**Daniel Fernandes Duarte**
Licenciatura em Química Aplicada

# Structural characterization of the urea-unfolded state of Colicin Immunity Protein Im7 and Im9

Dissertação para obtenção do Grau de Mestre em
Bioquímica Estrutural e Funcional

Orientador: Prof. Eurico José da Silva Cabrita, Professor Auxiliar, Faculdade de Ciências e Tecnologia

Co-orientador: Ângelo Miguel Mendes de Figueiredo, Doutor, Faculdade de Ciências e Tecnologia

Júri:

Presidente: Prof. Doutor José Ricardo Ramos Franco Tavares
Arguente: Prof. Doutor Doutor Carlos Alberto Gomes Salgueiro
Vogal: Prof. Doutor Doutor Eurico José da Silva Cabrita

**Março 2013**

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

**Daniel Fernandes Duarte**
Licenciatura em Química Aplicada

# Structural characterization of the urea-unfolded state of Colicin Immunity Protein Im7 and Im9

Dissertação para obtenção do Grau de Mestre em
Bioquímica Estrutural e Funcional

Orientador: Prof. Eurico José da Silva Cabrita, Professor Auxiliar, Faculdade de Ciências e Tecnologia

Co-orientador: Ângelo Miguel Mendes de Figueiredo, Doutor, Faculdade de Ciências e Tecnologia

**Março 2013**

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

**“Copyright”**

Daniel Fernandes Duarte
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

## Acknowledgements

I would like to start by expressing my gratitude to my supervisors Prof. Eurico Cabrita and Dr Ângelo Figueiredo, who allowed me this fantastic opportunity to perform my thesis, and also for providing me all the conditions and the support for a successful work.

To all my colleges from the laboratory, Dr Marta Corvo, Dr Filipa Marcelo, Dr João Sardinha and Rui Soares, thank you for all the help, disponibility and support.

To the group of Professor Carlos Salgueiro, for all the help offered and for allowing me to perform some of my experimental work there, thank you.

To Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa and Associate Laboratory REQUIMTE for providing the resources that allowed me to carry out my work.

To Fundação para a Ciência e a Tecnologia through grant no PEst-C/EQB/LA0006/2011 and Project no. PTDC/QUI-QUI/098892/2008. The NMR spectrometers are part of the National NMR Network (RNRMN) and are funded by Fundação para a Ciência e a Tecnologia.

To my family, without whom I wouldn't be here, thank you for all the support, motivation and encouragement through all this process.

Finally to the person who was always by my side and without her help nothing would be possible, thank you Joana.

## Resumo

A maioria das proteínas contendo um único dominio, têm a capacidade de se rearranjar (fold) espontaneamente numa estrutura tridimensional funcional num segundo ou menos. A compreensão destas transições, não só ajuda a revelar a maneira pela qual uma sequência de aminoácidos codifica a estrutura tridimensional correspondente, mas também pode fornecer informações sobre as transições de *folding/unfolding* que muitas proteínas sofrem como parte do seu funcionamento normal, bem como os processos *misfolding* que estão na base de uma série de doenças, como Alzheimer e Parkinson.

Neste trabalho, duas proteínas já bem estudadas e caracterizadas, Colicin immunity proteins Im7 e Im9, foram usadas como modelo para estudos estruturais envolvendo duas diferentes metodologias para promover a sua desnaturação/unfolding. As proteínas apresentam uma elevada homologia a nível de sequência primária e estrutural, divergindo no mecanismo de folding.

Utilizando espectros $^{1}H$-$^{15}N$ HSQC como ferramenta principal, foi realizado um estudo comparativo, de modo a identificar quais os resíduos mais pertubados por um efeito físico, Temperatura, e por um agente químico, Urea. Com este estudo espera-se detectar possíveis semelhanças que possam fornecer novas pistas sobre os aspectos que controlam as transições entre folding/unfolding.

Os resultados relativos à temperatura, evidenciam que os resíduos mais perturbados estão localizados maioritariamente em zonas de loop (random coil) entre hélices, enquanto que com a ultilização de ureia, verifica-se uma maior perturbação em resíduos acessíveis e expostos ao solvente.

Com este estudo, conseguimos verificar que as regiões de loop ou terminais de estrutura secundária, funcionam como pontos de partida no processo de unfolding, e contribuiem para a estabilização de características semelhantes ao estado nativo, mesmo em condições desnaturantes, permitindo que zonas mais hidrofóbicas e presentes no interior da proteína, estejam menos expostas ao solvente.

## Abstract

Most single domain proteins have the ability to fold spontaneously into a precise, functional three-dimensional structure in seconds or less. Understanding how this transition occurs will not only help to uncover the way in which an amino acid sequence encodes the corresponding structure but is also likely to provide insight into the folding/unfolding transitions that many proteins undergo as part of their normal functioning. The characterization of these states is particularly important because they often play crucial roles in folding and misfolding processes, responsible for many human neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases.

In this work, two well characterized proteins, Colicin immunity proteins Im7 and Im9, where used as model for a structural study involving two different approaches to promote their denaturation/unfolding. Im7 and Im9 share a high sequence and structural homology, but despite that fact they fold with different kinetic mechanism *in vitro*.

By using $^{1}H$-$^{15}N$ HSQC spectra as a main tool, we have undertaken a comparative study to identify the residues more affected during the denaturation process of Im7 and Im9 promoted by a physical effect, temperature increase, and by a chemical agent, urea. Our aim was to detect possible similarities that could give insight into the aspects that govern folding/unfolding transitions.

The results from the temperature study show that the residues most perturbed with increasing temperature are mostly located in loop regions between helices, while urea targets preferably residues that are accessible and solvent exposed.

Our study, points out that the ends of well-structured helices can concertedly unfold without entering the mid region residues in the same unfolding process. There seems to be a correlation between dynamic residues (most affected by temperature) and the residues in the regions most perturbed by urea. The results shown that entire loop regions on both proteins may act as concerted units during the unfolding process, and contribute for favorable interactions that delimit and stabilize native-like structural features on the urea-unfolded state, allowing buried regions to be less solvent exposed.

## Contents

## List of Figures

## List of Tables

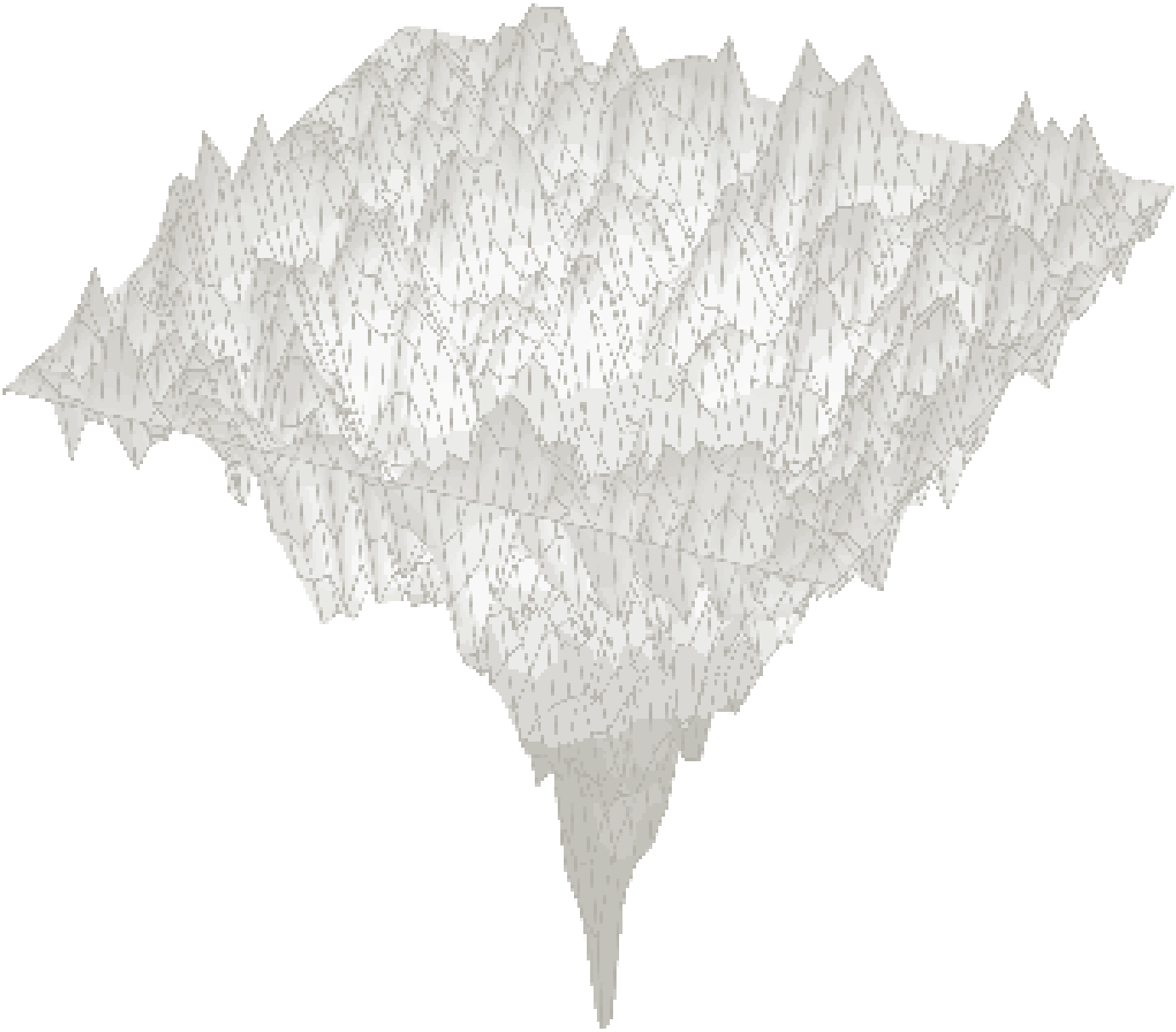
## Abbreviations and symbols

| | |
|---|---|
| 1D | One dimensional |
| 2D | Two dimensional |
| BLAST | Basic Local Alignment Search Tool |
| BMRB | Biological magnetic resonance data bank |
| DNases | Endonucleases |
| DSE | Denatured state ensemble |
| ESI-MS | Electrospray ionization mass spectroscopy |
| GuHCl | Guanidine hydrochloride |
| HCl | Hydrochloric acid |
| HSQC | Heteronuclear single-quantum coherence |
| MW | Molecular weight |
| Im7 | Colicin Immunity Protein Im7 |
| Im9 | Colicin Immunity Protein Im9 |
| MWCO | Molecular Weight Cut-Off |
| MS | Mass spectroscopy |
| NMR | Nuclear magnetic resonance spectroscopy |
| PDB | Protein data bank |
| PMSF | Phenyl-methylsulphonyl fluoride |
| pI | Isoelectric point |
| ppm | Parts per million |
| RF | Radio frequency pulse |
| RNases | Ribonucleases |
| rpm | Rotations per minute |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| TCI | Triple-resonance cryoprobe |
| TSE | Transition state ensemble |
| TOF | Time-of-Flight |
| UV-visible | Ultraviolet-visible |
| δ | Chemical shift |
| ε | Extinction absorption coefficient |

**Amino acid abbreviations**

| | | |
|---|---|---|
| Alanine | Ala | A |
| Arginine | Arg | R |
| Asparagine | Asn | N |
| Aspartate | Asp | D |
| Cysteine | Cys | C |
| Glutamate | Glu | E |
| Glutamine | Gln | Q |
| Glycine | Gly | G |
| Histidine | His | H |
| Isoleucine | Ile | I |
| Leucine | Leu | L |
| Lysine | Lys | K |
| Methionine | Met | M |
| Phenylalanine | Phe | F |
| Proline | Pro | P |
| Serine | Ser | S |
| Threonine | Thr | T |
| Tryptophan | Trp | W |
| Tyrosine | Tyr | Y |
| Valine | Val | V |

# 1 – Introduction



*Newly formed polypeptide sequence must be able to find the way to its correct fold rather than the countless alternatives. To discover how this happens is one of the greatest challenges in modern structural biology.*

*Dobson, 2000*

## 1 – Introduction

One of the defining characteristics of a living system is the ability of even the most intricate of its component molecular structures to self-assemble with precision and fidelity, being one of the most outstanding examples the proteins.

The word protein comes from the Greek word "*protos*", meaning first element. Proteins are the workhorses of biological macromolecules, essential elements for growth and repair, functioning and structure of all living cells, having the ability to self-assemble into a precise three dimensional structure in seconds or less in order to perform their function.

