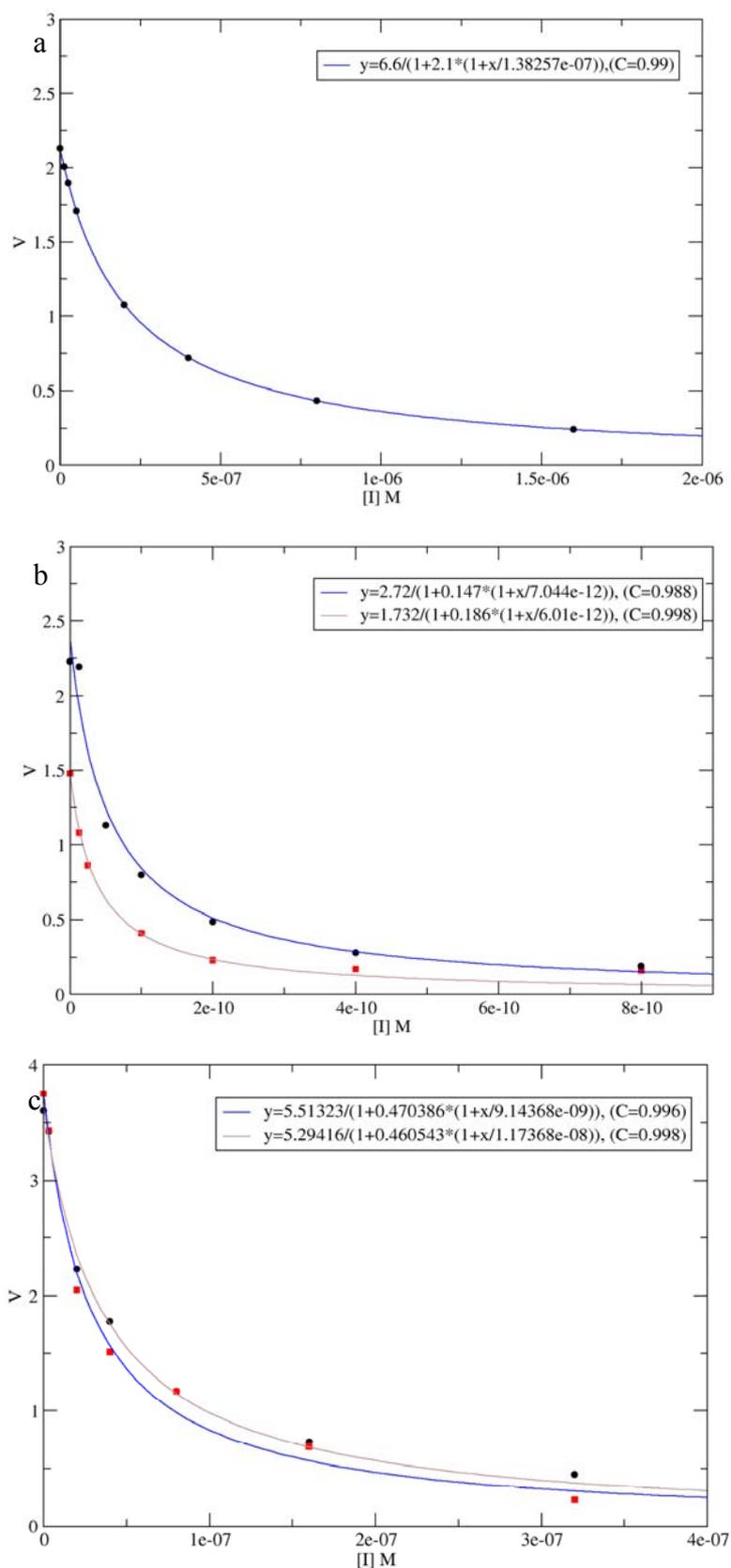
Fig 3.16 K_i determination of PA-ICP to cathepsin B (a), L (b) and H (c). Two sets of independent sets of measurements are shown for PA-ICP to cathepsin L and H. The fitting functions are listed in the legend boxes with the fitting correlation coefficients C.

The K_i values of PA-ICP toward cathepsin L, B and H fell into range of nM to pM (table 3.2). Like its homologs, chagasin and *L. mexicana* ICP, PA-ICP appeared to be more competent against cathepsin L-like enzymes than cathepsin B.

	Cathepsin L	Cathepsin H	Cathepsin B
PA-ICP	6.527 ± 0.517 pM	10.43 ± 1.3 nM	138 nM

Table 3.2   The K_i values of PA-ICP toward cathepsin L, B and H

The K_i determination for cathepsin S was less satisfactory (fig 3.17). The experimental points were more likely to fit to a straight line, indicating that under the assay conditions, PA-ICP interacts with cathepsin S pseudo-irreversibly.
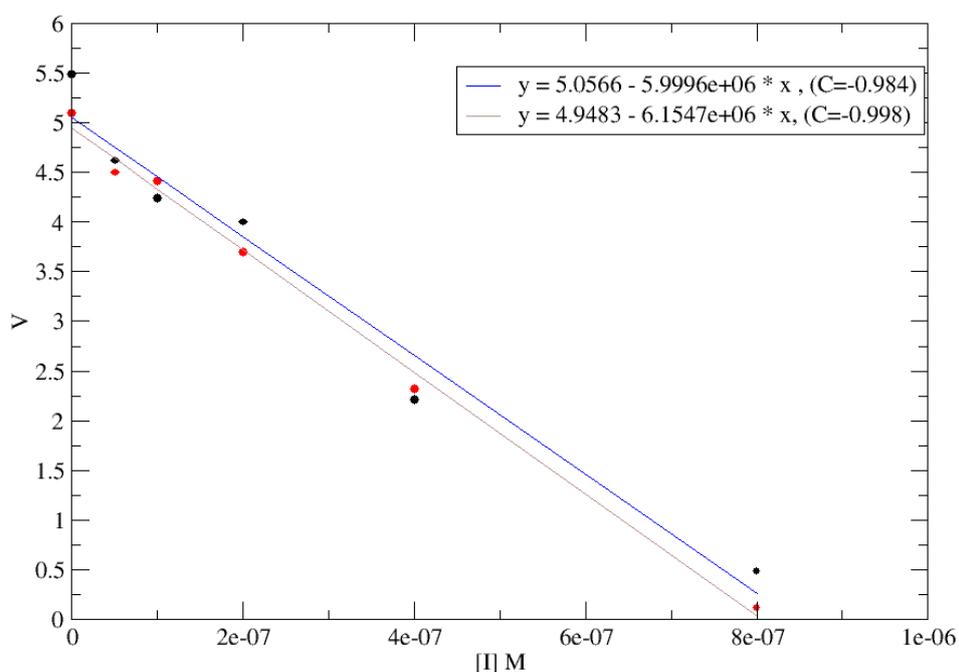
Fig 3.17 K_i determination of PA-ICP to cathepsin S. The fitting functions are listed in the legend boxes with the fitting correlation coefficients C.

3.6 Conclusion

The N-terminal hydrophobic extension of PA-ICP has been predicted to be a potential signal peptide. It is possible that the protein is synthesized as a proprotein and the signal peptide directs the export of the protein from *Pseudomonas*. This possibility, together with the finding that *Pseudomonas aeruginosa* appears to lack clan CA family C1 cysteine peptidase genes, supports the hypothesis that the real targets of PA-ICP, rather than being endogenous, are exogenous peptidases. Although PA-ICP was not detected in

the over night growth culture supernatant implies the protein is not secreted from bacteria into its surrounding environment, this discovery is limited by the sensitivity of the experiment as the concentration of the released protein in the medium may be too low to be detected. Nevertheless, since osmotic shock did not successfully liberate the periplasmic material from *Pseudomonas aeruginosa*, whether the protein is transported into periplasm remains unknown.

The redox states of the two cysteines in native PA-ICP are not yet determined. 3 forms of PA-ICP with different redox properties were present in the purified recombinant protein sample— the intermolecular disulphide bonded dimer, the reduced and the intramolecular disulphide bonded monomers. The intramolecular disulphide bonded monomer was chosen for structure determination because since the protein is believed to be translocated across the cytoplasmic membrane *in vivo*, it is more likely to end up in the periplasmic space and/or outside the bacteria, so, the functional protein is more likely to be disulphide bonded. Further more, the intramolecular disulphide bond is rather stable and cannot be dissociated with 10 mM DTT and the ^{15}N HSQC spectra of the intramolecular disulphide bonded and reduced monomers showed no significant changes in protein folding, suggesting the protein structure is not disturbed by the redox states of the two cysteines. The intermolecular dimer is believed to be a side-effect of over expression

The K_i values of PA-ICP to potential target cysteine peptidases fall into the pM to nM range. A more detailed comparison of the interaction of PA-ICP with different cysteine peptidases will be described in section 7.3. The K_i value of PA-ICP toward cathepsin L is very different from that determined by Sanderson, et al. (Sanderson et al., 2003b), which is 500 times greater than that determined in this project. This could be due to the different lengths in the enzyme-inhibitor incubation time (15 mins in Sanderson's experiments and 1 min in this project). As spotted for PA-ICP-cathepsin L interaction, the inhibitory activity of PA-ICP declined with increasing length of the incubation time. This may result from the hydrolysis process of the inhibitor and gives rise to a greater value in K_i determination. The K_i determination of cathepsin S failed because the small catalytic constant of the enzyme towards the substrate used in the assay. This can be improved by using a substrate towards which cathepsin S has a greater turnover number that can give rise to sensitive readings when working at low enzyme's concentrations.

CHAPTER 4

CHEMICAL SHIFT ASSIGNMENT OF PA-ICP

4.1 Overview

With the aid of advanced isotopic labelling techniques (uniform or fractional isotopic labelling with ^{13}C, ^{15}N and ^{2}H, selective labelling of amino acid types) and modern NMR instruments (powerful superconducting magnets and cryogenic probes), structure and dynamics investigations of a protein of less than 25 kDa is now more or less routine using standard NMR methods. NMR studies of PA-ICP were conducted in an effort to determine how the structure and dynamics of PA-ICP related to its inhibitory function toward different papain-like cystene peptidases. The work described in this chapter shows how the chemical shifts of the resonances of PA-ICP were assigned as a first step in this process. Triple resonance experiments were particularly advantageous in improving chemical shift resolution by exploiting ^{13}C and ^{15}N stable isotope labelling. A near complete assignment of PA-ICP has been obtained using this type of experiment and other heteronuclear NMR methods. The resonance assignments were subsequently used to derive interproton distances from NOESY spectra and these distance restraints were used to calculate the structure of PA-ICP (Chapters 5). The resonance assignments of the backbone amides were also used for dynamics studies (Chapter 6).

4.2 Optimization of the experimental conditions for NMR studies of PA-ICP

The protein samples for NMR studies were expressed as described in section 3.2 using labelled medium. After cell lysis, the protein was purified using a Ni^{2+} column followed by gel filtration chromatography. The samples were buffer exchanged into suitable NMR buffer and concentration to obtain a final volume of 570 µl. 30 µl D_2O was added to give a final concentration of 5% used for field frequency locking. NMR spectroscopy is a relatively insensitive technique. The protein sample needs to be very concentrated to obtain high-quality spectra in a practical length of time. For structural studies of PA-ICP, a sample concentration of 1 mM was used. Temperature is another important factor. A higher temperature results in faster molecular tumbling in solution and better sensitivity. Also, to avoid signal broadening caused by amide proton-water exchange, a low pH condition is preferable. The use of a cryogenic probe to increase the signal to noise ratio is subject to the condition that the sample under study has very low electrical conductivity and thus a low ionic strength is required. Having satisfied all the abovementioned conditions, a good NMR sample should be soluble and stable during NMR data acquisition which can typically take weeks. Therefore, different experimental temperatures and buffer conditions have to be tested beforehand. This can be done conveniently using ^{15}N-labelled sample. A series of ^{15}N HSQC experiments recorded

under different conditions were monitored to determine which conditions gave the best signal sensitivity without disturbing the protein structure. For PA-ICP, eleven different temperatures varying from 288 to 310K (15-37 ^{0}C) were investigated. The pH of 4.5, 6, 6.8, 7.2 and 8 and NaCl concentrations of 50, 100, 150 and 200 mM were also tested. The best of the buffer and temperature conditions tested for NMR studies of PA-ICP were 25 mM NaH_2PO_4, 100mM NaCl, pH 7.2 at 298 K (25 ^{0}C) for the His-tagged PA-ICP protein used for structural investigation and pH 6 at 308 K (35 ^{0}C) for the non-His-tagged protein used for backbone dynamics characterization. After determination of the optimal buffer and temperature conditions, a uniformly ^{15}N-^{13}C labelled sample was produced for the structural and dynamics studies of PA-ICP using NMR. A 1D ^{1}H spectrum acquired on an unlabelled, His-tagged sample is displayed in figure 4.1. The ^{1}H linewidths and the overall dispersion of chemical shifts suggest that the sample contains folded protein.
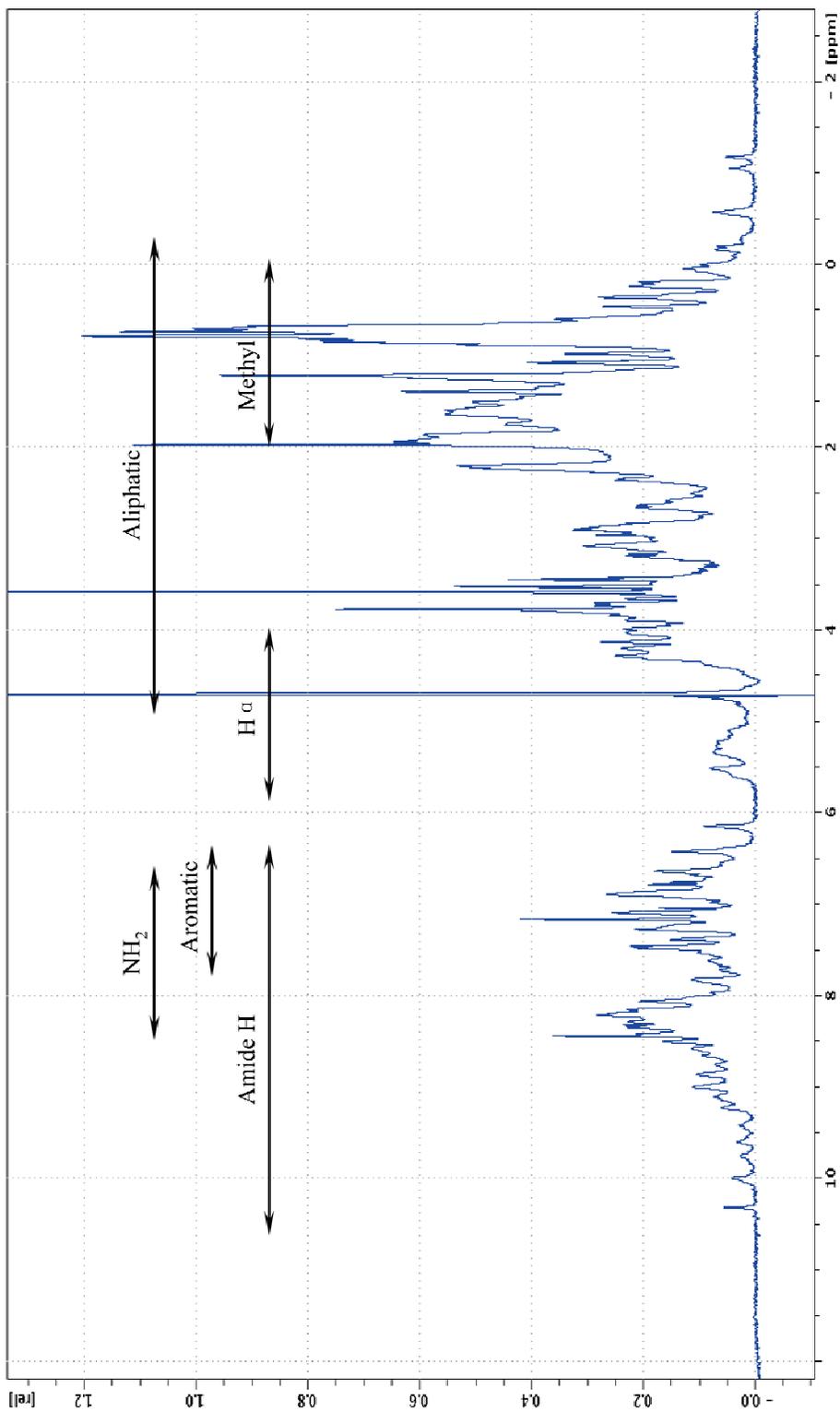
Fig 4.1   1D ^{1}H spectrum of PA-ICP acquired on an unlabelled, His-tagged sample (25 mM Na$_2$HPO$_4$, 75 mM NaCl, pH 7.2, 298K). Water suppression was achieved using a method described by Hwang and Shaka (Hwang and Shaka, 1995). The chemical shift distribution of different proton groups is displayed above.

4.3 Sequence specific assignment

4.3.1 Chemical shift assignment of the backbone resonances

The sequence specific resonance assignment of PA-ICP was achieved relying mainly on a pair of 3D NMR triple resonance experiments, HNCACB (Wittekind and Mueller, 1993)and CBCA(CO)NH (Grzesiek and Bax, 1993)acquired using a nonlinear sampling scheme (see section 4.3.2). The experiments are designed to transfer the magnetisation through the backbone and C_β atoms of adjacent residues. In the HNCACB, the magnetization starts from amide proton, is transferred to amide nitrogen and $^{13}C_{\alpha/\beta}$ and then transferred back to the amide proton for acquisition (fig 4.2). For CBCA(CO)NH, the magnetization starts at $^1H_\alpha/H_\beta$ and transferred to $^{13}C_{\alpha/\beta}$ before transferred to the amide proton where the signal is recorded (fig 4.2). As can be seen in fig 2.1, the scalar coupling constants between amide nitrogen and the $^{13}C_\alpha$ nuclei of the i and of the i-1 amino acid are close (11 Hz and 7 Hz respectively). Therefore, in the HNCACB experiment, the magnetization can be transferred from the nitrogen to $^{13}C_{\alpha/\beta}$ of the same and of the preceding amino acid, making it useful for providing intra and interresidue connectivities (fig 4.3). However, it is still difficult to distinguish the intra and interresidue $^{13}C_{\alpha/\beta}$ chemical shifts with only the HNCACB experiment. Hence, a CBCA(CO)NH spectrum giving the correlations between only the $^{13}C_{\alpha/\beta}$ of the preceding amino acid and the amide moiety, was also recorded (fig 4.3). By comparison of the two spectra, the $^{13}C_{\alpha/\beta}$ signals from residue i and i-1 were distinguishable and a unique assignment along the backbone of the protein could be completed. With this method, the resonances of the side chain $^{13}C_\beta$ nuclei are also assigned, from which the assignments can be extended to other side chain carbons and protons. An advantage of this method is that the backbone amide ^{15}N and ^1H atoms are correlated to the $^{13}C_\alpha$ and $^{13}C_\beta$ pairs which assists residue type assignment because the amino acid type of certain residues can be determined by the chemical shifts of the $^{13}C_{\alpha/\beta}$ pairs (Grzesiek and Bax, 1993), allowing the spin system to be located in the primary structure of the protein. Another pair of 3D triple resonance experiments, HNCO (Clubb et al., 1992) and HN(CA)CO that correlates the amide NHs and backbone carbonyls, provides an alternative route for backbone assignment. They were also acquired in case of overlaps of the crosspeaks of $^{13}C_\alpha$ and $^{13}C_\beta$ pairs in the HNCACB and CBCA(CO)NH spectra. Using this method, a total of 94% of backbone amide resonances of the PA-ICP sequence (appendix A) were assigned, excluding the proline residues and His tag (fig 4.4).
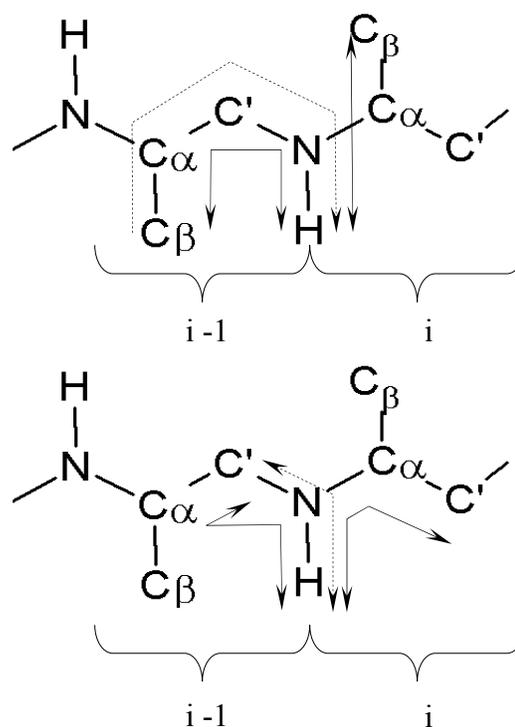
Fig 4.2   Magnetization transfer schemes of HNCACB (solid lines), CBCA(CO)NH (dashed lines) in figure a and HN(CA)CO (solid lines), HNCO (dashed lines) in figure b. The experiments correlate the backbone amide NH with related carbonyls and side chain C_α and C_β of the intra and proceeding residues and the signals were acquired on amide [1]H.
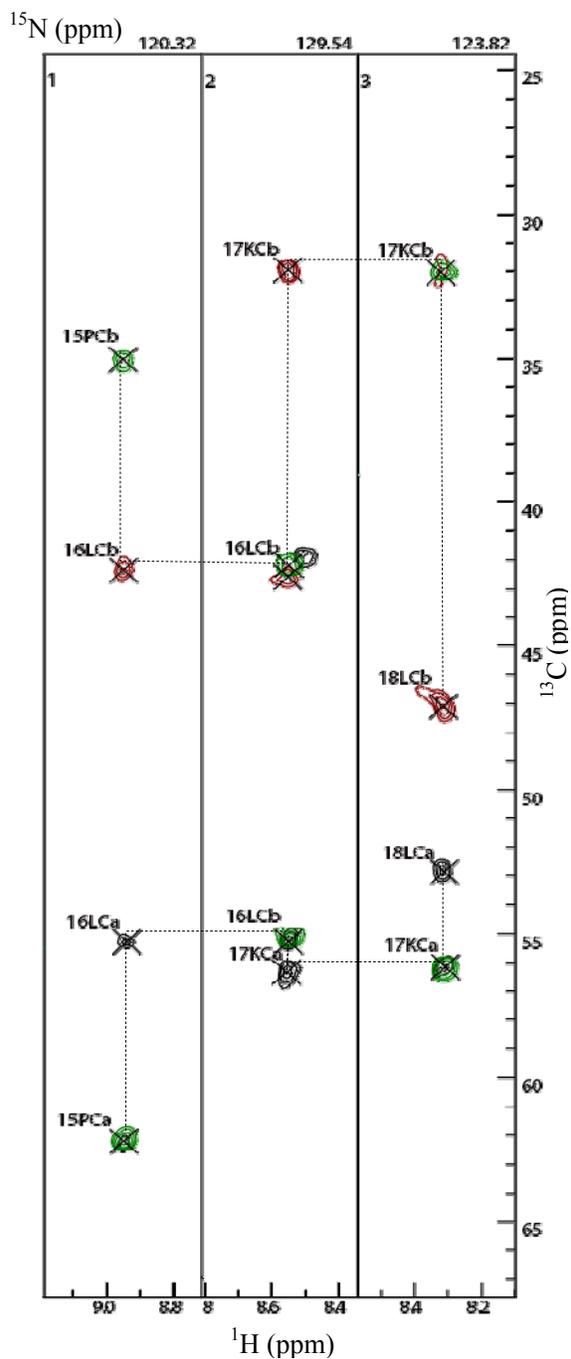
Fig 4.3   Backbone $^{13}C_{\alpha/\beta}$ chemical shift assignment of residues P15-L18 of PA-ICP (25 mM Na$_2$HPO$_4$, 75 mM NaCl, pH 7.2, 298K) using 3D HNCACB (black and brown) and CBCA(CO)NH (green) spectra with ^{1}H in x, ^{13}C in y and ^{15}N in z dimension . The connectivity was made by matching the $^{13}C_{\alpha/\beta}$ chemical shifts in the HNCACB which correlate to the HN chemical shifts of residue i with $^{13}C_{\alpha/\beta}$ chemical shifts in CBCA(CO)NH which correlate to the HN chemical shifts of residue i-1.
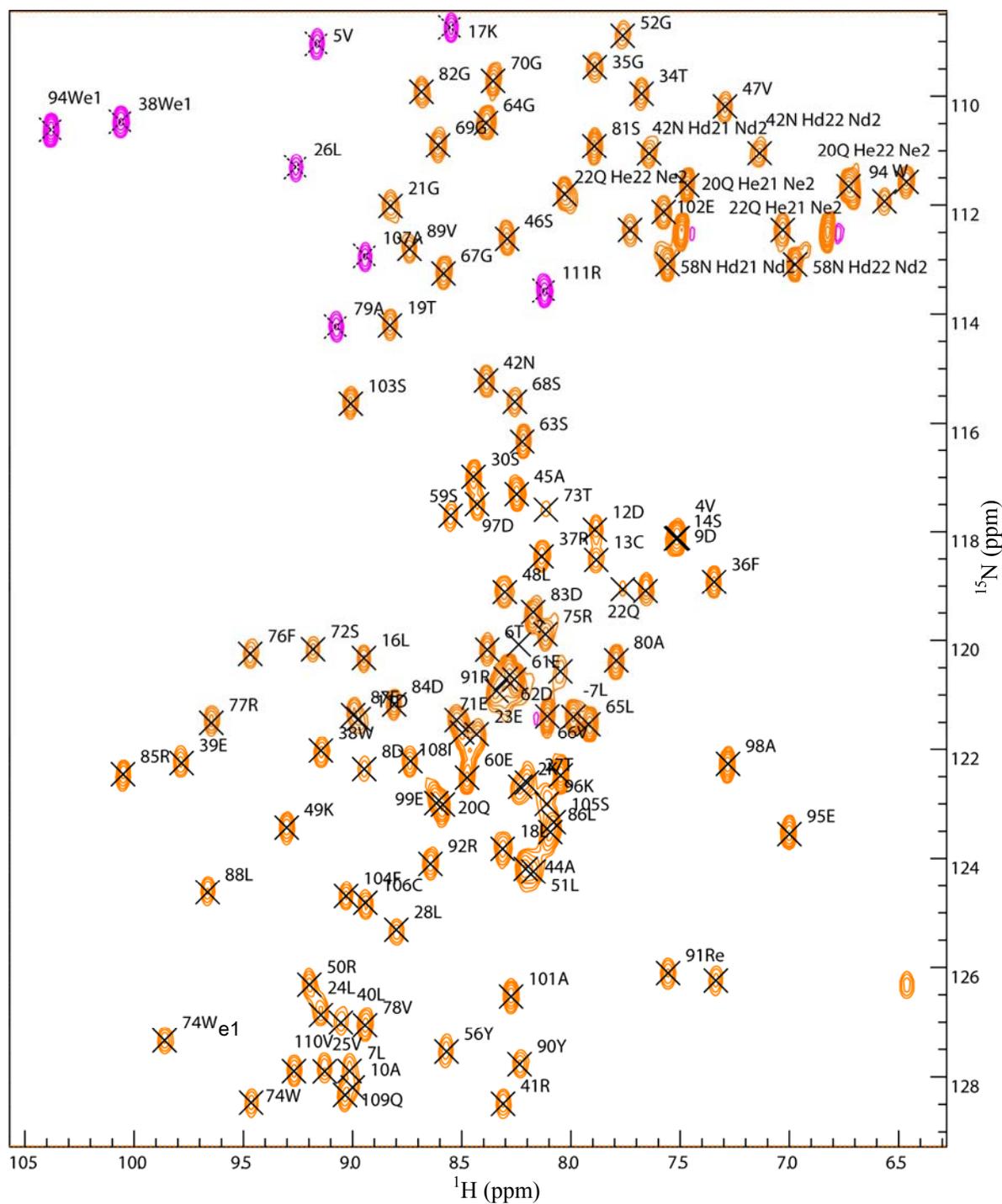
Fig 4.4   A ^{15}N HSQC spectrum of a ^{15}N labeled PA-ICP sample (25 mM Na$_2$HPO$_4$, 75 mM NaCl, pH 7.2, 298K) annotated into the residue specific assignment.

4.3.2 Nonlinear sampling

Multidimensional NMR experiments have been widely used for structural study of proteins in solution. In order to overcome the inherent spectral overlap problems that rise, the digital resolution along each dimension should be maximized to establish unambiguous data. This can result in a long experimental time of weeks to obtain all the multidimensional experiments required for protein structure determination. In particular, in the case of studying unstable or partially unfolded proteins, a significant reduction in experimental time is desired. Many efforts have been focused on accelerating the acquisition of the experiments to shorten the experimental time and/or to afford a greater number of scans for better sensitivity (Freeman and Kupce, 2003). Unlike conventional triple resonance experiments which utilize uniform time intervals for sampling chemical shift evolution, the nonlinear sampling scheme relies on the fact that the actual determinant of resolution is the longest sampling interval, and therefore the desired resolution can be achieved by sampling a selection of data points provided that the longest sampling interval is extended to retain the required resolution (Marion, 2005, Schmeider et al., 1994, Schmieder et al., 1993, Rovnyak et al., 2004). This results in the same length of required sampling period but with a reduced number of time-domain data points. As the time to acquire the spectra is mainly dictated by the desired number of sampling points in the indirectly detected dimensions, the nonuniform sampling scheme is only applied along ^{13}C and ^{15}N dimensions of the 3D triple resonance experiments. The sampling schedule needs to be designed with an awareness of the different properties between constant time and non-constant time experiments. In a non-constant time experiment, the intensity of the signal decays exponentially due to relaxation during chemical shift evolution. Therefore the effective signal to noise ratio is higher for the initial data points than for the last ones. Hence, in this type of experiments, fast and slow sampling rates are used to record the early data points and the later ones respectively to obtain the maximum signal to noise ratio. However, in a constant time experiment, the chemical shift evolution occurs in a fixed delay and the signal to noise ratio remains the same for all points. Therefore, the sampling points can be selected arbitrarily, provided that the last point is included to satisfy the longest sampling interval allowed by the constant time period (Schmeider et al., 1994).

The pulse programs of nonlinear sampling versions of HNCACB and CBCA(CO)NH were modified by Dr. B Smith (Division of Biochemistry and Molecular Biology, University of

Glasgow) based on the standard protocols described by Wittekind (1993) and Grzesiek (1993). The HNCACB experiment uses constant time in the ^{15}N dimension and non-constant time in the ^{13}C dimension whereas the CBCA(CO)NH experiment uses constant time in both dimensions. To favour both cases, the nonlinearly sampled HNCACB and CBCA(CO)NH experiments were recorded with 16 complex points (0, 1, 2, 3, 4, 5, 6, 7, 9, 11, 13, 15, 19, 23, 27, 30), selected out of 31 complex points in the nitrogen dimension and 27 complex points (0, 1, 2, 3, 4, 5, 6, 7, 9, 11, 13, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63, 67, 71, 75) selected out of 76 complex points in the carbon dimension (fig 4.5). Sweep widths were 1265 Hz in the ^{15}N and 11312 Hz in the ^{13}C dimension and maximum acquisition times of 11.8 and 3.3 ms respectively. A comparison of the non-linear and linear sampled HNCACB spectra is displayed in figure 4.6.

The nonlinear sampling method resulted 5-fold reduction in acquisition time which was partially invested in better signal to noise by doubling the number of scans while both indirect dimensions had effectively twice the digital resolution compared with standard practice. The proton dimension of the spectra was processed with the standard Fourier transform method while both indirect dimensions were reconstructed using Maximum Entropy (MaxEnt) to generate 256 points in the ^{13}C and 64 points in the ^{15}N dimension.

a

^{15}N Sample Index

b

^{15}N Sample Index

^{13}C Sample Index
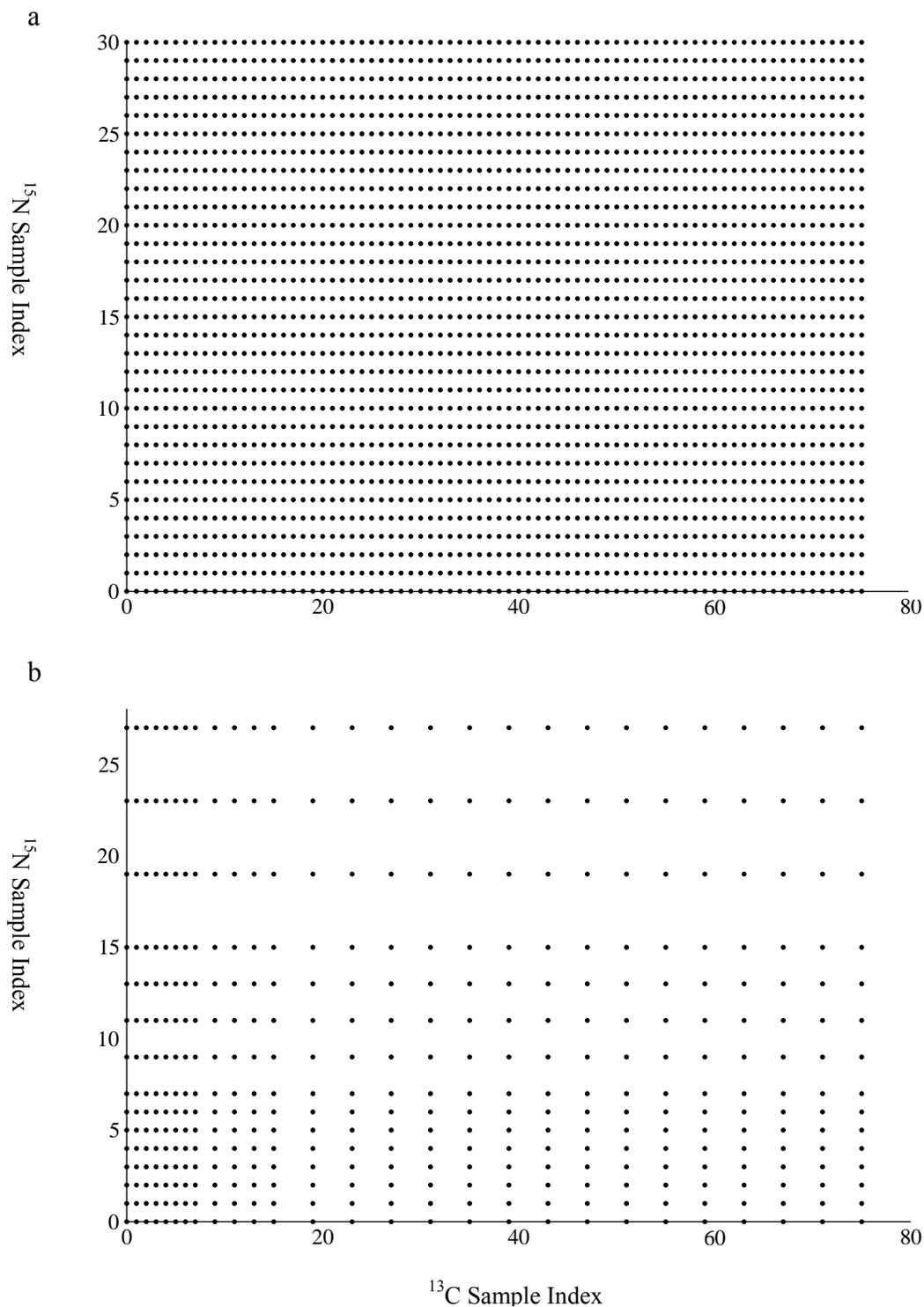
Fig 4.5   A representation of linear (a) and non-linear (b) sampling schemes employed in the HNCACB and CBCA(CO)NH experiments as a pattern of dots falling the ^{13}C and ^{15}N evolution periods, each of which represents four 1D spectra phase cycled for quadrature detection.
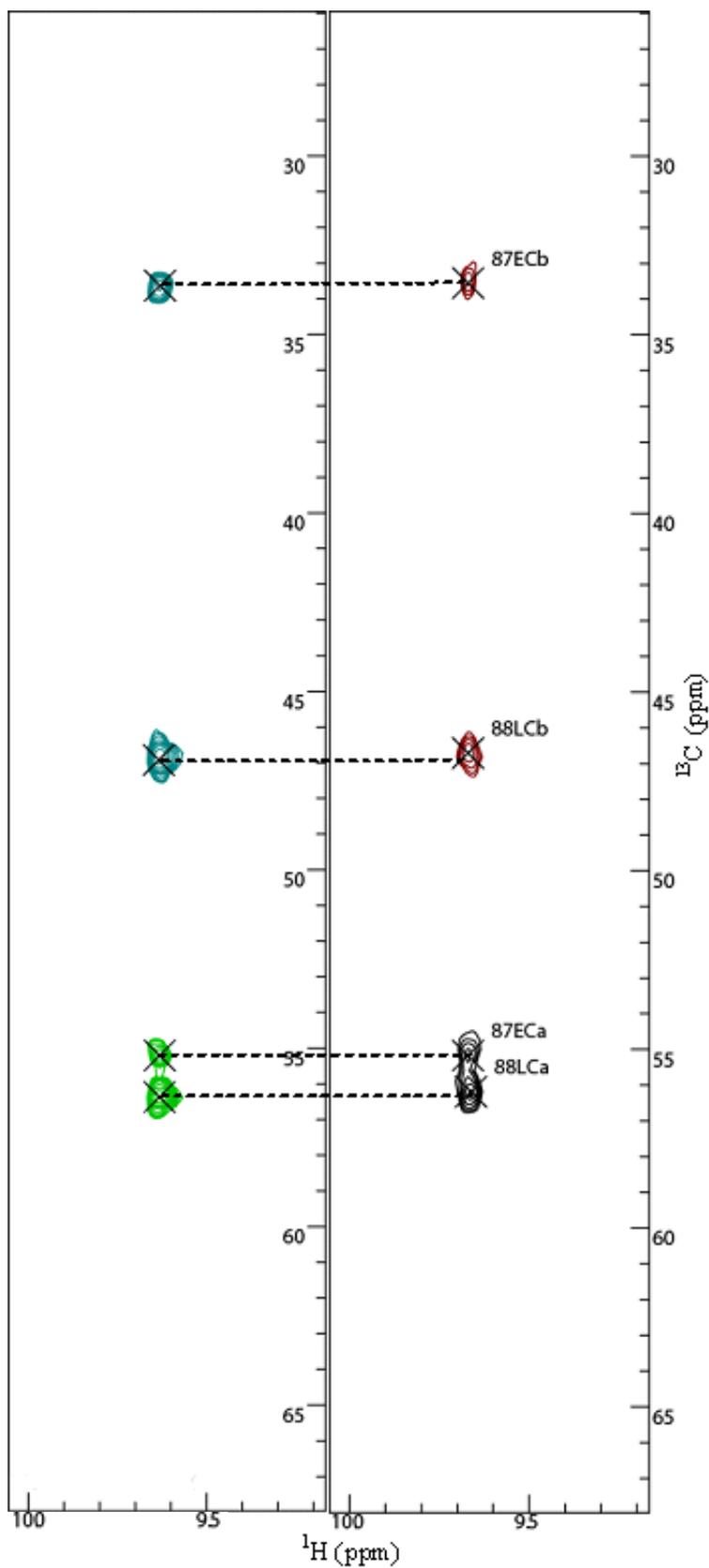
Fig 4.6   Comparison of the non-linear (left) and linear sampled HNCACB spectra (right). The strips illustrate the chemical shift assignment of side chain carbons of L88 and E87, extended from the amide HN crosspeaks of L88 on the HN planes of the two HNCACB spectra.

4.4 Side chain assignment

4.4.1 Chemical shift assignment of the aliphatic side chains

The aliphatic side chain carbons of PA-ICP were assigned using 3D CC(CO)NH-TOCSY (Montelione et al., 1992), which correlates the backbone NH chemical shifts with the preceding side chain ^{13}C chemical shifts (fig 4.7). Because the backbone NHs and the side chain $^{13}C_{\alpha/\beta}s$ had been assigned using HNCACB and CBCA(CO)NH, it was easy to navigate from the assigned amide of residue i to the correlated side chain carbons of residue i-1 and complete the assignment. On the basis of the side chain carbon assignments, the side chain protons were assigned using 3D HCCH-TOCSY (Kay et al., 1993), an experiment correlating each side chain ^{13}C with all ^{1}Hs in a spin system (fig 4.8).

For residues which do not have backbone amide chemical shift assignment (e.g., proline residues and whose backbone amide HNs were unable to be observed in the ^{15}N HSQC spectrum), their side chain ^{13}C chemical shifts could still be assigned from the assigned amides of the adjacent residues using 3D CBCA(CO)NH and CC(CO)NH experiments and the assignment can be extended to side chain ^{1}H using 3D HCCH-TOCSY. This allowed the side chain ^{13}C and ^{1}H chemical shift assignment of residue Q1, N31, T33, N58 and all prolines.

At this stage, the methyl groups of leucine, valine and isoleucine were assigned non-stereospecifically.

4.4.2 Chemical shift assignment of the aromatic side chains

The aromatic side chain ^{13}C and ^{1}H of PA-ICP were assigned using 2D and 3D ^{13}C-NOESY spectra with the aid of 2D (HB)CB(CGCD)HD and 2D (HB)CB(CGCDCE)HE (Yamazaki et al., 1993) which correlate the side chain C_{β} chemical shifts with the ring $^{1}H_{\delta}$ / $^{1}H\epsilon$ resonances of the aromatic residues (fig 4.9).
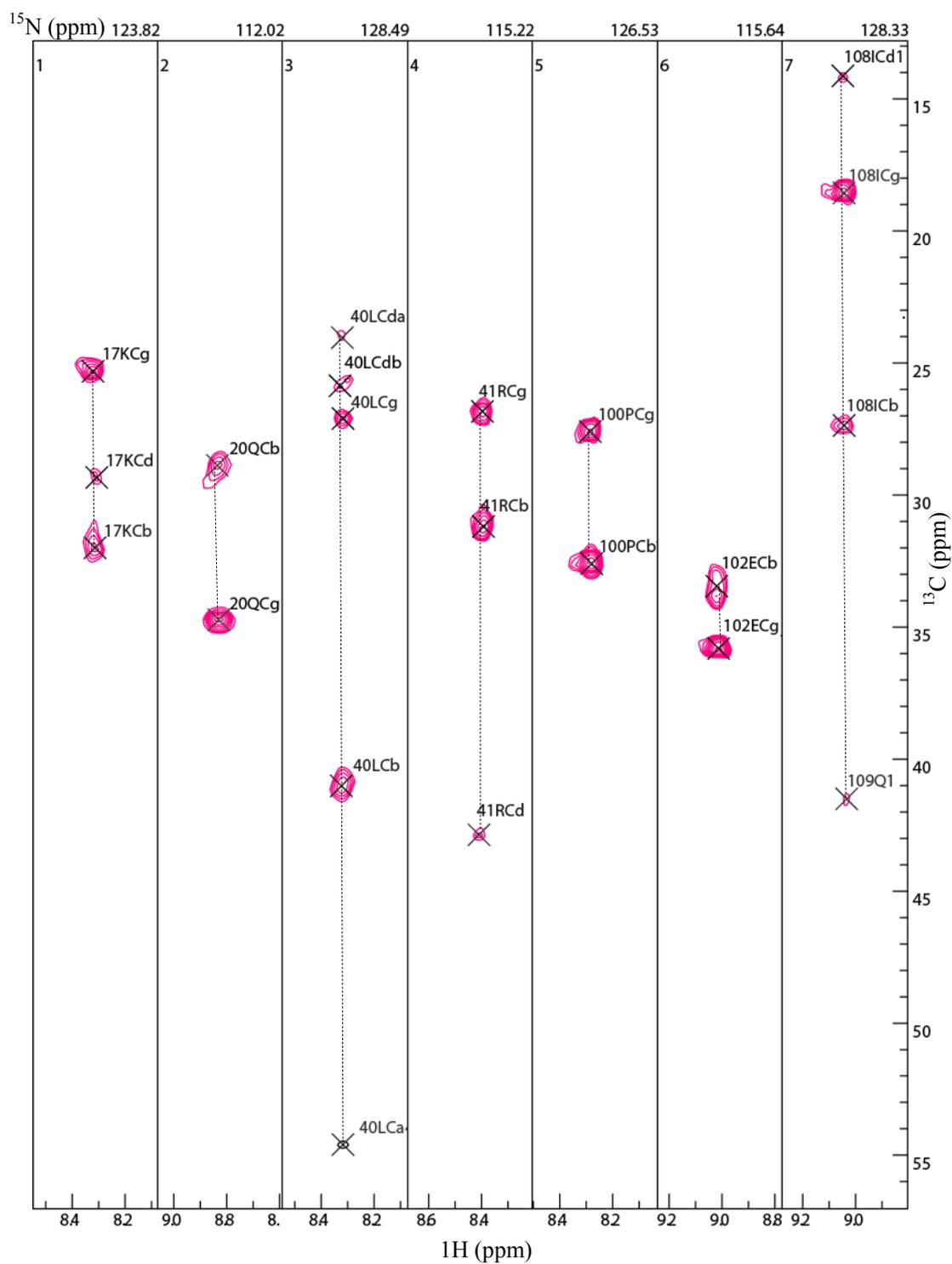
Fig 4.7   Strips taken from the 3D CC(CO)NH of PA-ICP, illustrating the side chain carbon chemical shift assignment of residues R41, Q20, L40, K17, I108, E102 and P100.
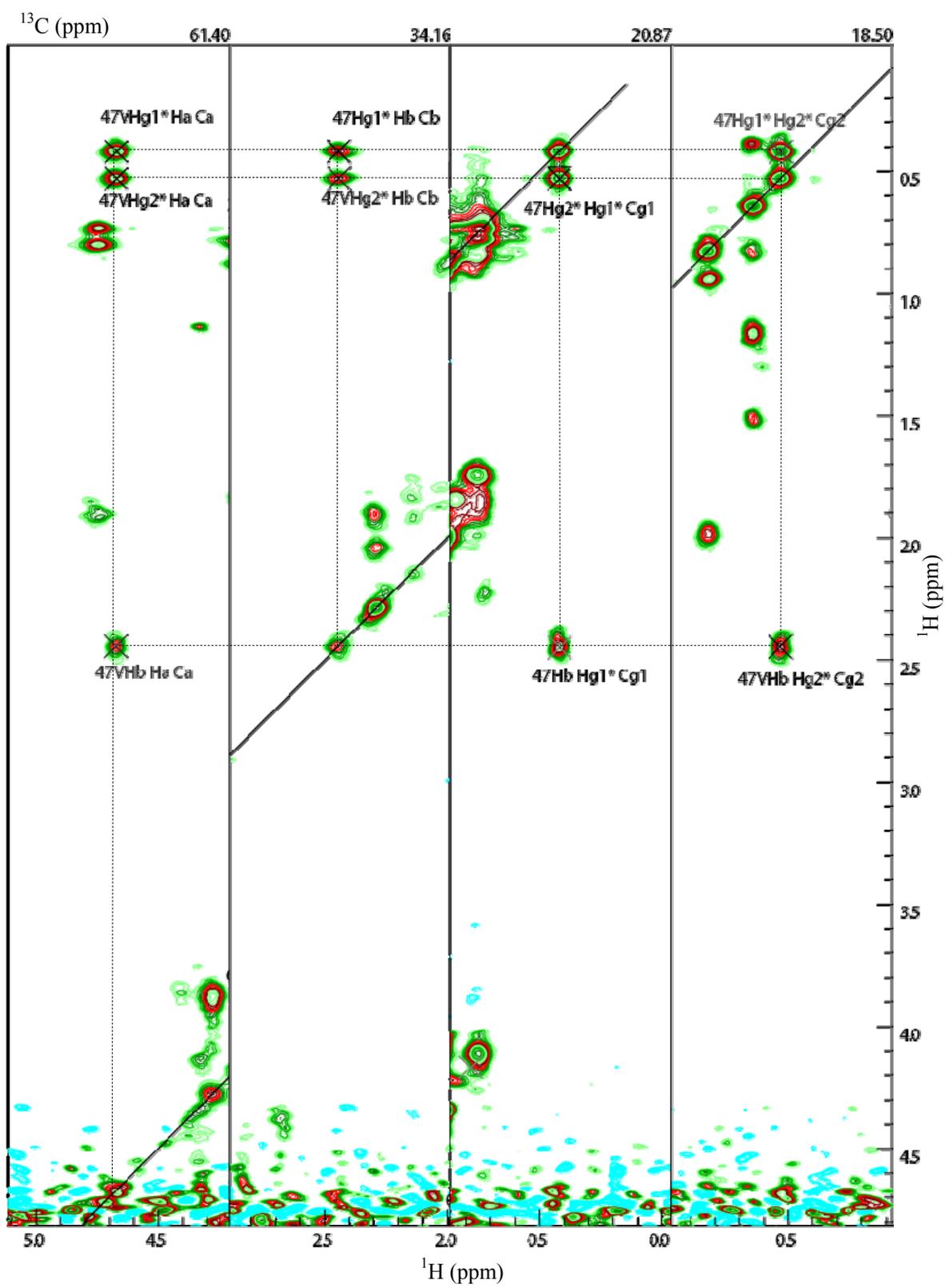
Fig 4.8   Strips taken from 3D HCCH-TOCSY spectrum of PA-ICP, demonstrating side chain ^{1}H and ^{13}C chemical shift assignment of residue V47.
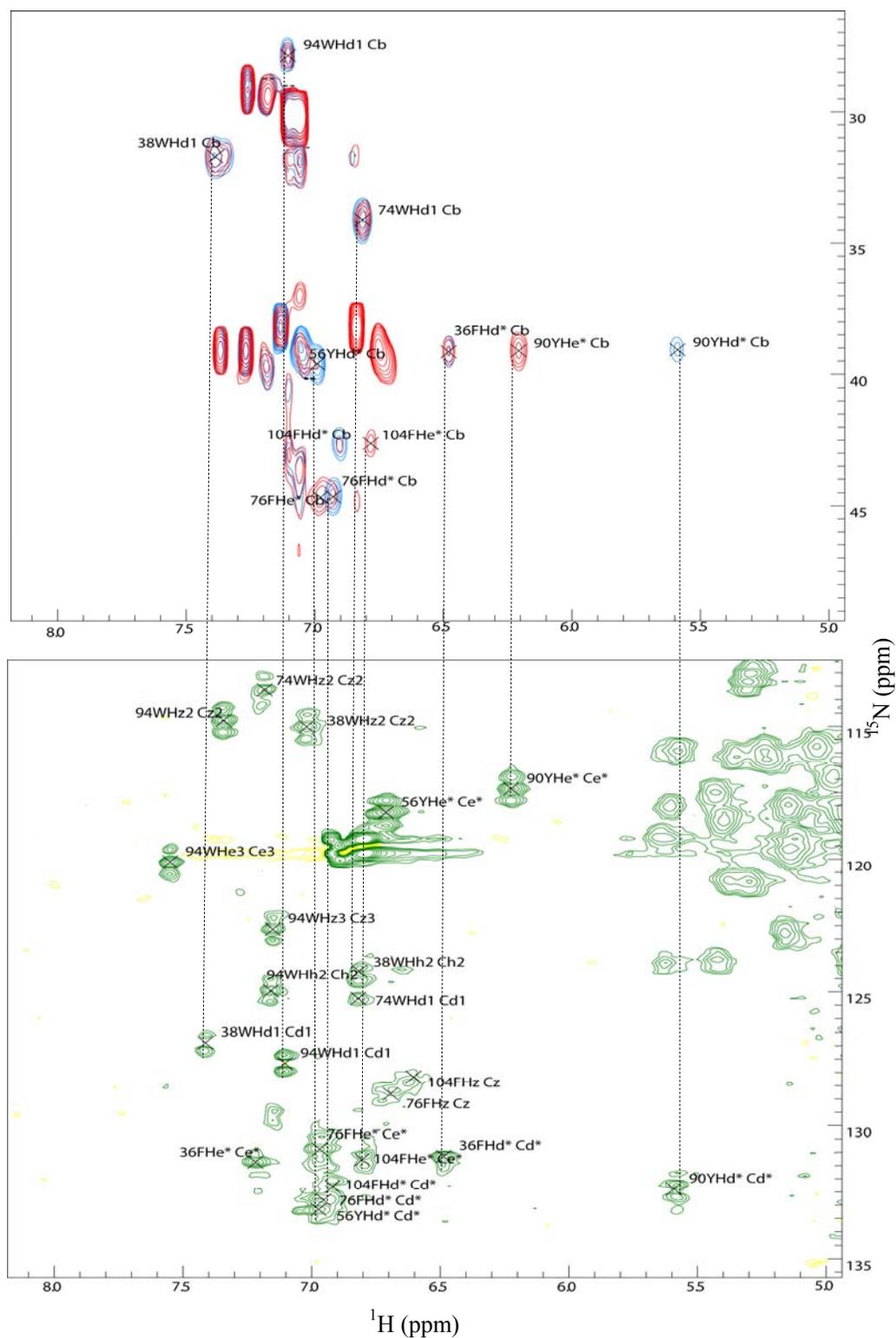
Fig 4.9   Illustration of aromatic side chain assignment of PA-ICP. The 2D (HB)CB(CGCDCE)HE is coloured in red, 2D (HB)CB(CGCD)HD in blue and 2D ¹³C-NOESY in green.

4.5 Secondary structure prediction

Having finished the chemical shift assignment, the secondary structure of PA-ICP could be predicted from H_α, C_α and C_β chemical shifts using CSI - Chemical Shift Index, a program for determining secondary structure in proteins from the chemical shift indices of ^{1}H and ^{13}C nuclei (http://www.bionmr.ualberta.ca/bds/software/csi/latest/csi.html). The secondary structure elements of PA-ICP were predicted based on the deviations of H_α, C_α and C_β chemical shifts of PA-ICP from database random coil values. The random coil chemical shifts are collected from peptides too short to possess any structural features and thus are chemical -environmentally independent. As the amino acids in folded protein experience different structural patterns, so do their chemical shifts. It is known that H_αs and C_βs experience a downfield shift in helices and an upfield shift in β-strands, whereas C_αs shift the opposite way. The protocol and analytical rules are described by D. Wishart *et al* (Wishart and Sykes, 1994) (fig 4.10). It can be seen that the PA-ICP protein, like its homologs, is mainly composed of β strands.
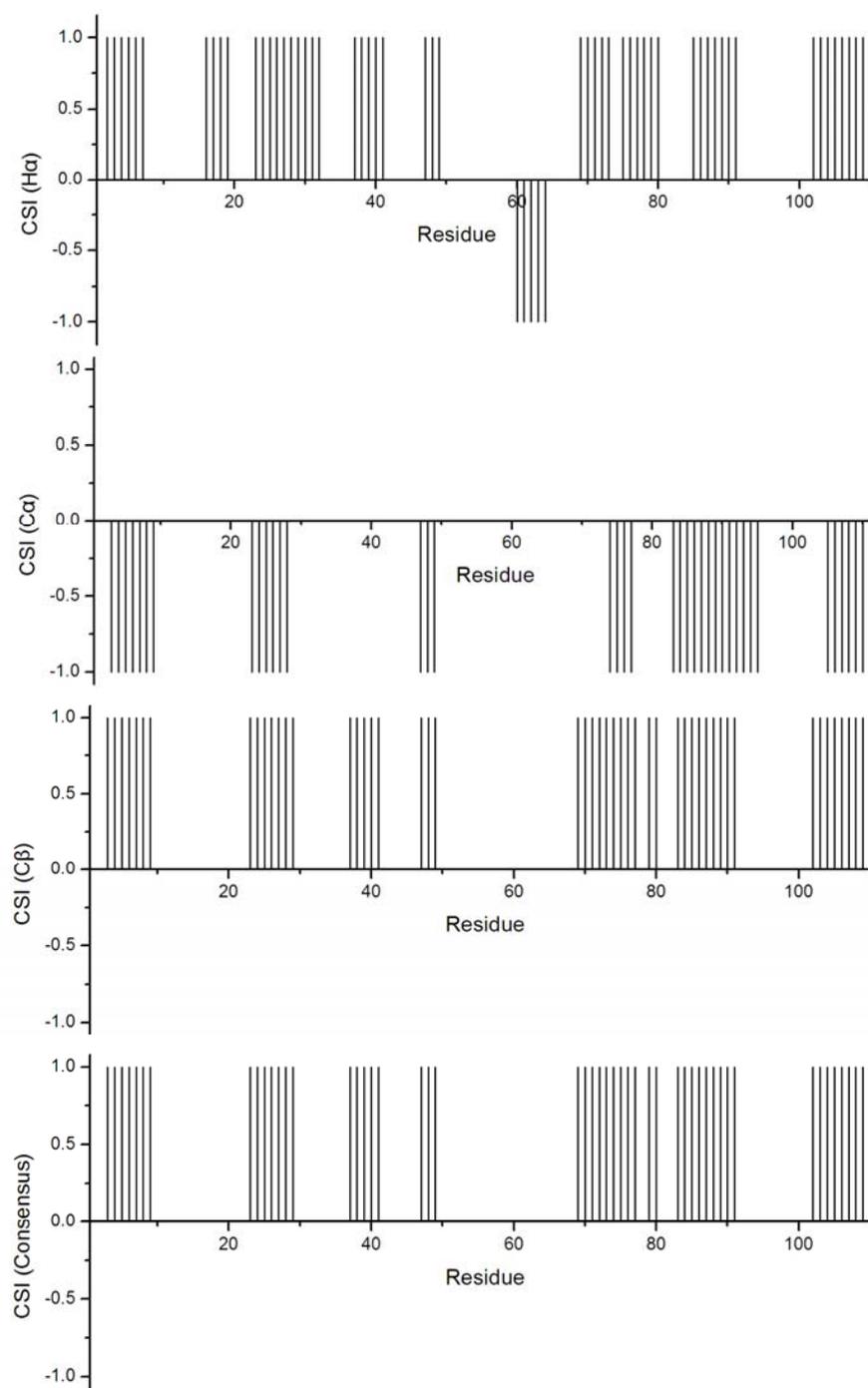
Fig 4.10   Secondary shift analysis of PA-ICP resonance assignments. Chemical shifts were analyzed using CSI. A value of +1 for the CSI consensus indicates β strand conformation.

4.6 Redox states of the cysteines

Statistical analysis has revealed that the C_β chemical shifts of cysteines are highly sensitive to whether the residue is involved in disulphide bond (Sharma and Rajarathnam, 2000). This is because the chemical environment of the C_β atom of the cysteine is profoundly affected by the redox state of the neighbouring S_γ. Therefore, by measuring the C_β chemical shifts, the redox states of cysteines in a protein sample can be determined using the rule suggested by D. Sharma et al, 2000 (Sharma and Rajarathnam, 2000). The C_β chemical shifts of the two cysteines in PA-ICP, C13 and C106, were assigned to 44.865 and 48.460 ppm respectively. Both values are greater than 35 ppm, suggesting that they are in oxidized form. Examining the NMR sample using SDS-PAGE under non-reducing conditions showed the major protein band at about 14 kDa which corresponded to the molecular weight of the monomeric PA-ICP (section 3.2.2). Therefore, it was inferred that the sample under study was an intramolecular disulphide bonded monomer.

4.7 Conclusion

With the exception of the His tag, there were seven residues whose backbone amide HNs were unable to be observed in the ^{15}N HSQC spectrum. They are Q1, N31, T33, E54, V55, S57 and N58. Considering the pH of the sample under investigation (pH 7.2), this is probably because these residues are located in unstructured loop regions that are more accessible to water and thus, the amide protons of these residues undergo rapid exchange with water resulting in an intermediate exchange phenomenon such that the linewidth of the peaks would be broad and the signals decay too fast to be detected.

The nonlinear sampling scheme used in 3D HNCACB and CBCA(CO)NH significantly reduced the acquisition time required to record these experiments. It will be extremely useful when dealing with unstable samples to shorten the time of acquisition and samples that have to be studied at low concentrations to provide higher sensitivity. The Maximum Entropy algorithm has proved to be a good way of reconstruction of incomplete data for chemical shift assignment as the backbone assignment can be completed unambiguously with the use of the nonlinear sampled spectra and the data agreed well with the uniformly sampled CCCONH and HCCH-TOCSY for assignment of the side chain ^{13}Cs and ^{1}Hs.

The chemical shift index showed that the protein is composed mainly of β strands. As the

sequence alignment indicated, PA-ICP adopts the same fold as its homologs, suggesting that PA-ICP may inhibit target CPs in a similar manner (fig 4.11).

```
 1        10        20        30        40        50
QKPVVTLDDADDCSPLKLTQGQELVLTLPSNPTTGFRWELRNPAASVLKRLGPEVY
            ---->    ------->        ---->        ---------

        ------->        ------->        ---->        ---->

    60        70        80        90        100       110
SNSEEDSGLVGSGGESTWRFRVAASGDDRLELVYRRPWEKDAEPAESFSCAIQVR
 >              ----------->    ---------->        -------->

                ----------->    ---------->        ------------>
```
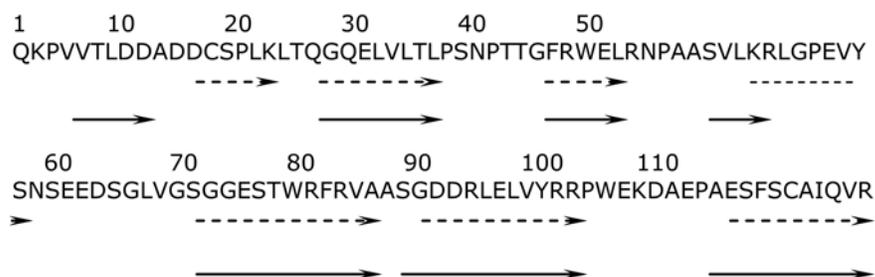
Fig 4.11   Likely secondary structure elements predicted by CSI (solid lines) and inferred from amino acid sequence alignment of ICP family from the known structures of Chagasin and *L. mexicana* ICP (dashed lines). The location of the β -strands is represented by arrows.

CHAPTER 5

STRUCTURE CALCULATION AND VALIDATION OF PA-ICP

5.1 Overview

The solution structure of PA-ICP was calculated from NMR data using the restrained molecular dynamics (rMD) protocol (section 2.3.2). Simulated annealing was used in the calculation to overcome local minima traps in potential energy and to ensure the calculation samples conformational space adequately (section 2.3.3). The experimental restraints used in this work included ambiguous distance restraints and hydrogen bond restraints. The calculation started from a set of random starting coordinates and incorporated ambiguous NOE data. The structures were calculated iteratively and the ambiguity of the NOE restraints was decreased based on the calculated structures. The iteration ended at the point where a good fraction of calculated structures achieved good agreement with both the experimental data and the theoretical covalent and non-covalent restraints. The hydrogen bond restraints were included at the later stages of the calculation. RDC restraints and χ_1 torsion angle restraints were used for structure validation and assessment of the quality of the stereospecific assignment made by the prochiral swapping protocol.

5.2 Structural restraints from NMR experiments

5.2.1 NOE restraints

5.2.1.1 NOESY spectra

The NOE restraints can be collected from a 3D ^{13}C-NOESY-HSQC and 3D ^{15}N-NOESY-HSQC (Sklenar et al., 1993) spectra. These experiments are a combination of ^{1}H homonuclear NOESY and HSQC experiments. The NOESY experiment detects the nuclear Overhauser effect between neighbouring spins (protons in this case). This part of the pulse sequence consists of 3 steps: preparation, evolution and mixing (fig 5.1) (Jeener et al., 1979). After preparation the spins are tilted to the transverse plane and chemical shift evolves during time t_1 followed by being returned to the z-axis for a mixing time τ_m during which the NOE builds up. The intensities of the crosspeaks in the spectra are used for derivation of distance information using equation 2.1. However, this derivation is made based on the isolated spin-pair approximation with the assumption that the cross relaxation occurs only in isolated spin pairs and that the NOEs buildup linearly with increasing mixing time. In practice, indirect NOE magnetization transfer can often occur. The NOE enhancement through direct cross relaxation of spin i to spin j can be transferred to spin k

that is also close to spin j, resulting in mixing-time dependent attenuation in NOE intensity of the crosspeak between spin i and j and increase in NOE intensity of the crosspeak between i and k. This effect is called spin diffusion. It can give rise to potential error in structure calculation if not interpreted properly. In theory, the intensity of the NOE reflects the genuine interproton distances only when $\tau_m=0$, i.e., at the initial NOE build up. One way of correcting the spin diffusion effect is by measuring NOEs with a series of increasing mixing time (e.g. at 50, 100, 200, 300 ms) and the mixing time is calculated from the initial slope of the curve fitted against the experimental points. This method gives very reliable results. However, it is rather time-consuming. Alternatively, spin diffusion can be tolerated when a short τ_m is used, assuming that the intensities of the NOE crosspeaks are proportional to the cross relaxation rate constant within this period. For the same reason, the mixing time cannot be made arbitrarily short, otherwise the intensities of the crosspeaks will be too low to detect. A mixing time between 50 to 150 ms is recommended to provide sensitive crosspeak intensities with tolerable spin diffusion (Cavanagh, 1996). The actual value for each individual molecule is dependent on its rotational correlation time. In general, large molecules need shorter mixing times to limit spin diffusion as the cross-relaxation is more efficient in large systems (Cavanagh, 1996).

Preparation | Evolution | Mixing time | Acquisition
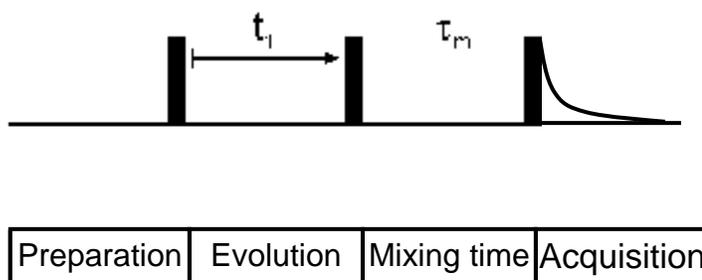
Fig 5.1   Anatomy of the ^{1}H homonuclear NOESY part of the 3D NOESY-HSQC experiment. The cross relaxation that arises from dipolar couplings occurs during the mixing time for NOE buildup, after which the spectra is extended to a third dimension of ^{13}C or ^{15}N by attaching a standard HSQC pulse sequence before detection.

For PA-ICP, 100 ms was used as mixing time for both ^{13}C and ^{15}N NOESY-HSQC experiments as this value has been used successfully to obtain NOE restraints from chagasin and *L. mexicana* ICP that have a molecular weight similar to PA-ICP.

5.2.1.2 Crosspeak picking

The structure calculation was carried out using the ARIA method (Ambiguous Restraints for Iterative Assignment) (Linge et al., 2003a). This method can handle ambiguous NOE restraints and does not require assignment of the chemical shifts of the NOE crosspeaks to specific atoms for input restraint lists. One would assume that the restraint lists can be generated by simply picking all peaks in the spectra. However, the real case is more complicated as there are many artifacts that should be taken into account. t_1 noise, for instance, caused by incomplete recovery of the longitudinal relaxation of spins before starting the succeeding scan due to their long T_1 relaxation time, is a typical source of artifacts in the NOESY spectra. It usually appears as stripes of peaks along the indirect proton dimension that can be mistaken as genuine crosspeaks. Although these artifacts may be identified later as violations during iterative structure calculation, they can introduce large errors at the initial stage of the structure calculation, making the interpretation of the calculated structures for further iterations rather difficult. For this reason, it is recommended to pick crosspeaks manually with care taken to look for symmetry-related crosspeaks. Because cross relaxation occurs between a pair of protons, the NOE enhancement of one spin results from its cross relaxation partner and vice versa. As a result, a pair of crosspeaks is expected to appear reflecting about the diagonal of a 2D NOESY spectrum, i.e., if there is a NOE crosspeak spotted at position (F_i, F_j), there will be a symmetrically related NOE crosspeak showing at (F_j, F_i) (fig 5.2). The symmetry of the NOE enhancement also holds for multi-dimensional heteronuclear NOESY spectra and for two protons i and j with chemical shifts F_{Hi} and F_{Hj} covalently connected to two heteronuclear spins I that resonance at the frequency F_{Ij} and F_{Ii} respectively, two symmetry-related crosspeaks (F_{Hi}, F_{Hj}, F_{Ij}) and (F_{Hj}, F_{Hi}, F_{Ii}) are expected to be seen if the two protons are close in space. As the noise peaks do not possess this symmetry property, they can be filtered out manually if they have no symmetry-related counterparts (fig 5.3). However, this method has to be used with caution as some artifacts can still survive if the noise peaks are too many and some of them happen to be symmetrical. Some processing errors may also generate artifacts that are symmetry-related.

There are also artifacts that result from incompletely suppressed water signals and from some water exchanging amide protons. Therefore, the crosspeaks close to the water frequency were not selected. Overlapped crosspeaks were not picked because the centers of

the peaks may be placed incorrectly, resulting in a failure to match the correct chemical shift assignment possibilities.
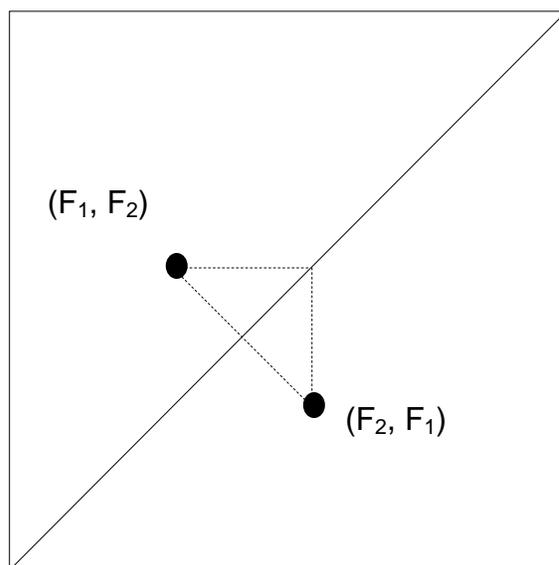
Fig 5.2   An illustration of symmetry-related peaks. The pair of peaks are reflected about the diagonal, indicating that proton A resonating at frequency F_1 is relaxed by dipolar coupling to its neighbour proton F_2 resonating at frequency b and vice versa.
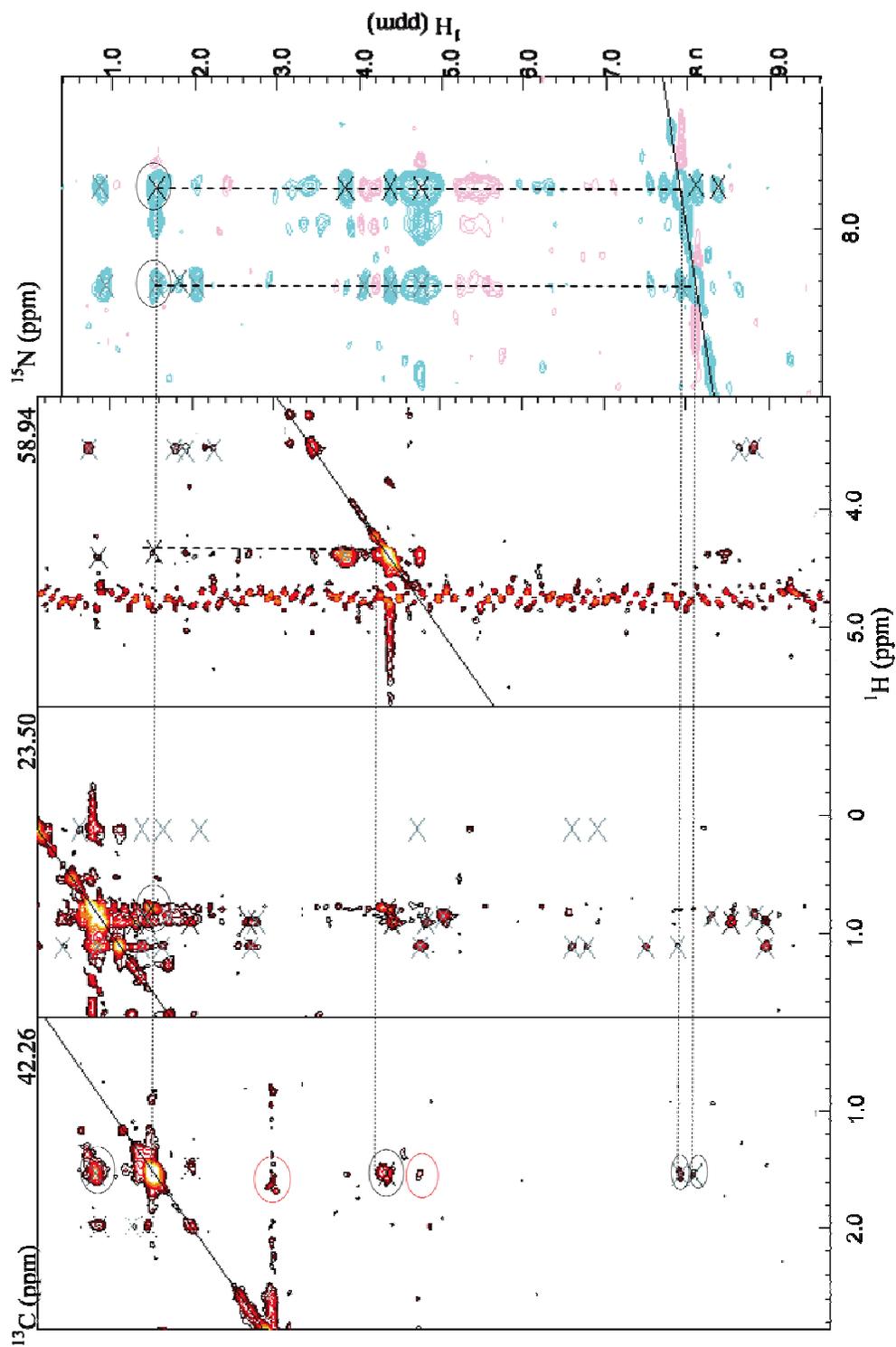
Fig 5.3   Strips taken from the ^{13}C NOESY-HSQC (strips 1-3) and ^{15}N-NOESY (strip 4) spectra of PA-ICP, illustrating the picking of the symmetry-related crosspeaks. In strip 1, only the peaks that have symmetry-related partners in strips 2-4 (black circles) are selected for structure calculation and the ones that do not have are considered to be artifacts (red circles).