When we think about a protein, is inherent to think about its three-dimensional structure and the routes or paths that drive to such conformation. As proteins undergo a series of events and reorganization (fold), they are characterized in four distinct conformations (**Figure 1**):

**Primary structure** - the linear order of amino acid sequence of the peptide chains.

**Secondary structure** - highly regular sub-structures; α-helix and strands of β-sheet, which are locally defined.

**Tertiary structure** - spatial arrangement of the secondary structures, describing the completely folded and compacted polypeptide chain.

**Quaternary structure** - complex of several protein molecules or polypeptide chains, usually called protein subunits.

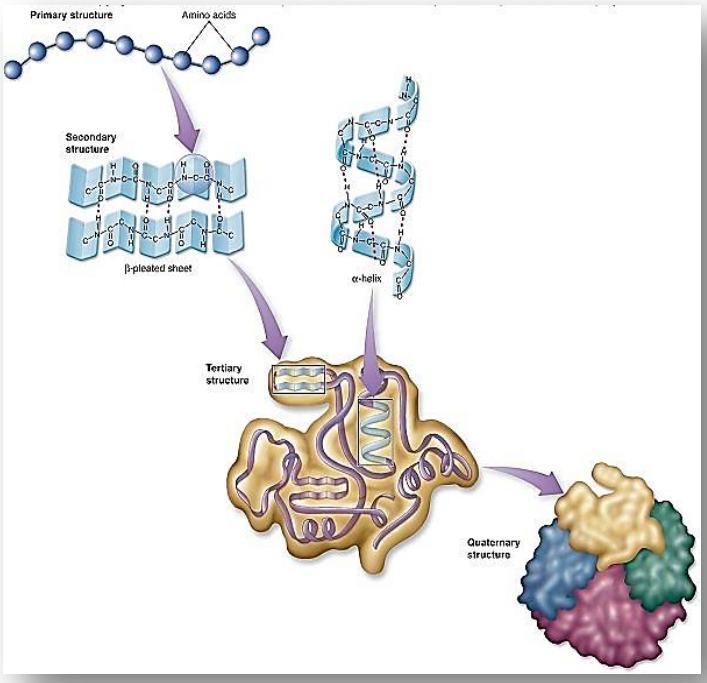



**Figure 1: Schematic organization levels of proteins [1]**

Understand and uncover the mechanism though such modifications and reorganization occur during the folding phenomena, is one of the challenges of science nowadays.

### *1.1 – Protein Folding*

The fulfillment of biological function requires the protein has correct three-dimensional structure. Proteins as explained before, fold from a random chain (usually called denatured or unfolded state) to a specific three-dimensional structure (native state), under suitable circumstances, like pH, temperature, solvent, via various kinds of folding intermediates or transitions states. Life requires in general that proteins achieve a reliable and functional fold, so the native fold of each protein has evolved to be more stable than all other possible folds. Since native fold are highly complex and irregular, their stability is one of the most impressive examples, where weak forces work together to originate an intricate and reproducible structures[*1*]. Nevertheless, a so complex mechanism can have failures, and some proteins are not able to remain with the correct fold or fail to fold correctly, leading to a wide variety of pathological conditions, such as Alzheimer's, cystic fibrosis, prion diseases and many neurodegenerative diseases[*2*].

For more than 50 years, scientists have become more and more interested into this unique field in biological investigation, birthed with two major observations during the 60's that led to two distinct views of the folding phenomena; First, Anfinsen demonstrated that fully denatured proteins require nothing more than dilution of denaturant to return to their active form. This observation identified each protein's unique amino acid sequence as sufficient for the determination of its structure, and this self-organization led to the formulation of one of the first hypothesis about protein folding; the thermodynamic Hypothesis, according to which the native state is the minimum of Gibbs energy in the folding path[*3, 4*]. Despite this fact, is still unclear how a simple primary sequence of amino acids can code for the three-dimensional structure and define the folding path.

The second observation was made by Cyrus Levinthal and coworkers [*5*], where they noted that a simple random search do not occur during the folding process, since it would take more than the age of the universe to exhaustively search for the native fold, with the observed fact that small proteins typically fold within microseconds to seconds in the time scale. This has become known as Levinthal's paradox, and brought up the idea that somehow a mechanism must guide an unfolded polypeptide to achieve the correct fold.

Understand in detail the mechanism of protein folding is a very important step and with huge importance in the field of protein science, but a conclusive answer wasn't yet achieved, due the fact that all proteins don't undergo the same unique mechanism, and in the last years several models have been proposed.

### *1.1.1 – Protein Folding Mechanisms*

Based on above two fundamental observations, several models have been proposed to elucidate the mechanisms of protein folding.

The hierarchic model of protein folding, also known as framework or diffusion-collision model, suggests that local secondary structural elements such as α-helix or β-sheet are the starting point in folding process, independently [6]. These structural elements would then diffuse and collide, ultimately coalescing into native tertiary structure. At the other end of the spectrum is the hydrophobic collapse model. This proposal asserts that folding is initiated with an extensive collapse of hydrophobic residues, followed by formation of secondary structure. The nucleation model took up residence somewhere in the middle and stated that partial native secondary structure would serve as a folding nucleus from which further structure would propagate [7].

All three of the proposed models implied the population of semi-folded structures (intermediates) along the pathway. Jackson and Fersht, however, reported that folding of chymotrypsin inhibitor 2 occurred in a two-state manner, with no detectable intermediates [8]; a mechanism termed nucleation-condensation.

Subsequent ϕ-value analysis, which probes the effects of amino acid substitution on the transition state stability, confirmed the two-state folding. Recently Fersht and coworkers [9, 10] have proposed a unifying mechanism, that consists of three basic steps: 1) formation of a nucleus with native-like topology, 2) polypeptide chain collapse, and 3) achievement of the native conformation through structural consolidation. Under the posited unifying mechanism proteins may appear to fold via the nucleation-condensation or framework models depending on the stability of their relative folding intermediates and the height of the corresponding transition barriers [11].

The current generally accepted view of protein folding extends Levinthal's pathway notion by incorporating the presence of many possible routes to the native state [12, 13]. Support for this view came from studies on ultrafast folding proteins, and the finding that the

folding rate is proportional to the number of folding routes [14]. The presence of many possible pathways leading to the native structure forms the basis of the folding funnel model [15, 16]. This model envisions folding as a conformational search on a rugged free energy landscape which is biased towards the native state.

The population of folding intermediates, or lack thereof, can then be explained by the degree of smoothness of the landscape. The energetic bias of the funnel explains, at least partially, how proteins can adopt complex structures on a biologically relevant timescale. Another factor that affects the folding rate is the amount of native-like structure present in the unfolded state. The denatured state ensemble (DSE) of a protein can be highly heterogeneous. However, many recent accounts suggest the retention of some residual structure [17, 18]. This would lighten the difficulties envisioned by Levinthal by limiting the conformations populated in the DSE to a small fraction of all possible conformations.

### *1.1.2 – Native Protein Structure*

Called the native fold, this structure is stabilized by chemical forces both within the protein and between the protein and its surrounding environment. These contain the hydrogen bonds, electrostatic interactions, disulfide bridges, and van der Waals forces.

The native structure of a protein is only marginally stable compared to the unfolded state. There is a large entropic penalty associated with the transition from a disordered chain to one of highly defined structure, as well as an enthalpic penalty for disrupting many interactions with the solvent. The protein must overcome these unfavorable contributions through intramolecular contacts. Some important examples are discussed below:

Hydrophobic interactions: Globular proteins exist in an aqueous environment however; many amino acid side chains are hydrophobic. The clustering of residues in the protein interior results in a stabilizing contribution that is proportional to the buried side chain surface area [19]. This leads to the formation of a hydrophobic core and leaves the hydrophilic residues mostly on the surface where they interact with the solvent. Recent experiments have demonstrated that for small-to-medium sized proteins, the hydrophobic effect contributes approximately 60% to the overall stability [20].

Hydrogen bonding: Hydrogen bonds require both a donor group and an acceptor group. Donors have a hydrogen atom bound to an electronegative heteroatom, eg. -NH, -OH, - SH, and acceptors have lone electron pairs, eg. C=O. The peptide backbone, as well as side chains, of an unfolded protein makes numerous hydrogen bonds with the solvent. Upon folding, loss of those contacts must be compensated for with intramolecular hydrogen bonds, most of which are found in α-helices and β-sheets. It has been estimated that the average hydrogen bond stabilizes a protein by ~2-5 kJ $mol^{-1}$ [21, 22].

Salt bridges: The positive charge on the side chains of Arg and Lys, as well as the negative charge on those of Asp and Glu, often result in their localization on the protein surface. Electrostatic pairing of these side chains is stabilizing. Occasionally such salt bridges can also occur within the protein core. Charge-charge interactions on the surface are weaker due to solvent screening. While perhaps not as important as other intramolecular forces, it has been shown that optimizing the charged interactions on the protein surface can increase stability over the wild-type molecule [23].

Cation-π interactions: While the positively-charged side chains of Arg and Lys can be involved in salt bridges, they are also often found interacting with the π-electrons of the aromatic amino acids Phe, Tyr, and Trp. A survey of the structures in the Protein Data Bank revealed ~25% of all Trp residues involved in a significant cation-π interaction[24]. A bias for Arg-Tyr interactions has been shown to exist at protein-protein interfaces, with the stabilizing electrostatic energy estimated to be around 10 kJ/mol[25]. A recent study found a specific cation-π interaction to be vital for the regulation of integrin affinity and function [26]. Similar interactions occur between aromatic residues. The hydrogen on the aromatic rings is slightly positive. This allows the aromatic side chains to interact favorably in an edge-to-face orientation, with the hydrogen of one in close proximity to the π-electrons of another. Face-to-face side chain stacking, akin to base stacking in nucleic acids, is also a possibility albeit with a slight offset to avoid electrostatic clash between the electron clouds.

Disulfide bonds: The thiol side chains of cysteine residues can be oxidized to form a disulfide (S-S) bond. These bonds can be intrachain as in lysozyme, or interchain as for insulin. Disulfides serve to stabilize the native state by reducing the conformational entropy of the

unfolded state. Also, residue interactions required in the native state can be promoted by disulfide linkages [27].

### *1.1.3 – Transition State Ensemble*

The transition state is an unstable state, never significantly populated and hardly obtained from direct measurements, due the fact that the transition state ensemble (TSE) is always associated with a high energy point on the folding pathway (**Figure 2**). Hence any characterization of structural and energetic properties must be performed indirectly, through the determination of rate constants and the ways in which these rates are altered by specific perturbations. [28]

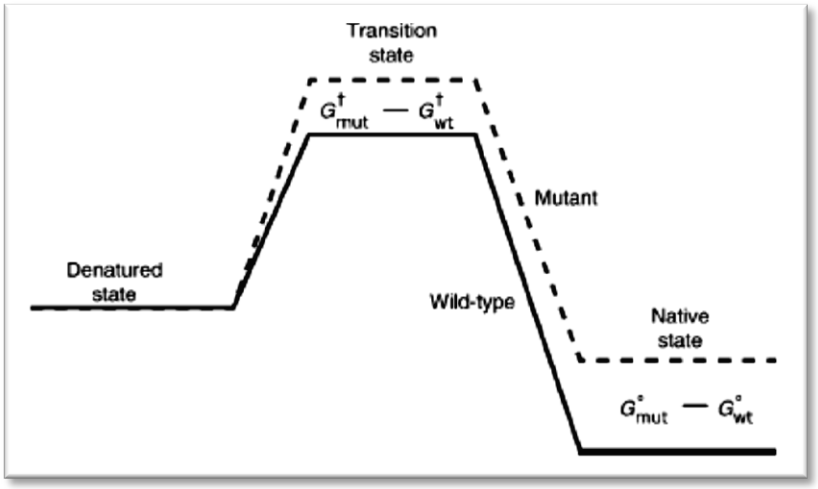



**Figure2: Schematic representation of the energy diagram for the protein folding process.**

The way to measure such rates is by using a technique developed by Fersht and coworkers, named φ-analysis[29]. This technique quantifies the energy of the TSE in response to mutations in the primary protein structure, comparative to that of the folded state.[30] This method provides means to identify interactions by specific amino acid side chains that stabilize the folding transition state [31]. This approach has been applied in a large number of proteins, such as Chymotrypsin inhibitor, myoglobin, SH3 domain and colicin immunity proteins – Im7 and Im9. In order to probe the energy landscape, the addition of chemicals such as denaturants and/or through changes in temperature so that the enthalpy, entropy, and heat capacity differences among native state, transition state and unfolded or denatured state are obtained.