5.2.1.3 Classification of distance restraints

The measured NOE intensities were calibrated to generate meaningful interproton distances for structure calculation. The equations 2.1 and 2.2 are valid provided that the molecule is tumbling isotropically and holding a single rigid conformation. The presence of internal motion and possibly, chemical exchange will lead to a diminution of the NOE intensity and the calculated distances will be longer than the real distances. Therefore, rather than giving them specific values, the NOE-derived distances were restricted to allowed distance ranges. The upper bound of a distance restraint range was set according to the NOE intensity and the lower bound set to zero. Further approximation was made by grouping the distances into different classes, i.e., strong, medium, weak and very weak. To do this, the intensity of each NOE crosspeak was normalized against the average intensity of all NOE crosspeaks and the relative ratios were grouped into different distance classes with the upper bounds defined empirically and no reference distance was required. For proteins, this approximation is acceptable as the quality of the final structure is dependent on the number and distribution of the restraint sets rather than the precision of the distances. The NOE distance classes used for PA-ICP are listed below.

Min. normalized NOE Intensity	Max Distance
3.00000	2.00000
1.00000	2.50000
0.25000	3.10000
0.10000	4.00000
0	6.00000

Table 5.1   NOE bins for structure calculation.

5.2.1.4 Treatment of the averaging groups and ambiguous assignment using r^{-6} summation

When more than one assignment possibilities contribute to a NOE restraint and we cannot rule out any one contributor to the restraint, the ambiguous distance restraints is used to calculate an effective distance, D_{eff}, involving all the nuclei by r^{-6} summation, assuming that

the r^{-6} summed distances should not exceed the upper distance bounds according to the measured NOE intensities of the ambiguous crosspeaks.

$$D^{-6}_{eff} = \sum_{i,j} D^{-6}_{ij} \quad (5.1)$$

This method can also be used to treat groups of protons that undergo rapid conformational exchange. Good examples are methyl groups and symmetrical related pairs of aromatic protons on fast-flipping rings. The resonances of these spins are not able to be resolved as distinct peaks. Instead, a single averaged signal is observed.

5.2.2 Hydrogen bond restraints

The backbone amide protons engaged in hydrogen bonds can be identified by dissolving a ^{15}N labelled sample into D_2O. The experiment relies on the fact that the solvent accessible protons can be readily replaced by deuterons while the ones donated to hydrogen bonds or buried in hydrophobic core are more difficult to substitute. As the ^{15}N HSQC experiment does not detect the deuterium signals from ^{15}N-^{2}H groups, the exchange rate can be measured using a series of ^{15}N HSQC experiments recorded at a series of time intervals to follow the decrease in intensities of the amide proton peaks. For PA-ICP, the experiments were recorded over 24 hours immediately after a lyophilised sample was dissolved in D_2O. The experimental time for each HSQC was limited to 17 mins so that the spectra could be acquired rapidly before the majority of the amide protons were exchanged with D_2O. 20 peaks that survived 2 h after dissolution in D_2O were considered to come from slowly exchanging amide protons. Care should be taken to distinguish whether these protons are hydrogen bonded or lie in the hydrophobic core of the protein where they are inaccessible to the deuterons. The purpose of this experiment is to identify the presence of hydrogen bonds while there is no need of measuring the precise hydrogen exchange rate. The hydrogen bond restraints were included at the later stage of structure calculation and the hydrogen acceptors were determined by inspection of the structures calculated using just NOE restraints.
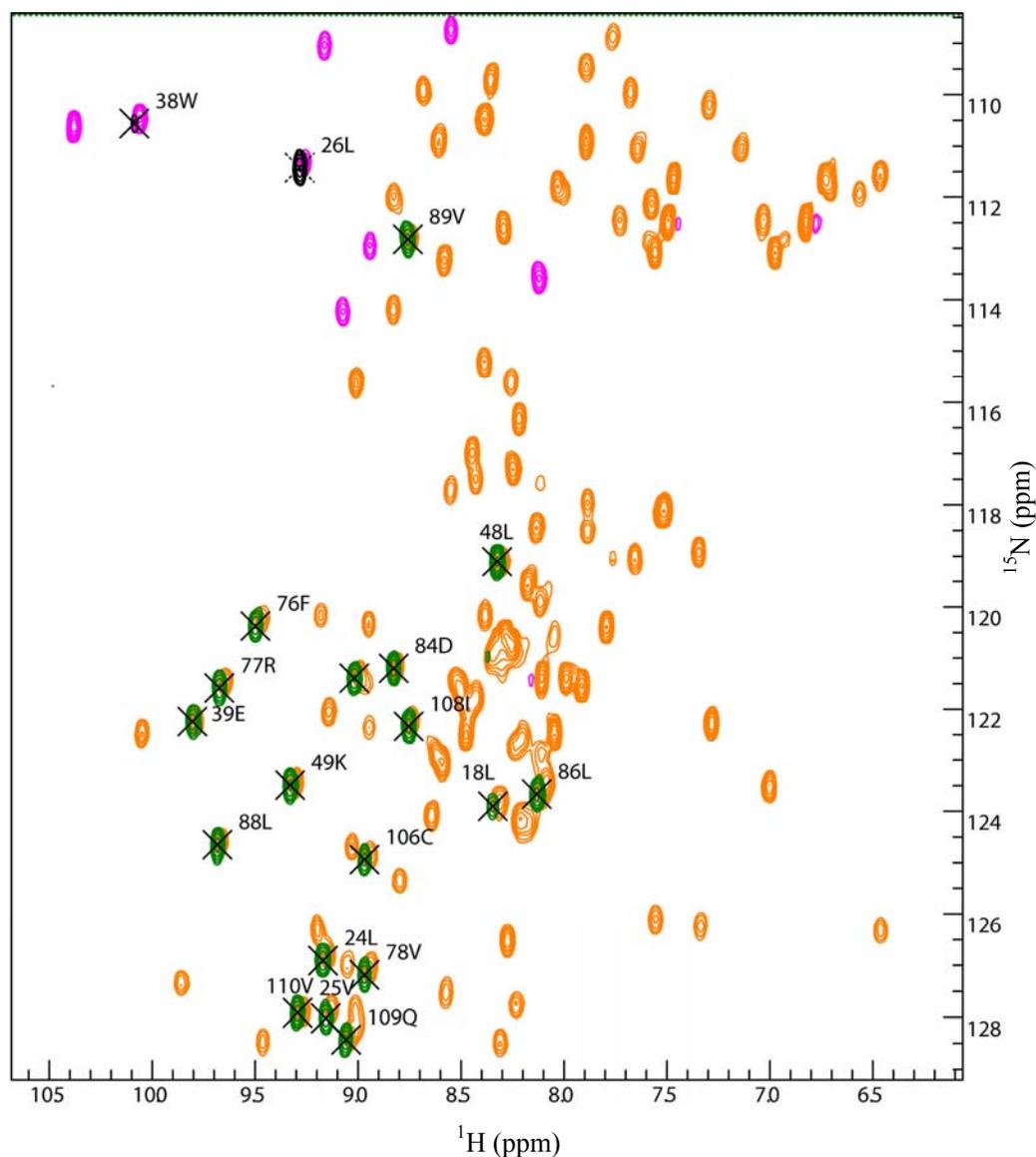
Fig 5.4   The superimposition of the HSQC spectra acquired on a ^{15}N-labelled PA-ICP before (orange and red for positive and negative crosspeaks respectively) and 2 hours after (green and black for positive and negative crosspeaks respectively) deuterium substitution. The protons giving rise to NH crosspeaks 2 hours after the protein was buffer exchanged into D_2O were considered to come from candidate hydrogen bond donor amide protons.

5.2.3 Residual dipolar coupling restraints

5.2.3.1 Weak alignment

RDC restraints were gathered on a uniformly ^{13}C, ^{15}N labelled PA-ICP sample which was weakly aligned with respect to the applied magnetic field. The partial alignment is induced by dissolution of the protein sample into a diluted aqueous medium that is strongly oriented with respect to the external magnetic field but interacts weakly with the protein. Many alignment media that can fulfill this mission have been identified, including dilute liquid crystals and anisotropically strained gels (Fleming and Matthews, 2004), each one of which has its own properties and optimal application conditions. The highly magnetically susceptible liquid crystal elements interact weakly with the biomolecule, resulting in a partial orientation of the molecule relative to the static magnetic field (Gabriel et al., 2001, Prosser et al., 1998). Unlike liquid crystals whose alignments are induced by the external magnetic field, the alignment of the compressed gels is dictated by compressing or stretching the gel away from an initial isotropic state (Tycko et al., 2000). In this project, a liquid crystal formed by the negatively charged filamentous bacteriophage pf1 (pI~4) was used as the alignment medium to resolve RDCs from backbone HN, H_nCO and NCO interactions in PA-ICP. The medium can be used in the temperature range of 5~30^{0}C, with pH from 6 to 8 and at salt concentration up to 100mM NaCl. One advantage of using pf1 phage is that the degree of the alignment of the D_2O, reflecting the degree of the alignment of the pf1 phage, can be tuned by simply adjusting pf1 phage concentration (Hansen et al., 1998). This is because not only the biomacromolecule is aligned by the phage, but also the D_2O added for locking. The quadropolar coupling of deuterium is no longer averaged to zero and the deuterium signal appears as a doublet and the amplitude of the splitting is dependent on the degree of alignment. The resultant quadropolar deuterium splitting in the 1D deuterium spectra is approximately linearly correlated to the phage concentration and therefore the alignment can be easily optimized (Hansen et al., 1998). A separation of 10 Hz in the ^2H doublet was observed with 20 mg/ml pf1 phage present in the PA-ICP sample (fig 5.5). This allowed accurate measurement of RDCs in a range of 0~ 30 Hz for dipolar couplings between backbone HN, H_nCO and NCO on high-resolution spectra using in-phase, anti-phase (IPAP) experiments.
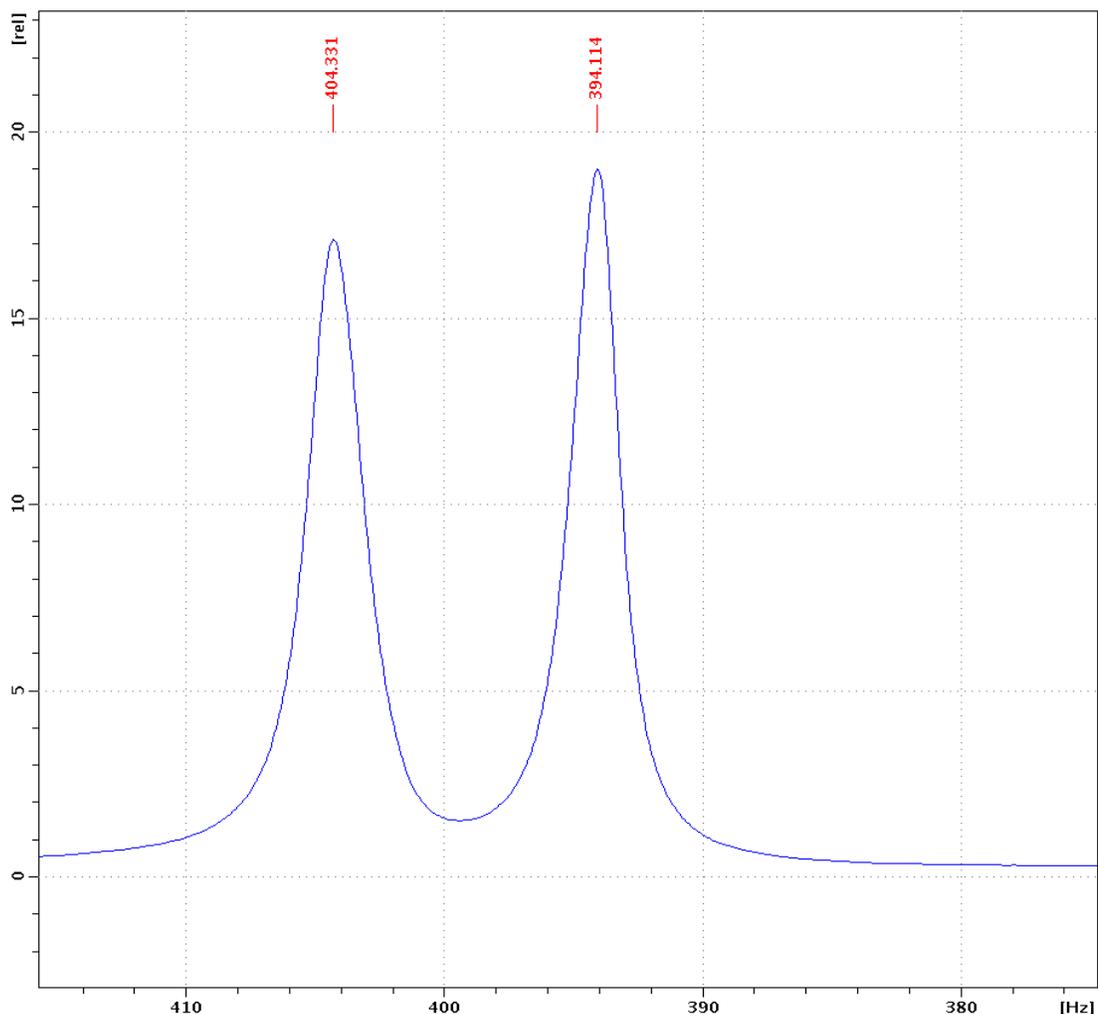
Fig 5.5   The quadropolar deuterium splitting of 10 Hz in the 1D ^{2}H spectrum collected on a ^{13}C-^{15}N double labeled PA-ICP sample weakly aligned in 20 mg/ml pf1 phage at 298K.

5.2.3.2 Measurement of RDC restraints using in-phase anti-phase (IPAP) experiments

RDCs from backbone HN, NCO and H_nCO interactions of PA-ICP were measured from simplified ^{15}N HSQC type spectra (Ottiger et al., 1998, Wang et al., 1998, Permi et al., 1999). The desired residual dipolar and scalar couplings were resolved in the ^{15}N frequency domain for HN and NCO interactions and ^{1}H frequency domain for H_nCO interactions.

Resulting spectra crowding could be reduced using IPAP method. Two separate spectra were recorded for each experiment to produce in-phase and anti-phase crosspeaks. By adding or subtracting one from the other, spin-state separated crosspeaks were generated and the splittings could be accurately measured (Fig 5.6). One set of spectra was collected on a sample dissolved in isotropic solution to allow the measurement of the scalar couplings and another set was collected on a sample in aligned pf1 phage to resolve scalar couplings and residual dipolar couplings (see Fig 5.7). RDC values are commonly normalized to the HN dipolar couplings and when dealing with dipolar couplings from NCO and H_nCO, they can be easily scaled with scaling factors -3.0 and 8.3 respectively (Wang et al., 1998).
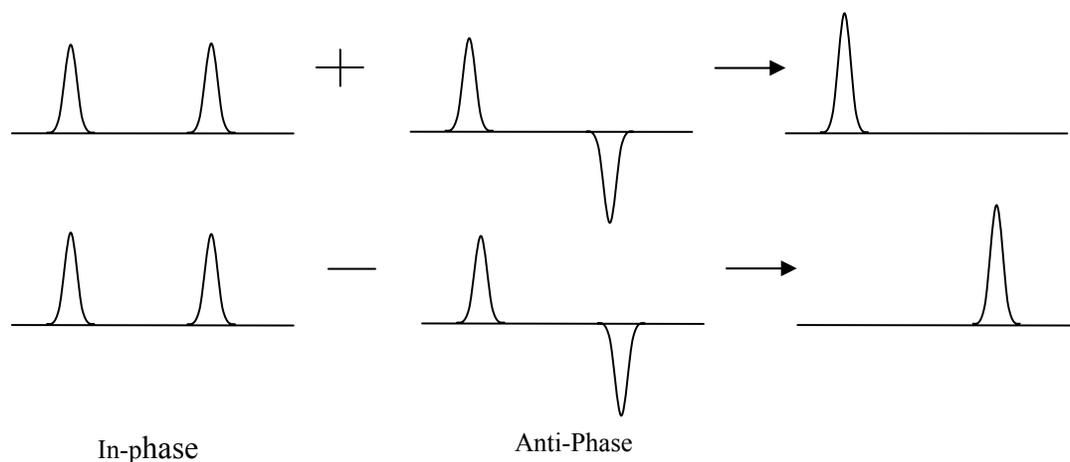
In-phase   Anti-Phase

Fig 5.6   Schematic representation of calculating the RDCs from the IPAP experiment. The experiments are designed to record the in-phase and anti-phase patterns of the scalar splittings. The two doublet components can be separated into different spectra by adding or subtracting the anti-phase on the in-phase crosspeaks.
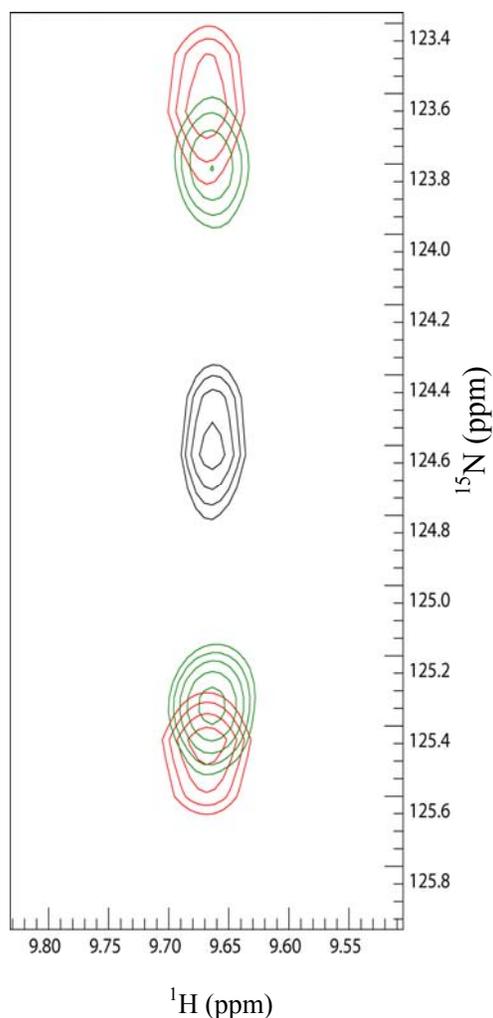
Fig 5.7   The overlay of amide HN crosspeaks of residue L88 in the IPAP ^{15}N HSQC spectra acquired on a uniformly [13]C –[15]N enriched PA-ICP sample in pf1 phage (red) and in water (green). The standard ^{15}N HSQC spectrum is in black.

5.2.4 χ_1 angle determination and stereospecific assignment

It has long been appreciated that three-bond J coupling constants can be used to extract local backbone and side chain conformations in proteins. The side chain χ_1 angles of valine, isoleucine and threonine residues can be obtained by the measurement of $^3J_{C\gamma CO}$ and $^3J_{C\gamma N}$ coupling constants. With the use of 2D spin-echo difference ^{13}C constant time heteronuclear single quantum (^{13}C CT-HSQC) experiments (Vuister et al., 1993, Grzesiek et al., 1993), the measurements quantify the coupling constants from the intensity ratio of the crosspeaks on the spectra recorded when the J coupling is either eliminated or activated during the

carbon chemical shift evolution using equation 5.2.

$$\frac{S_a - S_b}{S_a} = 1 - \cos\left(2\pi\, J_x T\right) = 2\sin^2\left(\pi\, J_x T\right) \quad (5.2)$$

where S_a and S_b refers to the crosspeak intensities with and without J coupling elimination, respectively, J_x, the corresponding three-bond coupling constants between side chain carbon and carbonyl or amide nitrogen and T, half of the duration of the constant time evolution period defined by $1/^1J_{CC}$ (27.5 ms). In general the $^3J_{C\gamma CO}$ values are less than 4 Hz. Large values (> 3 Hz) indicate trans orientation while small values (< 1.5 Hz) indicate gauche orientation of $CO\text{-}C_\alpha\text{-}C_\beta\text{-}C_\gamma$. Similarly, large (> 1.7 Hz) and small (< 1 Hz) values of $^3J_{C\gamma N}$ imply trans and gauche conformation of $N\text{-}C_\alpha\text{-}C_\beta\text{-}C_\gamma$. With this knowledge, $C_{\gamma 1}$ and $C_{\gamma 2}$ of the valine residues was then designated stereospecifically and the χ_1 angles of valine and threonine side chains determined according to the recommendations of IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (Markley et al., 1998) (fig 5.8). This allowed determination of the χ_1 angle of 4 out of a total of 10 valine, all 6 threonine and isoleucine residues in PA-ICP. The χ_1 angles of some valines were not determined due to crosspeak overlap.

RESIDUE TYPE	RESIDUE NUMBER	$C_{\Gamma X}$	$^3J_{C\Gamma CO}$	CONFORMATION	$^3J_{C\Gamma N}$	CONFORMATION	X_1
Valine	5	$C_{\gamma 1}$	1.17	Gauche	2.07	Trans	180^0
		$C_{\gamma 2}$	3.58	Trans	Small	Gauche	
	47	$C_{\gamma 1}$	3.81	Trans	Small	Gauche	-60^0
		$C_{\gamma 2}$	0.96	Gauche	0.05	Gauche	
	78	$C_{\gamma 1}$	Small	Gauche	2.40	Trans	180^0
		$C_{\gamma 2}$	3.87	Trans	0.78	Gauche	
	89	$C_{\gamma 1}$	3.58	Trans	0.75	Gauche	-60^0
		$C_{\gamma 2}$	1.17	Gauche	0.38	Gauche	
Threonine	6	$C_{\gamma 2}$	2.02	Rotamer averaging	Small	Gauche	n/a
	19	$C_{\gamma 2}$	2.15	Rotamer averaging	0.59	Gauche	n/a
	27	$C_{\gamma 2}$	Small	Gauche	1.89	Trans	-60^0
	33	$C_{\gamma 2}$	2.23	Rotamer averaging	0.24	Gauche	n/a
	34	$C_{\gamma 2}$	2.87	Rotamer averaging	1.85	Trans	n/a
	73	$C_{\gamma 2}$	1.02	Gauche	0.31	Gauche	180^0
Isoleucine	108	$C_{\gamma 2}$	1.25	Gauche	1.65	Trans	-60^0
Small- the difference between S_a and S_b is too small to be measured							
n/a- can not be estimated due to rotamer averaging							

Table 5.2   The χ_1 angles of valine, threonine and isoleucine residues in PA-ICP determined by $^3J_{C\gamma CO}$ and $^3J_{C\gamma N}$
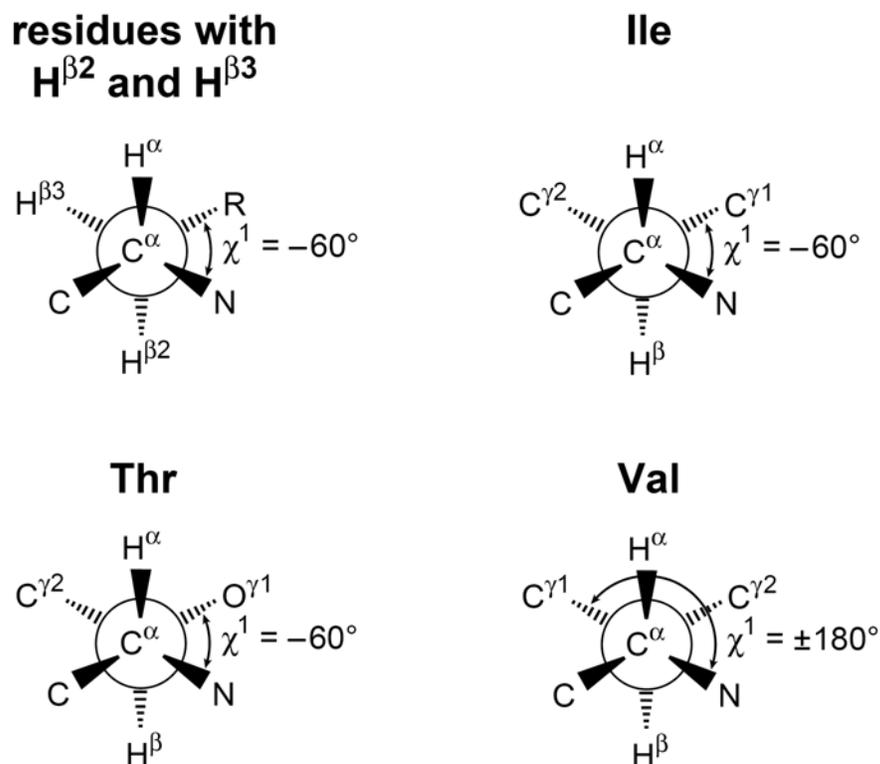
Fig 5.8   Definition of side chain χ_1 angle in residue isoleucine, threonine, valine and residues with $H_{\beta2}$ and $H_{\beta3}$ (Markley et al., 1998).

5.3 Structure calculation of PA-ICP

5.3.1 Calculation strategy

The calculation strategy used for calculation of PA-ICP structures was based on simulated annealing using highly ambiguous NOE-derived distance restraints. It consisted of three steps: randomize, regularize and refine, as suggested by Nilges (Nilges, 1995). Prochiral swapping took into account the stereo specificity of the prochiral centers in the molecule during the regularize and the refine phases without actually performing explicit stereospecific assignment. The algorithm of the calculation is stated in the following outline.

Input: random initial coordinates, ambiguous NOE restraints, force field

Randomize
1. Initial energy minimization
    -   selected atoms (C_α and single atoms representing the side chains)
    -   large soft atoms (selected)
    -   NOE, covalent and non-bonded forces weak
2. Simulated annealing at 2000 K
    -   selected atoms as above
    -   2000 K constant temperature
    -   bond and angle force constants increased
    -   other force kept weak
3. Simulated annealing at 1500 K
    -   all atoms
    -   1500 K constant temperature
    -   bond, angle and VdW forces increased
    -   NOE force strong
    -   atoms medium-sized and soft

Regularize
1. create local ideal geometry by template fitting and topology mirror image generated
2. energy minimization
    -   all atoms
    -   atoms soft and small
    -   non-bonded forces strong
    -   bonds, VdWs and NOEs minimized before minimizing angles
3. simulated annealing at 2000 K
    -   2000 K constant temperature
    -   very small and soft atoms
    -   dihedral angle and improper force introduced gradually
    -   covalent weights increased while non-covalent weights decreased
4. correct the handedness of the structure by rejection of the mirror image with higher potential energy
5. prochiral swapping
6. simulated annealing 2000 K – 100 K
    -   increased VdW interaction and NOE forces
    -   temperature decreased gradually from 2000 K to 100 K
7. energy minimization

Refine
1. simulated annealing at 2000 K
    -   selected atoms small and hard
    -   other forces increased
    -   2000 K constant temperature
    -   NOE forces strong
2. simulated annealing 2000 K – 1000 K
    -   temperature decreased slowly from 2000 K to 1000 K
    -   atoms small and hard
    -   other forces increases to final values
3. simulated annealing 1000 K – 100 K
    -   temperature decreases from 1000 K to 100 K slowly
    -   forces kept unchanged
4. energy minimization

output final structure

Fig 5.9   Outline of the protocol used for calculating PA-ICP structures.

5.3.2 Prochiral swapping

Prochiral groups in a protein have carbon centers with two different and two identical substitutes, for example, β-methylene protons and the methyl groups of leucines and valines. Identical substitutents with non–degenerate chemical shifts are typically arbitrarily stereospecifically assigned during chemical shift assignment. The arbitrary assignment may result in the wrong orientation of the prochiral pair being chosen. To overcome this problem, a floating chirality assignment routine was introduced. This allows the two non-degenerate chemical shifts to be assigned to either member of the prochiral pair by swapping their stereospecific assignments during the calculation with the lower energy conformer being chosen for each prochiral center. An adaptation of this method that swaps the prochiral centers in a random order and that accepts or rejects the swap with a metropolis style criterion related to the temperature was used in this study (Folmer et al., 1997).

5.3.3 Iterative structure calculation with automated assignment of ambiguous distance restraints

The structure of PA-ICP was calculated preliminarily based on ambiguous NOE restraints using the program CNS (Crystallography & NMR System) (Brunger et al., 1998) in the ARIA manner. Once a good consistent ensemble of structures was achieved based solely on NOE restraints, the resulting structures were then subjected to further refinement by introducing hydrogen bond restraints.

5.3.3.1 ARIA method

As NOE restraints yield the major distance information for structure calculation, the NOESY crosspeaks have to be assigned in order to extract distance restraints from the spectra. The knowledge of the assignment is derived from known chemical shift values determined by chemical shift assignment, in which several protons may share the same chemical shift due to the large number of protons in biological macro molecules and the limited chemical shift range. Furthermore, overlapping NOESY crosspeaks often occur and the assignment of NOESY crosspeaks is therefore inherently ambiguous. Hence, manual NOESY assignment can be time-consuming and mis-assignment may lead to the failure of the structure calculation.

Alternatively, the structure can be calculated using ambiguous NOE restraints and the ambiguity of the chemical shift assignment of the NOE crosspeaks can be reduced indirectly by eliminating unlikely assignment possibilities using ARIA method (Linge et al., 2001, Nilges and O'Donoghue, 1998, Nilges, 1995).

1219 and 872 NOE crosspeaks were collected from the ^{13}C and ^{15}N NOESY-HSQC experiments respectively. The restraints were supplied as a list of these selected crosspeaks with intensity related internuclear distances calculated using the method as described in section 5.2.1.3 and the superposition of the possibilities of chemical shift assignments. It was exported from CCPN analysis to ARIA format for structure calculation.

The initial input distance restraints files contain lists of NOE crosspeaks grouped into different distance classes with all possible chemical shift assignments. The frequency ranges in which chemical shifts can be potentially assigned to a given NOE crosspeak are defined by the chemical shift tolerances Δ. In a 3D NOESY spectrum, if a NOE crosspeak resonances at frequency {F_1, F_2, F_3}, the chemical shifts between F-Δ and F+Δ are all assigned to the given peak. If the tolerances are set too tight, the assignment may be incomplete. On the other hand, a large tolerance will lead to degeneration of the initial assignment and make the calculated structures less likely to converge. For PA-ICP, the tolerances for the direct, the indirect proton and the heteronuclear dimensions were set to 0.05, 0.05 and 0.4 ppm respectively.

At the end of each iteration, incorrect assignments and noise peaks are identified by violation analysis of the restraints used to calculate the structure, the assignment possibilities are reduced by filtering the chemical shift assignment, the filtered distance restraints are calibrated at the later stages of structure calculation, all based on the analysis of the ensemble of the calculated structures with the lowest potential energies.

In most cases, the initial structure ensembles do not exactly agree with all the input experimental restraints used for structure calculation due to assignment errors, the presence of the mis-picked noise peaks or as a result of improper calibration of the NOE restraints. The violated restraints can be identified by comparison of the distance bounds determined by the experimental data with the corresponding distances measured in the calculated

structure ensemble or their r^{-6} summed distances. If a distance found in a calculated structure lies outside the distance bounds by more than a user-defined violation threshold t (0.3Å for structure calculation of PA-ICP), the restraint is considered violated. For structure calculation of PA-ICP, all structures calculated after each calculation cycle were analyzed to identify the restraints that were systematically violated. The violated restraints appearing in more than 50% of the total calculated structures were then manually inspected with respect to the original NOESY spectra so as to classify the origins of the violations. Noise peaks were removed from the restraint lists. The wrongly assigned NOE crosspeaks were either corrected or removed from the restraint lists. The distance bounds may be adjusted for the next calculation iteration. The rest of the violated restraints were still used in the further structure calculation without any change although it may be reported as problem.

After all the violations in the current structures had been properly treated, the ambiguity of the NOE assignment could be reduced by discarding the possible chemical shift assignments if their corresponding internuclear distances in the minimal potential energy structures contribute less to the total peak intensity than a given fraction. The assignments were then re-evaluated before another structure calculation round started using the ambiguity-reduced restraints. The filtered data were used to calculate a new set of structures. At the end of the assignment/structure calculation cycle, a well defined biomacromolecule structure could be obtained as well as a comprehensive set of NOE assignments provided that a sufficient number of distance restraints are available.

In the later stages of the structure calculation when a good convergence of the calculated structures was achieved, an additional distance restraint calibration step was introduced based on the distance calculated from the structure ensemble. A calibration factor was employed to calibrate all experimental NOEs with corresponding internuclear distances smaller than a cutoff of 6Å,

$$C = \sum_{NOEs} \frac{D^{-6}}{I} \quad (5.3)$$

where I was the observed NOE intensity. D was the arithmetic average of the internuclear

distance d_{ij} in the ensemble of n calculated lowest energy structure, which was defined by,

$$D = \frac{1}{n}\sum_{n=1}^{n} d_{ij,n} \quad (5.4)$$

The calibrated NOE-derived distance d_{cal} was calculated as

$$d_{cal} = (C \times I)^{-\frac{1}{6}} \quad (5.5)$$

During the next round of structure calculation, the distance was allowed to deviate from target distance d_{cal} by no more than εd^2_{cal} (ε=0.125 Å^{-1}). This calibration was applied only at the later stage of the structure calculation because in the early iterations, the ensemble may be poorly defined and the average calculation would be biased towards the shortest distance in the ensemble.

At the end of each round, a total of 50 structures were calculated and a subset of the lowest NOE potential energy structures were selected (typically 30 out of 50) for violation analysis, assignment filtration and distance calibration to prepare modified restraint lists for the next round.

5.3.3.2 Inclusion of hydrogen bond restraints

The 20 slowly exchanging amide protons resolved by hydrogen-deuterium exchange experiment described in section 5.2.2 were examined against the NOE-based structures for hydrogen-bonding possibilities based on the property of the secondary structure of the protein calculated using the NOE restraints only. The potential hydrogen bond acceptors O_{CO} were selected by assuming that the hydrogen bonds are formed between them and the HN groups in the adjacent β strand. 19 O_{CO} atoms were found to be in the position favorable for hydrogen bonding with 19 out of 20 observed slowly exchanging amide protons (Fig 5.10). The hydrogen bonding information were supplied as an additional distance restraint list by assuming that the distance between the H and the O atoms in a hydrogen bond are 1.7~ 2.2 Å. To ensure the bond angle to be between 120^0~ 180^0, the distance between the N and O atoms are set to 2.7~ 3.2 Å (Roberts, 1993).

Residue
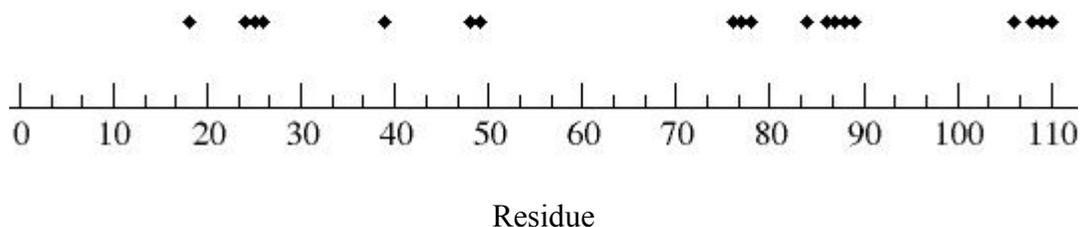
Fig 5.10　　The location of the 19 potentially hydrogen bonded amide protons determined by D_2O exchange in PA-ICP sequence.

5.3.3.3 Iterative calculation

The iterative calculation strategy used for the PA-ICP structure calculation is set out in table 5.3

The structure was first calculated based only on NOE-derived distance restraints. The hydrogen bond restraints were not incorporated until the very late stages of the calculation. Clearly, the total and experimental NOE potential energies of the calculated structures experienced noticeable decreases at the end of round 5 with the addition of the calibration step while they did not have distinguishable differences between the NOE-based and hydrogen bond refined structures (fig 5.11 b).

Iteration	Restraints	Analysis	Filter proportion	Calibration	Check	Hbonds
1	Original restraints	Original restraints	0.99	n/a	Filtered restraints	n/a
2	Original restraints	Original restraints	0.99	n/a	Filtered restraints	n/a
3	Checked restraints from round 2	Checked restraints from round 2	0.98	n/a	Filtered restraints	n/a
4	Checked restraints from round 3	Checked restraints from round 3	0.95	n/a	Filtered restraints	n/a
5	Checked restraints from round 4	Checked restraints from round 4	0.95	Filtered restraints	Calibrated restraints	n/a
6	Checked restraints from round 5	Checked restraints from round 5	0.95	Filtered restraints	Calibrated restraints	n/a
7	Checked restraints from round 6	Checked restraints from round 6	0.95			Introduced

Table 5.3   The calculation of PA-ICP structure using CNS in an ARIA manner. The filter step filters the original experimental restraints against currently calculated structures.
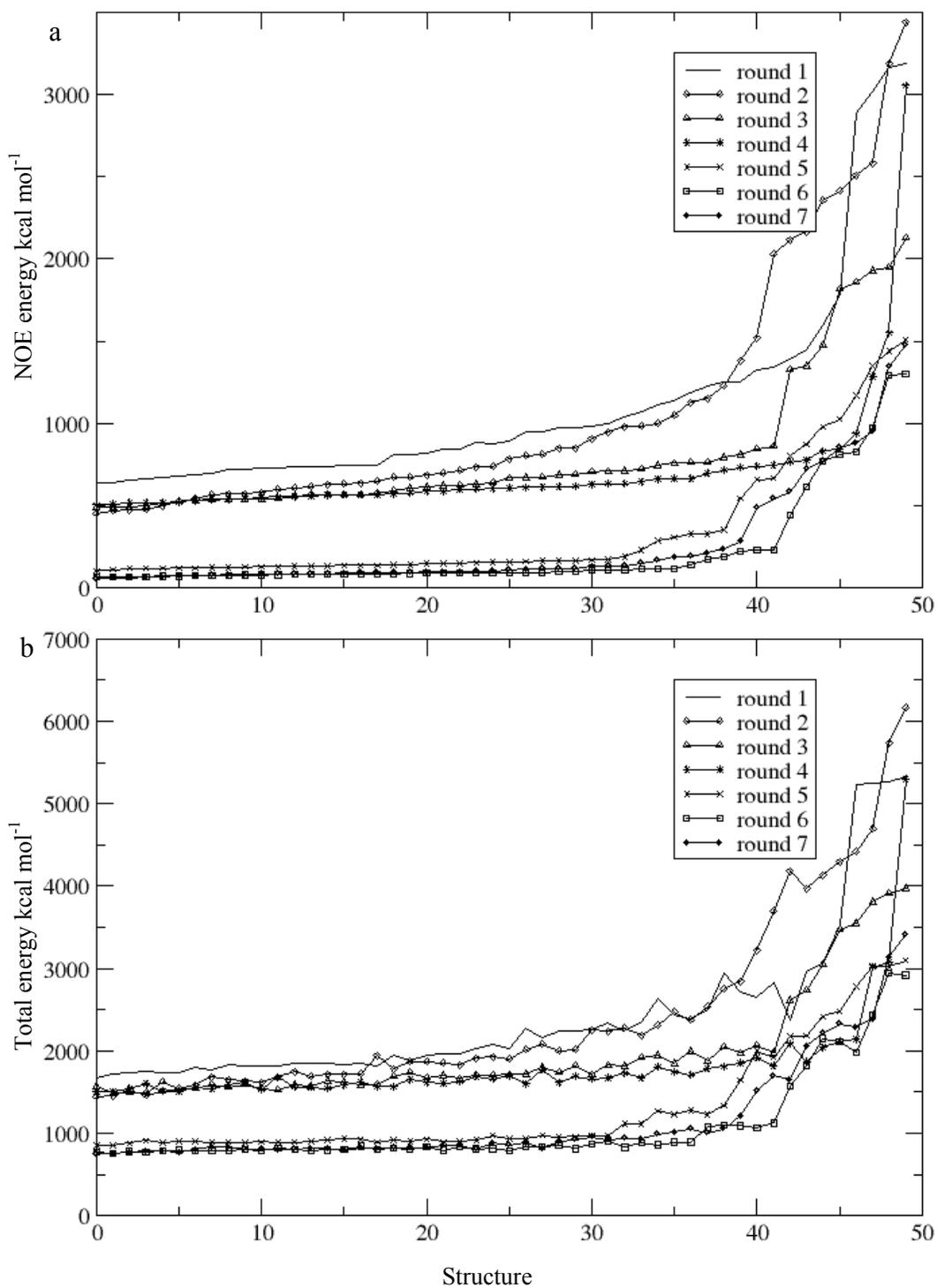
Fig 5.11 The convergence of the calculated structures from different iterations. a: NOE energy. b: total potential energy. The filter level decreased from 0.99 to 0.95. Calibration was carried out from round 5 onwards. Hydrogen bond restraints were introduced in iteration 7.

5.4 Validation of calculated structures

5.4.1 The precision of the ensemble of PA-ICP NMR structures

The precision of the determined structures is directly related to the number of experimental restraints used in the structure calculation. The more restraints are included, the higher resolution can be obtained (Nabuurs et al., 2004). The statistics of the experimental restraints is summarized in table 5.4

NMR restraints	
Total NOE	1703
Ambiguous	502
Unambiguous	1201
Intra-residue	635
Inter-residue	
Sequential (i-j=1)	245
Medium-range (i-j<4)	56
Long-range (i-j>5)	265
Hydrogen bonds	19

Structural Statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.0385±0.002
Deviations from idealized geometry	
Bond lengths (Å)	1.56e-3 ±8.68e-5
Bond angles (^{0})	0.294±8.626e-3
Impropers (^{0})	0.156±1.11e-2

Table 5.4   NMR structural statistics for PA-ICP

When using solution NMR to calculate molecular structures, the spatial arrangement of the atoms in the molecule is restrained in allowed ranges defined by the experimental data. As a result, the three-dimensional structure of a molecule cannot be defined uniquely. Thus, it is

difficult to represent the actual structure using a single model obtained using solution NMR method. This is particularly true for less-defined regions that undergo a high level of dynamics. A meaningful representation of the NMR solution structure of a protein can be produced using a structural ensemble consisting of a set of calculated structures that are each consistent with the experimental data. The structures with the lowest total energies and no more than 2 NOE restraint violations were included in the ensemble. The final ensemble (fig 5.13) of PA-ICP comprising 25 out of 50 structures was superimposed using THESEUS (Theobald and Wuttke, 2006) which makes use of a maximum likehood algorithm.

The similarity of the structures in the ensemble was assessed by calculation of the root mean square deviation (RMSD) of the atomic coordinates of each structure to an unbiased mean structure generated from the ensemble. The mean structure was calculated using an unweighted mean program (UWMN) written by M. Hartshorn and L. Caves, the University of York, UK. A matrix, M, was constructed containing the average distance M_{ij} between atoms i and j in the ensemble of the structures. The internal inconsistency of the mean distances made it impossible to directly project the matrix into three-dimensional space to reconstruct the structure. Alternatively, the matrix was projected into a multiple-dimensional space which was then oriented so that the matrix can be projected back into three dimension to give the mean structure with the least loss of structural information. The results are summarized in table 5.5.

RMSD to the unbiased mean structure (Å)	
Heavy	1.29+/-0.23
Backbone	0.80+/-0.20

Table 5.5   RMSD of the structures in the ensemble to an unbiased mean structure. This was calculated over residues 16-44, 74-95 and 101-111 among 25 refined structures.

5.4.2 Geometric quality of the ensemble of PA-ICP NMR structures

The geometric quality of the calculated structures can be evaluated by the distribution of the backbone (φ and ψ) and side chain (χ_1) torsion angles of residues in a protein (Morris et al., 1992). This information for the ensemble of structures of PA-ICP was plotted using

Procheck-NMR (Laskowski et al., 1996) (fig 5.12), as shown in appendix B. The Ramachandran statistics for the residues (excluding all prolines and glycine residues and residues at the N-terminus) of the ensemble of PA-ICP structures is summarised in table 4.6. More than 90% of the residues fall within the range of most and additionally allowed regions. This is consistent with high quality structures with good covalent geometry. Most residues located in the well-structured β strands lie within (or close to) the favourable regions of the Ramachandran plot for the β sheet across the ensemble. The residues in less well defined regions, such as the N-terminus and the loop region between residue 51 and 70, show a greater scatter of φ and ψ angles than other residues. This is due to the lack of restraints to define these regions of the protein. Residue D8, K17, L18, Q20, G21, E23, T34, G35, R37, V47, P43, A44, A79, S81, D84, R85 and A101-S105 are clustered in unfavourable regions. Most of these residues are located in the flexible loop regions of the molecule except for residues D84-R85, A101-S105, which reside on two adjacent β strands. The unfavorable main chain conformation of these residues may result from being distorted by nearby large side chains.

Regions of Ramachandran plot	% residues
Residues in most favoured regions	62.5%
Residues in additional allowed regions	33.2%
Residues in generously allowed regions	3.4%
Residues in disallowed regions	0.9%

Table 5.6   A summary of the Ramachandran statistics for the ensemble of PA-ICP structures as determined by Procheck-NMR.

The distribution of side chain torsion angles χ_1 are another good indicator of the geometric quality of a calculated structure. χ_1 is the torsion angle of N-C_α-C_β-$X_{\gamma 1}$ about C_α-C_β bond. The distribution of χ_1 for a residue across the ensemble of the structures indicates how well defined the position of the side chain is. The positions of the majority of the side chains in PA-ICP are relatively well defined. No residue in PA-ICP displays a χ_1 distribution in unfavourable regions.
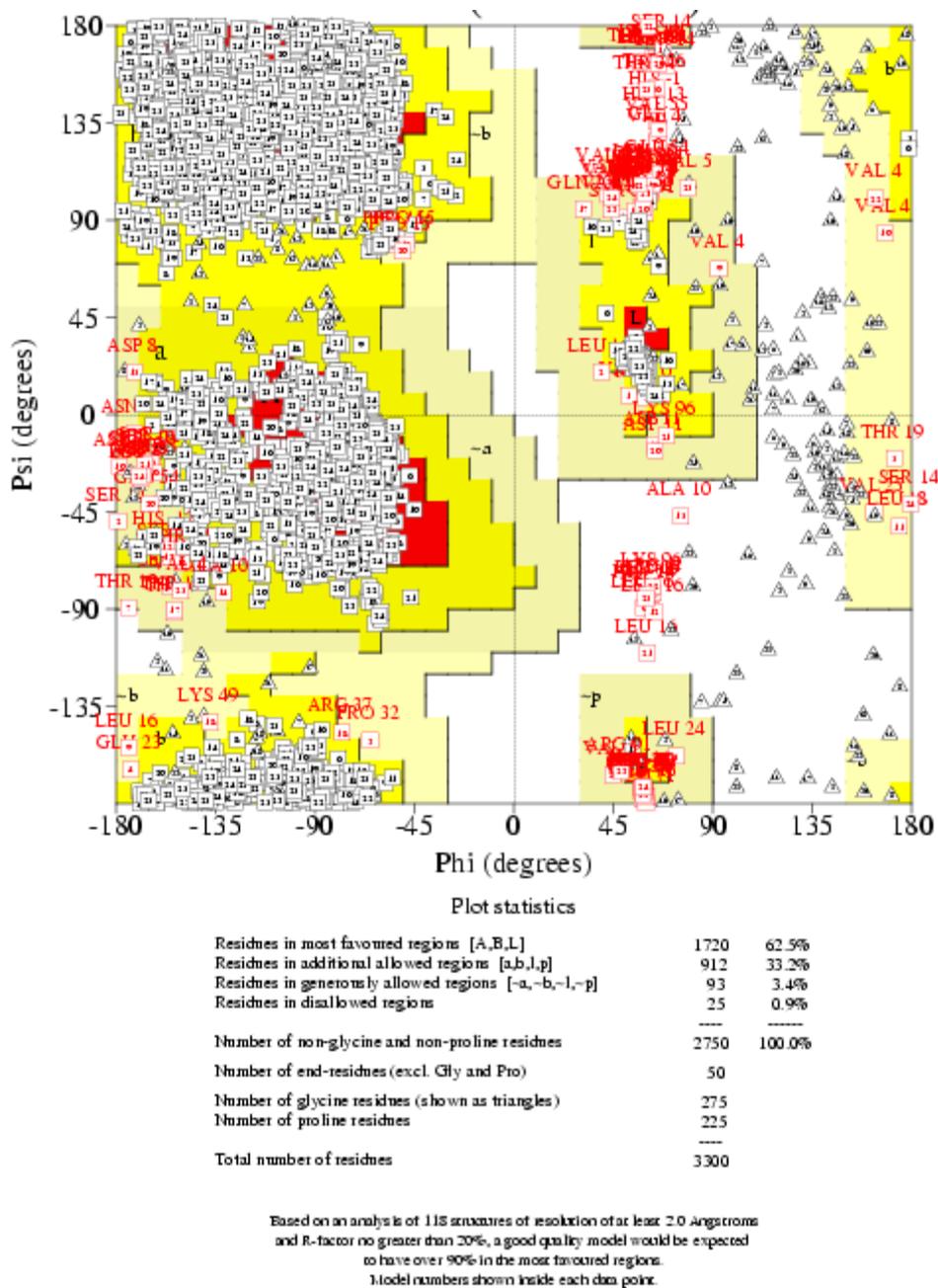
Fig 5.12   Graphical summary of Ramachandran statistics the final ensemble of 25 PA-ICP structures. Structures were analysed using Procheck-NMR.

5.4.3 Stereospecific assignment of valine, threonine and isoleucine residues

As mentioned in section 5.3.2, the stereospecific assignment of the prochiral groups in the protein was achieved using prochiral swapping during structure calculation. To assess whether this method was appropriate for this purpose, the calculated χ_1 angles of valine and isoleucine residues derived from the ensemble of the structures were compared with the χ_1 angles of five residues determined by the experimental data (section 5.2.4). The determined χ_1 angles of V5, V47, V78, V89 and I108 are consistent with the calculated values, suggesting that the prochiral groups in PA-ICP can be stereospecifically assigned using prochiral swapping during structure calculation without explicit assignment.

5.4.4 Violated restraints

None of the NOE restraints used in the final structure calculation were consistently violated by more than 0.5 Å. NOE restraints violated by more than 0.3 Å in more than 10% of the ensemble structures are listed in table 5.7 with possible assignments. All these violated restraints involve the flexible N-terminal residues, implying a calibration error resulting from the assumption that the flexible and rigid parts of the protein have the same correlation time τ_c.

No. of violations	Atom i	Atom j
14	3 PRO HD2	3 PRO HB2
14	3 PRO HA	4 VAL H
13	23 GLU H	2 LYS HA
11	3 PRO HB3	3 PRO HD2
7	103 SER H	102 GLU HG2, HG3
5	4 VAL H	4 VAL HG21, HG22, HG23
5	20 GLN HE22	110 VAL HG11, HG12, HG13
3	1 GLN HA	1 GLN HG2, HG3
3	2 LYS HA	2 LYS HD2, HD3
3	2 LYS HA 37 ARG HA	3 PRO HG2, HG3 93 PRO HG2
3	16 LEU HG 24 LEU HG 79 ALA HB1, HB2. HB3	4 VAL HG21, HG22, HG23 110 VAL HG11, HG12, HG13

Table 5.7   Violated NOE restraints in the final ensemble of 25 PA-ICP structures. Restraints violated by more than 0.3 Å in 3 or more structures are displayed.

5.4.5 Validation of PA-ICP with RDC restraints

The accuracy of the calculated PA-ICP structure was assessed using RDC restraints. The RDCs measured from the flexible regions were excluded since these would be averaged and inconsistent with any individual structure. Overlapped crosspeaks in the IPAP spectra were also excluded. 42 HN RDC restraints, 43 NCO RDC restraints and 43 H_nCO RDC restraints representing the residual dipolar coupling among the backbone amide ^1H, ^{15}N nuclei from residue i and the carbonyl ^{13}C from residue i-1 were collected as described in chapter. The restraints were fitted to the structure with the lowest RMSD to the unbiased mean structure of the ensemble using the program of module 2 (Dosset et al., 2001). The fitted data display a reasonable fit between the measured and back-calculated RDC values (fig 4.13), even without refinement with the RDC restraints, validating the structures calculated based solely on NOE and hydrogen bond restraints.
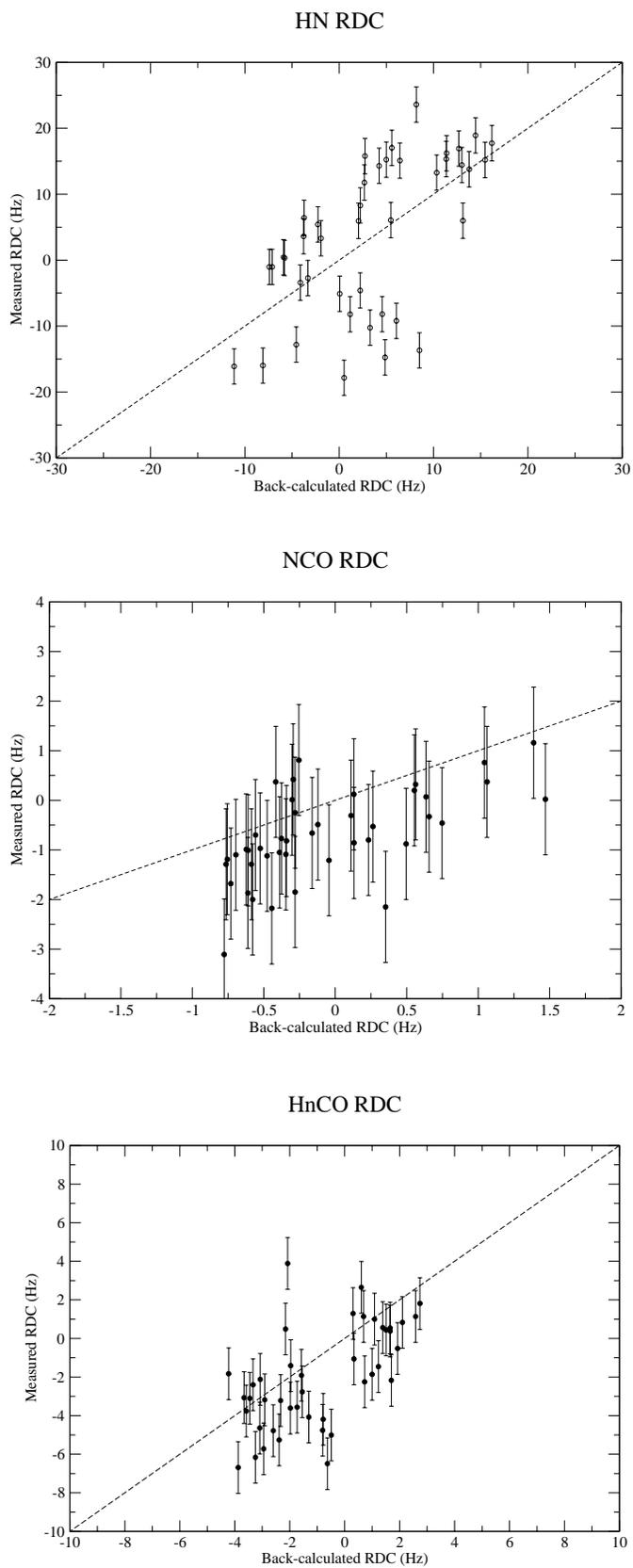
Fig 5.13   Plot of the back-calculated RDC restraints against experimental RDC restraints. The distribution of data points are scattered along the diagonal.
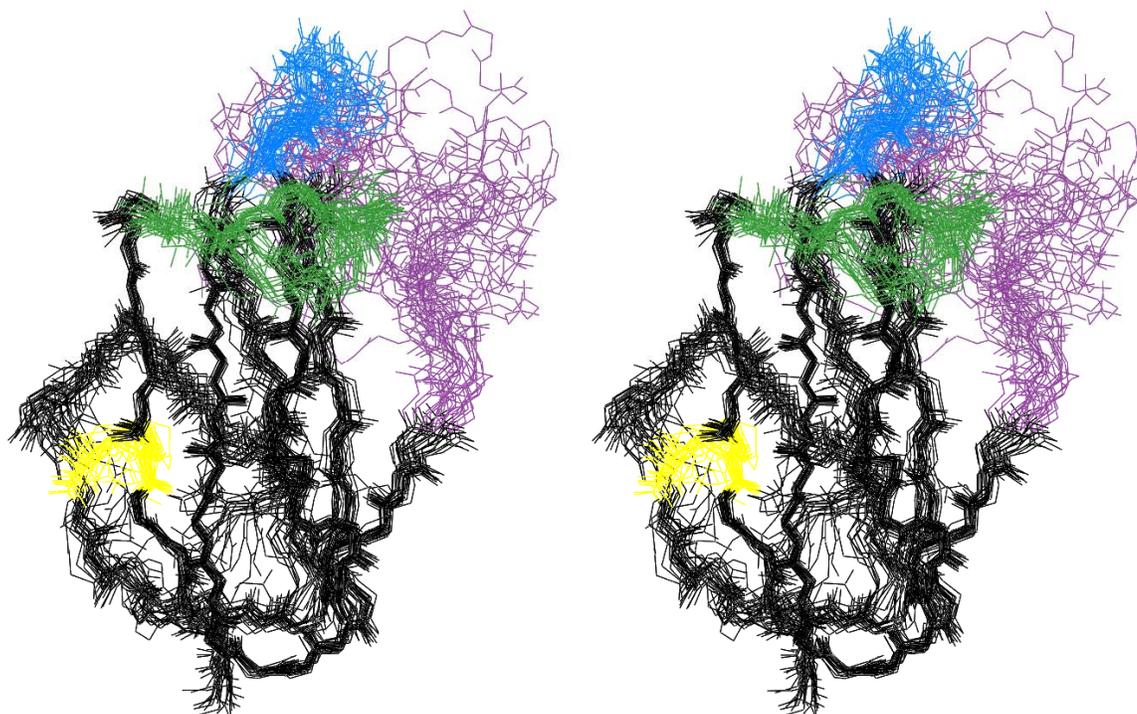
Fig 5.14   A stereo-view of the ensemble of the final 25 PA-ICP solution structures superposed on backbone C_α of residues in well-defined regions (residue 16-44, 74-95 and 101-111). The three less convergent inter-strand loops are coloured in blue (BC), magenta (DE) and green (FG) respectively. The disulphide bond was highlighted in yellow.

5.5 Conclusion

The PA-ICP structures were already reasonably well converged at the early iterations of the calculation, owing to the careful NOE crosspeak picking strategy. Inclusion of hydrogen bond restraints did not have appreciable impact on the potential energy of the calculated structures, implying that the NOE restraints on their own were already able to provide a well refined PA-ICP structure although the RDC restraints may help the further refinement of the overall quality of the ensemble.

CHAPTER 6

DYNAMICS STUDY OF PA-ICP

6.1 Overview

Having assigned the chemical shifts of the backbone amide ^{1}H and ^{15}N atoms, dynamics information of the backbone NH vectors can be extracted by studying the ^{15}N relaxation rates. In this thesis, the longitudinal and transverse relaxation time T_1 and T_2 and steady state {^{1}H, ^{15}N} heteronuclear NOE were measured. The internal motions of the backbone HN vectors were modelled using Lipari-Szabo model free formalism.

6.2 Measurement of T_1, T_2 relaxation rate and heteronuclear NOE

T_1 and T_2 spectra were based on the ^{15}N HSQC experiment collected as a pseudo 3D ^{15}N HSQC fashion with the time points selected on an extra dimension for T_1 and T_2 measurements (Kay et al., 1989, Farrow et al., 1994, Farrow et al., 1995). Both experiments utilize a refocused INEPT sequence to achieve an in-phase amide ^{15}N magnetization along the x axis transferred from the ^{1}H spin. An additional 90^0 pulse is applied in the T_1 relaxation experiment to flip the magnetization along the z axis before the relaxation period. In the T_2 relaxation experiment, the relaxation modulates ^{15}N coherence stored in the transverse plane by incorporation of a CPMG pulse train. For PA-ICP, the heights of 78 backbone amide cross peaks could be reliably quantified out of 96 assigned backbone HN cross peaks. For T_1 relaxation, the delays were set to 51, 301, 601, 901, 1201 ms with a duplicate spectrum at 301 ms for error estimation. For T_2 relaxation, the delays were set to 16, 32, 84, 80, 96 and 128 ms with duplicates at 16 and 48 ms. Highly overlapped signals were omitted from relaxation analysis.
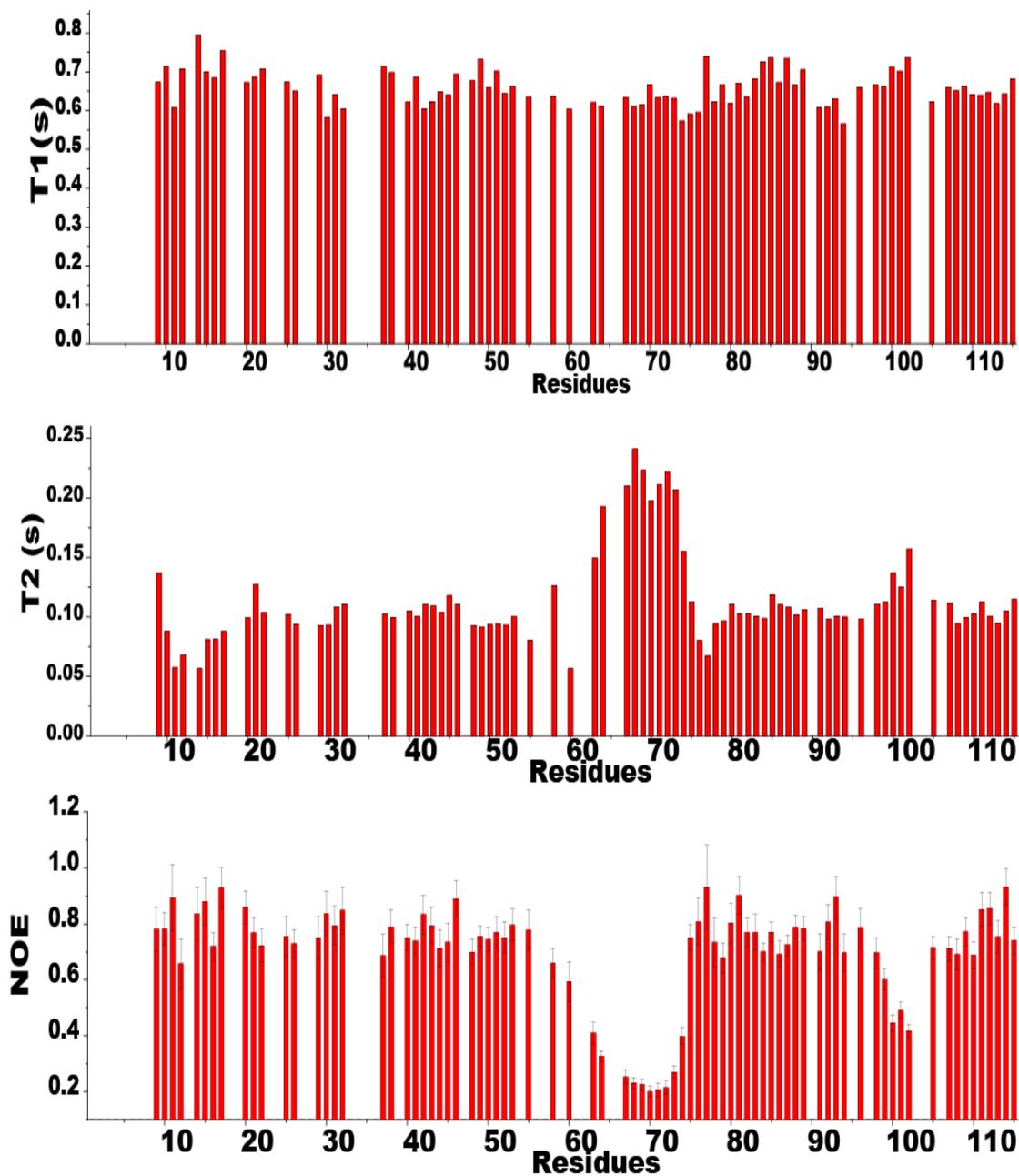
Fig 6.1   The ^{15}N T_1, T_2 and {^{1}H-^{15}N} heteronuclear NOE values for each resolved backbone amide in PA-ICP detected at 60.8 MHz (^{15}N) and 308K.

6.3 Estimation of overall correlation time and rotational diffusion tensor

As mentioned in section 2.4, the overall molecular tumbling can be either isotropic or anisotropic depending on the shape of the molecule. How the molecule tumbles needs to be estimated before extracting the internal dynamics information because underestimation of the anisotropy of the overall rotational motion may result in an overestimation of R_{ex}. The correlation time of the molecule can be extracted from the T_1/T_2 ratio (Kay et al., 1989). A plot of T_1 against T_2 allows a preliminary approximation of the rotational correlation time (fig 6.2).
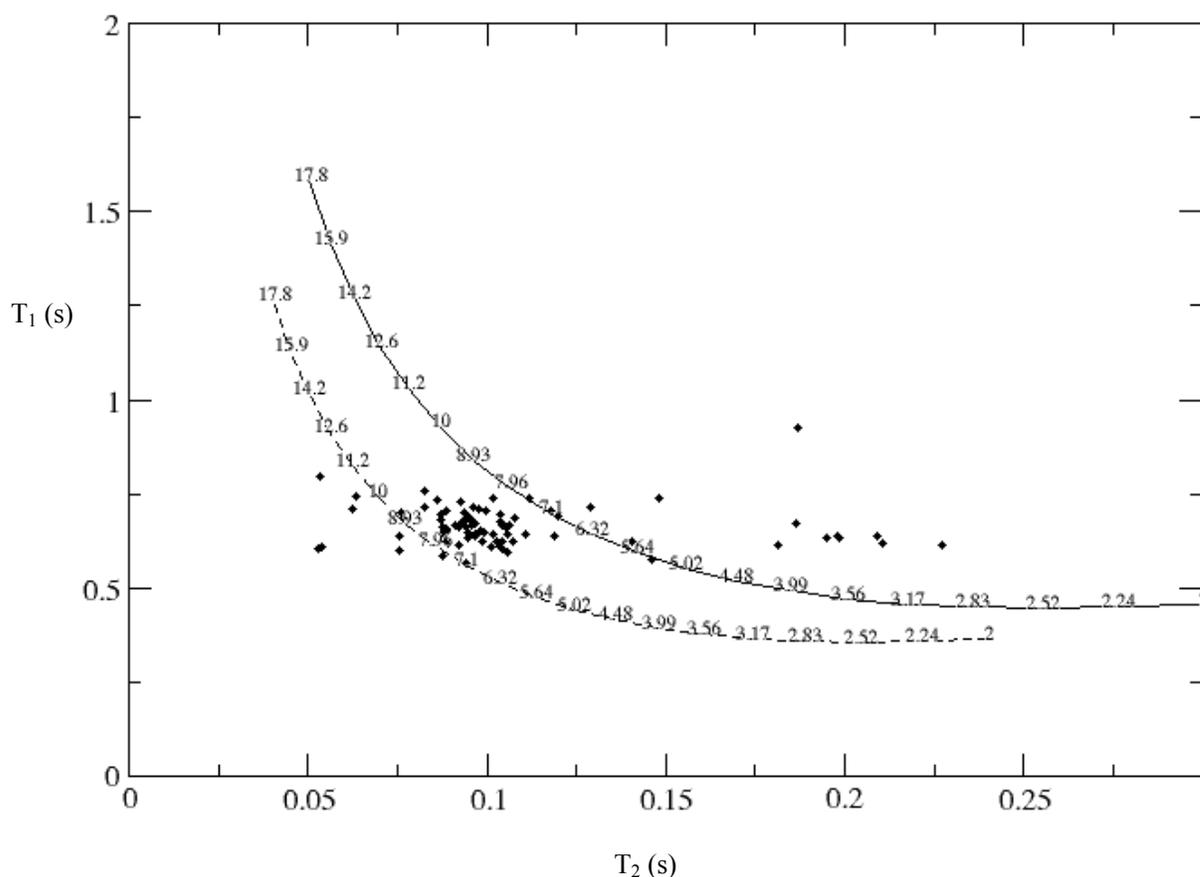
Fig 6.2   A plot of ^{15}N T_1 against T_2 for PA-ICP recorded at 60.8 MHz on a uniformly ^{13}C, ^{15}N labelled sample. Each data point corresponds to an individual residue of the protein. The residues of the His- tag and the residues giving rise to overlapped amide cross peaks were omitted. The solid line represents simulated S^2 of 0.8 and dashed line 1.0 with estimated correlation time for isotropic tumbling corresponding to expected T_1/T_2 values.

For most residues, the implied correlation times fall into a tight range between 6.3 to 9.0 ns, indicating that PA-ICP does not exhibit highly anisotropic rotational diffusion. The majority of the points are clustered between the two lines, corresponding to residues in relatively well-defined regions in the 3D structure. In the T_1/T_2 plot, there are several points lying at $S^2<0.8$ and $S^2>1.0$. The points with a general order parameter less than 0.8 are primarily contributed by residues 59-70 and 95-98, the highly disordered regions as suggested by the heteronuclear NOE data (fig 6.1). In theory, no residue should appear outside the $S^2=1.0$ boundary. However, the presence of slow conformational or chemical exchange R_{ex} will result in an increase in R_2 and therefore a shorter T_2. It consequently moves the associated T_1/T_2 points further to the left such that they lie outside the $S^2=1.0$ boundary. The residues occurring on the left of $S^2=1.0$ in figure 6.2 are mainly located on three β strands forming one of the β sheets with residue L7 and Y56 having the biggest deviation from $S^2=1.0$ together with residue S72, L26 and D8 lined up across one face of the β sheet formed by strands (fig 6.3). The co-localization of the residues undergoing slow exchange may be due to the conformational exchange experienced by N-terminal residue L7 and the movement is propagated among the above residues aligned across the same β sheet, while residue Y56 is water-exposed and therefore undergoes conformational exchange.
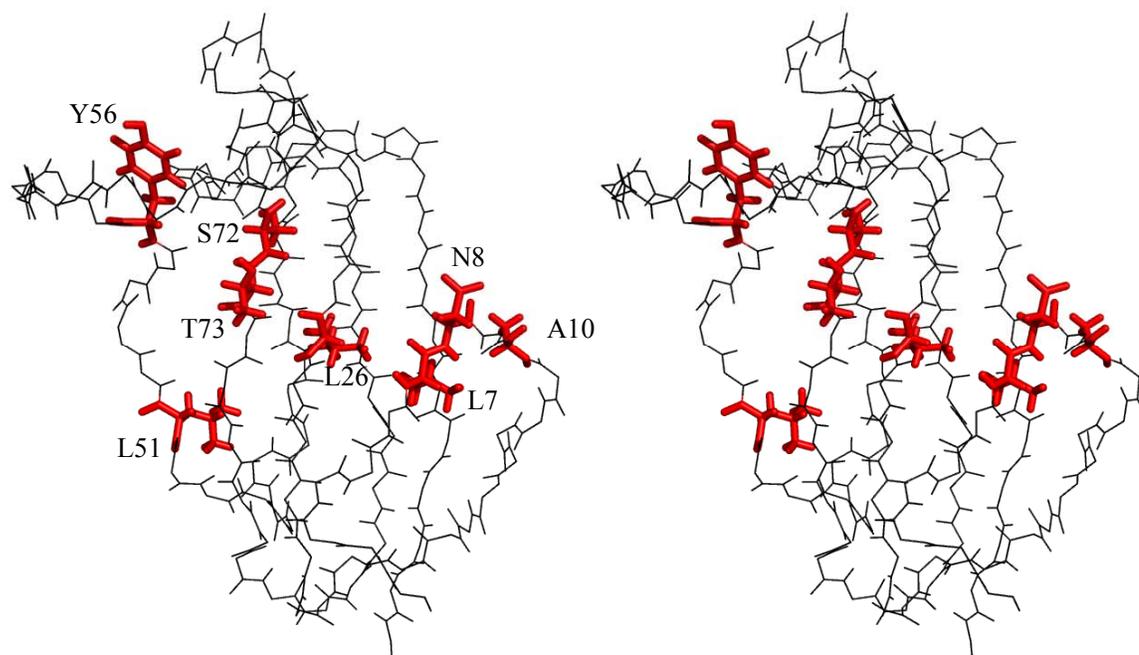
Fig 6.3   A stereo-view of the structure closest to the mean structure of the 25 calculated structures. The location of the residues in PA-ICP experiencing slow-exchange residues (red) as indicated by T_1/T_2 plot (fig 6.2). Their co-localization suggests a possible presence of propagated conformational change.

On the basis of the spins in PA-ICP with fast and limited internal motions, the average τ_m value was estimated to be 7.62±0.5 ns. This value is similar to that determined for monomeric *L. mexicana* ICP (7.4 ns). With the availability of the PA-ICP solution structure, the rotational diffusion tensor of the molecule was estimated from the T_1/T_2 ratio of the 45 backbone ^{15}N spins in PA-ICP that do not experience chemical exchange and undergo restricted internal motions. The analysis carried out using a program provided by Palmer (1991) indicated that the majority of the residues showed similar correlation times close to the average value with a weakly anisotropic axially symmetrical diffusion tensor of D_{par}/D_{perp}=1.21. The anisotropy is small enough to treat the motion of PA-ICP as isotropic (Hall and Fushman, 2003).