In a simple and brief description, φ-analysis is described by Fersht and coworkers [31], as being the ratio between the difference in activation free energy change for unfolding ($\Delta\Delta G_{f\text{-}\ddagger}$)

upon mutation and the difference in equilibrium free energy change upon mutation ($\Delta\Delta G_{f-u}$) (Equation 1).

**Equation 1:** $$\Phi = (\Delta\Delta G_{f-*}) / (\Delta\Delta G_{f-u})$$

### *1.1.4 – Folding Intermediates*

Determining the mechanism by which a protein attains its unique native state requires the structural characterization of the entire folding pathway. As mentioned above, ϕ-value analysis has been useful for characterizing transition states. However, it is critical to also assess the partially folded states that become populated in the pathway to the native state [32]. The formation of intermediates for multidomain proteins is well established [33] but the case for small proteins remains unclear [34].

Many intermediates are formed via productive folding events, off-pathway species can become populated, and parallel routes are available to reach the native state [35].

Protein folding transitions are highly cooperative, implying that intermediate species cannot be detected. Some proteins establish populate equilibrium intermediates that are similar to their kinetic counterparts [36], but this does not appear to be universal. It is postulated that intermediates are always present [*37*]. However, due to their transient nature time-resolved experiments are required for their characterization.

Detection of folding intermediates has become a fairly common practice, but their detailed structural characterization continues to present a challenge.

### *1.1.5 – Denatured state ensemble*

Unfolded proteins are disordered under physiological conditions, being its characterization crucial for understanding protein stability, the mechanism of protein folding and gain insights of the early beginning of protein folding [*38*]. Although the unfolded state or Denatured state ensemble (DSE) were traditionally assumed to have a randomly fluctuating structure, mounting evidence suggests that it can contain significantly more well-defined structure, which can have either native or non-native character, with implications for protein stability, folding mechanism and rate [*39*]. The characterization of unfolded state is hampered by

the low solubility, dynamic nature and the diversity of the unfolded ensembles as compared to the native state.

Regarding the characterization these states, chemical and thermal denaturation are the primary methodologies used to reach the unfolded state in solution. Guanidine hydrochloride (GuHCl) and urea are chemical agents widely used to perform protein denaturation, once it can have profound effects on protein stability and structure.

#### *1.1.5.1 – Denaturation of Proteins by Urea*

The ability of urea to denature proteins has been known since 1900. Tanford was the first to study quantitatively the unfolding of proteins by urea and showed that when a protein unfolds many nonpolar side chains and peptide groups that were buried in the folded protein are exposed to solvent in the unfolded state [*40*]. Since then, several works regarding the mechanism and the role of urea in the denaturation process have been performed. Despite the interest and effort to understand the role of urea, the molecular basis for its ability to denature proteins remains unknown. Experimental and theoretical studies have improved our understanding of urea solutions and how urea and water interact with peptide groups and nonpolar side chains. It has been proposed that urea may interact directly, by binding to the protein, or indirectly, by altering the solvent environment [*41*].

### *1.2 – E colicin Immunity Proteins*

Colicins are proteins produced by some strains of *Escherichia coli,* as part of the SOS stress response to nutrient-limited environments. These proteins are plasmid encoded antibacterial proteins that kill sensitive strains, by binding to specific outer membrane receptors of target cell, controlling and manipulating the entry of specific nutrients such as ionophores, vitamins and nucleosides.

Until now nine E-group colicins (E1-9) have been reported and characterized, exhibiting three domains, each corresponding to one step of colicin action. The N-terminal domain is involved in translocation through the membrane; the central domain is involved in binding to the target cell receptor and the C-terminal cytotoxic domain. The nine E-group colicins do not exhibit the same cytotoxicity mechanism, and can be divided in three different subgroups; pore-forming colicins, endonucleases (DNases) and ribonucleases (RNases).

In order to overcame, the survival problem of expressing a colicin with nuclease activity, *E. coli* expresses a specific antagonist protein called the immunity protein, which will bind to the colicin nuclease domain, protecting against the cytotoxicity activity. Each of the known E-group colicins (E1-E9) has it cognate inhibitor (Im1-Im9).

### *1.2.1 – Im7 and Im9 proteins*

As pointed before, one of the most important questions to make regarding protein folding is how the sequence of a protein will determine the structural and thermodynamic characteristics of the native state and what is the mechanism underlying the folding phenomena? Answering those questions is not a straightforward process, and several authors had tried to address them by studying homologous proteins, among them two proteins of the family of DNase E-colicin immunity proteins, Im7 and Im9 were characterized in detail.

These two proteins have the same four helical fold (**Figure 3 A**) and exhibits 60% sequence homology, as can be observed in **Figure 3 B** that shows the sequence alignment of Im7 and Im9 determined using the Basic Local Alignment Search Tool (BLAST) [*42*].

Despite their similarity in structure and sequence, however, Im7 and Im9 fold with different kinetic mechanism *in vitro;* Im7 folding with a three-state mechanism, in which an on-pathway intermediate is populated early during folding, whilst Im9 folds with a two-state transition without populating intermediate.

The folding problem is currently addressed from two perspectives. The first, thermodynamic-centric, line of research is mainly concerned with the stability and uniqueness of the native fold and how it is encoded by the protein sequence. The second, kinetics-centric, line of research is concerned with the mechanisms and time scales for folding. The aim of this Thesis is the elucidation of the early events of folding phenomena in small class of immunity proteins (Im7 and Im9) using NMR spectroscopy as a tool.

Characterize unfolded ensembles in aqueous solution is the starting point for understanding the folding process, as well as to uncover the way in which amino acid sequence encodes for a specific structure and provide insights about folding/unfolding transitions that many proteins suffer as part of their function. This work will be important for the understanding,

in atomic detail, of the beginning of Im's folding and the disclosure of the full folding landscape. Another aspect that would be clarified would be the role of urea during the unfolding events.

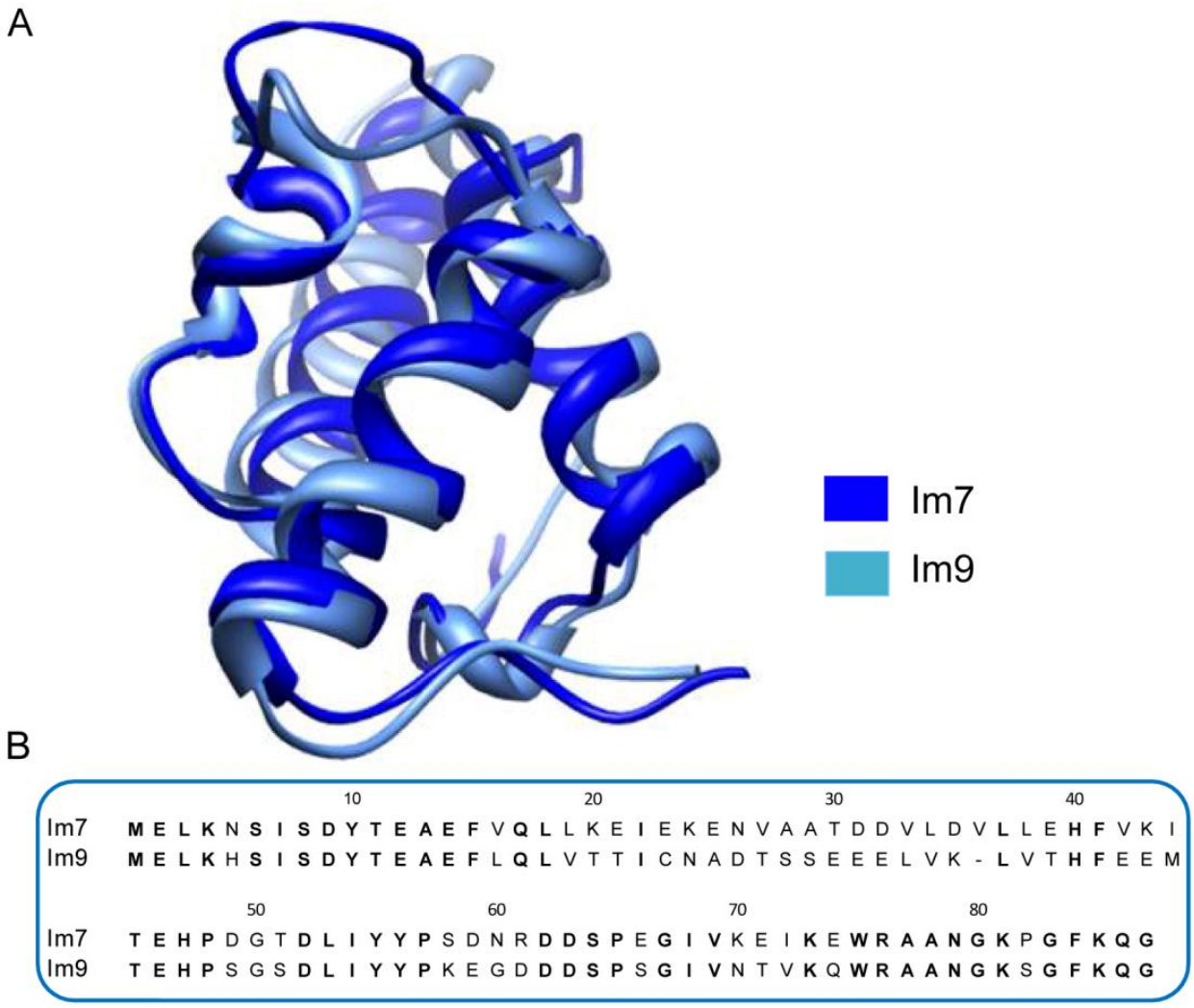



**Figure 3 – Comparison of Im7 and Im9 protein structures and sequences.** A) Ribbon diagram of the homologous E type immunity proteins Im7 (pdb code, 1AYI) and Im9 (pdb code, 1IMQ). B) Alignment of Im7 and Im9 amino acid sequences. The conserved residues in the proteins are in bold. The sequence identity is 60%.

# 2 – Material and Methods

The technique that I have mainly used during my studies was NMR spectroscopy applied to the study of urea and temperature denatured states of colicin immunity proteins. NMR spectroscopy relies in good sample preparation, thus high yields of proteins with high degree of purity are essential to perform the technique.

## 2.1 – Molecular Biology

All proteins (Im7 and Im9) were previously produced and were available at the beginning of this Thesis. These proteins were over-expressed using plasmid pJR347, based on the expression vector pTrc99A and were kindly provided by Dr. Angelo Figueiredo (FCT-UNL) and Prof. G.R Moore (Univ. of East Anglia, UK). The system and strategy that they have used relies on *E.coli* expression system in minimal media.[*43*] The use of minimal media for protein overexpression is essential for isotopic-labeling of proteins, in which a source of $^{15}N$ ammonium is provided to the bacterial strain in order to, overproduced isotopic label protein. Since, this step was previously prepared I will describe in detail the purification step of Im9 protein.

### 2.1.1 – Purification of Im9

In this work, Im9 protein was purified according to the protocol published in the literature. [*44*]. Overproduced Im9 protein was purified from 5L culture of *E. coli* cells, where the cells were harvested by centrifugation and resuspended in a 40ml solution (20 mM Tris/HCl, 5mM Imidazole and 500mM Sodium Chloride, pH 7.5), containing the protease inhibitor phenyl-methylsulphonyl fluoride (PMSF) at 1mM. The cells were broken using a French-press (French® Pressure Cell Press, Thermo scientific) operating at 10000-12000 psi.

Subsequent, cell debris was removed by centrifugation, using an Avanti J-26XPI Centrifuge (Beckman Coulter) at 18500 rpm, during 45min at 4 ºC, with a JA-10 fixed angle rotor.

The Im9 protein was precipitated by Ammonium Sulphate, and then dialyzed using a Snakeskin 7,000 MWCO dialysis tubing against a solution containing 50mM Tris, 1mM EDTA and 1mM 2-mercaptoethanol at pH 7.5, 4ºC overnight.

#### *2.1.1.1 – Anion-exchange Chromatography*

A 5ml HiTrap Q Sepharose HP column (Amersham Biosciences) coupled to an AKTA chromatographic system was used in the anion exchange chromatography step. The column was equilibrated with 5 column volumes of 50mM Tris-HCl, pH 7.5 (1mM EDTA, 1mM DTT, 50mM KCl) and the protein eluted with 100ml of elution buffer: 50mM Tris-HCl pH 7.5, 500mM KCl, 1mM DTT, 1mM EDTA. A flow rate of 5 ml/min was applied during all chromatographic process; 2.5 ml fractions size samples were collected.