6.4 Internal motions of PA-ICP backbone HN vector

The internal motion of the PA-ICP backbone was modelled from the three relaxation parameters of T_1, T_2 and heteronuclear NOE by 5 possible models with no more than 3 model-free parameters. The analysis was carried out using the program ModelFree4.20 (Mandel et al., 1995, Palmer et al., 1991). The 5 models were model 1 (S^2), 2 (S^2, τ_e), 3 (S^2, R_{ex}), 4 (S^2, τ_e, R_{ex}) and 5 (S^2, τ_e, S^2_f). For model 1-4, S^2_f= 1.0 and S^2= S^2_s. The selection of the models was performed based on Mandel's protocol (1995) with small modifications suggested by Chen (Chen et al., 2003). The flowchart of the protocol used for modelling dynamics of PA-ICP backbone is shown in figure 6.4. Having selected the best-fit dynamical model for each nuclear spin, the overall rotational diffusion of PA-ICP and the internal motion parameters were further optimised. The final optimisation gives rise to a global τ_m of 7.725 ns. Among the analysed 79 residues, 17, 16, 7, 2 and 19 residues were sufficiently fitted by model 1, 2, 3, 4 and 5 respectively. The dynamics of residues L7, D12, L16, L18, L28, T34, A44, S46, L51, T73, A80, D84, V89, D97, F104, I108 and V110 could not be adequately fitted by any model (see the flowchart for SSE criterion) . The optimised modelfree parameters of S^2, τ_e, R_{ex} and S^2_f for the backbone HN vector of each residue are listed in appendix C together with its sum square error (SSE) and F-statistic value.

The square of order parameter S^2 is typically between 0.75–1 for most fitted residues, except for residues 54-70 whose S^2s are between 0.25-0.5 and residues 96-101 whose S^2 between 0.55-0.78, suggesting that these two regions are flexible in solution with the first region experiencing more flexibility (smaller S^2 values). The residues in these two loops are

best modelled with two order parameters, S^2 and S^2_f, with distinct timescales faster than τ_m and internal correlation time of order of a nanosecond. The only exception is residue Y56, which is best modelled by an S^2 of 0.9, internal correlation time of 0.5 ns and a millisecond timescale chemical exchange term. The residues in the N-terminal region (residue 6-13) are best fitted by S^2 and R_{ex} parameters, indicating the presence of a slow motion in an order of milliseconds. Most residues in the other loops are best fitted with 2 or 3 modelfree parameters, whereas the residues in the well-structured strands are best modelled with one (S^2) or two (S^2 and τ_e) parameters. Residues D8, A10, Y56 and S72 can be modeled by 2 or 3 model free parameters each including a chemical exchange term, which is consistent with the qualitative analysis of the T_1/T_2 plot (fig 6.2) in the last section.

For PA-ICP, certain residues can not be fitted with any one of the five models. These residues primarily reside at the junctions between the β strands and the loops. Given the location of these residues, the combination of the different dynamic properties of the well structured and the highly disordered regions are likely to influence their dynamics and result in complex motions which can not be sufficiently modeled by the limited number of model free parameters. Therefore, acquiring extra experimental data points at different fields or sampling different slow exchange time scales may be useful for fitting the relaxation data to models.

Fit experimental data
with model 1

Is
$SSE_i<SSE_i(0.05)$?

Y → Use model 1 for spin i

N

N

Fit experimental data
with model 2

Fit experimental data
with model 3

Is
$SSE_i<SSE_i(0.1)$?
And
$F_i>F_i(0.3)$?

Is
$SSE_i<SSE_i(0.1)$?
And
$F_i>F_i(0.3)$?

Y

N

N

Y

Use model 2 for spin i

Use model 3 for spin i

Use the model with the
lowest SSE_i for spin i
among model 1,2,3

N

Is
$SSE_i>SSE_i(0.05)+20$ for
Model 1,2,3?

Y

Y

Fit experimental data
with model 4

Fit experimental data
with model 5

Is
$SSE_i<0.1$?

Is
$SSE_i<0.1$?

Y

N

N

Y

Model 4 sufficiently
fits spin i

Model 5 sufficiently
fits spin i

No model is
appropriate for spin i

Fig 6.4   The flowchart of model selection strategy based on Mandel's method with small modifications made by Chen.
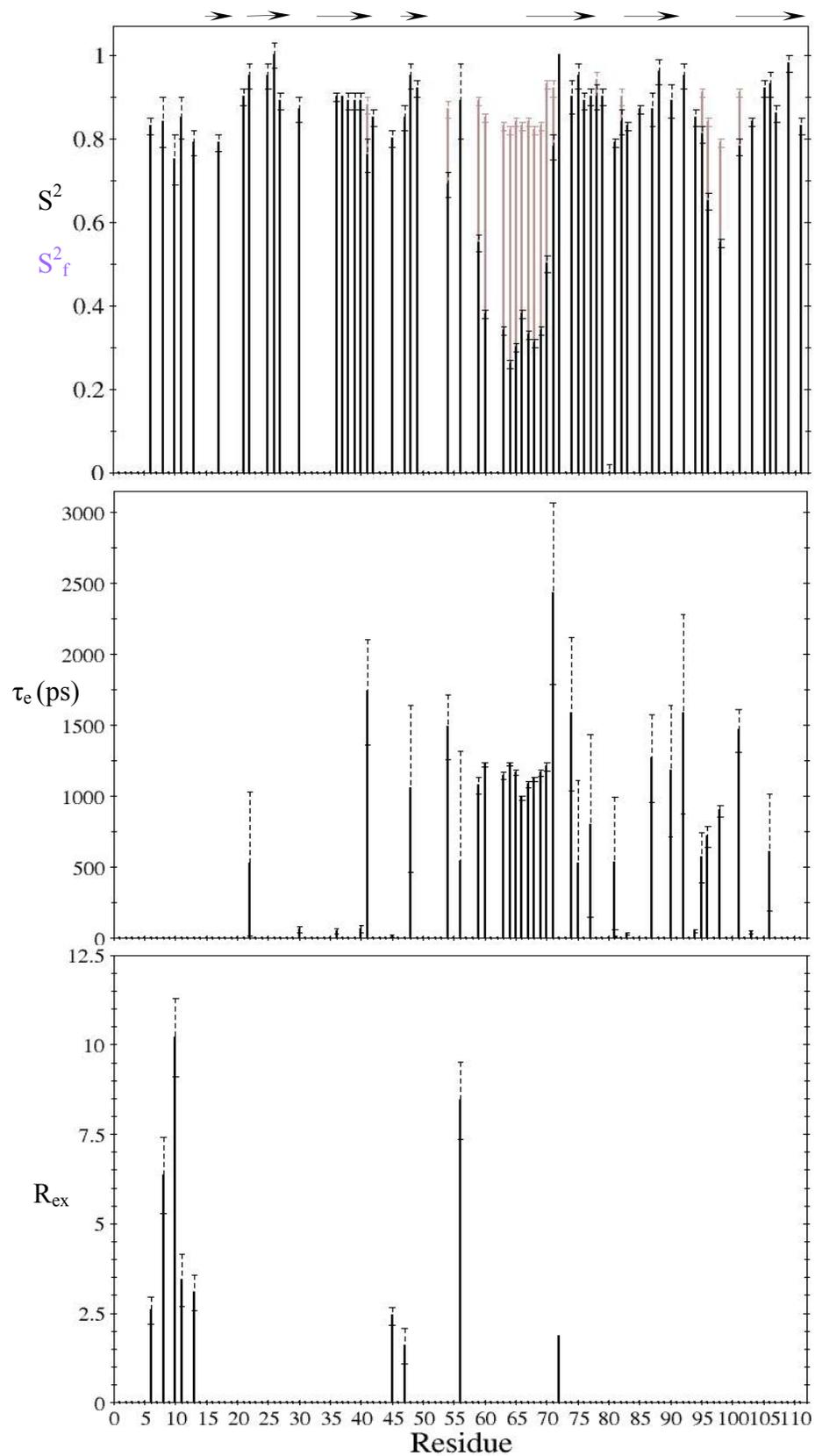
Fig 6.5 ^{15}N backbone dynamics of PA-ICP. Model free parameters are plotted as a function of residue number. The location of the β strands are indicated by arrows.
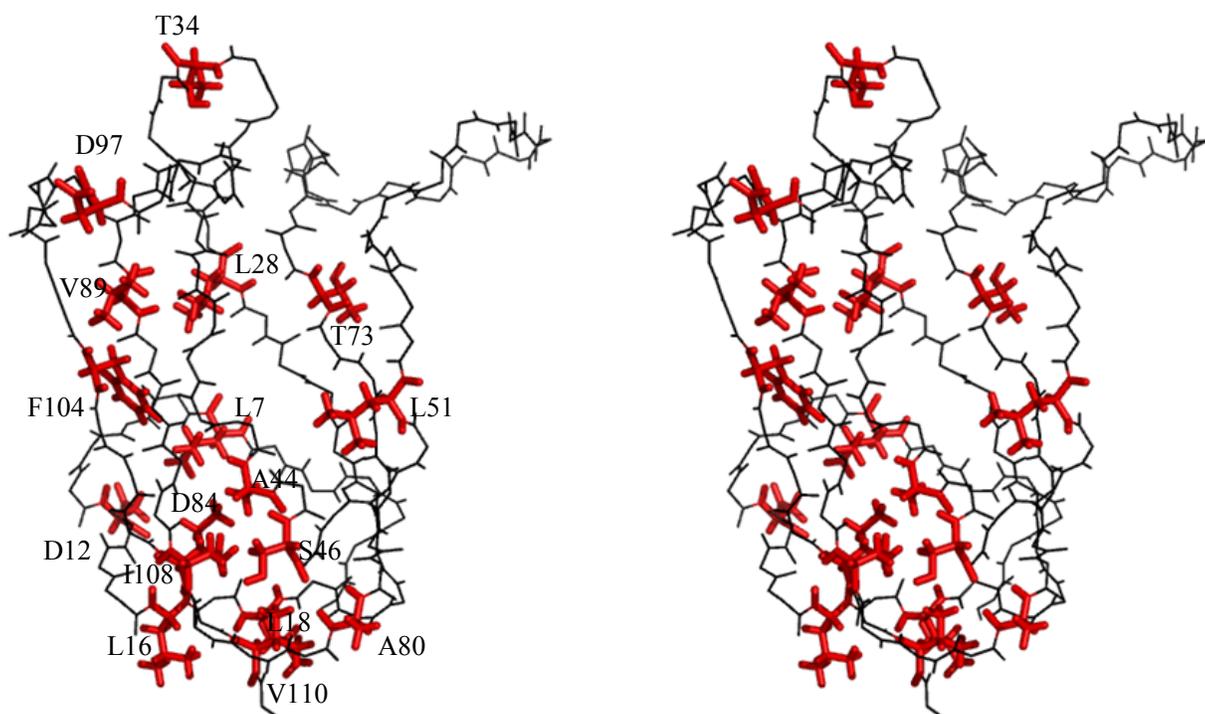
Fig6.6   A stereo-view of the structure closest to the mean structure of the 25 calculated structures. The residues which can not be fitted by any model are highlighted with red sticks.

6.5 Conclusion

The relatively longer T_2s and smaller NOEs implied a highly flexible region around residues 60-75. Another less flexible region from residue 96 to 108 was also identified (fig 6.1). The motion of the residues in these regions were modelled with three model free parameters S^2, S^2_f and τ_m. The backbone dynamics study of PA-ICP revealed that two loop regions have higher degree of mobility than the rest of the protein. The molecule undergoes isotropic rotational diffusion with a good proportion of the residues whose motions can be modelled by 5 model free models. However, 23% of the analysed HN vectors can not be adequately fitted by any models. This inadequate fit phenomenon has also been reported (Katahira et al., 2001, Pang et al., 2002) although there are less unfitted residues. These residues in PA-ICP are primarily located at the junctions connecting the β strands and the loops and may experience complicated dynamics. This may be improved by acquiring additional experimental data points.

CHAPTER 7

SOLUTION STRUCTURE OF PA-ICP

7.1 Overview

In chapter 5, the structure of PA-ICP was calculated using NOE and hydrogen bond restraints. This chapter presents the structural features of PA-ICP and compared them with its homologs, chagasin and *L. mexicana* ICP. A homology model of the PA-ICP-cathepsin L complex was also built using the crystal structure of chagasin and cathepsin L complex as a template on the basis of the calculated PA-ICP structures.

7.2 The structure of PA-ICP

PA-ICP adopts an immunoglobulin-like beta-sandwich fold. The protein consists of seven β strands forming two β sheets. The first β sheet is formed by a very short strand A (residue L16 and K17) parallel to C (residue R37 to N42), which is in turn anti-parallel to strand F (residue D85 to R91) with anti-parallel orientation to strand G (residue S101 to Q109). In the other β-sheet, strand B (residue E23 to L 28) contacts strand E (residue S72 to V78) in anti-parallel orientation with strand D (residue L48 to R50) anti-parallel to strand E. The two β sheets are well-packed enclosing a hydrophobic core comprising residues L24, L26, L28 on strand B; W38, L40 on strand C; W74, F76 on strand E, L86, L88 on strand F and F104, I108 on strand G. These residues are principally hydrophobic residues conserved along the strands in an alternating manner (fig 7.1). The conformation of the N-terminal loop region is maintained via hydrophobic interactions among residues L7, A10 and L26, F104 together with the disulphide bridge between residues C13 and C106. There are five inter-strand loop regions in PA-ICP. The DE loop was poorly defined by the NMR data and had the least structural convergence. This is usually circumstantial evidence of high mobility and the dynamics of this loop was confirmed by ^{15}N relaxation experiments. This loop region together with other two less flexible loops lie on one end of the protein bearing three highly conserved groups of residues reported for all members in ICP family (Rigden et al., 2002). The NPTTG motif is located in the loop BC flanked by the GXGG motif lying in the highly flexible loop DE and the RPW/F motif on loop FG. The studies of PA-ICP homologs have identified them as forming the CP-binding site.
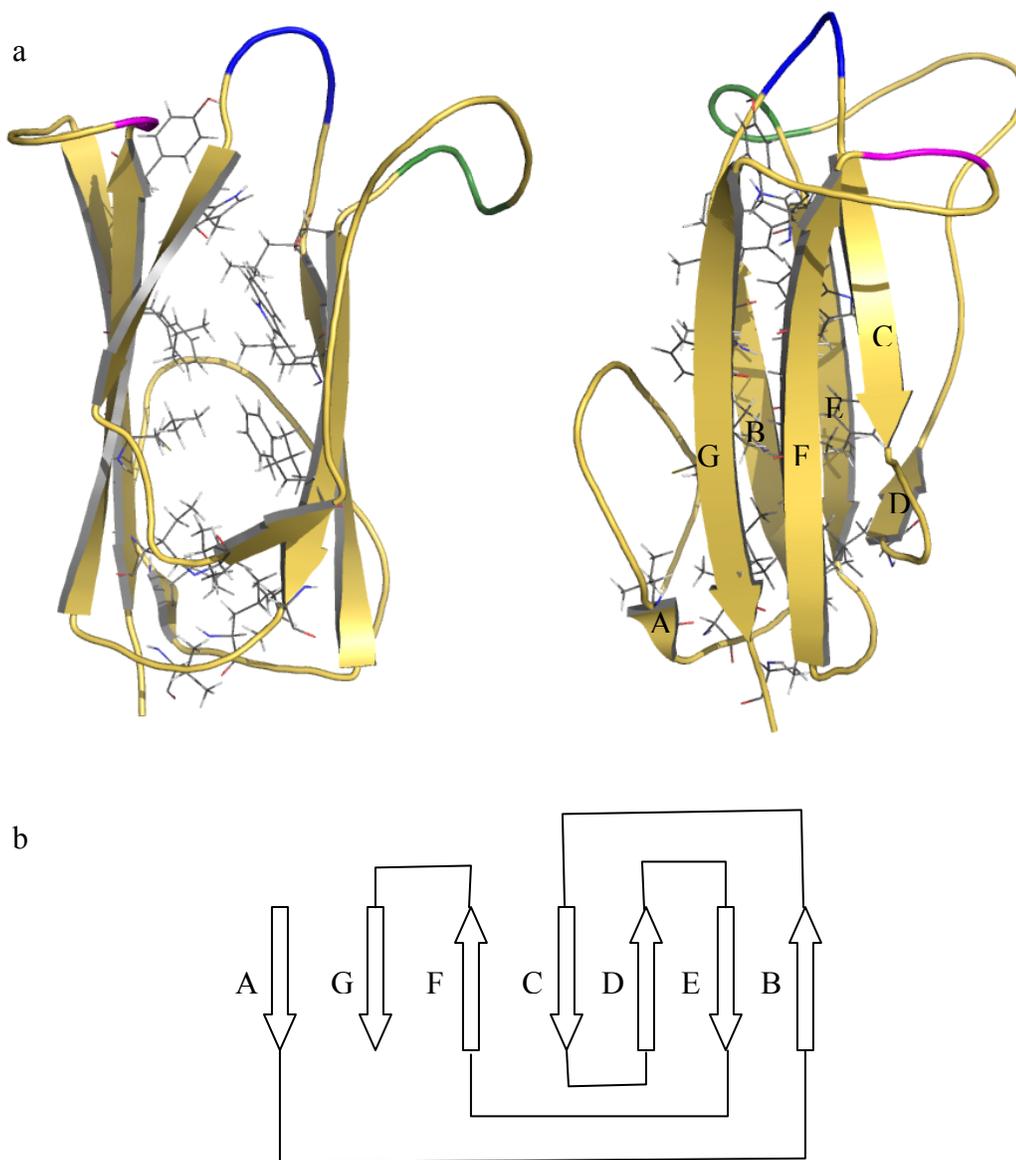
Fig 7.1   a. Cartoon representations of PA-ICP structure. The residues forming the hydrophobic core are shown as gray lines and the conserved BC, DE and FG loop regions are colored in blue, green and red respectively. b. Topological representation of PA-ICP.

7.3 Comparison of the global fold of PA-ICP with those of chagasin and *L. mexicana* ICP

Sequence alignment of the ICP family indicates that all ICP proteins adopt a similar immunoglobulin like fold. The solution structures of PA-ICP, chagasin and *L. mexicana* ICP confirm that this is the case as these three proteins consist of two β-sheets formed by 7 to 8 β–strands folded to a β sandwich. In general, these proteins fold in a similar way. The SCOP database has classed the proteins to a ICP-like family with similarity to the Cupredoxin-like fold (http://scop.mrc-lmb.cam.ac.uk/scop/). However, the first two β–strands and the 3_{10} helix observed in chagasin solution structure are less well defined in PA-ICP. Comparison of the structures of these ICP proteins also reveals that strand C in PA-ICP is longer than the corresponding strands of 4 in chagasin and C in *L. mexicana* ICP. This is because the following loop region in PA-ICP is shorter than the equivalent loop regions in chagasin (loop 3) and *L. mexicana* ICP (the CD loop) which are pinned back to the outer face of one β-sheet. The neutral residues that serve to pin the loops, M35, T37 and T93 in chagasin and A35, Y37 and M90 in *L. mexicana* ICP, are replaced by charged residues in PA-ICP (R37, E39 and R91). In contrast, strand D of PA-ICP is shorter than the corresponding strands 5 in chagasin and D in *L. mexicana* ICP while the following loop DE is significantly longer than the equivalent loop regions in chagasin (loop 4) and *L. mexicana* ICP (DE loop). The extension of the strand is disrupted by the large side chain of R50, which appears to displace the side chain of L51 to the surface of the protein in PA-ICP. What is more, the signals from the backbone amide groups of residue E54, V55, S57 and N58 were absent from the ^{15}N HSQC spectra, indicating fast backbone amide proton exchange with water for these residues at pH 7.2. This suggests that this region is more accessible to the solvent than other parts of the protein, as a consequence of being less well structured.

Despite the structural differences between PA-ICP and chagasin and *L. mexicana* ICP, its global fold still places the three highly conserved motifs on one end of the molecule to form the binding site to interact with target cysteine peptidases. The dynamic data indicates that the DE loop of PA-ICP experiences the greatest degree of mobility as can be seen for the equivalent loop 4 in chagasin and loop DE in *L. mexicana* ICP. In addition, a depressed S^2 value was modelled for the FG loop of PA-ICP which was not observed from the dynamics data for the corresponding loops in chagasin (loop 6) and *L. mexicana* ICP (FG loop),

suggesting that this loop is more dynamic in PA-ICP than in chagasin and *L. mexicana* ICP. The differences in the specificity of ICPs to different cysteine peptidases may be influenced by the flexibility of these two loops.

7.4 Inhibitory activity of PA-ICP to cathepsin L, H and B, compared to Chagasin

The dissociation constants K_i for chagasin binding to cathepsins have been determined by dos Reis (A) (dos Reis et al., 2008), Wang (B) (Wang et al., 2007)and Redzynia (C) (Redzynia et al., 2009). The results along with K_i values determined for PA-ICP are listed in table 7.1. The K_i values for chagasin to the same peptidases are considerably different. The possible cause of the inconsistency may be variations in assay protocols and conditions. The protocols for A and B differ from C in that the result A and B were determined with an additional pre-equilibration of the inhibitor and the enzyme prior to the addition of the substrate for assay. The dissociation constant measures the ratio of the free enzyme and inhibitor concentrations to the enzyme–inhibitor complex concentration at steady state and the equilibrium is not established instantaneously after mixing the inhibitor with the enzyme. Failure to reach the steady state before the enzyme assay may lead to an underestimation of the inhibitory activity. Therefore, an optimal length of pre-incubation procedure is essential for accurate measurement. This is especially true for cathepsin B, which, when assayed with PA-ICP, took 15 mins to reach the equilibrium (data not shown). This may explain why in the assay of chagasin activity toward cathepsin B, it appeared to have a K_i 100 times lower in result C than in B. The assay condition of pH could also have an impact on the K_i determination, as it affects the electrostatic potential on the protein surfaces, which could be particularly important at the molecular interface. The active site cleft of cathepsin L has a large negative potential and accommodates several acidic residues that make contacts with the inhibitor (fig 7.2). In the complex, residue E141e, for instance, is in close proximity to K43 of the inhibitor. It seems to interact with K43 competing with E71 of chagasin. A lower pH will make the carboxyl group of this residue more likely to be protonated and weakens the interaction with the positively charged residue. Therefore, the K_i determined at pH 6.5 may be less than determined at pH 5.5, as seen in results A and B.

Enzyme	K_i (nM)			
	Chagasin (A)	Chagasin (B)	Chagasin (C)	PA-ICP
Cathepsin L	0.007±0.0012	0.35±0.1	0.039	0.0065±0.0005
Cathepsin H	N/A	15±4.8	N/A	10.43±1.3
Cathepsin B	N/A	100±9.5	0.93	138

Table 7.1   Dissociation constants of chagasin and PA-ICP to cathepsin L, H and B determined using different protocol and experimental conditions.

A sensible evaluation of inhibitory activity of chagasin and PA-ICP to cathepsin L is assessed by comparison with the K_i value determined by dos Reis because the protocols and pH conditions used for the determination were identical. While for cathepsin B, the dissociation constant of PA-ICP is compared with that of chagasin determined by Wang as the pH was adjusted to 6.0 for assaying the chagasin-cathepsin B complex. The result indicates PA-ICP inhibits cathepsin L, H and B to a degree highly comparable to chagasin. The inhibitory activities are nearly indistinguishable against cathepsin L. When interacting with cathepsin B, PA-ICP shows a slightly lower inhibitory activity than chagasin wherase, it inhibits cathepsin H slightly better than chagasin.
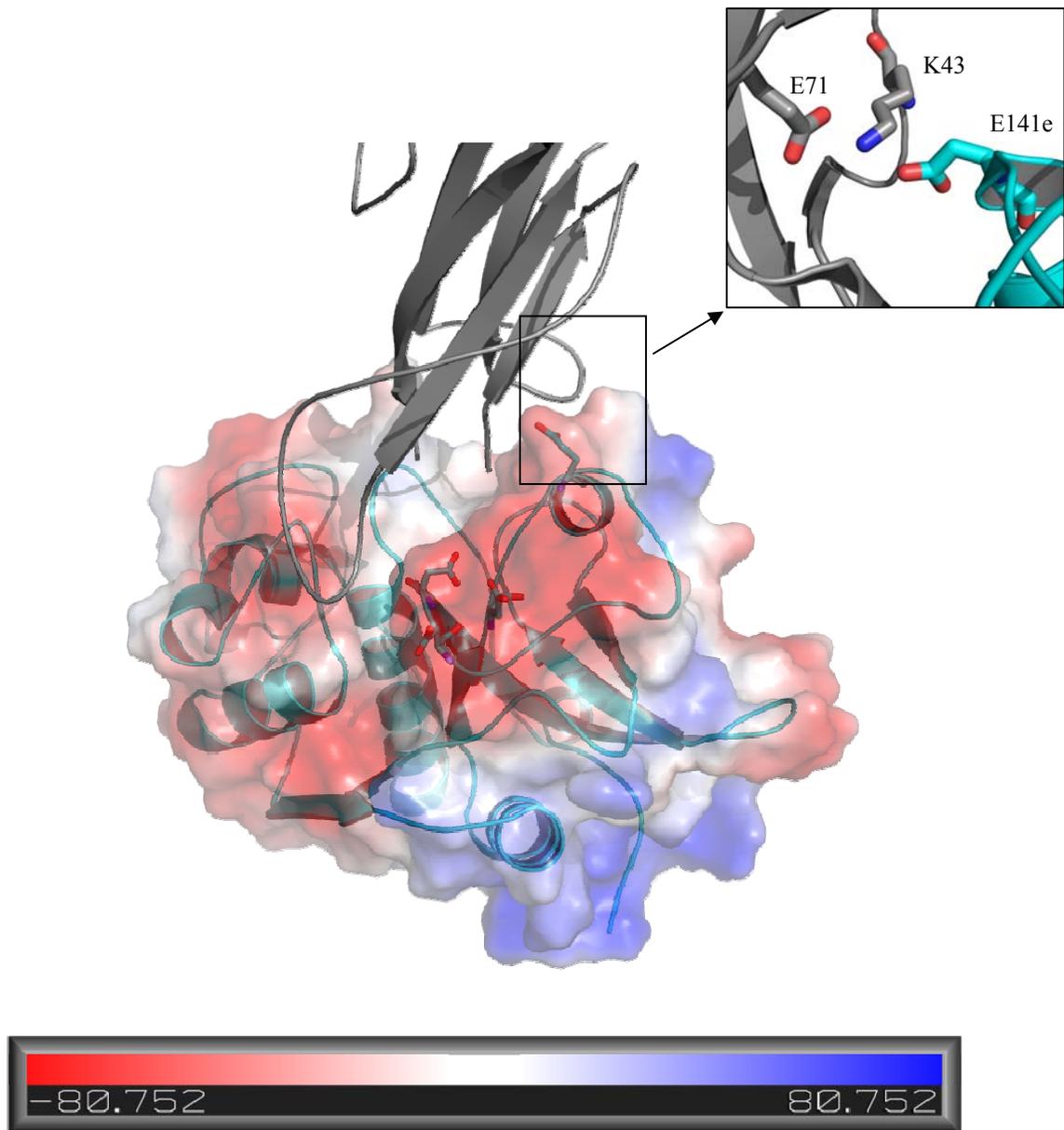
Fig 7.2   A cartoon representation of the predicted binding of PA-ICP (gray) to the active site cleft of cathepsin L (blue). The surface of the enzyme is modelled according to its electrostatic potential from red (negative) to blue (positive) using PYMOL. The enzyme's active site is highly negatively charged and contains several Glu and Asp residues (highlighted with fatter sticks) that interact with chagasin (grey). For example, E141e is in close proximity to K43 of chagasin, which is also close to residue E71.

7.5 Predicted PA-ICP-Peptidases Interactions

Given the similarity of the structure and inhibitory activity between PA-ICP and chagasin, PA-ICP is expected to interact with model CPs in an analogous fashion. In order to shed light on its inhibitory function, models of the PA-ICP-cathepsin L (fig 7.3-6) complex were created using modeller 9.5 based on the coordinates of the chagasin-cathepsin L complex (Sali and Blundell, 1993). The intermolecular contacts discovered in chagasin in complex with cathepsin L may be imitated by PA-ICP (Ljunggren et al., 2007). The important hydrogen bond between the side chains of R92 (PA-ICP numbering) and N18e is conserved (fig 7.4). The interaction is supported by the recognition of the hydrophobic cluster close to the enzyme active site by W94 through packing against W189e, W193e and F143e. The shape of loop FG is maintained by P93 via hydrophobic interaction with the side chains of F145e and L144e. Loop DE, that experiences the greatest degree of dynamics, contains the GXGG motif. The conformational flexibility of glycine residues is proposed to be crucial in this interaction as a mutated *L. mexicana* ICP with these glycines replaced by more rigid proline residues diminished the inhibition of *L. mexicana* ICP to the parasite CP CPB (Smith et al., 2006). The measured backbone dynamics revealed that loop DE in PA-ICP experiences a greater degree of mobility than the equivalent loop in chagasin. This may help the loop of PA-ICP more readily adopt the optimal binding conformation but, on the other hand, would be subject to a greater loss of entropy upon binding. The hydrophobic contacts between L65 and Y72e and between V66 and L69e are conserved (fig 7.5), while the polar contact of K63 and Y72e in chagasin-cathepsin L complex is not present in the PA-ICP-cathepsin L interaction because the lysine residue is replaced by a glycine at the equivalent position. The NPTTG motif directly blocks the catalytic triads of the enzyme, with the residue T33 making contact with C25e. The backbone conformation of T33 places the backbone carbonyl oxygen instead of the carbon atom facing the side chain of the enzyme active site cysteine so as to protect the loop from being cleaved by the peptidase (fig 7.6 a). Intriguingly, the hydrophobic interaction between Y57 and P30 in chagasin that serves to define the conformation of loop 2 no longer exits in PA-ICP as the equivalent tyrosine Y56 in PA-ICP is instead highly mobile. In the chagasin-cathepsin L interaction, mutation of the proline residue to an alanine resulted in weaker binding due to the increase in the mobility of this loop (dos Reis et al., 2008) causing looser packing. The dynamics of the tyrosine residue may have a similar impact on the packing quality of the BC loop. This effect may be counteracted by additional possible interactions between the side chains of E42e and R37

(fig 7.6 b) which is replaced by hydrophobic residues of alanine in chagasin and methionine in *L. mexicana* ICP at equivalent positions. Although it is not obvious in the modeled complex's structure, given the proximity of these two residues, they could form either a salt bridge between the guanidinium group of R37 and the carboxyl group of E42e or two hydrogen bonds between part of the guanidinium group of R37 and part of the carboxyl group of E42e and between the other part of the carboxyl group and H_ε of R37. This interaction was not modeled using modeller's structure based alignment algorithm because the default algorithm did not take into account the electrostatic interactions.

Chagasin binds cathepsin L more tightly than cathepsin H and B. This is also true for PA-ICP due to the steric hindrance of the occluding loop in cathepsin B and the mini chain in cathepsin H which restrict access to the active site clefts of the enzymes. The crystal structures of chagasin in complex with cathespin B demonstrate that the inhibitor interacts with the enzyme in a mode similar to interacting with cathepsin L with loop 4 affording additional interactions with cathespin B by forming direct or water-mediated hydrogen bonds involving residues N55-Y57 in strand 5 of chagasin (Redzynia, 20008). The structure-based sequence alignment suggests that these interactions may be mimicked in the PA-ICP-cathepsin B complex. The exceptions are residues E54-Y56 (equivalent to residues N55-Y57 in chagasin) which experience highly mobility in PA-ICP while in chagasin their conformations are rather fixed by the well defined secondary structure. The flexibility of loop region E54-Y56 in PA-ICP would be expected to introduce a greater entropic penalty in becoming rigid when binding to the enzyme and this may consequently give rise to a somewhat lower inhibitory activity for PA-ICP than for chagasin.
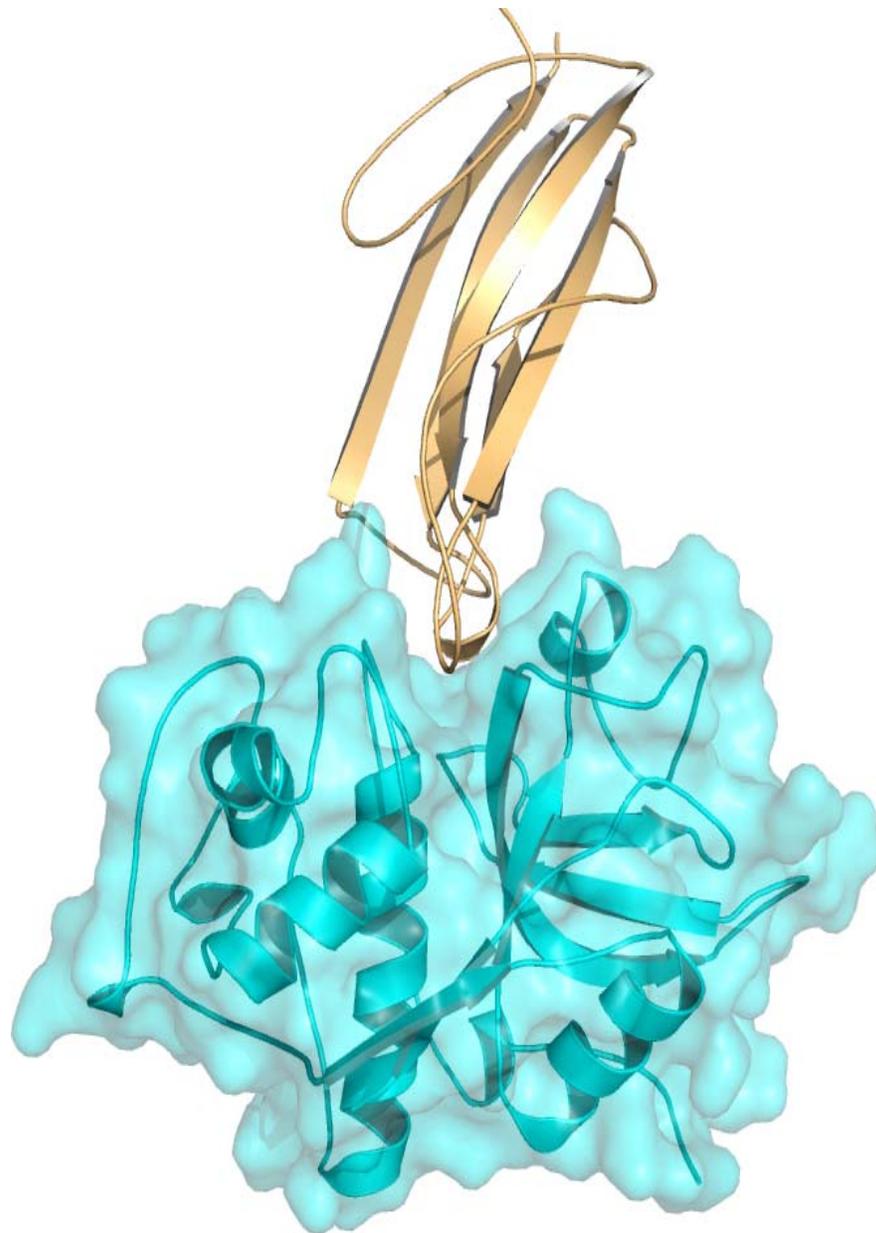
Fig 7.3   Model of PA-ICP in complex with cathepsin L. the model was built based on the crystal structure of the chagasin-cathepsin L complex. The loops bearing the highly conserved sequences NPTTG, GSGG and RPW compose the inhibitory epitope binding to the active site cleft of cathepsin L in a manner similar to chagasin.
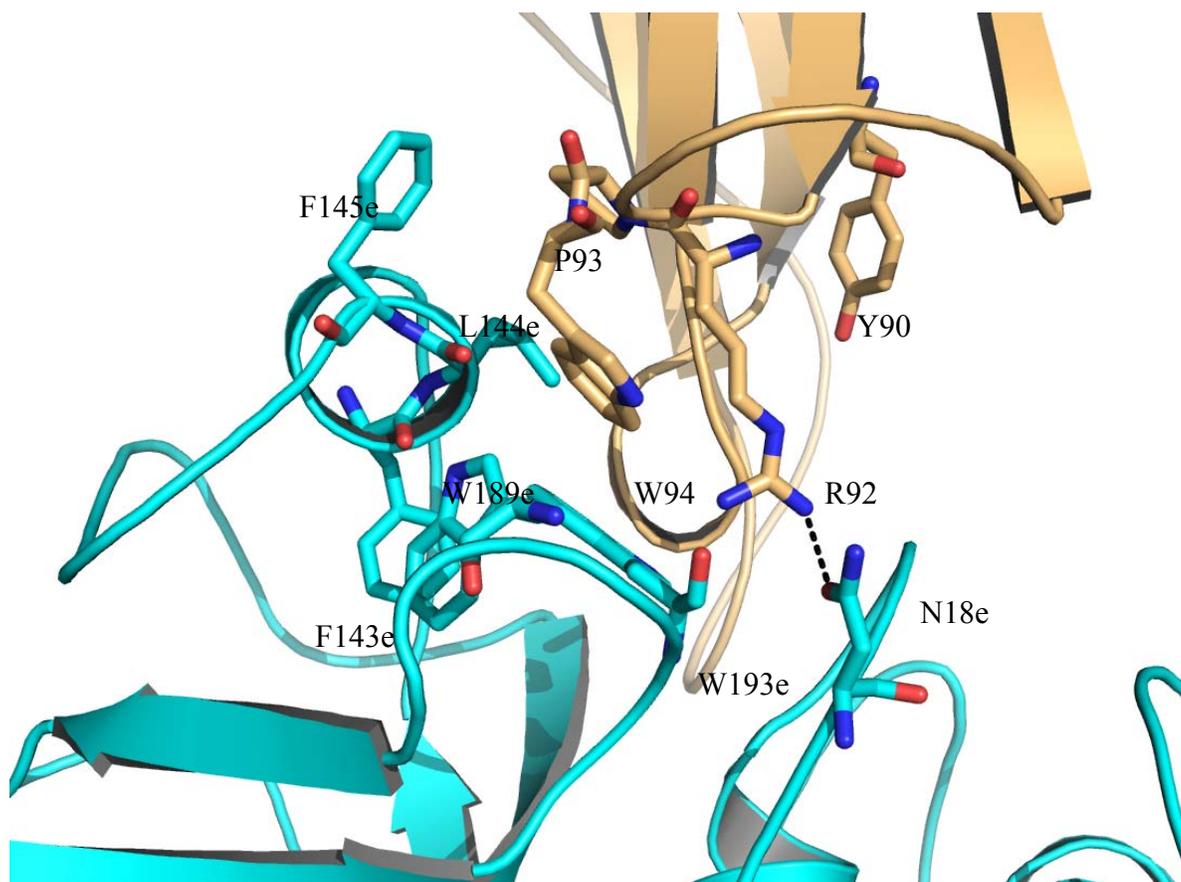
Fig 7.4   PA-ICP-cathepsin L interactions of the RPW motif. The side chain of R92 is anchored by the aromatic side chains of Y90 and W94 and forms an important hydrogen bond with the carbonyl side chain of N18e. W94 recognizes the enzyme's hydrophobic cluster via π interactions with W189e, W193e and F143e. The conformation of the FG loop is maintained by P93 packing against F145e and L144e.
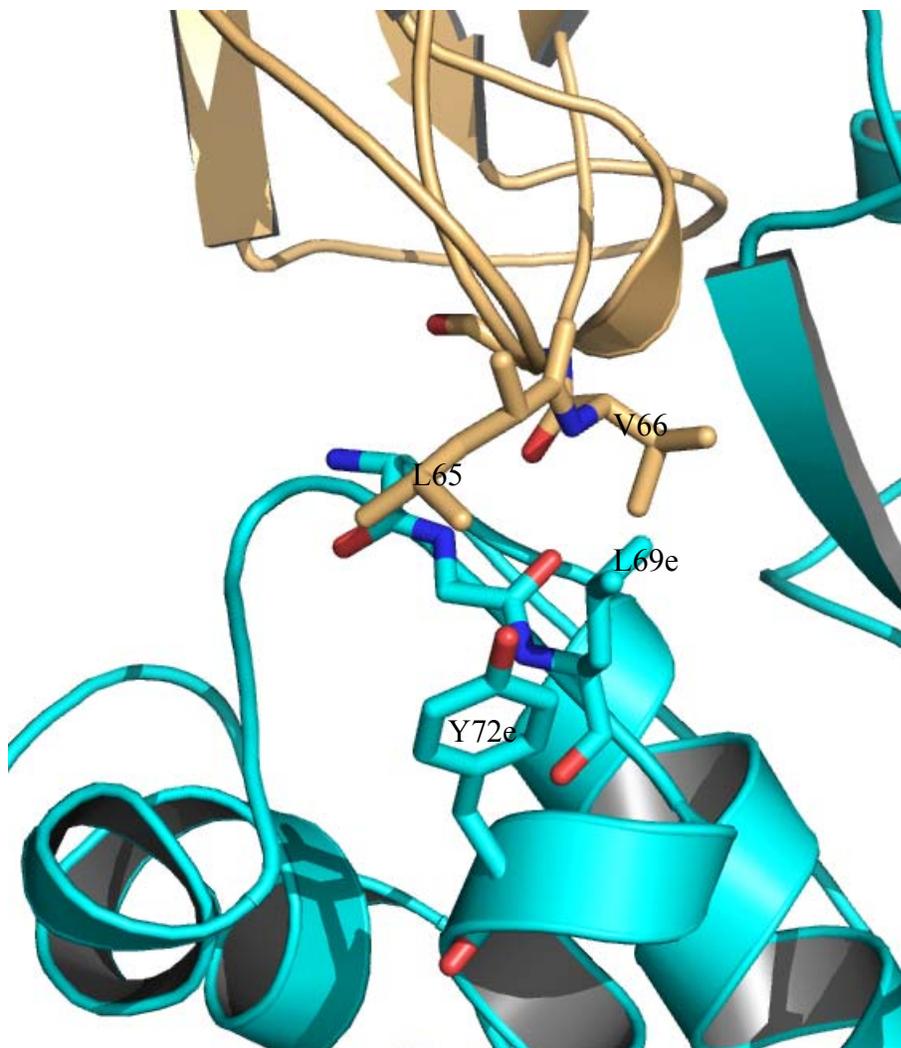
Fig 7.5   Interaction of the highly dynamic DE loop of PA-ICP with cathepsin L. Hydrophobic interactions are seen between the side chains of V66 and L69e and between L65 and Y72e.