The fractions containing the desired protein were pooled together and concentrated using a Stirred Cell 8050 from Millipore with a 3,000 MWCO porous membrane. Concentrated proteins were dialyzed overnight (using a Snakeskin 7,000 MWCO dialysis tubing) against a buffer solution containing 50mM Tris, 1mM EDTA and 1mM 2-mercaptoethanol at pH 7.5, 4ºC overnight.

The presence and purity of the protein was confirmed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Samples of 20µL of each collected fraction were boiled at 100 ºC with 10µL of 5x sample buffer for 5 min before loading 20µL of each into the gel. The gel was stained with Coomassie brilliant blue for 15 min and then distained with a mixture of 10% methanol/10% acetic acid in water

#### *2.1.1.2 – Gel Filtration*

A Superdex 75 gel filtration column with a bed volume of 300 ml was used for the final step of Im9 purification. The column was equilibrated with 1 liter of 50mM phosphate buffer pH 7.0, 200mM KCl. The sample loop (5ml) was partial filled with protein, i.e., 2.5ml of Im9 for high recovery. The sample volume loaded should be at maximum 50% of the loop total volume. The flow rate was setup to 5 ml/min.

A 3ml proteins fraction were collected and identified by SDS-PAGE gel. All purified protein was dialyzed using a Snakeskin 7,000 MWCO dialysis tubing membrane overnight, against the same buffer used in section 3.1.1.2. After dialyze, protein was concentrated in a centrifuge 5804R with fixed angle (Eppendorf), using an Amicon Ultra-15 Centrifugal Filter Unit with a 10 MWCO membrane (Millipore). The final step of the purification protocol was to lyophilize the pure protein, and store at -10 ºC.

#### *2.1.1.3 – Mass Spectroscopy*

Mass spectroscopy was used to confirm the purity and identity of Im9. Time-of-Flight mass spectroscopy (TOF-MS) was performed and indicated that the final sample was highly pure, but evidences of dimerization was also found.

The molecular mass determined was 9691.33 Da for the $^{15}N$ labeled sample. This value is in good agreement with the theoretical molecular weight, 9689.71, value calculated from the amino acid sequence, plus the isotope effect of $^{15}N$. The theoretical values were extracted from the Protein Calculator website: http://www.scripps.edu/~cdputnam/protcalc.html, accessed February 22, 2012.

### *2.2 – NMR Spectroscopy studies*

All NMR experiments were performed on a Bruker Avance II+ 600 MHz spectrometer at the Chemistry Department of Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, part of the National NMR Network (PTNMR), equipped with a triple resonance probehead cryoprobe.

All spectra acquired, $^{1}H$ chemical shifts were referenced directly to external DSS, and the $^{15}N$ chemical shifts indirectly using the protocol described by Wishart and coworkers [*45*].

NMR data, were processed using NMRPipe [*46*] and analyzed using CCPNMR [*47*].

In all cases, data was processed with a skewed cosine-bell apodization function and zero-filling (one time) was used in all dimensions.

#### *2.2.1 – Urea titration*

Protein NMR samples containing 0.4mM protein were dissolved in 50mM potassium phosphate buffer 90% $H_2O$/10% $^{2}H_2O$ at pH 7. For equilibrium urea-unfolding experiments, a 8M urea stock solution were always freshly prepared in phosphate buffer and the urea final concentration determined by a refractometer, using a Urea concentration calculator available free on internet (http://sosnick.uchicago.edu/gdmcl.html) [*48, 49*]. Starting from the stock solution, a set of different urea concentrations solutions were prepared: 0.5M, 1M, 2M, 3M, 4M, 5M, 6M and 7M, and confirmed by a refractometer.

Urea-protein samples were then allowed to reach the equilibrium and the full dissolution after 1 hour before NMR acquisition. NMR experiments were acquired at 10 ºC at different urea concentrations, ranging from 0 to 6M for Im7 wild-type protein, and from 0 to 8M urea for Im9 wild-type protein.

$^{1}H$-$^{15}N$-HSQC spectra were acquired using a fast HSQC pulse sequence, where the great advantage, comparing to the traditional HSQC, is the absence of additional radio frequency (RF) pulses to avoid water saturation, which leads to a more rapid optimization of the pulse sequence. [*50*]

The $^{1}H$-$^{15}N$ HSQC spectra were recorded with the following parameters: 2048 ($^{1}H$) and 256 ($^{15}N$) complex points with 32 scans for each fid for Im7 and 16 scans for Im9. For Im7, the spectral widths were 7690 Hz for $^{1}H$ and 1680 Hz for $^{15}N$, where the offset carrier frequencies were set to 2783.2 Hz ($^{1}H$) and 7175.66 Hz ($^{15}N$). For Im9, the spectral widths were 8013 Hz for $^{1}H$ and 2000 Hz for $^{15}N$, centered at 2802 Hz ($^{1}H$) and 7175.66 Hz ($^{15}N$).

### *2.2.2 – Temperature Denaturation*

Protein NMR samples contained 0.4mM protein were dissolved in 50mM sodium phosphate buffer and 90% $H_2O$/10% $^{2}H_2O$ at pH 7. All samples were equilibrated during 15min at the experiment temperature before NMR acquisition. As described in section 2.2.1, a fast HSQC pulse sequence was used for all spectra.

NMR experiments were acquired at different temperatures, ranging from 283K to 323K for Im7 and Im9 wild-type proteins. The $^{1}H$-$^{15}N$ HSQC spectra were recorded with the following parameters: 2048 ($^{1}H$) and 256 ($^{15}N$) complex points with 16 scans for each fid for Im7 and Im9. For Im7, the spectral widths were 7944.9 Hz for $^{1}H$ and 1659.4 Hz for $^{15}N$, where the offset carrier frequencies were set to 2788.4 Hz ($^{1}H$) and 7175.7Hz ($^{15}N$). For Im9, the spectral widths were 7692.3 Hz for $^{1}H$ and 1680.0 Hz for $^{15}N$, centered at 2807.9 Hz ($^{1}H$) and 7175.66 Hz ($^{15}N$).

### *2.2.3 – Combined Chemical Shift*

For the evaluation of the behavior of individual amino acids upon addition of increasing amounts of urea, and/or increased temperature was calculated the combined amide proton and nitrogen chemical shift differences using the following equation [*51*]:

**Equation 2:** $$\Delta\delta comb = \sqrt{(\Delta\delta\mathrm{H})^2 + (\mathrm{wi}\Delta\delta\mathrm{N})^2}$$

where $\Delta\delta$H and $\Delta\delta$N are the chemical shifts of proton and nitrogen, respectively and wi is a weighting factor which accounts for differences in sensitivity of different resonances in an amino acid (e.g. amide $^{1}$H and $^{15}$N). When chemical shifts are expressed in ppm a suitable estimate for the weighting factors is given by [*51*]:

**Equation 3:** $$wi = \frac{|\gamma \mathrm{i}|}{|\gamma \mathrm{H}|}$$

with γi and γH the magnetogyric ratio of nucleus i and the proton, respectively.

In order to calculate the combined chemical shift, the first step is to perform the pick picking of the $^{1}$H-$^{15}$N-HSQC spectrum. After all resonances picked, the next step is to assign each resonance to a specific residue. In this work, a file with the chemical shifts already deposited in BMRB was used. The file code for protein Im7 and Im9 was 7188 and 4115 respectively.

In order to decide whether a given residue belongs to the class of interacting or non-interacting residues was calculated a cutoff value. In a first approximation, the chemical shift distributions of the non-interacting residues can be assumed to follow a normal distribution with a mean of zero. Therefore, the standard deviation to zero, $\sigma_o$, for the class of non-interacting residues is a reasonable measure to predict if a residue belongs to the class of interacting residues or not . Nevertheless, if the values of all residues are used, the obtained result will be strongly biased by the large chemical shift changes of the interacting residues. Therefore, it was used an iterative procedure that successively removes outliers to calculate a corrected standard deviation to zero $\sigma_0 corr$ that is used in the following as cutoff value.

# 3 – Results

## 3.1 – Purification and biochemical characterization of Im7 and Im9

Im7 was previously produced and purified, and was available at the beginning of this Thesis. In the first stage of this work, Im9 was purified by two chromatographic methods: Anion-Exchange chromatography and Gel Filtration.

The protein Im9 exhibits a theoretical isoelectric point of 4.53, calculated using the bioinformatics tool ProtParam from Expasy server [*52*]. After this purification step the majority of the contaminants were removed. Despite highly efficient, this chromatographic step needs to be complemented with a Gel Filtration chromatographic step. The molecular weight of Im9 is 9.7 kDa (theoretical value determined with Compute pI/Mw tool [*53*]). After each chromatographic method the purity of this protein was evaluated by SDS-PAGE (**Figures 4 and 5**) and one intense band was obtained in the expected MW region (~11kDa), after the final purification step (**Figure 5**).

The purity of Im9 was further confirmed by MALDI-TOF-MS (**Figure A1, Appendix A**). The signal of the pure protein in the mass spectrum corresponds to a molecular mass of 9691.33 Da, which is in excellent agreement with the predicted MW for Im9 [*54*]. Protein yield were determined by UV-visible spectroscopy using the extinction absorption coefficient characteristic of the proteins ($\varepsilon_{280nm}$ = 9530 $M^{-1}cm^{-1}$).The yield obtained for Im9 was 1.12 mg/mL.

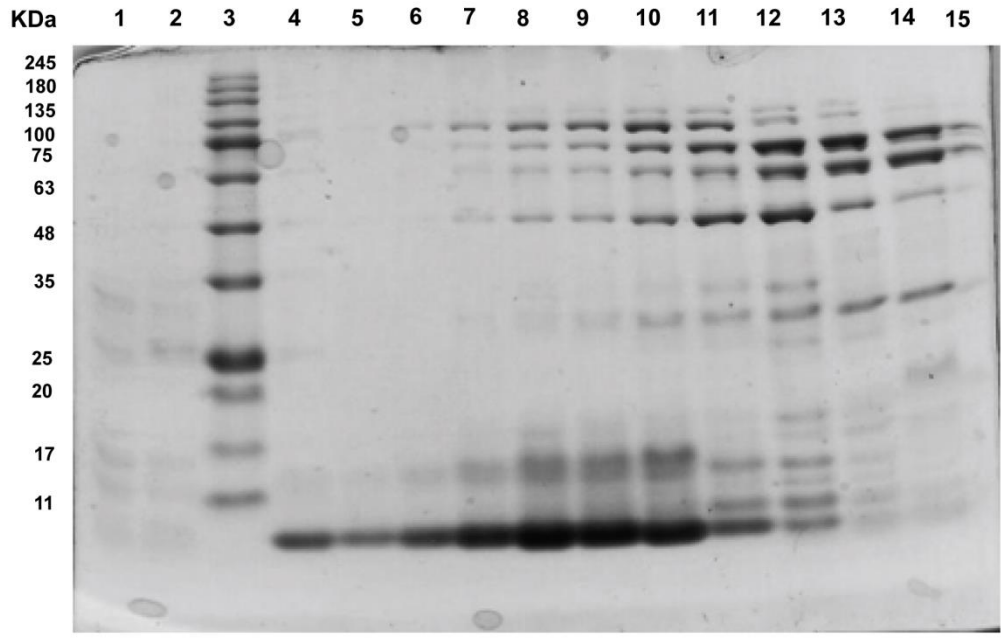



**Figure 4:** SDS-PAGE 15% gel of Im9 fractions after Anion-Exchange Chromatography. Lane 3 – MW markers; Lane 1-2 and 4-15: fractions obtained after Anion-Exchange Chromatography.