a

b
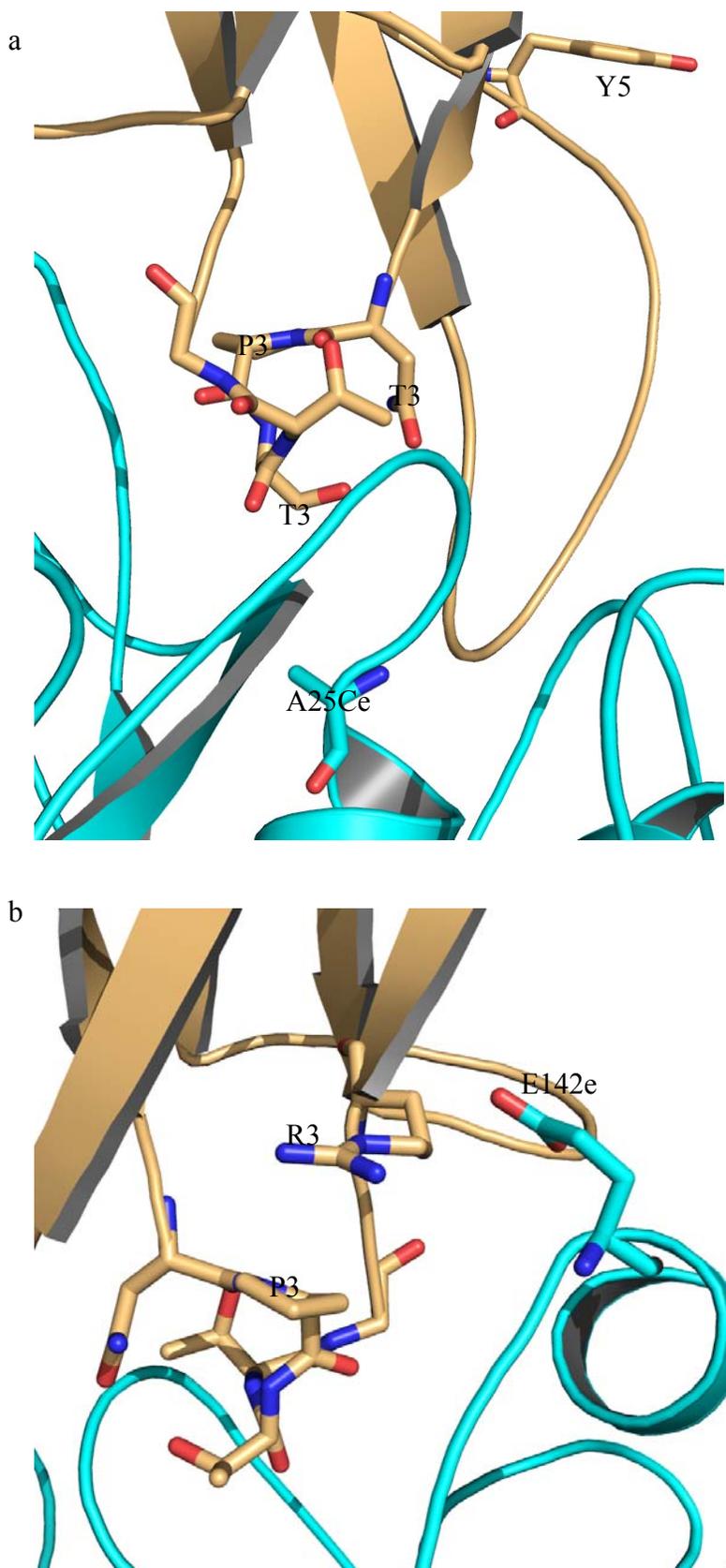
Fig 7.6 Interaction of the BC loop of PA-ICP with cathepsin L. a: The BC loop interacts directly with the active site residues with T33 making contact with the active site cysteine (mutated to an alanine). The main chain conformation of T33 places the carbonyl oxygen atom instead of the carbon atom toward the nucleophilic cysteine thiolate so as to protect the protein from being cleaved. b: A possible interactions between side chains of R37 and E142e may be present to compensate for the looser packing of the BC loop to achieve high affinity binding.

7.6 Conclusion

PA-ICP adopts an immunoglobulin fold placing the three mobile loops containing the highly conserved motifs on one end of the protein to interact with the peptidases. The interactions of PA-ICP with CPs were investigated by homology modeling. The interactions were predicted to be similar to those between chagasin and cathepsins. The dynamics of the DE and FG loops and the electrostatic properties of certain residues near the binding site may contribute to the specificity of the ICP proteins.

CHAPTER 8

MATERIALS AND METHODS

8.1 Molecular biology

8.1.1 Purification of plasmid DNA

Plasmid DNA was purified using QIAprep spin miniprep kit (Qiagen, Cat No. 27104) according to the manufacturer's instructions.

8.1.2 Agarose electrophoresis of DNA

130ml of TBE buffer containing 1% multiple purpose agarose (Roche, Cat No. 1388991) was microwaved at 900 W until the agarose was completely melted. The gel was pre-stained with 3 μl of ethidium bromide (10 mg/ml in ethanol). After cooling the solution to about 60 ^{0}C, it was poured into the casting tray of the electrophoresis apparatus (BioRad, Cat No. 164-0310) containing a sample comb and allowed to solidify at room temperature. The comb was removed carefully and the rest of the gel tank was filled with TBE buffer until the gel was covered. The samples containing DNA mixed with loading buffer were pipetted into the wells alongside 2-log DNA markers (BioLabs, Cat No. N3200S). Electrophoresis was carried out under conditions of constant voltage at 110V for over half an hour. The migration of the DNA in the gel was detected by ethidium bromide fluorescence upon exposure to UV light.

8.1.3 Transformation of recombinant DNA into *E.coli* strains

Tranformation competent cells BL21(DE3) (Novagen, Cat No. 69450), BL21(DE3) codonplus (Stratagene, Cat No. 230240), DH5α (Invitrogen, SKU NO. 11319-019), were transformed according to the manufacturers' instructions. Briefly, 50 μl of competent *E. coli* cells from -80 ^{0}C storage were defrosted on ice briefly prior to addition of 1 μl plasmid DNA. The transformation reaction was mixed by gentle flicking of the tube. The tube was kept on ice for 5 min followed by heat-shocking at 42 ^{0}C for exactly 20 seconds. The tube was returned immediately to ice for a further 2 min. 250 μl of pre-warmed SOC medium was added and the reaction was incubated at 37 ^{0}C for 1 h with vigorous shaking at 200 rpm. The recovered cells were plated onto LB-agar containing appropriate selection antibiotics and incubated at 37 ^{0}C overnight. The colonies formed were used to inoculate suitable medium for protein expression or DNA extraction.

8.2 Protein expression and purification

Unless stated elsewhere, any protein solution was kept on ice or stored at 4 ^{0}C. The plasmids for expression of PA-ICP protein were kindly provided by Prof. G. Coombs, division of infection and immunity, Institute of Biomedical and Life Science, University of Glasgow. The genes were inserted into the *Nde*1 and *Xho*1 sites of pET 28 (a)$^+$ vector, giving rise to a N-terminal (His)$_6$-tag to facilitate the purification step.

8.2.1 Cultivation of *E. coli* cells containing target protein

LB Medium	1% (w/v) tryptone, 0.5%(w/v) yeast extract, 1% NaCl (w/v)
5XM9 stock	3.4% (w/v) Na_2HPO_4, 1.5% KH_2PO_4, 0.25% (w/v) NaCl, autoclaved prior to use
Salt Mixture	4 mM $ZnSO_4$, 1 mM $MnSO_4$, 0.7 mM H_3BO_3, 0.7 mM $CuSO_4$, filtered prior to use
M9 minimal medium	100 ml 5XM9, 500 μl salt mixture, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.3% D-glucose (w/v), 8 mM $(NH_4)_2SO_4$, 0.004% thiamine
Labeled medium	for ^{13}C labelling, replace 0.3% glucose with 0.2% D-glucose (U^{13}C$_6$, 99%) (Cambridge Isotope laboratories, Inc, Cas 110187-42-3), for ^{15}N labelling, replace 8 mM $(NH_4)_2SO_4$ with 16 mM ^{15}NH$_4$Cl (Isotec, 11186AE). Appropriate selection antibiotics were added in all media prior to use.

A single colony of freshly transformed *E. coli* cells was transferred from a selective agar plate into appropriate medium. The culture was incubated on a rotatory shaker (200 rpm) at 37 ^{0}C overnight, which was used to inoculate fresh medium at 15:500 (v/v) ratio. The culture was then grown in the shaker at 37 ^{0}C and 200 rpm until the optimal logarithmic growth phase (OD_{600}=0.6~0.8) was reached. The cells were induced with 1 mM IPTG for a further 4 h at the chosen temperature before harvest by centrifugation (4300 g) at 10 ^{0}C for 20 min and the cell pellet was stored at -20 ^{0}C.

To examine the protein expression level, culture samples were taken before and after IPTG induction for subsequent analysis. The cell density (OD_{600}) of the samples were measured and the sample volumes were determined using the formula,

$$\frac{0.6}{OD_{600}} \times 500 (\mu l)$$

The cells were isolated by microcentrifugation and resuspended with 30 μl lysis buffer together with 10X bugbuster (Novagen, Cat No. 70921-4) and 1 μl Benzonaze (Novagen, Cat No. 70746-3). The samples were assayed using SDS-PAGE under denaturing conditions.

8.2.2 Lysis of *E. coli* cells

lysis buffer   50mM NaH$_2$PO$_4$, 300mM NaCl, 0.01% NaN$_3$, pH 8.0

The defrosted cells were resuspended thoroughly into lysis buffer (5 ml buffer/1 g wet cell paste) by pipetting. The cells were lysed with 10X Bugbuster protein extraction reagent. Benzonase Nuclease was added (1:3000 v/v) at this point to reduce the viscosity of the extract by digestion of chromosomal DNA. 1 mg/ml of lysozyme (Novagen, Cat No. L-6876) was added to enhance the extraction efficiency for the hosts which did not express T7 lysozyme. A protease inhibitor cocktail (Roche, Cat. No. 11836170001) was also added at a concentration of 1 tablet / 30 ml lysis buffer to prevent unwanted proteolysis of PA-ICP. Following incubation at room temperature for over 30 min to allow complete lysis, the insoluble cell debris was removed by centrifugation at 10 ^{0}C 19872 g for 20 min. The supernatant was taken on to the next purification step immediately.

8.2.3 Immobilized metal affinity chromatography

binding/wash buffer   50 mM NaH$_2$PO$_4$, 300 mM NaCl, 0.01% NaN$_3$, pH 8.0
elution buffer   50 mM NaH$_2$PO$_4$, 300 mM NaCl, 250 mM imidazole, 0.01% NaN$_3$, pH 8.0

Ni-NTA resin was packed into a disposable column (BioRad, Cat No. 195-6586) and equilibrated with 20 bed volumes of binding buffer prior to applying the cell lysate onto the column. The lysate was passed through by gravity flow and the flow-through was collected using a clean tube. The unbound proteins were washed off with 20 bed volumes of wash buffer. The bound proteins were eluted with elution buffer until no further protein appeared in the eluate, which was determined by examination of the eluate with Bradford reagent. In case of saturation of the resin, the procedure was repeated with the flow through, until no

protein of interest was found in the eluate. Samples were taken from each step for SDS-PAGE.

8.2.4 Cleavage of His-tag

The protein concentration of the eluate was determined using Bradford assay. The eluate and thrombin (Novagen, Cat No. 69671-3) were combined at a ratio of 0.8 unit thrombin per 1 mg protein. The reaction was incubated at 15 ^{0}C overnight. To assess the efficiency of the cleavage, samples for SDS-PAGE were removed before and at various points after adding thrombin. Once the cleavage was complete, the residual thrombin was removed by passage through benzamidine sepharose.

8.2.5 Ion exchange chromatography

low salt buffer   50 mM NaH_2PO_4, 0.01% NaN_3, pH 8.0

high salt buffer   50 mM NaH_2PO_4, 1 M NaCl, 0.01% NaN_3, pH 8.0

The salt concentration of the protein solution was adjusted to be less than 30 mM using low salt buffer. The protein solution was past through a 0.2 μm cellulose nitrate membrane filter (Whatman International Ltd, Cat No. 7182-004) to filter out any precipitate to avoid the damage to the column. Both buffers were filtered with the membrane filter and dissolved air removed by degassing. Fast flow anion exchange ($-CH_2N^+(CH_3)_3$) Q sepharose (Sigma, Cat No. Q1126) was packed into a column, resulting in a bed volume of 18.85 ml. After assembly of the anion column onto an AKTA chromatography system (Amersham) at 4 ^{0}C, The column was pre-equilibrated with 30 ml of low salt buffer. The protein sample was loaded through a pump inlet and the column was washed with 3 column volumes of low salt buffer. The bound protein was eluted off the column with a linear 0 to 1M salt gradient over 10 column volumes. The flow rate and the pressure limit were set to 5 ml/min and 0.8 MPa respectively throughout the whole experiment. The elution profile was monitored by UV absorbance at wavelengths 280 nm and 220 nm. 5 ml of eluate was collected for each fraction.

8.2.6 Size exclusion chromatography

running buffer   25mM Na_2HPO_4, 75mM NaCl, pH 7.2, 0.01% azide, filtered and
                                  degassed before use

The protein sample was concentrated to 500 µl using a Vivaspin20 concentrator, molecular weight cutoff 5,000 kDa (VIVASCIENCE, Product No. VS2012). Any precipitate was removed by microcentrifugation. A pre-packed Superdex 75 10/300 GL (GE healthcare, Product code. 17-5174-01) gel filtration column was connected to the AKTA chromatography system housed in a coldroom. After equilibration with 30 ml of running buffer, the protein sample was loaded onto the column through the injection valve. The flow rate was set to 0.5 ml/min to ensure the pressure of the system was below 1.8 MPa during the experiment. The UV absorbance at 280 nm was used to monitor the experiment. The protein was eluted with 1.5 column volumes (35.343 ml) of the running buffer with 0.5 ml of eluate collected in each fraction. The eluate samples were examined using SDS-PAGE. The molecular weights of proteins in the eluate were estimated by comparison of the UV trace with a reference created by calibration of the column using LMW gel filtration calibration kit (Amersham biosciences, Product No. 17-0442-01). The calibration procedure was as stated on the manufacturer's instructions.

8.2.7 Buffer exchange and concentration

A   buffer exchange using PD10 desalting column

A PD10 desalting column (Amersham biosciences, Product No. 52-1308-00) was pre-equilibrated with 10 ml of the desired elution buffer. 2.5 ml of protein solution was loaded onto the column and then eluted off with 3.5 ml elution buffer.

B   buffer exchange using Vivaspin concentrator

The protein sample buffer was exchanged to desired buffer in a 20 ml Vivaspin centrifugal concentrator (Sartorius, Product No. VS15RH11) with an appropriate molecular weight cut-off at 3056 g and 10 ^{0}C. The sample in original buffer was concentrated to about 500 µl by centrifugation. 19.5 ml desired buffer was added and the cycle repeated 3 times to complete buffer exchange. The protein solution was then concentrated to the required volume.

8.2.8 Disulphide bond formation

A   glutathione disulphide shuffling

The protein solution was incubated with different ratios (1:2, 1:4, 1:6, 1:8, 1:10) of reduced glutathione and oxidized glutathione at 4 ^{0}C for 2 days. 1 µM of $CuSO_4$ was added to the solution to catalyse the oxidation of the two cysteines. The pH of the solution was adjusted

to 8.5.

B   peroxide oxidation

The protein was oxidized in the presence of 1 M H_2O_2 at 4 ^0C for 2 days.

8.3 Protein assays

8.3.1 SDS-PAGE

Coomassie Stain   45% methanol, 45% dH_2O, 10% acetic acid, 0.25% (w/v) coomassie
                       R250

Destain Buffer   5% methanol, 10% acetic acid, 85% dH_2O

The samples were mixed with 4X NuPAGE LDS sample buffer (Invitrogen, Cat No. NP0007) prior to being loaded into pre-cast NuPAGE 4~12% bis-tris gels (Invitrogen, Cat No. NP0321BOX). The reduced samples were heated to 85 ^0C for 10 min with 90 mM β-Mercaptoethanol before loading. Protein markers (BioRad, Cat No. 161-0373) were diluted with two volumes of water and one volume of NuPAGE LDS sample buffer and 10 μl was loaded. The gel was run at 200 V constant voltage for 45 minutes. The protein bands were visualized after staining the gel with approximately 100 ml of coomassie stain for 5 to15 min followed by destaining in 100 ml of destaining buffer overnight at room temperature.

8.3.2 Antibody purification

Coupling buffer   0.1 M Na_2HPO_4, 0.05% NaN_3, pH 7.0

Wash buffer   1 M NaCl, 0.05% NaN_3

PA-ICP was expressed and purified as stated before. 2 mg of purified protein was immobilized covalently to AminoLink coupling gel (Pierce, No. 20381) as bait for purification of PA-ICP antibody from polyclonal antisera produced. 1 ml of coupling gel was washed with 10 ml coupling buffer 3 times. The gel was centrifuged at 173 g for 5 min and the supernatant removed after each wash. 1 ml of protein solution was diluted to 3 ml with coupling buffer. The diluted protein solution was mixed with coupling gel and the coupling procedure was performed according to the manufacturer's instructions. To bind antibody, PA-ICP coupled gel was pre-equilibrated with 10ml IgG binding buffer (Pierce,

Product No. 21001). 5 ml of PA-ICP rabbit antisera (final bleed) was mixed with 2.5 ml IgG binding buffer and incubated with the coupling gel overnight at 4 ^{0}C. The gel and the sera were transferred into a disposable column. After washing with 20 ml of IgG binding buffer, the bound antibody was then eluted with 5 ml of IgG elution buffer (Pierce, Product No. 21004). 10 fractions were collected with each fraction containing 500 μl eluate plus 25 μl of 1 M tris (pH 9.0) to neutralize the eluate. The concentration of the purified antibody in each fraction was measured at 280 nm using 5 μl sample plus 95 μl H$_2$O on biophotometer. Fractions 2-4 were buffer exchanged using a PD10 column into PBS, pH 7.55 and concentrated to 350 μl. The antibody solution was mixed with an equal volume of 2 mg/ml BSA and stored at -20 ^{0}C in 50% glycerol. The column was washed with 20 ml wash buffer and stored at 4 ^{0}C.

8.3.3 Western immunoblotting

Transfer Buffer	5% 20X NuPAGE transfer buffer (Invitrogen, Cat No. 1222517), 20% methanol, 75% dH$_2$O
Ponceau S Solution	0.5% (w/v) Ponceau S, 1% glacial acetic acid, 99% dH$_2$O
10X TBST Buffer	1.21% tris, 4% NaCl, 1% Tween 20, pH7.6
Block Buffer	1X TBST, 5% (w/v) nonfat milk
Wash Buffer	1X TBST, 1% (w/v) nonfat milk

The proteins of interest were electrophoretically separated using SDS-PAGE. After electrophoresis, the pre-soaked nitrocellulose transfer membrane (Anderman & co. ltd., Conv No. 7001623) was placed on the surface of the gel. The membrane and the gel were sandwiched in between the filter papers (BioRad, Cat No. 1703932) and blotting pads. The whole assembly was then carefully put into the Xcell II blot module (Invitrogen, Cat No. E19051). The gel transfer tank was assembled. The blot apparatus was filled with transfer buffer until the gel/membrane sandwich was covered and the outer buffer chamber was filled with ~650 ml deionized water. The transfer was performed at 30 V constant for 1 h. To check the efficiency of the transfer, the membrane was stained in Ponceau S solution for 2 min. After destained in deionized water, the membrane was blocked with 50 ml of block buffer for 2 h at room temperature and then incubated with primary antibody (1:10,000 (v/v)) in 50 ml of wash buffer at 4 ^{0}C overnight. The unbound antibody was washed off in wash buffer for 4X 30 min at room temperature. The membrane was incubated with 50 ml

of wash buffer plus 1:5,000 (v/v) secondary antibody for 2 h at room temperature followed by another wash. The membrane was rinsed in 1X TBST buffer and incubated with ECL reagent for 3.5 min before detection.

8.3.4 Bradford assay

The protein concentration was measured using Bradford assay kit brought from Pierce (Product No. 23236). The measurement was performed according to the standard "test tube" protocol in the manufacturer's instructions. For samples in buffers containing imidazole, the concentration of the imidazole was adjusted to be lower than 200 mM.

8.3.5 Enzyme assay

Assay buffer A: 0.1 M sodium acetate, 2 mM EDTA, 10 mM DTT pH 5.5

Assay buffer B: 0.1 M NaH_2PO_4, 2 mM EDTA, 10 mM DTT

Chromogenic substrate stock: 0.1 M in ethanol

Fluorogenic substrate stock: 10 mM in DMSO

8.3.5.1 Active site titration of papain and cathepsins

Papain   Active site titration of papain (Sigma, Product No. P4762) with irreversible inhibitor E64 was performed using z-PHE-ARG-pNA (Biochem, Cat No. L-1242) as chromogenic substrate. E64 was diluted serially to span the concentration from 0 to 0.2 μM. 30 µl of diluted inhibitors was incubated with an equal volume of enzyme at 33 ^{0}C for 5 min. 50 µl of reaction mixture was assayed with 500 µl of assay buffer A containing 300 µM substrate at 33 ^{0}C on a spectrophotometer at 410nm, for 60 sec, with data recorded at 0.5 sec intervals. The initial rate of each reaction was calculated and a plot was produced of velocity (y axis) versus concentration of inhibitor (x axis). The plot should be linear with the x intercept indicating the concentration of inhibitor required to completely inhibit the enzyme. The functional molarity of the enzyme can then be calculated. Papain was also titrated with PA-ICP in a similar pseudo-irreversible manner with a previously determined active enzyme concentration to estimate the active concentration of PA-ICP.

Cathepsins L and B   The assays of active site titration of cathepsin L (Calbiochem, Cat No. 219402) and cathepsin B (Calbiochem, Cat No. 219362) with E64 were performed as for papain.

Cathepsin S   The active site titration for Cathepsin S (Calbiochem, Cat No. 219343) was performed using the fluorogenic substrate z-FR-AMC (Bachem Cat No. I-1160). The enzyme was diluted with assay buffer B, 50 µl of which was incubated with 50 µl of E64 for 1 min on ice. The mixtures were then added in pre-warmed assay buffer B containing 200 µM substrate. Fluorescence of the released amino-methylcoumarin was measured immediately in a fluorimeter with the exitation wavelength set to 380 nm and emission monitored at 465 nm. The measurements were carried out at 30 ^{0}C for 1 min, measuring interval 0.1 s. The initial rates of the reactions were calculated and a linear plot was created as described above.

8.3.5.2 K$_i$ determination

Cathepsin L   The enzyme was diluted with assay buffer B to a concentration of 8 nM and activated on ice for 1 h. PA-ICP was also diluted with assay buffer B to a range between 0 nM to 1.6 nM. 50 µl of each dilution was pre-incubated with 50 µl of diluted enzyme at 30 ^{0}C for 1 min to allow the inhibitor binding to approach equilibrium. The mixture was assayed at 30 ^{0}C in 900 µl assay buffer B, pH 5.5 containing 15 µM fluorogenic substrate z-FR-AMC. The measurements were carried out for 1 min in a fluorimeter. The initial velocity of the reaction with and without inhibitor were calculated and a plot of velocity (v) versus inhibitor concentration([I])was generated.

Cathepsin B   The assay was as for cathepsin L, except that the enzyme concentration was 2 nM and PA-ICP concentration used was in a range of 0 to 32 µM. The incubation time for cathepsin B and PA-ICP was 15 min.

Cathepsin S   The assay was as for cathepsin L, except that the enzyme concentration was 100 nM and pH was 6.0. The substrate concentration was 200 µM.

Cathepsin H (Calbiochem, Cat No. 219404) The assay was as for cathepsin L, except that the enzyme concentration was equal or less than 200 nM. The substrate was H-ARG-AMC (Biomol, Cat No. P-135) and pH 6.8

8.3.6 Osmotic shock

Wash buffer   10 mM tris-HCl, 30 mM NaCl (pH 7.1)

P. aeruginosa PAO1 was cultivated in 1 L LB medium overnight (kindly provided by Prof. T. Evans, Institute of Biomedical and Life Science, University of Glasgow). After harvest at 3500 g at 25 ^{0}C for 20 min, the cells were washed with 40 ml of wash buffer at 5311 g for 30 min. The pellet was resuspended in 40 ml 33 mM tris-HCl (pH 7.1) and then incubated with 40 ml 40% (w/v) sucrose at 25 ^{0}C for 30 min, followed by centrifugation at 6200 g for 20 min. The pellet was quickly resuspended in 40 ml 0.5 mM $MgCl_2$ on ice. The sudden change in the osmolarity should lead to the release of periplasmic material. The location of PA-ICP was detected using western blot and the efficiency of the osmotic shock was assessed by assay of the activity of acid phosphatase as periplasmic marker protein using 4-Nitrophenyl phosphate disodium salt hexahydrate（Sigma, Product Code: S 0942）

8.4 NMR spectroscopy

8.4.1 Sample preparation

NMR buffer   25 mM Na_2HPO_4, 75 mM NaCl, 0.01% azide (w/v)

^{15}N labelled or ^{15}N-^{13}C double labelled samples were expressed as described above using labelled media. After purification, the sample buffer was exchanged for NMR buffer (pH 6 for non-his-tagged samples and 7.2 for his-tagged samples) using a 20 ml Vivaspin concentrator with molecular weight cut-off of 5 kDa. The concentrated sample was then microcentrifuged at 13,000 g at 4 ^{0}C for 10 min to remove any fine particles prior to addition of 5% D_2O. The sample was stored at 4 ^{0}C for a week to allow sufficient formation of the intra-molecular disulphide bond.

8.4.2 NMR data acquisition

All data were collected on Bruker Avance 600 MHz NMR spectrometer, equipped with a TCI 5 mm cryoprobe, at 298 K unless stated otherwise.

8.4.2.1 Resonance assignment

Resonance assignments of PA-ICP have been achieved using ^{13}C-^{15}N double labelled samples and triple resonance heteronuclear 3D NMR spectroscopy methods. The chemical shifts of the protein backbone were assigned using two pairs of complementary 3D experiments, HNCO+HN(CA)CO and HNCACB+CBCA(CO)NH of which the HNCACB and CBCA(CO)NH were non-linear sampled on both indirect dimensions to reduce the acquisition time. On the basis of backbone assignment, the assignments were then expanded to the side chains using CCCONH, HCCONH and HCCH-TOCSY experiments for aliphatic residues. Aromatic side chains were assigned using HBCBCGCDHD and HBCBCGCDCEHE.

8.4.2.2 Structure restraints gathering

Distance Restraints

NOE restraints were collected using 3D ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC experiments with a mixing time of 100 ms. After processing, the spectra were loaded into CCPN analysis and NOE peaks were picked manually and corroborated by looking for symmetry correlated peaks. The picked peaks were used to generate ambiguously assigned distance restraints for structure calculations. The detailed calculation method is set out in chapter 5.

8.4.2.3 RDCs

RDCs of the protein backbone were resolved by weak alignment of a double labeled sample with 20 mg/ml of Pf1 filamentous phage (Profos AG, cat No, 311059). 12 mg phage was buffer exchanged into alignment buffer and concentrated to 180 μl using a 20 ml Vivaspin concentrator (molecular weight cut-off 10 kDa) before the addition of 420 μl protein sample. 5% D_2O was added and the sample was mixed carefully to avoid the formation of fine bubbles, resulting in a final phage concentration of 20 mg/ml. Due to the linear relationship between the quadrople splitting resulted from aligned deuterium and Pf1 phage concentration, the degree of the alignment was tuned by adjustment of the concentration of the Pf1 phage according to the magnitude of quadropolar deuterium splitting using 1D deuterium experiment. A splitting of 10 Hz in the quadropolar ^{2}H peak pair was observed. RDC data were collected using modified IPAP J-coupled ^{15}N HSQC experiments with the incorporation of spin-state selection. ^{1}J (HN), ^{1}J (NCO) and ^{2}J (H$_n$CO) splittings were

measured on both oriented and isotropic samples and dipolar couplings obtained from differences of the splittings.

8.4.2.4 Hydrogen Bonds

A ^{15}N labelled PA-ICP sample was purified as stated above. After lyophilization the sample was redissolved in D_2O. ^{1}H-^{15}N HSQC experiments were recorded immediately before and after sample preparation at several times. Cross peaks remained after 2 h were considered to represent the hydrogen bonded amide protons.

8.4.2.5 ^{15}N relaxation experiments

Relaxation experiments for protein dynamics study were performed on a non-his-tagged ^{15}N labelled sample at pH 6 and 308K. ^{15}N T_1 and T_2 relaxation rate were measured using non-sensitivity enhanced pseudo 3D experiments with 6 T_1 and 8 T_2 relaxation points selected to cover the possible T_1 and T_2 durations for PA-ICP. ^{15}N heteronuclear NOEs were measured by comparing the intensity of signal transferred from amide ^{15}N to ^{1}H in the absence and presence of ^{1}H saturation. The CurveFit program (http://biochemistry.hs.columbia.edu/labs/palmer/software/curvefit.html) was used to fit the T_1 and T_2 relaxation data.

8.4.2.6 NMR data processing and viewing

All multi-dimensional NMR data were processed using the AZARA suite of programs (http://www.bio.cam.ac.uk/azara/, Wayne Boucher, unpublished). Window functions and zerofilling were routinely applied to optimize the signal resolution before the raw FID was Fourier transformed from time domain to frequency domain data. For those spectra with poor baseline and water suppression, baseline correction and convolution to remove the residual water signal were also employed. In particular, for all 3D experiments, a Maximum Entropy scheme was used for processing of the indirect dimensions.

The processed data were assigned and analyzed manually using the CcpNmr suite of programs (http://www.ccpn.ac.uk/) (Fogh et al., 2006).

Exp	Dimension 1 (direct)					Dimension 2 (indirect)					Dimension 3 (indirect)					RD	NS	WS
	Nuc	TD	SW	AQ	PL	Nuc	TD	SW	AQ	QD	Nuc	TD	SW	AQ	QD			
HNCO	^{1}H	896	8389	107	11.5	^{15}N	24	1265	18.9	ST	^{13}C	60	2717	22	ST	1	8	WG
HNCACO	^{1}H	896	8389	107	11.5	^{15}N	24	1265	18.9	ST	^{13}C	60	2717	22	ST	1		WG
HNCACB	^{1}H	896	8389	107	13.18	^{15}N	16/31	1265	11.8	ST	^{13}C	28/80	11312		ST	1	24	WG
CBCACONH	^{1}H	896	8389	118	13.23	^{15}N	16/31	1265	11.8	ST	^{13}C	27/76	11312	3.3	ST	1	24	WG
CCCONH-TOCSY	^{1}H	896	8389	107	12.75	^{15}N	27	1265	21	E-Anti	^{13}C	55	11312	48	ST	1	32	
HCCCONH-TOCSY	^{1}H	896	8389	107	11.5	^{15}N	18	1265	14.2	ST	^{1}H	106	8403	13	ST	1	16	WG
HCCH-TOCSY	^{1}H	896	8993	99.7	13.38	^{13}C	31	4753	6.5	ST	^{1}H	70	5403	12.9	ST	1	24	
HBCBCGCDCEHE	^{1}H	896	8993	99.7	11.5	^{13}C	40	6002	6.7	ST						1	640	
HBCBCGCDHD	^{1}H	896	8993	99.7	11.5	^{13}C	40	6002	6.7	ST						1	768	
^{15}N-NOESY-HSQC Mixing time 100 ms	^{1}H	896	8993	99.7	13.38	^{15}N	35	1265	27.6	ST	^{1}H	75	7800	9.6	ST	1	24	WG
^{13}C-NOESY-HSQC Mixing time 100 ms	^{1}H	896	8389	107	13	^{13}C	41	4753	8.6	E-Anti	^{1}H	101	7800	12.9	ST	1	24	
^{13}C-NOESY 2D Projection	^{1}H	896	8389	107	13.23	^{13}C	128	4753	27	E-Anti						1	24	
J$_{COCme}$	^{1}H	896	8993	99.7	12.75	^{13}C	400	4753	84.1	E-Anti						1	32	
J$_{NCme}$	^{1}H	896	8993	99.7	12.75	^{13}C	400	4753	84.1	E-Anti						1	32	
IPAP ^{15}N HSQC	^{1}H	896	8389	107	11.25	^{15}N	152	1265	120	ST						1	64	
IPAP ^{15}N HSQC	^{1}H	896	8389	107	11.25	^{15}N	80	1265	63	ST						1	240	
IPAP ^{15}N HSQC _AB_CACO	^{1}H	896	8389	107	11.25	^{15}N	64	1265	50.6	ST						1	400	
IPAP ^{15}N HSQC _AB_CACO$^\Delta$	^{1}H	896	8389	107	11.25	^{15}N	75	1265	59.2	ST						1	400	
IPAP ^{15}N HSQC _AB_NCO	^{1}H	896	8389	107	11.25	^{15}N	80	1265	63	ST						1	240	
IPAP ^{15}N HSQC _AB_NCO	^{1}H	896	8389	107	11.25	^{15}N	80	1265	63	ST						1	240	
Hetero-nuclear NOE	^{1}H	896	8389	107	13.5	^{15}N	86	1265	67.9	ST						2	80	FB
T$_1$ relaxation	^{1}H	896	8389	107	13.5	^{15}N	128	1265	101	ST						1.5	32	FB
T$_2$ relaxation	^{1}H	896	8389	107	13	^{13}C	86	1265	68	ST						2	32	FB

Table 8.1   Acquisition parameters of NMR experiments used for resonance assignment, structural restraints collection and backbone dynamics investigation. Exp: experiment, Nuc: Nucleus, TD: time domain points (complex), SW: sweep width in Hz, AQ: acquisition time in ms, PL:90^0 high power pulse length in µs, QD: quadrature detection mode, RD: relaxation delay, NS: number of scans, WS: water suppression, ST: states-TPPI, E-Anti: Echo-antiEcho, WG: water gate, FB: flip-back, Δ: experiments run on an aligned sample.