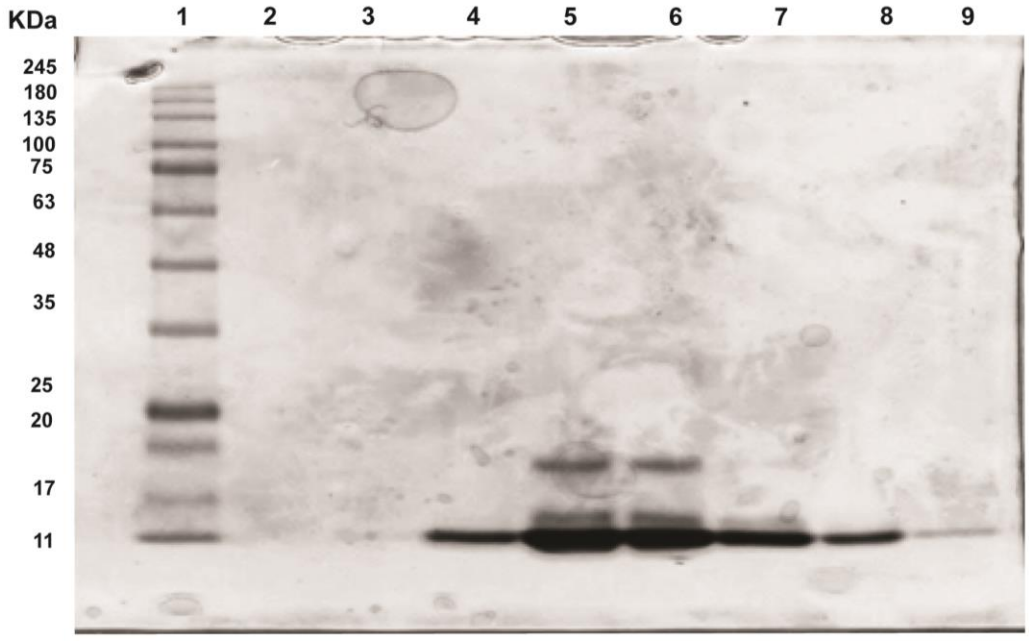



**Figure 5:** SDS-PAGE 15% gel of Im9 fractions after Gel Filtration. Lane 1 – MW markers; Lanes 2-9 – Fractions obtained after Gel Filtration.

## *3.2 – NMR experiments*

### 3.2.1 – Protein denaturation process

A series of NMR experiments were acquired to evaluate the residues involved in the Im7 and Im9 unfolding process induced by increasing amounts of urea or increasing temperature. In order to achieve this, $^{1}$H-$^{15}$N HSQC spectra for the Im7 and Im9 protein were acquired and compared to the corresponding spectra described in the literature, as a starting point for the following approaches:

- The chemical shift files deposited in BMRB for Im7 and Im9, 7188 and 4116 respectively, were used to perform the assignment.

- The protein denaturation process was evaluated by probing the chemical shift perturbation of the backbone signals of Im7 and Im9, either:

a) with increasing temperature, or

b) in presence of increasing amounts of urea.

The NH backbone chemical shift perturbations were monitored by recording a series of $^{1}$H-$^{15}$N HSQC NMR spectra (**Figures 6** and **9**, **Appendix A2** and **A3**).

After the assignment of the NH backbone signals, a table with $^{1}$H and $^{15}$N resonances of each residue is created for each set of experiments, as exemplified in **Table 1** for condition b. (all data referent to the chemical shifts are presented in **Appendix B1** and **B2).**

**Table 1: Example of the tables created and the results treatment preformed.**

| | 0M Urea | | 0.5M Urea | | 0.5 - 0M Urea |
|---|---|---|---|---|---|
| | **$\delta$ $^{1}$H (ppm)** | **$\delta$ $^{15}$N (ppm)** | **$\delta$ $^{1}$H (ppm)** | **$\delta$ $^{15}$N (ppm)** | **$\Delta\delta_{comb}$** |
| **7 Ile** | 9.38 | 125.87 | 9.40 | 125.88 | 0.016 |
| **12 Glu** | 9.18 | 125.87 | 9.20 | 122.00 | 0.012 |
| **18 Leu** | 7.59 | 122.81 | 7.60 | 122.80 | 0.010 |
| **28 Ala** | 7.49 | 124.74 | 7.53 | 124.83 | 0.049 |

In order to better represent the distribution of perturbed and not perturbed residues the combined chemical shift perturbation, $\Delta\delta_{comb}$, were calculated and a cut-off line determined, as explained in section 2.2.3, for each step of urea or temperature gradient. The results are summarized below for the two conditions under study.

**a) Protein denaturation with increasing temperature**

The majority of resonances from the native conformation assigned for both proteins in the thermal denaturation process are represented in Table 2. (All data referent to the chemical shifts are presented in **Appendix B3** and **B4)**

**Table 2: Number of residues assigned during thermal denaturation.**

| | Im7 | Im9 |
|---|---|---|
| **Temperature (K)** | | |
| **283** | 79 | 80 |
| **288** | 76 | 80 |
| **293** | 73 | 80 |
| **298** | 73 | 80 |
| **303** | 71 | 79 |
| **308** | 70 | 79 |
| **313** | 63 | 79 |
| **318** | - | 78 |
| **323** | - | 77 |
| **328** | - | - |

Figure 6 and Appendix A2 show $^{1}H$-$^{15}N$ HSQC spectra of Im7 and Im9 respectively at various temperatures. With increasing temperature, the signals were distributed in a narrower region in the spectrum, leading to a loss of tertiary structure at 318K for Im7 and 328K for Im9.

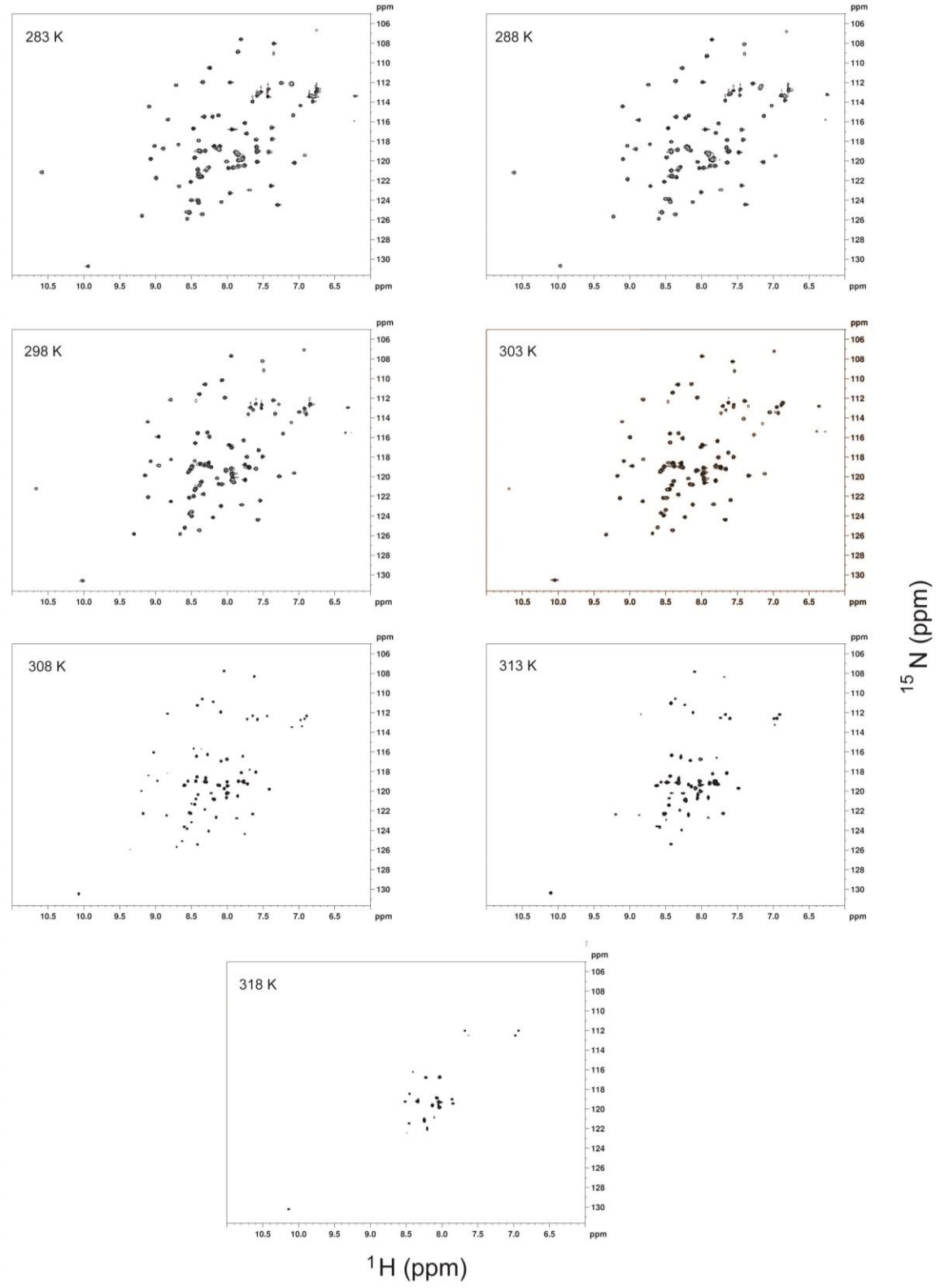



**Figure 6: $^{1}$H-$^{15}$N HSQC spectra of Im7 at different temperatures.** The temperatures used in each HSQC of are indicated inside each panel.

The perturbation induced by temperature clearly shows that most changes occur for residues Glu12; Gln17; Leu18; Glu21; Glu23; Lys24; Ala28; Thr30; Asp35; Val36; Thr45; Glu46; His47; Gly50; Leu53; Asp59; Asn60; Arg61; Asp62-63; Ser64; Gly67; Val69; Ala77-78; Asn79 and Gly80 for protein Im7 (**Figure 7 A** and **B**).

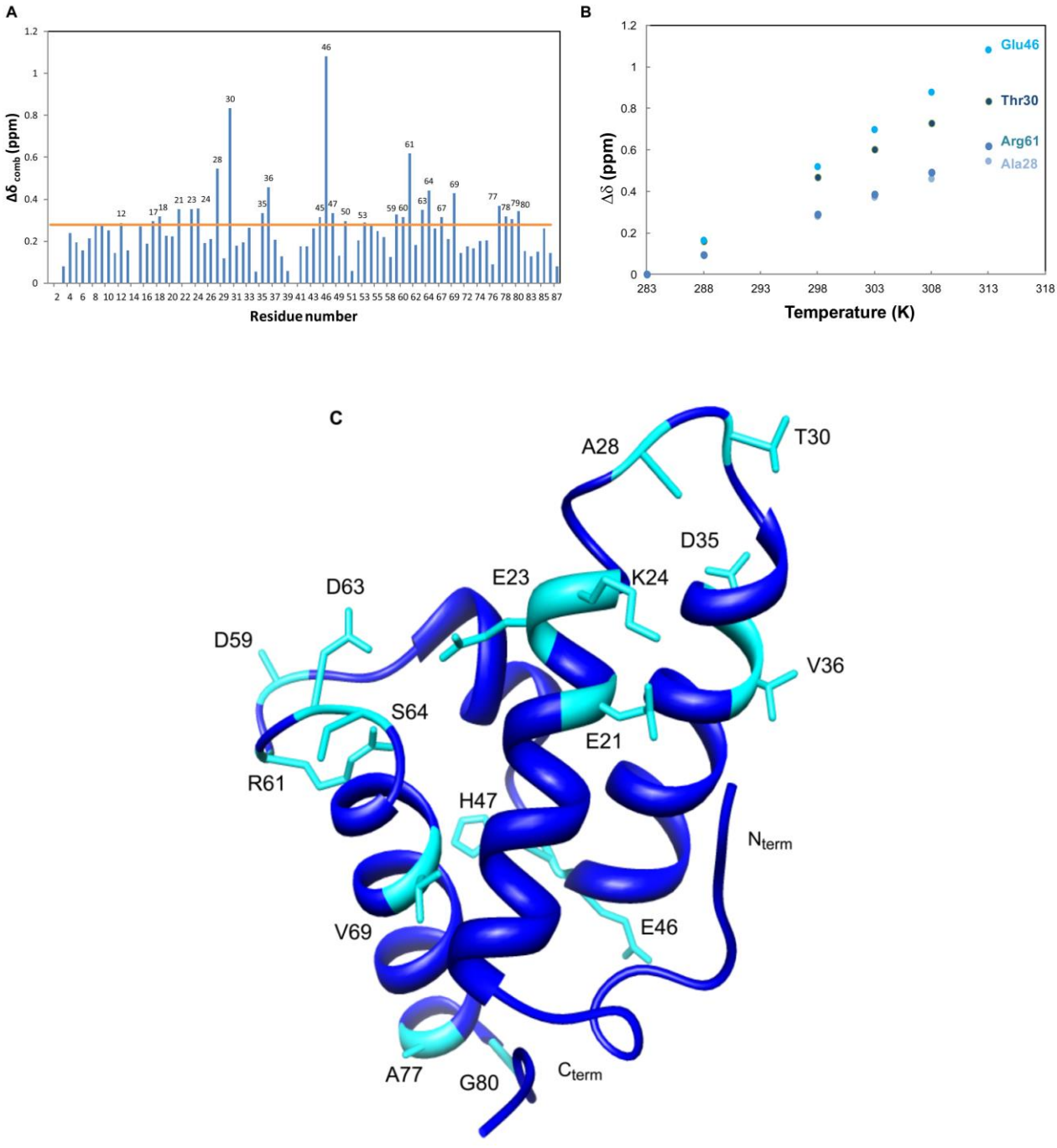



**Figure 7:** *A*) Im7 backbone amide chemical shift variations between 283K and 313K; *B*) Evolution of Δδ with temperature for the most affected residues; *C*) Structure of Im7 (PDB: 1AYI), which highlights the most affected residues.