8.5 Homology modeling

The PA-ICP-cathepsin L complex was modeled using modeller 9.5 based on the coordinates of the chagasin-cathepsin L complex. The sequence alignment that directs the modeling was generated on the basis of a structural comparison, with loop region G52-G64 of PA-ICP not aligned to the template structures to allow for its flexibility.

CHAPTER 9

REFERENCES

Alphey, M. S. & Hunter, W. N. (2006) High-resolution complex of papain with remnants of a cysteine protease inhibitor derived from Trypanosoma brucei. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications,* 62**,** 504-508.

Baneyx, F. (1999) Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology,* 10**,** 411-421.

Bardwell, J. C. A. (1994) Building bridges - disulfide bond formation in the cell. *Molecular Microbiology,* 14**,** 199-205.

Barrett, A. J. (1994) Classification of peptidases. *Proteolytic Enzymes: Serine and Cysteine Peptidases,* 244**,** 1-15.

Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M. & Hanada, K. (1982) L-trans-epoxysuccinyl-leucylamido (4-guanidino) butane (E-64) and its analogs as inhibitors of cysteine proteinases including cathepsins B, H and L. *Biochemical Journal,* 201**,** 189-198.

Bax, A. (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Science,* 12**,** 1-16.

Bax, A. & Grishaev, A. (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Current Opinion in Structural Biology,* 15**,** 563-570.

Bax, A., Kontaxis, G. & Tjandra, N. (2001) Dipolar couplings in macromolecular structure determination. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B.*

Besteiro, S., Coombs, G. H. & Mottram, J. C. (2004) A potential role for ICP, a leishmanial inhibitor of cysteine peptidases, in the interaction between host and parasite. *Molecular Microbiology,* 54**,** 1224-1236.

Bromme, D. & Mcgrath, M. E. (1996) High level expression and crystallization of recombinant human cathepsin S. *Protein Science,* 5**,** 789-791.

BRUNGER, A. T., ADAMS, P. D., CLORE, G. M., DELANO, W. L., GROS, P., GROSSE-KUNSTLEVE, R. W., JIANG, J. S., KUSZEWSKI, J., NILGES, M., PANNU, N. S., READ, R. J., RICE, L. M., SIMONSON, T. & WARREN, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography,* 54**,** 905-921.

BRUSCHWEILER, R., LIAO, X. B. & WRIGHT, P. E. (1995) Long-range motional

restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science,* 268**,** 886-889.

BUHLING, F., WALDBURG, N., REISENAUER, A., HEIMBURG, A., GOLPON, H. & WELTE, T. (2004) Lysosomal cysteine proteases in the lung: role in protein processing and immunoregulation. *European Respiratory Journal,* 23**,** 620-628.

BUTTLE, D. J., BONNER, B. C., BURNETT, D. & BARRETT, A. J. (1988) A CATALYTICALLY ACTIVE HIGH-MR FORM OF HUMAN CATHEPSIN-B FROM SPUTUM. *Biochemical Journal,* 254**,** 693-699.

CAVANAGH, J., FAIRBROTHER, W.J., III, A.G. PALMER AND SKELTON, N.J. (1996) *Protein NMR spectroscopy: Principles and practice,* London, Academic Press, Inc.

CHEN, J., BROOKS, C. L. & WRIGHT, P. E. (2003) Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation. *Journal of Biomolecular NMR***,** 14.

CHOWDHURY, S. F., JOSEPH, L., KUMAR, S., TULSIDAS, S. R., BHAT, S., ZIOMEK, E., MENARD, R., SIVARAMAN, J. & PURISIMA, E. O. (2008) Exploring inhibitor binding at the S ' subsites of cathepsin L. *Journal of Medicinal Chemistry,* 51**,** 1361-1368.

CLARK, E. D. B. (1998) Refolding of recombinant proteins. *Current Opinion in Biotechnology,* 9**,** 157-163.

CLUBB, R. T., THANABAL, V. & WAGNER, G. (1992) A CONSTANT-TIME 3-DIMENSIONAL TRIPLE-RESONANCE PULSE SCHEME TO CORRELATE INTRARESIDUE H-1(N), N-15, AND C-13(') CHEMICAL-SHIFTS IN N-15-C-13-LABELED PROTEINS. *Journal of Magnetic Resonance,* 97**,** 213-217.

DA SILVA, A. A., VIEIRA, L. D., KRIEGER, M. A., GOLDENBERG, S., ZANCHIN, N. I. & GUIMARAES, B. G. (2007) Crystal structure of chagasin, the endogenous cysteine-protease inhibitor from Trypanosoma cruzi. *Journal of Structural Biology,* 157**,** 416-423.

DEROME, A. (1987) *Modern NMR techniques for chemistry research*, Pergamon books, Ltd.

DOS REIS, F. C. G., SMITH, B. O., SANTOS, C. C., COSTA, T. F. R., SCHARFSTEIN, J., COOMBS, G. H., MOTTRAM, J. C. & LIMA, A. (2008) The role of conserved residues of chagasin in the inhibition of cysteine peptidases. *Febs Letters,* 582**,** 485-

490.

ESCHENLAUER, S. C. P., FARIA, M. S., MORRISON, L. S., BLAND, N., RIBEIRO-GOMES, F. L., DOSREIS, G. A., COOMBS, G. H., LIMA, A. & MOTTRAM, J. C. (2009) Influence of parasite encoded inhibitors of serine peptidases in early infection of macrophages with Leishmania major. *Cellular Microbiology,* 11**,** 106-120.

FARROW, N. A., MUHANDIRAM, R., SINGER, A. U., PASCAL, S. M., KAY, C. M., GISH, G., SHOELSON, S. E., PAWSON, T., FORMANKAY, J. D. & KAY, L. E. (1994) BACKBONE DYNAMICS OF A FREE AND A PHOSPHOPEPTIDE-COMPLEXED SRC HOMOLOGY-2 DOMAIN STUDIED BY N-15 NMR RELAXATION. *Biochemistry,* 33**,** 5984-6003.

FARROW, N. A., ZHANG, O. W., FORMANKAY, J. D. & KAY, L. E. (1995) COMPARISON OF THE BACKBONE DYNAMICS OF A FOLDED AND AN UNFOLDED SH3 DOMAIN EXISTING IN EQUILIBRIUM IN AQUEOUS BUFFER. *Biochemistry,* 34**,** 868-878.

FILLOUX, A., MICHEL, G. & BALLY, M. (1998) GSP-dependent protein secretion in Gram-negative bacteria: The Xcp system of Pseudomonas aeruginosa. *Fems Microbiology Reviews,* 22**,** 177-198.

FLEMING, K. & MATTHEWS, S. (2004) *Media for studies of partially aligned states*, Humana Press.

FOGH, R. H., VRANKEN, W. F., BOUCHER, W., STEVENS, T. J. & LAUE, E. D. (2006) A nomenclature and data model to describe NMR experiments. *Journal of Biomolecular Nmr,* 36**,** 147-155.

FOLMER, R. H. A., HILBERS, C. W., KONINGS, R. N. H. & NILGES, M. (1997) Floating stereospecific assignment revisited: Application to an 18 kDa protein and comparison with J-coupling data. *Journal of Biomolecular Nmr,* 9**,** 245-258.

FREEMAN, R. & KUPCE, E. (2003) New methods for fast multidimensional NMR. *Journal of Biomolecular Nmr,* 27**,** 101-113.

GABRIEL, J. C. P., CAMEREL, F., LEMAIRE, B. J., DESVAUX, H., DAVIDSON, P. & BATAIL, P. (2001) Swollen liquid-crystalline lamellar phase based on extended solid-like sheets. *Nature,* 413**,** 504-508.

GARDNER, K. H. & KAY, L. E. (1998) The use of H-2, C-13, N-15 multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure,* 27**,** 357-406.

GRZESIEK, S. & BAX, A. (1993) AMINO-ACID TYPE DETERMINATION IN THE SEQUENTIAL ASSIGNMENT PROCEDURE OF UNIFORMLY C-13/N-15-ENRICHED PROTEINS. *Journal of Biomolecular Nmr,* 3**,** 185-204.

GRZESIEK, S., VUISTER, G. W., BAX, A. & LR (1993) A SIMPLE AND SENSITIVE EXPERIMENT FOR MEASUREMENT OF JCC COUPLINGS BETWEEN BACKBONE CARBONYL AND METHYL CARBONS IN ISOTOPICALLY ENRICHED PROTEINS. *Journal of Biomolecular Nmr,* 3**,** 487-493.

GUNCAR, G., PODOBNIK, M., PUNGERCAR, J., STRUKELJ, B., TURK, V. & TURK, D. (1998) Crystal structure of porcine cathepsin H determined at 2.1 angstrom resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure,* 6**,** 51-61.

GUNTERT, P. (1998) Structure calculation of biological macromolecules from NMR data. *Quarterly Reviews of Biophysics,* 31**,** 145-237.

HALL, J. B. & FUSHMAN, D. (2003) Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *Journal of Biomolecular Nmr,* 27**,** 261-275.

HANSEN, M. R., MUELLER, L. & PARDI, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology,* 5**,** 1065-1074.

HORE, P. J. (1995) *Nuclear magnetic resonance,* USA, Oxford university press.

HWANG, T. L. & SHAKA, A. J. (1995) WATER SUPPRESSION THAT WORKS - EXCITATION SCULPTING USING ARBITRARY WAVE-FORMS AND PULSED-FIELD GRADIENTS. *Journal of Magnetic Resonance Series A,* 112**,** 275-279.

ILLY, C., QURAISHI, O., WANG, J., PURISIMA, E., VERNET, T. & MORT, J. S. (1997) Role of the occluding loop in cathepsin B activity. *Journal of Biological Chemistry,* 272**,** 1197-1202.

JEENER, J., MEIER, B. H., BACHMANN, P. & ERNST, R. R. (1979) INVESTIGATION OF EXCHANGE PROCESSES BY 2-DIMENSIONAL NMR-SPECTROSCOPY. *Journal of Chemical Physics,* 71**,** 4546-4553.

JENSCH, T. & FRICKE, B. (1997) Localization of alanyl aminopeptidase and leucyl aminopeptidase in cells of Pseudomonas aeruginosa by application of different

methods for periplasm release. *Journal of Basic Microbiology,* 37**,** 115-128.

KAY, L. E. (2005) NMR studies of protein structure and dynamics. *Journal of Magnetic Resonance,* 173**,** 193-207.

KAY, L. E., TORCHIA, D. A. & BAX, A. (1989) BACKBONE DYNAMICS OF PROTEINS AS STUDIED BY N-15 INVERSE DETECTED HETERONUCLEAR NMR-SPECTROSCOPY - APPLICATION TO STAPHYLOCOCCAL NUCLEASE. *Biochemistry,* 28**,** 8972-8979.

KAY, L. E., XU, G. Y., SINGER, A. U., MUHANDIRAM, D. R. & FORMANKAY, J. D. (1993) A GRADIENT-ENHANCED HCCH TOCSY EXPERIMENT FOR RECORDING SIDE-CHAIN H-1 AND C-13 CORRELATIONS IN H2O SAMPLES OF PROTEINS. *Journal of Magnetic Resonance Series B,* 101**,** 333-337.

KERR, J. R. (2000) Gram-negative bacterial protein secretion. *Journal of Infection,* 40**,** 121-126.

KIRKPATRICK, S., GELATT, C. D. & VECCHI, M. P. (1983) OPTIMIZATION BY SIMULATED ANNEALING. *Science,* 220**,** 671-680.

LASKOWSKI, R. A., RULLMANN, J. A. C., MACARTHUR, M. W., KAPTEIN, R. & THORNTON, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular Nmr,* 8**,** 477-486.

LECAILLE, F., KALETA, J. & BROMME, D. (2002) Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design. *Chemical Reviews,* 102**,** 4459-4488.

LINGE, J. P., HABECK, M., RIEPING, W. & NILGES, M. (2003a) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics,* 19**,** 315-316.

LINGE, J. P., O'DONOGHUE, S. I. & NILGES, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B.*

LINGE, J. P., WILLIAMS, M. A., SPRONK, C., BONVIN, A. & NILGES, M. (2003b) Refinement of protein structures in explicit solvent. *Proteins-Structure Function and Bioinformatics,* 50**,** 496-506.

LIPARI, G. & SZABO, A. (1982a) MODEL-FREE APPROACH TO THE INTERPRETATION OF NUCLEAR MAGNETIC-RESONANCE RELAXATION IN MACROMOLECULES .1. THEORY AND RANGE OF VALIDITY. *Journal of*

the American Chemical Society, 104**,** 4546-4559.

LIPARI, G. & SZABO, A. (1982b) MODEL-FREE APPROACH TO THE INTERPRETATION OF NUCLEAR MAGNETIC-RESONANCE RELAXATION IN MACROMOLECULES .2. ANALYSIS OF EXPERIMENTAL RESULTS. *Journal of the American Chemical Society,* 104**,** 4559-4570.

LJUNGGREN, A., REDZYNIA, L., ALVAREZ-FERNANDEZ, M., ABRAHAMSON, M., MORT, J. S., KRUPA, J. C., JASKOLSKI, M. & BUJACZ, G. (2007) Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease. *Journal of Molecular Biology,* 371**,** 137-153.

LO CONTE, L., AILEY, B., HUBBARD, T. J. P., BRENNER, S. E., MURZIN, A. G. & CHOTHIA, C. (2000) SCOP: a Structural Classification of Proteins database. *Nucleic Acids Research,* 28**,** 257-259.

LYCZAK, J. B., CANNON, C. L. & PIER, G. B. (2000) Establishment of Pseudomonas aeruginosa infection: lessons from a versatile opportunist. *Microbes and Infection,* 2**,** 1051-1060.

MANDEL, A. M., AKKE, M. & PALMER, A. G. (1995) BACKBONE DYNAMICS OF ESCHERICHIA-COLI RIBONUCLEASE HI - CORRELATIONS WITH STRUCTURE AND FUNCTION IN AN ACTIVE ENZYME. *Journal of Molecular Biology,* 246**,** 144-163.

MARION, D. (2005) Fast acquisition of NMR spectra using Fourier transform of non-equispaced data. *Journal of Biomolecular Nmr,* 32**,** 141-150.

MARKLEY, J. L., BAX, A., ARATA, Y., HILBERS, C. W., KAPTEIN, R., SYKES, B. D., WRIGHT, P. E. & WUTHRICH, K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids - (IUPAC Recommendations 1998). *Pure and Applied Chemistry,* 70**,** 117-142.

MASON, R. W. (1986) SPECIES VARIANTS OF CATHEPSIN-L AND THEIR IMMUNOLOGICAL IDENTIFICATION. *Biochemical Journal,* 240**,** 285-288.

MATSUMOTO, K., YAMAMOTO, D., OHISHI, H., TOMOO, K., ISHIDA, T., INOUE, M., SADATOME, T., KITAMURA, K. & MIZUNO, H. (1989) MODE OF BINDING OF E-64-C, A POTENT THIOL PROTEASE INHIBITOR, TO PAPAIN AS DETERMINED BY X-RAY CRYSTAL ANALYSIS OF THE COMPLEX. *Febs Letters,* 245**,** 177-180.

MCGRATH, M. E. (1999) The lysosomal cysteine proteases. *Annual Review of Biophysics*

and Biomolecular Structure, 28**,** 181-204.

MESAROS, N., NORDMANN, P., PLESIAT, P., ROUSSEL-DELVALLEZ, M., VAN ELDERE, J., GLUPCZYNSKI, Y., VAN LAETHEM, Y., JACOBS, F., LEBECQUE, P., MALFROOT, A., TULKENS, P. M. & VAN BAMBEKE, F. (2007) Pseudomonas aeruginosa: resistance and therapeutic options at the turn of the new millennium. *Clinical Microbiology and Infection,* 13**,** 560-578.

MONTELIONE, G. T., LYONS, B. A., EMERSON, S. D. & TASHIRO, M. (1992) AN EFFICIENT TRIPLE RESONANCE EXPERIMENT USING C-13 ISOTROPIC MIXING FOR DETERMINING SEQUENCE-SPECIFIC RESONANCE ASSIGNMENTS OF ISOTOPICALLY-ENRICHED PROTEINS. *Journal of the American Chemical Society,* 114**,** 10974-10975.

MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. & THORNTON, J. M. (1992) STEREOCHEMICAL QUALITY OF PROTEIN-STRUCTURE COORDINATES. *Proteins-Structure Function and Genetics,* 12**,** 345-364.

NABUURS, S. B., SPRONK, C., VRIEND, G. & VUISTER, G. W. (2004) Concepts and tools for NMR restraint analysis and validation. *Concepts in Magnetic Resonance Part A,* 22A**,** 90-105.

NAGLER, D. K., STORER, A. C., PORTARO, F. C. V., CARMONA, E., JULIANO, L. & MENARD, R. (1997) Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry,* 36**,** 12608-12615.

NEUHAUS, D. & WILLIAMSON, M. P. (2000) *The nuclear Overhauser effect in structural and conformational analysis,* Canada, Wiley-VCH, Inc.

NIELSEN, H., ENGELBRECHT, J., BRUNAK, S. & VONHEIJNE, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering,* 10**,** 1-6.

NILGES, M. (1995) CALCULATION OF PROTEIN STRUCTURES WITH AMBIGUOUS DISTANCE RESTRAINTS - AUTOMATED ASSIGNMENT OF AMBIGUOUS NOE CROSSPEAKS AND DISULFIDE CONNECTIVITIES. *Journal of Molecular Biology,* 245**,** 645-660.

NILGES, M., CLORE, G. M. & GRONENBORN, A. M. (1988) DETERMINATION OF 3-DIMENSIONAL STRUCTURES OF PROTEINS FROM INTERPROTON DISTANCE DATA BY DYNAMICAL SIMULATED ANNEALING FROM A RANDOM ARRAY OF ATOMS - CIRCUMVENTING PROBLEMS

ASSOCIATED WITH FOLDING. *Febs Letters,* 239**,** 129-136.

NILGES, M. & O'DONOGHUE, S. I. (1998) Ambiguous NOEs and automated NOE assignment. *Progress in Nuclear Magnetic Resonance Spectroscopy,* 32**,** 107-139.

NOSSAL, N. G. & HEPPEL, L. A. (1966) RELEASE OF ENZYMES BY OSMOTIC SHOCK FROM ESCHERICHIA COLI IN EXPONENTIAL PHASE. *Journal of Biological Chemistry,* 241**,** 3055-&.

OSBORNE, M. J., SCHNELL, J., BENKOVIC, S. J., DYSON, H. J. & WRIGHT, P. E. (2001) Backbone dynamics in dihydrofolate reductase complexes: Role of loop flexibility in the catalytic mechanism. *Biochemistry,* 40**,** 9846-9859.

OTTIGER, M. & BAX, A. (1998) Determination of relative N-H-N N-C ', C-alpha-C ', andC(alpha)-H-alpha effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *Journal of the American Chemical Society,* 120**,** 12334-12341.

OTTIGER, M., DELAGLIO, F. & BAX, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance,* 131**,** 373-378.

PALMER, A. G., RANCE, M. & WRIGHT, P. E. (1991) INTRAMOLECULAR MOTIONS OF A ZINC FINGER DNA-BINDING DOMAIN FROM XFIN CHARACTERIZED BY PROTON-DETECTED NATURAL ABUNDANCE C-12 HETERONUCLEAR NMR-SPECTROSCOPY. *Journal of the American Chemical Society,* 113**,** 4371-4380.

PERMI, P., HEIKKINEN, S., KILPELAINEN, I. & ANNILA, A. (1999) Measurement of (1)J(NC ') and (2)J(HNC ') couplings from spin-state-selective two-dimensional correlation spectrum. *Journal of Magnetic Resonance,* 140**,** 32-40.

PROSSER, R. S., LOSONCZI, J. A. & SHIYANOVSKAYA, I. V. (1998) Use of a novel aqueous liquid crystalline medium for high-resolution NMR of macromolecules in solution. *Journal of the American Chemical Society,* 120**,** 11010-11011.

RAWLINGS, N. D. & BARRETT, A. J. (1994) FAMILIES OF CYSTEINE PEPTIDASES. *Proteolytic Enzymes: Serine and Cysteine Peptidases,* 244**,** 461-486.

RAWLINGS, N. D., TOLLE, D. P. & BARRETT, A. J. (2004) Evolutionary families of peptidase inhibitors. *Biochemical Journal,* 378**,** 705-716.

REDZYNIA, I., LJUNGGREN, A., ABRAHAMSON, M., MORT, J. S., KRUPA, J. C., JASKOLSKI, M. & BUJACZ, G. (2008) Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin. *Journal*

of Biological Chemistry, 283**,** 22815-22825.

REDZYNIA, I., LJUNGGREN, A., BUJACZ, A., ABRAHAMSON, M., JASKOLSKI, M. & BUJACZ, G. (2009) Crystal structure of the parasite inhibitor chagasin in complex with papain allows identification of structural requirements for broad reactivity and specificity determinants for target proteases. *Febs Journal,* 276**,** 793-806.

RIEKENBERG, S., WITJES, B., SARIC, M., BRUCHHAUS, I. & SCHOLZE, H. (2005) Identification of EhICP1, a chagasin-like cysteine protease inhibitor of Entamoeba histolytica. *Febs Letters,* 579**,** 1573-1578.

RIGDEN, D. J., MOSOLOV, V. V. & GALPERIN, M. Y. (2002) Sequence conservation in the chagasin family suggests a common trend in cysteine proteinase binding by unrelated protein inhibitors. *Protein Science,* 11**,** 1971-1977.

ROBERTS, G. C. K. (1993) *NMR of macromolecules*, Oxford university express.

ROVNYAK, D., FRUEH, D. P., SASTRY, M., SUN, Z. Y. J., STERN, A. S., HOCH, J. C. & WAGNER, G. (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *Journal of Magnetic Resonance,* 170**,** 15-21.

RUDOLPH, R. & LILIE, H. (1996) In vitro folding of inclusion body proteins. *Faseb Journal,* 10**,** 49-56.

SALI, A. & BLUNDELL, T. L. (1993) COMPARATIVE PROTEIN MODELING BY SATISFACTION OF SPATIAL RESTRAINTS. *Journal of Molecular Biology,* 234**,** 779-815.

SALMON, D., DO AIDO-MACHADO, R., DIEHL, A., LEIDERT, M., SCHMETZER, O., LIMA, A., SCHARFSTEIN, J., OSCHKINAT, H. & PIRES, J. R. (2006) Solution structure and backbone dynamics of the Trypanosoma cruzi cysteine protease inhibitor chagasin. *Journal of Molecular Biology,* 357**,** 1511-1521.

SANDERSON, S. J., WESTROP, G. D., SCHARFSTEIN, J., MOTTRAM, J. C. & COOMBS, G. H. (2003a) Functional conservation of a natural cysteine peptidase inhibitor in protozoan and bacterial pathogens. *Febs Letters,* 542**,** 12-16.

SANDERSON, S. J., WESTROP, G. D., SCHARFSTEIN, J., MOTTRAM, J. C., COOMBS, G. H. & FK (2003b) Functional conservation of a natural cysteine peptidase inhibitor in protozoan and bacterial pathogens. *Febs Letters,* 542**,** 12-16.

SANTOS, C. C., SANT'ANNA, C., TERRES, A., CUNHA-E-SILVA, N. L.,

SCHARFSTEIN, J. & LIMA, A. (2005) Chagasin, the endogenous cysteine-protease inhibitor of Trypanosoma cruzi, modulates parasite differentiation and invasion of mammalian cells. *Journal of Cell Science,* 118**,** 901-915.

SATTLER, M., SCHLEUCHER, J., GRIESINGER, C. & TJ (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy,* 34**,** 93-158.

SCHECHTE.I & BERGER, A. (1967) ON SIZE OF ACTIVE SITE IN PROTEASES .I. PAPAIN. *Biochemical and Biophysical Research Communications,* 27**,** 157-&.

SCHMEIDER, P., STERN, A. S., WAGNER, G. & HOCH, J. C. (1994) IMPROVED RESOLUTION IN TRIPLE-RESONANCE SPECTRA BY NONLINEAR SAMPLING IN THE CONSTANT-TIME DOMAIN. *Journal of Biomolecular Nmr,* 4**,** 483-490.

SCHMIEDER, P., STERN, A. S., WAGNER, G. & HOCH, J. C. (1993) APPLICATION OF NONLINEAR SAMPLING SCHEMES TO COSY-TYPE SPECTRA. *Journal of Biomolecular Nmr,* 3**,** 569-576.

SCHWARTZ, W. N. & BARRETT, A. J. (1980) HUMAN CATHEPSIN-H. *Biochemical Journal,* 191**,** 487-497.

SHARMA, D. & RAJARATHNAM, K. (2000) C-13 NMR chemical shifts can predict disulfide bond formation. *Journal of Biomolecular Nmr,* 18**,** 165-171.

SKLENAR, V., PIOTTO, M., LEPPIK, R. & SAUDEK, V. (1993) GRADIENT-TAILORED WATER SUPPRESSION FOR H-1-N-15 HSQC EXPERIMENTS OPTIMIZED TO RETAIN FULL SENSITIVITY. *Journal of Magnetic Resonance Series A,* 102**,** 241-245.

SMITH, B. O., PICKEN, N. C., WESTROP, G. W., BROMEK, K., MOTTRAM, J. C. & COOMBS, G. H. (2006) The structure of Leishmania mexicana ICP provides evidence for convergent evolution of cysteine peptidase inhibitors. *Journal of Biological Chemistry,* 281**,** 5821-5828.

STORER, A. C. & MENARD, R. (1994) CATALYTIC MECHANISM IN PAPAIN FAMILY OF CYSTEINE PEPTIDASES. *Proteolytic Enzymes: Serine and Cysteine Peptidases,* 244**,** 486-500.

THEOBALD, D. L. & WUTTKE, D. S. (2006) THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics,* 22**,**

2171-2172.

TURK, B., TURK, D. & TURK, V. (2000) Lysosomal cysteine proteases: more than scavengers. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology,* 1477**,** 98-111.

TYCKO, R., BLANCO, F. J. & ISHII, Y. (2000) Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *Journal of the American Chemical Society,* 122**,** 9340-9341.

VASILJEVA, O., DOLINAR, M., TURK, V. & TURK, B. (2003) Recombinant human cathepsin H lacking the mini chain is an endopeptidase. *Biochemistry,* 42**,** 13522-13528.

VUISTER, G. W., WANG, A. C. & BAX, A. (1993) MEASUREMENT OF 3-BOND NITROGEN CARBON-J COUPLINGS IN PROTEINS UNIFORMLY ENRICHED IN N-15 AND C-13. *Journal of the American Chemical Society,* 115**,** 5334-5335.

WANG, S. X., PANDEY, K. C., SCHARFSTEIN, J., WHISSTOCK, J., HUANG, R. K., JACOBELLI, J., FLETTERICK, R. J., ROSENTHAL, P. J., ABRAHAMSON, M., BRINEN, L. S., ROSSI, A., SALI, A. & MCKERROW, J. H. (2007) The structure of chagasin in complex with a cysteine protease clarifies the binding mode and evolution of an inhibitor family. *Structure,* 15**,** 535-543.

WANG, Y. X., MARQUARDT, J. L., WINGFIELD, P., STAHL, S. J., LEE-HUANG, S., TORCHIA, D. & BAX, A. (1998) Simultaneous measurement of H-1-N-15, H-1-C-13 ', and N-15-C-13 ' dipolar couplings in a perdeuterated 30 kDa protein dissolved in a dilute liquid crystailine phase. *Journal of the American Chemical Society,* 120**,** 7385-7386.

WISHART, D. S. & SYKES, B. D. (1994) THE C-13 CHEMICAL-SHIFT INDEX - A SIMPLE METHOD FOR THE IDENTIFICATION OF PROTEIN SECONDARY STRUCTURE USING C-13 CHEMICAL-SHIFT DATA. *Journal of Biomolecular Nmr,* 4**,** 171-180.

WITTEKIND, M. & MUELLER, L. (1993) HNCACB, A HIGH-SENSITIVITY 3D NMR EXPERIMENT TO CORRELATE AMIDE-PROTON AND NITROGEN RESONANCES WITH THE ALPHA-CARBON AND BETA-CARBON RESONANCES IN PROTEINS. *Journal of Magnetic Resonance Series B,* 101**,** 201-

205.

WUTHRICH, K. (1990) PROTEIN-STRUCTURE DETERMINATION IN SOLUTION BY NMR-SPECTROSCOPY. *Journal of Biological Chemistry,* 265**,** 22059-22062.

WÜTHRICH, K. (1986) *NMR of Proteins and Nucleic Acids,* New York, Wiley-Interscience.

YAMAZAKI, T., FORMANKAY, J. D. & KAY, L. E. (1993) 2-DIMENSIONAL NMR EXPERIMENTS FOR CORRELATING C-13-BETA AND H-1-DELTA/EPSILON CHEMICAL-SHIFTS OF AROMATIC RESIDUES IN C-13-LABELED PROTEINS VIA SCALAR COUPLINGS. *Journal of the American Chemical Society,* 115**,** 11054-11055.

APPENDICES

Appendix A   Chemical shift assignments of PA-ICP. Chemical shifts are given in p.p.m. (-) indicates unassigned or unobserved resonances, (*) indicates atoms with degenerated chemical shift. The prochiral groups are assigned ambiguously.

		H	N	CA									
-20	M	-											
-19	G	-											
-18	S	-											
-17	S	-											
-16	H	-											
-15	H	-											
-14	H	-											
-13	H	-											
-12	H	-											
-11	H	-											
-10	S	-											
-9	S	-											
-8	G	-											
-7	L	7.99	121.4	55.12	4.30:Ha	1.49:Hba 42.40:Cb	1.52:Hbb	1.50:Hg 26.93:Cg	0.78:Hda* 23.45:Cda	0.83:Hdb* 24.75:Cdb			
-6	V	-		59.76	4.34:Ha	2.00:Hb 32.67:Cb	0.86:Hga* 20.46:Cga	0.89:Hgb* 21.03:Cgb					
-5	P	-		63.12	4.35:Ha	1.83:Hba 32.11:Cb	2.22:Hbb	1.89:Hga 27.44:Cg	1.98:Hgb	3.59:Hda 51.06:Cd	3.80:Hdb		
-4	R	-		56.43	4.26:Ha	1.74:Hba	1.81:Hbb	1.62:Hga	1.64:Hgb	3.14:Hda	3.14:Hdb		
-3	G	-											
-2	S	-											
-1	H	-											
1	Q	-		55.94	4.24:Ha	1.91:Hba 29.32:Cb	2.04:Hbb	2.29:Hga 33.84:Cg	2.29:Hgb				
2	K	8.24	122.69	54.14	4.58:Ha	1.66:Hba 32.77:Cb	1.81:Hbb	1.41:Hga 24.86:Cg	1.42:Hgb 29.45:Cd	1.70:Hda	1.70:Hdb	3.02:Hea 42.20:Ce	3.02:Heb
3	P	-		64.34	4.49:Ha	2.00:Hba 32.16:Cb	2.29:Hbb	2.03:Hga 27.41:Cg	2.04:Hgb	3.70:Hda 50.55:Cd	3.82:Hdb		
4	V	7.51	118.11	60.89	4.79:Ha	1.84:Hb 34.71:Cb	0.74:Hga* 21.33:Cga	0.84:Hgb* 21.09:Cgb					

#	AA	H	N	CA									
5	V	9.16	129.83	61.62	4.17:Ha	1.83:Hb 34.02:Cb	0.79:Hg1* 20.63:Cg1	0.87:Hg2* 21.57:Cg2					
6	T	8.38	120.15	60.71	4.92:Ha	4.11:Hb 69.79:Cb	1.04:Hg2* 21.39:Cg2						
7	L	9.02	127.88	53.78	4.77:Ha	1.16:Hba 46.29:Cb	1.63:Hbb	1.39:Hg 27.49:Cg	0.86:Hda* 27.44:Cda	1.12:Hdb* 23.63:Cdb			
8	D	8.95	122.35	53.23	4.73:Ha	2.61:Hba 42.64:Cb	2.98:Hbb						
9	D	7.52	118.12	52.99	4.83:Ha	2.58:Hba 42.76:Cb	2.85:Hbb						
10	A	9	128.2	54.01	4.06:Ha	0.95:Hb* 17.38:Cb							
11	D	8.97	121.44	56.66	4.34:Ha	2.53:Hba 39.98:Cb	2.71:Hbb						
12	D	7.89	117.95	55.27	4.55:Ha	2.69:Hba 40.53:Cb	2.79:Hbb						
13	C	7.88	118.49	55.66	4.85:Ha	3.64:Hba 44.87:Cb	4.29:Hbb						
14	S	7.52	118.12	57.16	4.65:Ha	3.59:Hba 65.03:Cb	4.12:Hbb						
15	P	-	-	62.28	4.33:Ha	1.61:Hba 35.16:Cb	2.01:Hbb	1.27:Hga 23.22:Cg	1.70:Hgb	3.40:Hda 52.22:Cd	3.56:Hdb		
16	L	8.95	120.37	55.26	4.44:Ha	1.46:Hba 42.37:Cb	1.98:Hbb	1.30:Hg 27.53:Cg	0.82:Hda* 27.92:Cda	0.91:Hdb* 23.45:Cdb			
17	K	8.55	129.52	56.28	5.11:Ha	1.75:Hba 32.03:Cb	1.75:Hbb	1.23:Hga 25.36:Cg	1.49:Hgb	1.65:Hda 29.29:Cd	1.65:Hdb	2.91:Hea 42.00:Ce	2.91:Heb
18	L	8.31	123.79	52.86	5.04:Ha	1.43:Hba 47.05:Cb	1.70:Hbb	0.66:Hg 27.29:Cg	0.85:Hda* 23.31:Cda	0.85:Hdb* 23.32:Cdb			
19	T	8.83	114.21	60.41	4.82:Ha	3.83:Hb 71.66:Cb	1.14:Hg2* 20.84:Cg2						
20	Q	8.62	122.96	58.76	3.48:Ha	1.77:Hba 28.99:Cb	1.92:Hbb	1.82:Hga 34.81:Cg	2.14:Hgb	6.73:He2a	7.47:He2b 111.65:Ne2		