For protein Im9 the same approach was taken, and the major perturbations occurred for Ser8; Ala13; Leu18; Val19; Ile22; Thr27; Ser29; Leu33; Leu36; His39; Glu42; Glu45;His46; Gly49; Leu52; Ile53; Gly59; Asp61; Asp62; Ser63; Gly66; Trp74; Ala76; Ala77; Gly79; Lys80 and Ser81 (**Figure 8 A** and **B**).

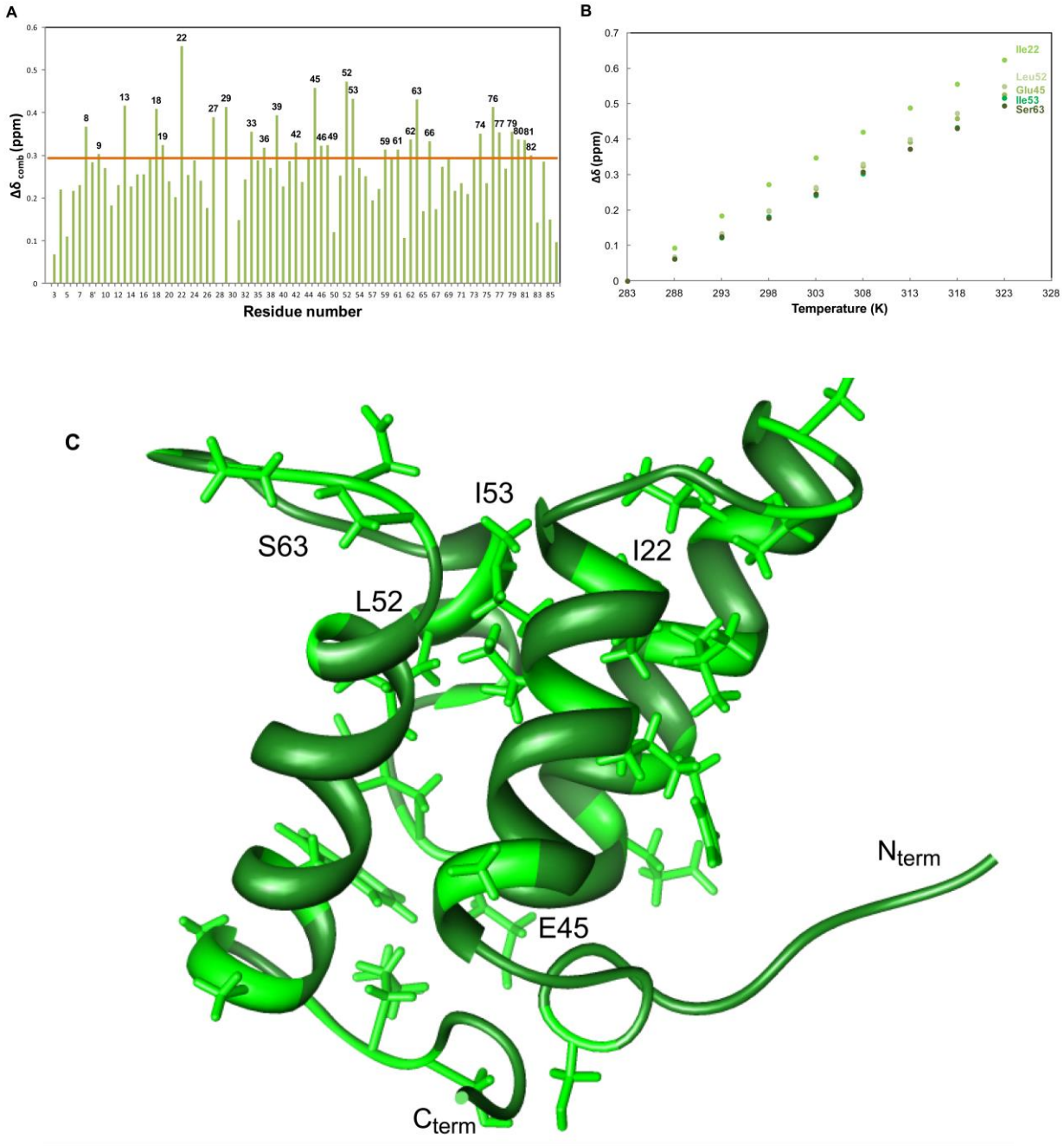



**Figure 8:** *A*) Im9 backbone amide chemical shift variations between 283K and 323K; *B*) Evolution of Δδ with temperature for the most affected residues; *C*) Structure of Im9 (PDB: 1IMQ), which highlights the most affected residues.

**b) Protein denaturation in presence of increasing amounts of urea.**

To check whether the proteins reach similar states at a given concentration of urea or if the denaturation pathway is identical, $^{1}H$-$^{15}N$ HSQC spectra were acquired with samples prepared starting from a folded protein, as well as with those prepared starting from protein denatured in 6 M and 8M of urea (Figure 9). In the absence of urea, the spectrum is typical of a well-folded protein with good dispersion of peaks. At 3 - 4 M urea, the HSQC spectra have signatures of both folded species (well-dispersed peaks) as well as of partially unfolded species. From 5 M to 6 M urea, the spectra are indicative of largely unfolded protein, with all the peaks appearing in a very narrow amide proton chemical shift range.

The majority of resonances from the native conformation assigned for both proteins in the chemical denaturation process are represented in Table 3.

**Table 3: Number of residues assigned during urea denaturation.**

| | Im7 | Im9 |
|---|---|---|
| **[Urea (M)]** | | |
| **0.0** | 80 | 77 |
| **0.5** | 80 | 77 |
| **1.0** | 80 | 77 |
| **2.0** | 79 | 70 |
| **3.0** | 79 | 70 |
| **4.0** | - | 68 |
| **5.0** | - | 68 |
| **6.0** | - | - |
| **7.0** | - | - |

Figure 9 and Appendix A3 show $^{1}H$-$^{15}N$ HSQC spectra of Im7 and Im9 respectively at different concentration of urea. In the absence of urea, $^{1}H$-$^{15}N$ HSQC spectrum presents a good signal dispersion which indicates a three dimensional structure in solution (folded). On the other hand, in the presence of increasing amounts of urea is observed a loss of tertiary structure followed by a narrower distribution of the signals in the $^{1}H$-$^{15}N$ HSQC spectra.

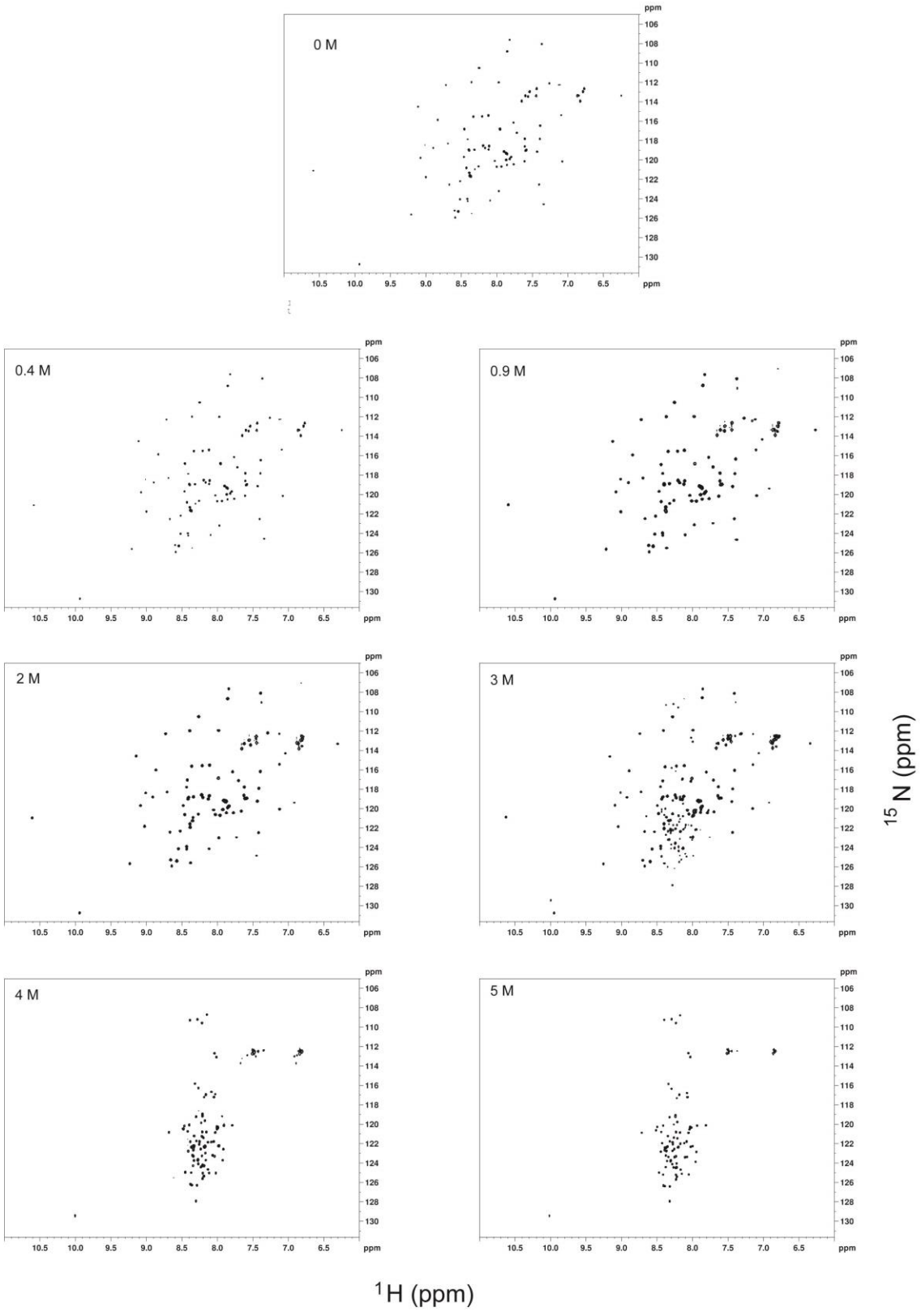



**Figure 9: $^{1}$H-$^{15}$N HSQC spectra of Im7 at different concentrations of urea.** The urea concentrations used in each HSQC of are indicated inside each panel.

The perturbation induced by urea shows that most changes occur for residues Glu2; Asn5; Ser8; Glu21; Asn26; Ala28; Ala29; Thr30; Asp32; Val33; Val36; Glu39; Thr51; Asp52; Ser58; Asp59; Arg61; Asp62; Glu66; Ala77 and Asn79 for Im7 (**Figure 10 A** and **B**).

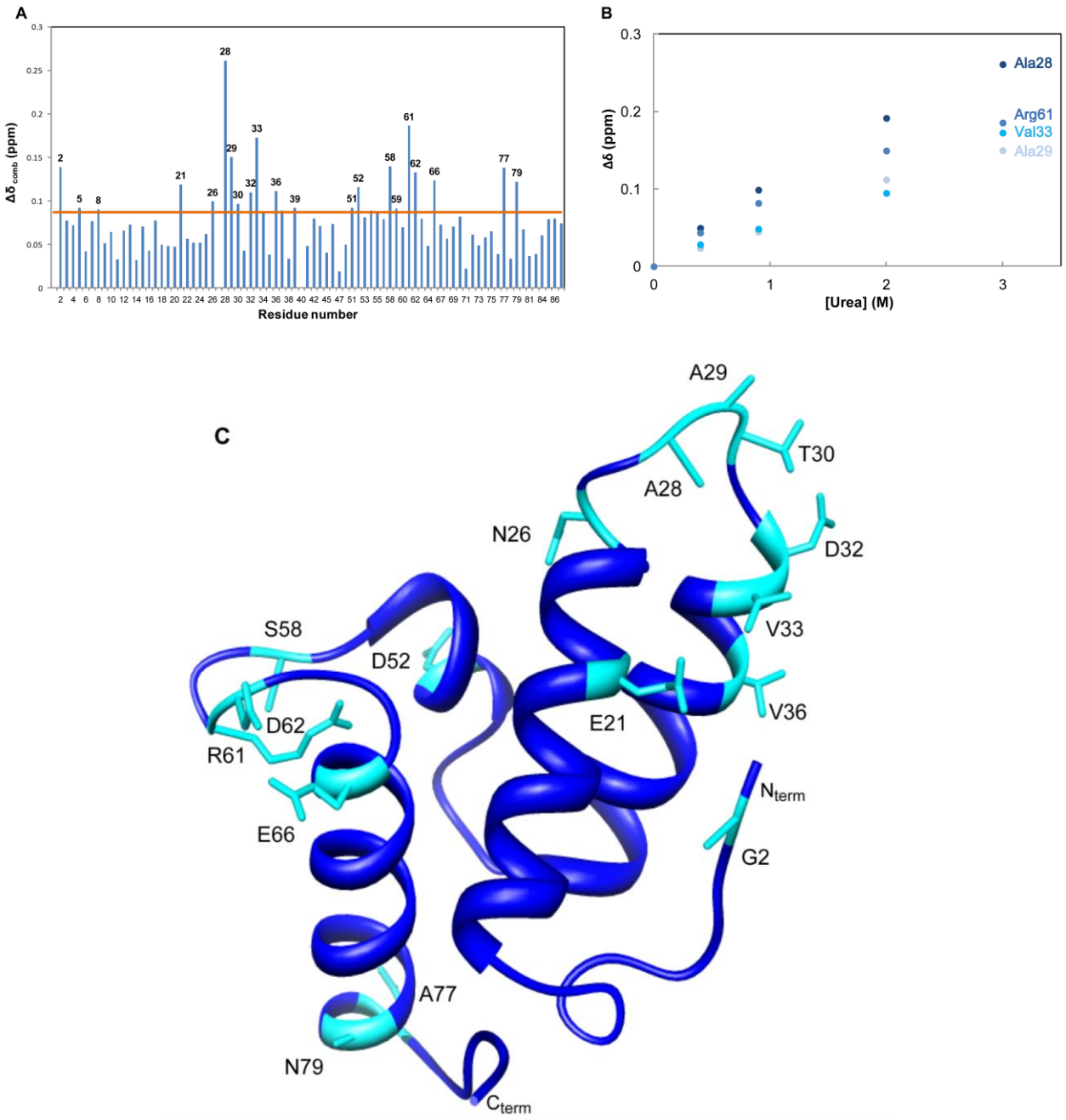
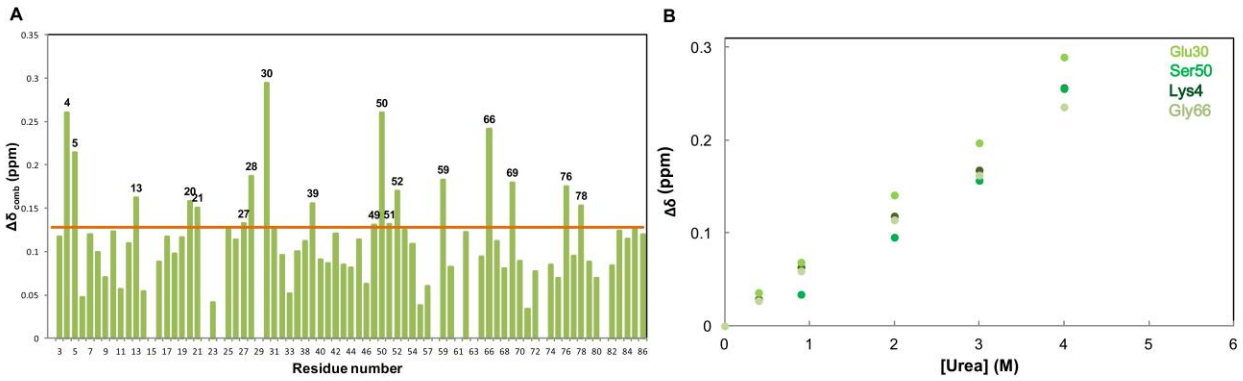



**Figure 10:** *A*) Im7 backbone amide chemical shift variations between native and 3M urea; *B*) Evolution of Δδ with urea for the most affected residues; *C*) Structure of Im7 (PDB: 1AYI), which highlights the most affected residues.

For protein Im9, the perturbations induced by urea occur for residues Lys4; His5; Ala13; Thr20; Thr21; Thr27; Ser28; Glu30; His39; Gly49; Ser50; Asp51; Leu52; Gly59; Gly66; Asn69; Ala76 and Asn78 (**Figure 11 A** and **B**).



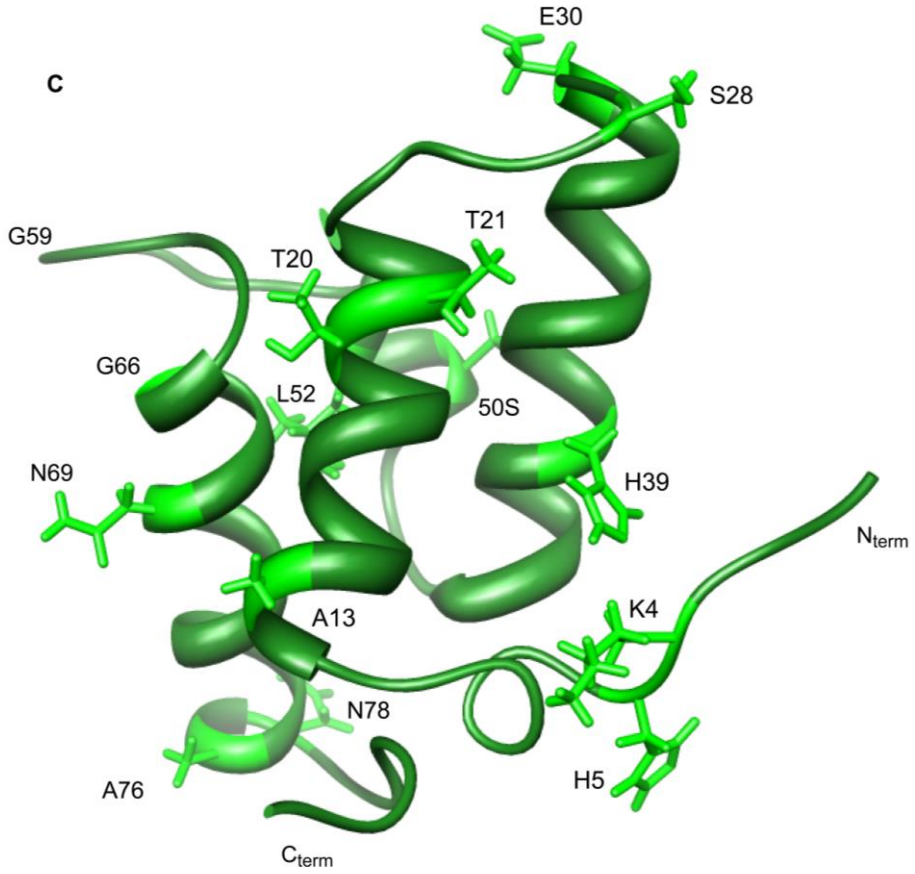




**Figure 11:** *A*) Im9 backbone amide chemical shift variations between native and 4M urea; *B*) Evolution of Δδ with urea for the most affected residues; *C*) Structure of Im9 (PDB: 1IMQ), which highlights the most affected residues.

The data obtained for thermal and chemical denaturation are summarized in **Table 4**, where is possible to see a comparison between the two proteins. The regions containing the most perturbed residues are indicated for both proteins, as well as the nature of each residue above the cut-off line.

**Table 4: Comparison of the chemical shift perturbation by thermal (T) and chemical (urea) denaturation of Im7 and Im9 proteins.** The aromatic, negatively and positively charge and the aliphatic residues are highlighted by symbols (Θ), (-), (+), (Υ) respectively. Chemical shifts variations larger than the corrected standard deviation to zero were considered as significant and marked with the "x".

| Im7 | | | | Im9 | | | |
|---|---|---|---|---|---|---|---|
| | **Position** | **T** | **Urea** | | **Position** | **T** | **Urea** |
| **E2 (-)** | N-term | | x | **E2 (-)** | N-term | | |
| **L3 (Υ)** | N-term | | | **L3 (Υ)** | N-term | | |
| **K4 (+)** | N-term | | | **K4 (+)** | N-term | | x |
| **N5 (Υ)** | N-term | | x | **H5 (+)** | N-term | | x |
| **S6 (Υ)** | N-term | | | **S6 (Υ)** | N-term | | |
| **I7 (Υ)** | N-term | | | **I7 (Υ)** | N-term | | |
| **S8( Υ)** | N-term | | x | **S8 (Υ)** | N-term | x | |
| **D9 (-)** | N-term | | | **D9 (-)** | N-term | x | |
| **Y10 (Θ)** | N-term | | | **Y10 (Θ)** | N-term | | |
| **T11 (Υ)** | N-term | | | **T11 (Υ)** | N-term | | |
| **E12 (-)** | Helix I | x | | **E12 (-)** | Helix I | | |
| **A13 (Υ)** | Helix I | | | **A13 (Υ)** | Helix I | x | x |
| **E14 (-)** | Helix I | | | **E14 (-)** | Helix I | | |
| **F15 (Θ)** | Helix I | | | **F15 (Θ)** | Helix I | | |
| **V16 (Υ)** | Helix I | | | **L16 (Υ)** | Helix I | | |
| **Q17 (Υ)** | Helix I | x | | **Q17 (Υ)** | Helix I | | |
| **L18 (Υ)** | Helix I | x | | **L18 (Υ)** | Helix I | x | |
| **L19 (Υ)** | Helix I | | | **V19 (Υ)** | Helix I | x | |
| **K20 (+)** | Helix I | | | **T20 (Υ)** | Helix I | | x |
| **E21 (-)** | Helix I | x | x | **T21 (Υ)** | Helix I | | x |
| **I22 (Υ)** | Helix I | | | **I22 (Υ)** | Helix I | x | |
| **E23 (-)** | Helix I | x | | **C23 (Υ)** | Helix I | | |
| **K24 (+)** | Helix I | x | | **N24 (Υ)** | Random coil | | |
| **E25 (-)** | Helix I | | | **A25 (Υ)** | Random coil | | |
| **N26 (Υ)** | Random coil | | x | **D26 (-)** | Random coil | | |
| **V27 (Υ)** | Random coil | | | **T27 (Υ)** | Random coil | x | x |
| **A28 (Υ)** | Random coil | x | x | **S28 (Υ)** | Random coil | | x |
| **A29 (Υ)** | Random coil | | x | **S29 (Υ)** | Random coil | x | |
| **T30 (Υ)** | Random coil | x | x | **E30 (-)** | Helix II | | x |
| **D31 (-)** | Helix II | | | **E31 (-)** | Helix II | | |
| **D32 (-)** | Helix II | | x | **E32 (-)** | Helix II | | |
| **V33 (Υ)** | Helix II | | x | **L33 (Υ)** | Helix II | x | |
| **L34 (Υ)** | Helix II | | | **V34 (Υ)** | Helix II | | |
| **D35 (-)** | Helix II | x | | **K35 (+)** | Helix II | | |
| **V36 (Υ)** | Helix II | x | x | **L36 (Υ)** | Helix II | x | |
| **L37 (Υ)** | Helix II | | | **V37 (Υ)** | Helix II | | |