21	G	8.83	111.97	44.72	3.55:Hsa	4.47:Hab					
22	Q	7.66	119.09	57.61	4.38:Ha	1.80:Hba 30.94:Cb	2.23:Hbb	2.38:Hga 36.01:Cg	2.38:Hgb	6.56:He2a	8.03:He2b	111.87:Ne2
23	E	8.43	121.72	55	5.16:Ha	1.93:Hba 32.46:Cb	2.04:Hbb	2.02:Hga 36.88:Cg	2.29:Hgb			
24	L	9.14	126.86	52.91	5.25:Ha	0.94:Hba 46.58:Cb	1.73:Hbb	1.31:Hg 27.39:Cg	0.44:Hda* 25.96:Cda	0.77:Hdb* 24.40:Cdb		
25	V	9.13	127.88	61.44	4.74:Ha	1.91:Hb 33.39:Cb	0.74:Hga* 21.29:Cga	0.80:Hgb* 21.41:Cgb				
26	L	9.26	132.1	53.21	5.36:Ha	1.03:Hba 45.71:Cb	2.09:Hbb	1.15:Hg 27.56:Cg	0.11:Hda* 23.21:Cda	0.15:Hdb* 27.30:Cdb		
27	T	8.2	122.61	60.77	5.43:Ha	3.68:Hb 71.41:Cb	0.95:Hg2* 21.53:Cg2					
28	L	8.8	125.3	50.95	4.80:Ha	0.67:Hba 43.91:Cb	0.67:Hbb	0.62:Hg 25.96:Cg	-1.11:Hda* 23.68:Cda	0.04:Hdb* 22.47:Cdb		
29	P	-	-	62.83	4.80:Ha	1.96:Hba 32.28:Cb	2.38:Hbb	2.06:Hga 27.48:Cg	2.19:Hgb	3.55:Hda 50.62:Cd	3.89:Hdb	
30	S	8.45	117.03	57.84	4.73:Ha	4.02:Hba 66.48:Cb	4.07:Hbb					
31	N	-	-	50.33	5.32:Ha	2.82:Hba 39.14:Cb	3.15:Hbb					
32	P	-	-	64.81	4.56:Ha	1.94:Hba 32.17:Cb	2.24:Hbb	1.58:Hga 26.66:Cg	1.67:Hgb	3.60:Hda 51.25:Cd	3.90:Hdb	
33	T	-	-	64.65	4.21:Ha	4.30:Hb 69.01:Cb	1.29:Hg2* 21.80:Cg2					
34	T	7.68	109.96	62.72	4.39:Ha	4.54:Hb 70.86:Cb	1.31:Hg2* 21.46:Cg2					
35	G	7.89	109.41	45.38	3.72:Haa	4.25:Hab						
36	F	7.34	118.88	58.31	3.49:Ha	0.71:Hba 39.18:Cb	2.26:Hbb	6.48:Hd* 131.15:Cd*	7.22:He* 131.37:Ce*	6.65:Hz		
37	R	8.14	118.44	54.37	4.59:Ha	1.89:Hba 33.74:Cb	1.58:Hbb	1.58:Hga 26.42:Cg	3.15:Hda 43.34:Cd	3.26:Hdb		

#	Res	H	N	CA	Ha	Hb	Hbb					Ne1	
38	W	9.15	122.04	57.82	5.39:Ha	3.13:Hba 31.80:Cb	3.48:Hbb	7.40:Hd1 126.89:Cd1	10.06:He1	7.02:Hz2 115.01:Cz2	6.81:Hh2 124.22:Ch2	131.27:Ne1	
39	E	-	-										
40	L	9.05	126.98	54.78	4.30:Ha	1.26:Hba 41.23:Cb	1.79:Hbb	1.30:Hg 27.15:Cg	0.60:Hda* 24.04:Cda	0.76:Hdb* 25.95:Cdb			
41	R	8.32	128.51	57.35	4.36:Ha	1.68:Hba 31.29:Cb	1.69:Hbb	1.58:Hga 27.17:Cg	1.72:Hgb	3.15:Hda 43.20:Cd	3.15:Hdb		
42	N	8.39	115.21	50.06	5.30:Ha	2.67:Hba 41.26:Cb	3.11:Hbb	7.14:Hd2a	7.64:Hd2b	111.05:Nd2			
43	P	-	-	62.57	4.39:Ha	1.94:Hba 31.18:Cb	2.38:Hbb	1.79:Hga 26.53:Cg	2.16:Hgb	3.57:Hda 50.53:Cd	4.05:Hdb		
44	A	8.21	124.23	52.78	3.82:Ha	0.78:Hb* 15.85:Cb							
45	A	8.25	117.26	54.08	4.38:Ha	1.45:Hb* 19.21:Cb							
46	S	8.3	112.6	61.37	4.28:Ha	3.88:Hba 62.65:Cb	3.88:Hbb						
47	V	7.29	110.24	61.39	4.67:Ha	2.45:Hb 34.15:Cb	0.42:Hg1* 20.91:Cg1	0.53:Hg2* 18.47:Cg2					
48	L	8.31	119.1	53.4	5.23:Ha	1.06:Hba 43.82:Cb	1.87:Hbb	1.59:Hg 26.48:Cg	-0.12:Hda* 25.23:Cda	0.53:Hdb* 23.09:Cdb			
49	K	9.31	123.41	54.26	4.78:Ha	1.66:Hba 35.39:Cb	1.83:Hbb	1.65:Hga 29.19:Cg	1.65:Hgb	1.37:Hda 24.85:Cd	1.39:Hdb	2.94:Hea 42.19:Ce	2.94:Heb
50	R	9.2	126.27	57.85	4.16:Ha	1.89:Hba 30.67:Cb	1.94:Hbb	1.69:Hga 27.33:Cg	1.93:Hgb	3.16:Hda 43.46:Cd	3.16:Hdb		
51	L	8.17	124.27	53.51	4.56:Ha	1.36:Hba 41.95:Cb	1.36:Hbb	1.47:Hg 26.47:Cg	0.71:Hda* 25.50:Cda	0.76:Hdb* 22.55:Cdb			
52	G	7.76	108.83	44.92	3.93:Hea	4.21:Hab							
53	P	-	-	62.6	4.41:Ha	1.95:Hba 32.66:Cb	2.21:Hbb	1.77:Hga 26.62:Cg	1.94:Hgb	3.46:Hda 49.60:Cd	3.54:Hdb		
54	E	-	-										
55	V	-	-	61.56	4.11:Ha	1.75:Hb	0.74:Hga*	0.77:Hgb*					

No	AA	HN	N	CA	HA					
56	Y	8.57	127.55	56.96	5.14:Ha	34.02:Cb 2.79:Hba 39.71:Cb	20.99:Cga 2.94:Hbb	20.99:Cgb 6.98:Hd* 133.16:Cd*	6.71:He* 118.26:Ce*	
57	S	·	·							
58	N	·	·	52.97	4.70:Ha	2.71:Hba 39.34:Cb	2.84:Hbb	6.97:Hd2a	7.56:Hd2b	113.08:Nd2
59	S	8.56	117.76	58.97	4.39:Ha	3.85:Hba 63.72:Cb	3.85:Hbb			
60	E	8.48	122.47	57.09	4.23:Ha	1.91:Hba 30.01:Cb	2.04:Hbb 2.22:Hga 36.29:Cg	2.22:Hgb		
61	E	8.26	120.73	57.24	4.18:Ha	1.90:Hba 30.14:Cb	2.01:Hbb 2.22:Hga 36.31:Cg	2.22:Hgb		
62	D	8.3	120.71	54.72	4.57:Ha	2.62:Hba 41.13:Cb	2.73:Hbb			
63	S	8.22	116.25	58.97	4.39:Ha	3.88:Hba 63.80:Cb	3.88:Hbb			
64	G	8.39	110.45	45.39	3.82:Haa	3.83:Hab				
65	L	7.92	121.5	54.94	4.38:Ha	1.53:Hba 42.25:Cb	1.53:Hbb 26.79:Cg	1.50:Hg 26.79:Cg 0.80:Hda* 23.40:Cda	0.85:Hdb* 24.83:Cdb	
66	V	8.11	121.38	62.88	4.07:Ha	2.02:Hb 32.38:Cb	0.91:Hga* 21.00:Cga	0.91:Hgb* 21.00:Cgb		
67	G	8.58	113.25	45.39	4.03:Haa	4.03:Hab				
68	S	8.26	115.56	58.76	4.37:Ha	3.79:Hba 63.88:Cb	3.91:Hbb			
69	G	8.61	111.12	45.5	4.12:Haa	4.12:Hab				
70	G	8.36	109.7	45.55	4.08:Haa	4.24:Hab				
71	E	8.5	121.69	55.51	4.94:Ha	1.69:Hba 32.72:Cb	1.81:Hbb 2.01:Hga 35.70:Cg	2.01:Hgb		
72	S	9.18	120.14	58.12	4.88:Ha	3.23:Hba 66.12:Cb	3.49:Hbb			
73	T	8.11	117.59	60.91	5.64:Ha	3.51:Hb 71.82:Cb	0.94:Hg2* 21.88:Cg2			

#	Res										
74	W	9.47	128.48	57.92	4.37:Ha	2.68:Hba 34.21:Cb	2.99:Hbb	6.82:Hd1 125.25:Cd1	9.86:He1	7.18:Hz2 113.64:Cz2	127.34:Ne1
75	R	8.12	119.86	54.55	5.44:Ha	1.51:Hba 33.58:Cb	1.74:Hbb	1.47:Hga 28.77:Cg	1.58:Hgb	3.08:Hda 43.10:Cd	3.08:Hdb
76	F	9.47	120.27	55.55	5.40:Ha	2.66:Hba 44.64:Cb	2.94:Hbb	6.94:Hd* 132.81:Cd*	6.97:He* 130.88:Ce*	6.69:Hz 128.81:Cz	
77	R	9.65	121.44	53.89	4.98:Ha	1.75:Hba 32.58:Cb	1.78:Hbb	1.42:Hga 27.07:Cg	1.58:Hgb	3.18:Hda 42.33:Cd	3.18:Hdb
78	V	8.95	127.06	63.83	4.09:Ha	2.27:Hb 30.10:Cb	0.75:Hg1* 25.28:Cg1	0.77:Hg2* 22.26:Cg2			
79	A	9.08	135.03	53.1	4.47:Ha	1.27:Hb* 22.20:Cb					
80	A	7.79	120.39	51.45	4.47:Ha	1.46:Hb* 21.99:Cb					
81	S	7.89	110.95	57.98	4.10:Ha	3.77:Hba 64.15:Cb	3.77:Hbb				
82	G	8.68	109.91	44.45	3.67:Haa	4.40:Hab					
83	D	8.17	119.47	52.94	5.58:Ha	2.40:Hba 43.76:Cb	2.56:Hbb				
84	D	8.81	121.13	52.88	4.60:Ha	2.26:Hba 45.78:Cb	2.34:Hbb				
85	R	10.05	122.42	55.52	4.90:Ha	1.48:Hba 33.93:Cb	1.56:Hbb	1.28:Hga 27.55:Cg	1.28:Hgb	3.03:Hda 43.38:Cd	3.05:Hdb
86	L	8.1	123.61	52.82	4.70:Ha	1.05:Hba 45.76:Cb	1.24:Hbb	1.06:Hg 26.73:Cg	0.27:Hda* 25.00:Cda	0.30:Hdb* 24.30:Cdb	
87	E	9	121.35	55	5.62:Ha	1.81:Hba 33.52:Cb	2.09:Hbb	2.08:Hga 36.31:Cg	2.16:Hgb		
88	L	9.67	124.61	56.21	5.64:Ha	1.31:Hba 46.88:Cb	1.95:Hbb	1.49:Hg 29.91:Cg	-0.51:Hda* 26.52:Cda	0.28:Hdb* 26.30:Cdb	
89	V	8.74	112.81	59.89	5.17:Ha	1.99:Hb 35.83:Cb	0.95:Hg1* 22.14:Cg1	0.83:Hg2* 18.82:Cg2			
90	Y	8.24	127.79	55.58	4.71:Ha	0.11:Hba	2.11:Hbb	5.59:Hd*	6.22:He*		

No	AA	H	N	CA	HA	39.12:Cb		132.38:Cd*	117.37:Ce*					
91	R	8.34	121	54.95	4.73:Ha	1.55:Hba / 33.71:Cb	1.72:Hbb	1.00:Hga / 25.63:Cg	1.29:Hgb	2.88:Hda / 43.89:Cd	2.99:Hdb	7.56:He	126.10:Ne	
92	R	8.64	124.05	53.32	4.11:Ha	0.26:Hba	0.51:Hbb	-0.19:Hga	0.02:Hgb	2.44:Hda	2.54:Hdb			
93	P	-	-	64.6	3.95:Ha	28.42:Cb / 1.87:Hba / 31.74:Cb	2.19:Hbb	26.10:Cg / 1.88:Hga / 27.40:Cg	2.01:Hgb	43.19:Cd / 3.35:Hda / 50.85:Cd	3.64:Hdb			
94	W	6.47	111.6	56.76	4.66:Ha	3.19:Hba / 27.88:Cb	3.42:Hbb	7.10:Hd1 / 127.66:Cd1	10.38:He1 / 120.11:Ce3	7.55:He3	7.34:Hz2 / 114.77:Cz2	7.15:Hz3 / 122.64:Cz3	7.16:Hh2 / 124.96:Ch2	131.41:Ne1
95	E	7	123.55	55.37	4.36:Ha	1.45:Hba / 29.14:Cb	1.77:Hbb	1.53:Hga / 36.56:Cg	1.88:Hgb					
96	K	8.1	122.96	58.05	3.95:Ha	1.74:Hba / 32.59:Cb	1.74:Hbb	1.63:Hga / 29.02:Cg	1.63:Hgb	1.36:Hda / 24.71:Cd	1.38:Hdb	2.98:Hea / 42.16:Ce	2.98:Heb	
97	D	8.43	117.47	54.36	4.55:Ha	2.57:Hba / 40.88:Cb	2.69:Hbb							
98	A	7.29	122.32	51.75	4.32:Ha	1.28:Hb* / 20.03:Cb								
99	E	8.59	123.03	54.55	4.44:Ha	1.83:Hba / 29.08:Cb	1.98:Hbb	2.30:Hga / 35.82:Cg	2.30:Hgb					
100	P	-	-	63.16	4.28:Ha	1.94:Hba / 32.67:Cb	2.10:Hbb	1.79:Hga / 27.65:Cg	2.04:Hgb	3.64:Hda / 50.60:Cd	3.85:Hdb			
101	A	8.28	126.55	54.05	4.18:Ha	1.28:Hb* / 20.07:Cb								
102	E	7.57	112.14	54.78	4.64:Ha	1.64:Hba / 2.19:Hgb	2.03:Hbb / 33.51:Cb	2.19:Hga / 35.85:Cg						
103	S	9.01	115.58	57.85	5.31:Ha	3.88:Hba / 66.01:Cb	3.88:Hbb							
104	F	9.03	124.66	57.92	4.67:Ha	2.71:Hba / 42.67:Cb	2.71:Hbb / 132.29:Cd*	6.91:Hd* / 131.27:Ce*	6.79:He* / 128.18:Cz	6.60:Hz				
105	S	8.08	123.33	56.65	5.16:Ha	3.55:Hba / 64.35:Cb	3.60:Hbb							

106	C	8.94	124.9	55.14	4.80:Ha	2.59:Hba 48.46:Cb	2.98:Hbb		
107	A	8.94	133.73	53.14	4.42:Ha	1.32:Hb* 18.68:Cb			
108	I	8.74	122.23	60.17	4.67:Ha	1.17:Hb 41.81:Cb	0.84:Hg1a 27.57:Cg1	1.51:Hg1b 18.59:Cg2	0.65:Hg2* 14.34:Cd1	0.39:Hd1*
109	Q	9.03	128.29	55.09	4.96:Ila	1.78:Ilba 34.02:Cb	1.79:Ilbb	1.80:Ilga 35.00:Cg	2.42:Ilgb
110	V	9.27	127.89	60.83	4.79:Ha	2.23:Hb 33.35:Cb	0.68:Hga* 22.39:Cga	0.73:Hgb* 21.35:Cgb	
111	R	8.12	134.38	56.15	4.35:Ha	1.70:Hba 32.63:Cb	1.83:Hbb	1.45:Hga 27.32:Cg	1.57:Hgb	3.06:Had 42.91:Cd	3.40:Hdb

Appendix B   Ramachandran plots and χ_1 torsion angle distributions for the final ensemble of PA-ICP NMR structures

Per-residue Ramachandran plots for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR.   Favourable φ, ψ dihedral angle combinations are indicated by a yellow box, unfavourable by a red box.

Per-residue Ramachandran plots for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable φ, ψ dihedral angle combinations are indicated by a yellow box, unfavourable by a red box.

Per-residue Ramachandran plots for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable φ, ψ dihedral angle combinations are indicated by a yellow box, unfavourable by a red box.

Per-residue Ramachandran plots for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable φ, ψ dihedral angle combinations are indicated by a yellow box, unfavourable by a red box.

Per-residue Ramachandran plots for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable φ, ψ dihedral angle combinations are indicated by a yellow box, unfavourable by a red box.

Per-residue Ramachandran plots for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable φ, ψ dihedral angle combinations are indicated by a yellow box, unfavourable by a red box.

Per-residue Ramachandran plots for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable φ, ψ dihedral angle combinations are indicated by a yellow box, unfavourable by a red box.

Per-residue χ_1 sidechain torsion angle values for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable χ_1 torsion angle values are indicated in yellow, unfavourable in red.

Per-residue χ_1 sidechain torsion angle values for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable χ_1 torsion angle values are indicated in yellow, unfavourable in red.

Per-residue χ_1 sidechain torsion angle values for structures in the final PA-ICP ensemble. Structures were analysed using Procheck-NMR. Favourable χ_1 torsion angle values are indicated in yellow, unfavourable in red.

Appendis C   The optimal modelfree parameters for backbone dynamics by residue

Resi	2cn	model	S^2	S^2 err	S^2_f	S^2_f err	τ_e (ps)	τ_e err (ps)	R_{ex}	R_{ex} err	SSE	F-stat
6		3	0.834	0.022					2.556	0.379	1.34	39
8		3	0.841	0.062					6.346	1.067	12.63	3.2
10		3	0.75	0.058					10.202	1.09	1.45	58
11		3	0.851	0.047					3.421	0.741	2.43	9.6
13		3	0.79	0.034					3.071	0.503	10.61	3.9
17		1	0.792	0.018							17.29	
21		1	0.903	0.023							3.22	
22		2	0.947	0.026			526.316	507.708			12.26	99
25	β	1	0.953	0.027							8.86	
26	β	1	1	0.034							16.9	
27	β	1	0.887	0.022							2.92	
33		2	0.871	0.027			60.781	24.943			2.74	2.7
36		2	0.904	0.014			47.95	16.902			4.31	3.9
38	β	1	0.892	0.022							9.92	
39	β	1	0.89	0.024							20.31	
40	β	2	0.892	0.019			63.49	23.598			0.03	251
41	β	5	0.755	0.039	0.876	0.024	1733.1	369.574			0	
42	β	1	0.848	0.019							5.56	
45		4	0.804	0.016			17.079	5.253	2.418	0.246	0	
47		3	0.849	0.034					1.58	0.498	2.75	4.5
48	β	2	0.947	0.029			1052.63	591.02			13.32	36
49	β	1	0.918	0.02							1.06	
54		5	0.69	0.027	0.868	0.019	1487.12	229.371			0	
56		4	0.891	0.086			540.326	774.594	8.435	1.078	0	
59		5	0.551	0.019	0.893	0.014	1075.57	57.296			0	
60		5	0.375	0.009	0.846	0.007	1221.34	14.704			0	
63		5	0.335	0.01	0.827	0.01	1144.38	24.28			0	
64		5	0.256	0.009	0.823	0.008	1227.71	11.902			0	
65		5	0.296	0.009	0.837	0.009	1166.45	19.407			0	
66		5	0.382	0.008	0.827	0.007	985.855	13.379			0	
67		5	0.334	0.01	0.84	0.01	1085.57	22.019			0	
68		5	0.308	0.01	0.823	0.01	1116.72	14.838			0	
69		5	0.342	0.01	0.827	0.009	1162.87	22.85			0	
70		5	0.504	0.016	0.928	0.012	1206.92	29.596			0	
71		5	0.776	0.025	0.921	0.017	2429.24	638.608			0	
72	β	3	1	0					1.838	0	2.97	4.7
74	β	2	0.895	0.044			1578.95	538.656			5.92	106
75	β	2	0.947	0.029			526.316	582.109			13.03	2.8

Resi	2cn	model	S^2	S^2 err	S^2_f	S^2_f err	τ_e	τ_e err	R_{ex}	R_{ex} err	SSE	F-stat
76	β	1	0.887	0.023							15.96	
77	β	1	0.897	0.02							6.88	
78	β	5	0.897	0.03	0.936	0.018	791.963	642.502			0	
79		1	0.903	0.024							3.2	
81		2	0.789	0.009			12.335	4.403			0.9	84
82		5	0.837	0.032	0.895	0.019	528.791	468.614			0	
83	β	2	0.832	0.01			28.882	6.337			10.01	1.5
85	β	1	0.865	0.014							3.39	
87	β	2	0.872	0.035			1266.39	305.45			3.41	328
88	β	1	0.957	0.025							0.41	
90	β	2	0.893	0.037			1176.84	462.313			0.66	7.4
92		2	0.947	0.027			1578.95	700.999			16.13	19
94		2	0.849	0.02			49.982	13.813			0.64	88
95		5	0.811	0.022	0.911	0.013	564.988	176.308			0	
96		5	0.647	0.018	0.84	0.012	715.641	72.495			0	
98		5	0.547	0.011	0.787	0.008	896.304	40.094			0	
101		5	0.782	0.019	0.909	0.012	1462.79	152.628			0	
103	β	2	0.844	0.013			42.107	9.284			3.65	12
105	β	1	0.92	0.017							4.16	
106	β	2	0.928	0.031			602.349	412.074			5.71	10834
107	β	1	0.864	0.02							8.09	
109	β	1	0.975	0.021							5	
111		2	0.828	0.016			26.78	8.506			2.33	8.4

Resi- residue number

2cn- secondary structure

model- fitted modelfree model

S^2 err, S^2_f err, τ_e err, R_{ex} err- errors for S^2, S^2_f, τ_e and R_{ex} respectively

SSE- sum of squared-error

F-stat- F-statistics

Appendix D   Structures of the side chains of the 20 amino acids
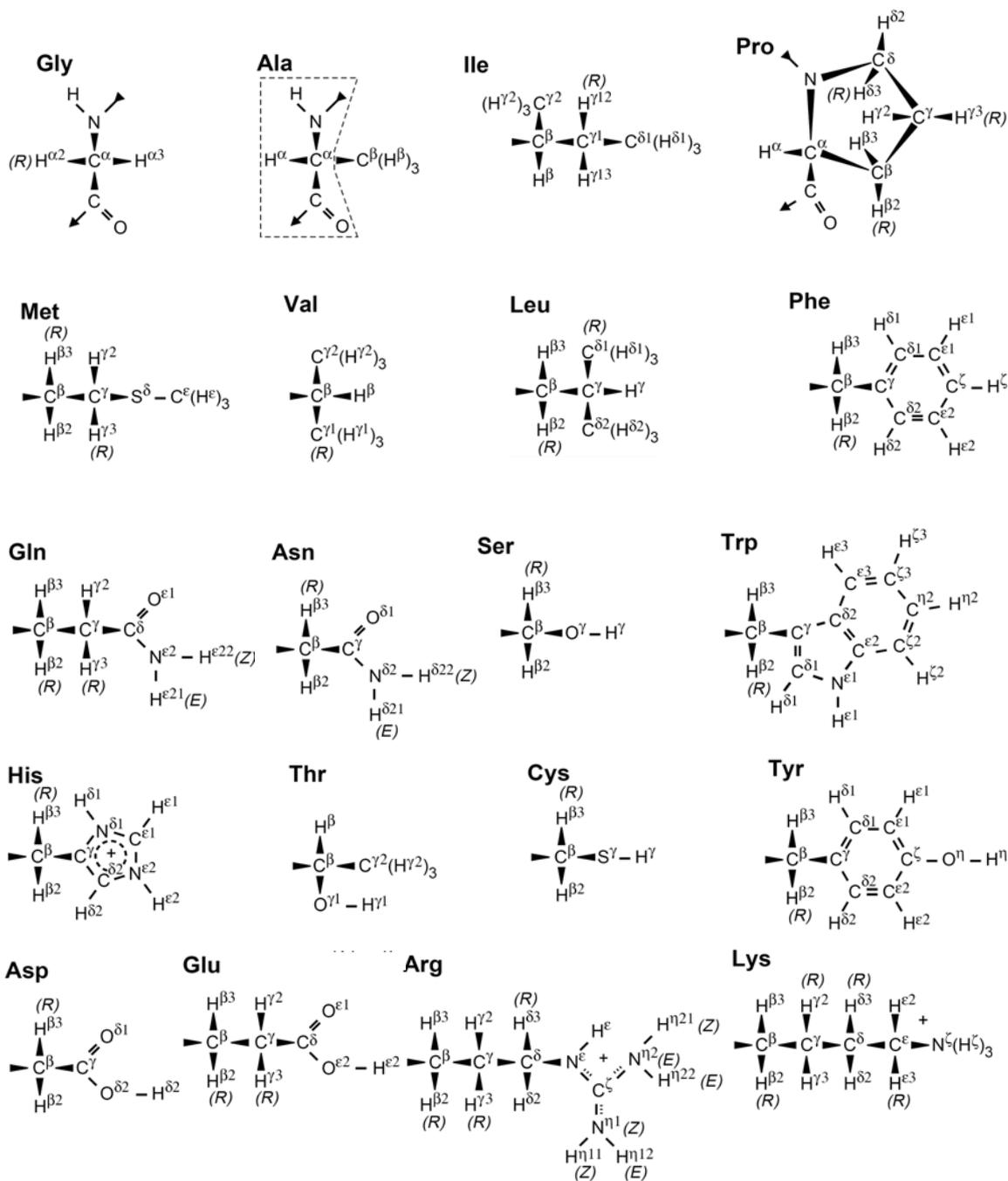
Appendix E   **Alignment of amino acid sequence of PA-ICP**

```
P_A_975/23-135      QQ-KPVVTIDLADDCSPIKITQGCEIVITILPSNF-TTGERWEIRNPAAASVIKR------IGPEVYSNSEED 63
A_V_691/69-180      DDGSSTVTIIVGDDDC-PLHIESRCHIVITILPSNF-STGYRWQVRESASRVIRS------IGPEVYSEAKE- 62
P_S_003/21-131      TP-KNIVSIDTQSDC-PLTITSGCTLIITILPSNF-TTGFRWLTCNPAQNIIRS------IGPEVYANAES- 61
P_S_353/21-131      TP-KNIVSIDTQSDC-PLAIKTGCTLIITILPSNF-TTGFRWLTCNPAQSIIRS------IGPEVYANAES- 61
P_P_175/81-191      QS-SQPVDIDAESEC-PKRIQVGCIITITILPSNF-TTGYRWLVCNPAPNVIQS------IGPEVYSAPED- 61
P_F_030/21-131      QP-KTNVTVEKQSEC-FWRITNGCNLIVMIPSNF-TTGYRWAIODSAGCVIRA------ISPEVYSNPED- 61
LMx_ICP/1-114       -----MIAPILSVKDNDKWWDTHVCKTTEIHIKGNF-TTGYMWTRVCFVCKDYISDEILEVVCKYTPTPSST 65
LMa_ICP/1-118       MQPKMTAFLTMKDNNKCLSVRVGSTLEIHIEGNF-TTGYIWTRVCFVGKEMISDEHLEVTSKYTPKPVSG 69
T_B_250/1-121       ----MSHNIFTEEDNNKTIRMVICETFTIEIESNF-TTGYIWLRSCLAC-TELSDCTFAIQSKFNNRAPHI 65
T_C_894/1-111       ----MSHKVTKAHNGATLIVAVGELVEIQLPSNF-TTGFAWYFECGTK-ESPNESMFTVENKYFP-PDSK 63
Ms_Ac_369/168-281   AMPPEQQVITBAGNGTTISIEKGFTFYIRIEBNF-TTGYSWELNLSQGISI------VSDDYYPPEAED 62
Ms_B_380/99-213     SISTKQQVITBADNGTETSIENGSTFYIKIKBNF-TTGYSWELNLTQGINN-------ISGEYYPPEQPD 62
Ms_B_479/34-146     DIESQNKMITBSDNGNTIYIKEGHAFFIKIKBNF-STGYSWQIRLSNGINQ-------ISDKYHPFESSD 62
Ms_Ac_451/81-191    ETVETGQVVTBADNGTSISINNGEKFTIQILGNF-TTGYSWELNVSEGISI-------ISEDYTQDAVPD 62
Ms_Ac_457/71-181    ETVVTGQVVTBADNGTNISINNGBKFTIQIRGNF-TTGYSWQLNVSEGISI-------ISEDHTQDPAPD 62
Ms_M_371/66-176     QTIETGQIITBAENGKSIRIKNGEIFIIQIRDNF-STGYSWELNVSEGINI-------ISDGYTQDQSPD 62
Ms_B_788/66-176     EIIAKGQVITBADSGKTISIKKGCNFTVSIRBDE-SAGYIWKLNLSGGISI-------IDDEYIEGLNPD 62
Mc_B_332/298-404    EKIPADHVFTEEDNGSIVSVNAGSIWRIDIDAST-ADTYAWNMTLTDGIQIK------RDQFIPQS-- 59
Ms_Ac_113/1-103     ----MVAQIFTNANCFDTVSWNLGITFVIKIRDEF-VAETVWKMETGEGIKIL------REQFTPDIPD 58
Ms_Ac_822/1-85      -------MIKKP-----------FTIQFPBNF-VSGFSWDIATSNGIQI------MRDRFVPEDED 41
Ms_M_986/1-85       -------MIKKP-----------FTIQFPBNF-VSGFSWDIATSNGIQV------MRBRFIPEDKD 41
E_H_990/9-102       STYAAIHIITBKEDHATIHISFNILIKIQIRTNF-STGYAWNIBYPTDTFSLS------QDTIKAEPHPS 63
E_H_923/9-118       STYAAIHIITBKEDHATIHISFNILIKIQIRTNF-STGYAWNIBYPTDTSSLS------QDTIKGEPHPS 63
E_H_178/1-103       -------MSITBDNNNTITIIAKGBNKEIIIHGNF-TTGYSWVVISCEGISN------TVEYYADQHAP 55
Ms_M_779/714-828    NVAQVIPBITBAENGKSINFKNGEIFYIILSBNF-GTFYSLQIKLCSGITILSE---KSIPRNPPLKID 65
Mc_M_201/634-746    YIEPVEDFADVEIBTIDEKVVIGBNITIKIDDBNF-TTGYIWNYSISIDDGKIEL---IFDEYIQDEVEI 64
C_A_732/1-108       -MTNNEGYCILLITNGINRVKLGEKTWIIINANS-STGYVWRITIDNSKVYR------IDKNFYLPSRTG 62
Th_F_578/1-102      -------MAEVCVTISGRVISRVGITVVVRILPBNA-ATGYVWSVASLGDGIIILE------EDSGSPVGS- 55
Ms_M_897/24-132     LLNNTQKMEMTNTKINDINIKVGERTKVEIEDNGGSTGYSWVIGHKPDSLWLI------ESYYIPPEDP 63
My_A_140/33-145     LQVPMNDWITGSDISQNTIIAVGNTIVWQIGSNY-TTFYRWTBDAKIGDSAIVK----QTSHBEVPPTSI 65
Ms_Ac_034/70-177    LSEKFAMIINEQDKGRTIBVEAQSLINVRINBNF-TTGYRWEIBTAGGIEMVG------DSFEKTGDA- 61
C_P_772/54-169      AIINVQDISSSDSIIYFITVKFGCTEITVNIKGNF-TTGYSQQMIIKPNDSIVKVID--AEPSYVPDPHPD 67
Co_B_575/34-149     AALAAPETEFTDFNKPVWVSADSPHIVLKIKSNF-TTGYSWFIVDYIYQLISP------ESHQYIRPSGN- 63


P_A_975/23-135      -------SGIVGSGGESTWRFRVAA------SGDIRIELVYRRPWBKDAEPAESDSCAIQWR- 112
A_V_691/69-180      -------SGIVCSAGQSTWRFCAVQ------AGQISIILLVYQRPWBAGREPDKVDDCAISVE- 111
P_S_003/21-131      -------KEMVGNGGQSVWRFKATD------AGIGRIMMVYQQPWAPEVAPECTBECAISVK- 110
P_S_353/21-131      -------KEMVGNGGQSVWRYKATD------AGIGRIVMTYQQPWAPEVPPECTBECAITVN- 110
P_P_175/81-191      -------TGIVGSSGISTWRFCARA------TCEGNIVLVYQQPWAPEVRPVCTBDCAIRVN- 110
P_F_030/21-131      -------AGIVGAAGISTWRFCAFA------TGIGRIRITSQQPWAPEVIPVETBDCAISVN- 110
LMx_ICP/1-114       -------PMVGVGGIYWVIWKPRK------RCHHTIELVYTRPFBGIKPENBRYTIHLNVK- 113
LMa_ICP/1-118       -------SMVGAGGSYTVFWKPLR------KCQHAVQLVYARPFBGPKPDNBRYTIHLNVE- 117
T_B_250/1-121       NHKNHRRIIVGAGGTMVLBWKAIK------ACKHTISLAYGRPWVGFNAAAKRYNIHVEATA 121
T_C_894/1-111       ---------LIGAGGTEHFHVTWKA------AGIHAVNIITYMRPWTGPSHDSDRDTVYIKAN- 110
Ms_Ac_369/168-281   -----SEQFLIGAGGVHIWBIKADS------BGKCQATAIYKRSWBNETCTEDKDIIDNVEWA- 113
Ms_B_380/99-213     G----IKQFLIGAGGVHIWBIKAIS------KGSCQVTGIYKRPWBKVTCBEEBHBMIKVEWV- 114
Ms_B_479/34-146     -----NNFVIGAGGFRIWRIBGVA------KGNCQVNAIYKRSWBPETCQECNBKINIVWV- 112
Ms_Ac_451/81-191    --------ESTGVFGTYTWIIGAVD------CGTCQVNGEYIRLWBNATCPEDNBTITVEWA- 110
Ms_Ac_457/71-181    --------GYTGVSGIYTWVIGAVD------TGTCQVNGIYKRIWBNTTGMEBNBTIIVEWS- 110
Ms_M_371/66-176     --------GCVGVFGTHSWTIBAVS------CGSCQVNGIYKRTWBNITCTEBKBTINVEWE- 110
Ms_B_788/66-176     --------NITGVFGTHWWVIESTA------FGSCKVNGIYKRTWBNITCTEBKBTINVEWE- 110
Mc_B_332/298-404    --------KGSIYGKHBWKIBAIA------AGBCNISAINMKQKVDITCYEQIFBDITVKWL- 106
Ms_Ac_113/1-103     --------TKTIPGIHBWEYBAVK------FGIWDVECT---YTIFRFGGBRKBKITWKWV- 102
Ms_Ac_822/1-85      --------QTCSGGGYRIWDIVTS------CGSCKITGTWRBGNCIAS-----CFBINIDAE- 84
Ms_M_986/1-85       --------QTSGGGGYRVWDIQVTC------CGSCKITGTWRBGNCIAS-----CFBINIDAE- 84
E_H_990/9-102       --------GMVGFPSIREIQIRFIK------VGITTIKVRLFKT--------------------- 93
E_H_923/9-118       --------GMVGSPRIRCNSIHTIKD----LEQBIIRSVLHBHRCTRQKSHHQVITSS----- 109
E_H_178/1-103       --------GICCGCGKYHIKITGTQ------TCEGKIVLVYRPBAPNAN-DRTBTIKVNWQ- 102
Ms_M_779/714-828    -------KFGGTMGWVHIWKIKVNS------ENDCEIKCTYIQEVTREIARTYTVYIKIGNAD 115
Mc_M_201/634-746    -------GIVGAGCVHBWTFNATE------SGEVEITFDVYRSWBGVGCSINTIIYKITVE- 112
C_A_732/1-108       --------SWGTFCKAVWIIKAIK------KCEASIVINYARVDKEAIKSIVYSIIVE--- 107
Th_F_578/1-102      --------LSGGAFCGHVRVCAEE------FGIWHIDIQLARGWBAGPDKRITVBYS--- 101
Ms_M_897/24-132     --------IFGKFGTRVBTFYGAE------KCQISIQFIHVQIWNKESAEIIDCAVKTS--- 108
My_A_140/33-145     --------AIGAFGTEVWTFAAIK------FGITTIITTSYSSFVIKNAKPACTYITSVTVR- 112
Ms_Ac_034/70-177    ---------IGAAGVRVBQFRAPG------TGEYKISIKNWRDWBGEGSIIIDIFYWTIILVK- 107
C_P_772/54-169      G--------MVGYGCKYTBRFSAVG------SGCTVSTIEYARYBDRPPKCIFKTEIQFKVID 116
Co_B_575/34-149     --------IVGAFGFEIWQFAWNADAFRVPQVTKISIQSIRPWTVPVNGRKLKFVVTIHSE 116
```

XXIII

Appendix F   The amino acid sequence of PA-ICP. The His tag is coloured in red.

<pre>
        -1<u>0</u>   <u>1</u>   1<u>0</u>   2<u>0</u>   3<u>0</u>
MGSSHHHHHH SSGLVPRGSH QKPVVTLDD ADDCSPLKLT QGQELVLTLP SNPTTGFRWE
</pre>

<pre>
4<u>0</u>   5<u>0</u>   6<u>0</u>   7<u>0</u>   8<u>0</u>   9<u>0</u>
LRNPAASVLK RLGPEVYSNS EEDSGLVGSG GESTWRFRVA ASGDDRLELV YRRPWEKDAE
</pre>

<pre>
10<u>0</u>   11<u>0</u>
PAESFSCAIQ VR
</pre>

		No H-tag	With H-tag
Number of amino acids:		111	131
Molecular weight:		12265.6	14428.9
Theoretical isoelectric point:		4.72	5.94
Amino acid composition:	A	8	8
	R	9	10
	N	3	3
	D	8	8
	C	2	2
	Q	5	5
	E	10	10
	G	8	11
	H	0	7
	I	1	1
	L	12	13
	K	4	4
	M	0	1
	F	3	3
	P	8	9
	S	11	16
	T	6	6
	W	3	3
	Y	2	2
	V	9	10