| | | | |
|---|---|---|---|
| **L38 (Y)** | Helix II | | |
| **E39 (-)** | Helix II | | x |
| **H40 (+)** | Helix II | | |
| **F41 (Θ)** | Helix II | | |
| **V42 (Y)** | Helix II | | |
| **K43 (+)** | Helix II | | |
| **I44 (Y)** | Helix II | | |
| **T45 (Y)** | Helix II | x | |
| **E46 (-)** | Random coil | x | |
| **H47 (+)** | Random coil | x | |
| **P48 (Y)** | Random coil | | |
| **D49 (-)** | Random coil | | |
| **G50 (Y)** | Random coil | x | |
| **T51 (Y)** | Helix III | | x |
| **D52 (-)** | Helix III | | x |
| **L53 (Y)** | Helix III | x | |
| **I54 (Y)** | Helix III | | |
| **Y55 (Θ)** | Helix III | | |
| **Y56 (Θ)** | Helix III | | |
| **P57 (Y)** | Random coil | | |
| **S58 (Y)** | Random coil | | x |
| **D59 (-)** | Random coil | x | x |
| **N60 (Y)** | Random coil | x | |
| **R61 (+)** | Random coil | x | x |
| **D62 (-)** | Random coil | | x |
| **D63 (-)** | Random coil | x | |
| **S64 (Y)** | Random coil | x | |
| **P65 (Y)** | Helix IV | | |
| **E66 (-)** | Helix IV | | x |
| **G67 (Y)** | Helix IV | x | |
| **I68 (Y)** | Helix IV | | |
| **V69 (Y)** | Helix IV | x | |
| **K70 (+)** | Helix IV | | |
| **E71 (-)** | Helix IV | | |
| **I72 (Y)** | Helix IV | | |
| **K73 (+)** | Helix IV | | |
| **E74 (-)** | Helix IV | | |
| **W75 (Θ)** | Helix IV | | |
| **R76 (+)** | Helix IV | | |
| **A77 (Y)** | Helix IV | x | x |
| **A78 (Y)** | Helix IV | x | |
| **N79 (Y)** | Helix IV | x | x |
| **G80 (Y)** | C-term | x | |
| **K81 (+)** | C-term | | |
| **P82 (Y)** | C-term | | |
| **G83 (Y)** | C-term | | |
| **F84 (Θ)** | C-term | | |
| **K85 (+)** | C-term | | |
| **Q86 (Y)** | C-term | | |
| **G87 (Y)** | C-term | | |

| | | | |
|---|---|---|---|
| **T38 (Y)** | Helix II | | |
| **H39 (+)** | Helix II | x | x |
| **F40 (Θ)** | Helix II | | |
| **E41 (-)** | Helix II | | |
| **E42 (-)** | Helix II | x | |
| **M43 (Y)** | Helix II | | |
| **T44 (Y)** | Helix II | | |
| **E45 (-)** | Random coil | x | |
| **H46 (+)** | Random coil | x | |
| **P47 (Y)** | Random coil | | |
| **S48 (Y)** | Random coil | | |
| **G49 (Y)** | Random coil | x | x |
| **S50 (Y)** | Random coil | | x |
| **D51 (-)** | Helix III | | x |
| **L52 (Y)** | Helix III | x | x |
| **I53 (Y)** | Helix III | x | |
| **Y54 (Θ)** | Helix III | | |
| **Y55 (Θ)** | Helix III | | |
| **P56 (Y)** | Random coil | | |
| **K57 (+)** | Random coil | | |
| **E58 (-)** | Random coil | | |
| **G59 (Y)** | Random coil | x | x |
| **D60 (-)** | Random coil | | |
| **D61 (-)** | Random coil | x | |
| **D62 (-)** | Random coil | x | |
| **S63 (Y)** | Random coil | x | |
| **P64 (Y)** | Random coil | | |
| **S65 (Y)** | Helix IV | | |
| **G66 (Y)** | Helix IV | x | x |
| **I67 (Y)** | Helix IV | | |
| **V68 (Y)** | Helix IV | | |
| **N69 (Y)** | Helix IV | | x |
| **T70 (Y)** | Helix IV | | |
| **V71 (Y)** | Helix IV | | |
| **K72 (+)** | Helix IV | | |
| **Q73 (Y)** | Helix IV | | |
| **W74 (Θ)** | Helix IV | x | |
| **R75 (+)** | Helix IV | | |
| **A76 (Y)** | Helix IV | x | x |
| **A77 (Y)** | Helix IV | x | |
| **N78 (Y)** | C-term | | x |
| **G79 (Y)** | C-term | x | |
| **K80 (+)** | C-term | x | |
| **S81 (Y)** | C-term | x | |
| **G82 (Y)** | C-term | x | |
| **F83 (Θ)** | C-term | | |
| **K84 (+)** | C-term | | |
| **Q85 (Y)** | C-term | | |
| **G86 (Y)** | C-term | | |

## 4 – Discussion

In this work, two different approaches to promote the denaturation/unfolding of both Im7 and Im9 where followed in order to allow a comparative perspective towards the residues more affected by a physical effect, such as temperature, and by a chemical agent, urea.

$^{1}$H-$^{15}$N HSQC spectra were used in order to map the urea or temperature induced perturbations into the protein. The dispersion of signals in the $^{1}$H-$^{15}$N HSQC spectrum can be used to map the changes of the chemical shift of the individual residues, when going from a folded to an unfolded protein as shown in **Figure 12** .

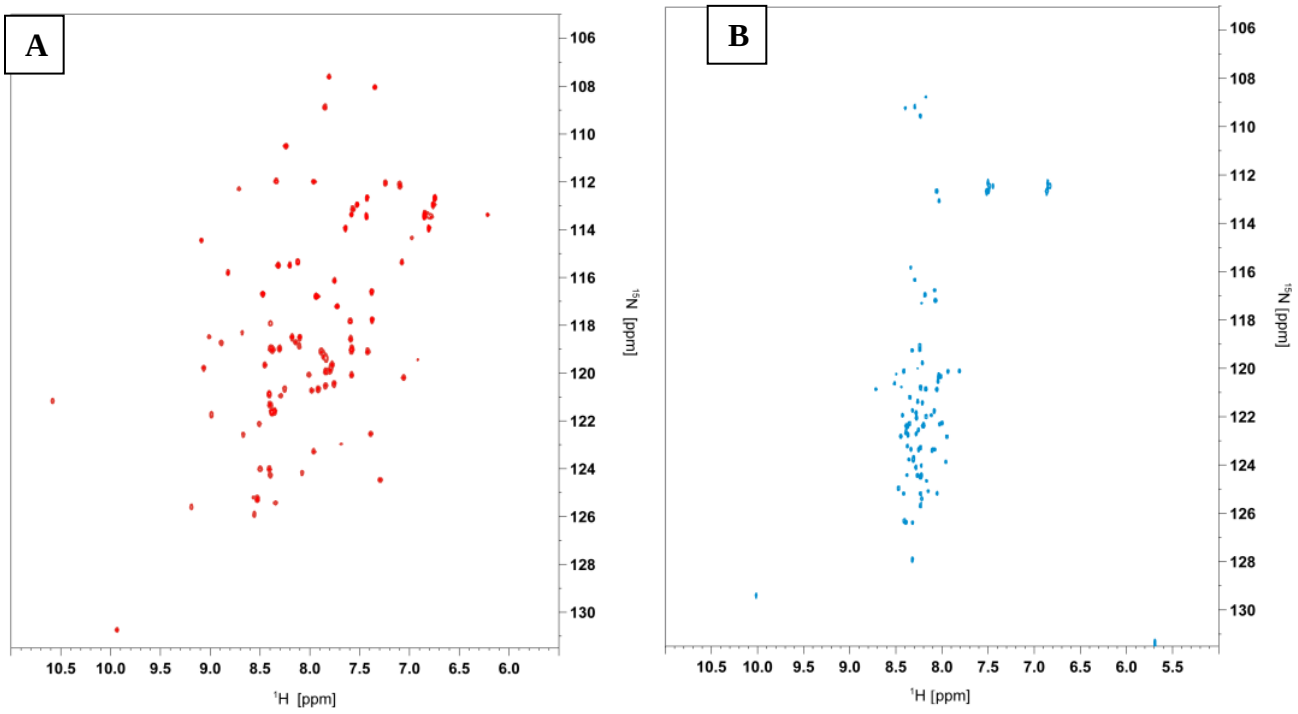



**Figure 12: $^{1}$H-$^{15}$N-HSQC spectra at 600 MHz of native Im7 in 50mM phosphate buffer, pH 7, in 10% $^{2}H_2O$, at 10 ºC. Spectrum A: protein Im7 without urea; Spectrum B: protein Im7 in 6M urea.**

A well dispersed spectrum, as shown in **Figure 12A**, evidences that the protein is well structured, organized in a three dimensional structure, reflecting individual and local environments for the different residues. In the other hand, **Figure 12B** exhibits a narrow distribution of signals in the spectrum, which is an indicator that the protein is in a

denatured/unfolded conformation, behaving like a random coil, leading to a more similar environment for all residues.

This change in the $^{1}H$, $^{15}N$ – HSQC is observed despite the denaturing method used, for both Im7 and Im9 proteins, leading to two questions; are the pathways and the forces implied in thermal denaturation the same in urea? Being two homologous proteins, why is the folding pathway different?

In **Table 4** (section 3.2), we can observe the distribution of the affected residues in the native structure of both proteins.

The temperature dependence of proton HN chemical shifts was observed long ago, but its mechanism is still not well understood. Temperature coefficients are almost always negative (a decrease in chemical shift value as the temperature is increased) and it has been shown that temperature coefficients in proteins are determined mainly by local melting of the structure, and not necessarily by the strength of hydrogen bonds to the amide. However, there is a tendency for (a) intramolecular hydrogen-bonded amides to have a smaller (less negative) coefficient, and (b) strongly hydrogen-bonded amides to have a larger coefficient, because the same fractional loss in structure gives rise to a larger change in shift.

Recently, by correlating temperature coefficients with $^{3h}J_{NC'}$ it has been suggested by computational studies that the hydrogen bond distance expansion is the main factor contributing to the $^{1}$HN chemical shift temperature dependence. [*55*]

In our work, instead of the amide proton temperature coefficient, we used the combined chemical shift. The combined chemical shift is commonly used to perform chemical shift mapping of protein-protein or protein-ligand interactions, we choose to use this method in order to evaluate the specific changes induced by temperature and to compare them with the ones induced by urea.

The results from the temperature study show that the residues most perturbed with increasing temperature are mostly located in loop regions between helices. In Figure 7A (chapter 3), is easy to observe that the relevant residues are spread among the sequence, but, Glu46, Thr30, Arg61, and Ala28, displayed a higher $\Delta\delta_{comb}$, when compared with the remaining residues.

Following the evolution of those residues (figure 7B), we can observe that the $\Delta\delta_{comb}/\Delta T$ is not the same for all residues and follows the order Glu46 > Thr30 > Arg61 > Ala28. This shows that temperature shifts can be used to measure of the capability of the residue to become unstructured as the temperature is raised. The largest change for Glu46 must be related with the position of this residue at the end of helix II, relatively unrestrained.

When comparing with Im9, the first observation is that the collapse temperature is different, being 318K for Im7 and 328K for Im9, which is in accordance with the fact that Im7 is significantly less stable than Im9 (ΔΔG 9.3 kJ/mol)[*56*]. Im9 protein also displays the most perturbed residues in the loop regions and at the edges of α-helices, but we can clearly observe a more global influence of temperature, that the one presented for Im7. The relevant residues considered for Im9 exhibits values for Δδcomb, between 0.3 and 0.6, while for Im7 we can obtain values for $\Delta\delta_{comb}$ between 0.3 and 1.2.

With the sets of experiments performed for temperature, we realized that a global effect, responsible for disruption of hydrogen bonds, will lead to the loss of the secondary and tertiary structure.

Using urea as a denaturant, the same approach was made for Im7 and Im9, and in figure 10 we have once more represented the relevant residues, as well the representation of those residues in Im7 three-dimensional structure. As described in the previous section Glu2; Asn5; Ser8; Glu21; Asn26; Ala28; Ala29; Thr30; Asp32; Val33; Val36; Glu39; Thr51; Asp52; Ser58; Asp59; Arg61; Asp62; Glu66; Ala77 and Asn79 are the most affected residues upon addition of urea, being located on loop regions and at the edges of the four helices. When we make a closer look about the nature of these residues (Table 3), we identify that most negatively charged residues, Glu21; Asn26; Glu39; Asp59; Asp62; Glu66, are located at the edges of helix I, II and IV, being the rest of relevant residues aliphatic and located at loop regions.

In the case of protein Im9, the interaction with urea occur for residues Lys4; His5; Ala13; Thr20; Thr21; Thr27; Ser28; Glu30; His39; Gly49; Ser50; Asp51; Leu52; Gly59; Gly66; Asn69; Ala76 and Asn78. A brief comparison between the profile for Im7 and Im9 displayed in **Figures 10** and **11** reveal a more located perturbation for Im7 located in the regions formed by residues 26-36 and 62-66, which is not observed for Im9 that exhibits an overall perturbation.

This study points out that the ends of well-structured helices can concertedly unfold without entering the mid region residues in the same unfolding process. There seems to be a

correlation between dynamic residues (most affected by temperature) and the residues in the regions most perturbed by urea. The results shown that entire loop regions on both proteins may act as concerted units during the unfolding process, and contribute for favorable interactions that delimit and stabilize native-like structural features on the urea-unfolded state, allowing buried regions to be less solvent exposed. This important contribution of the loops and their inherent dynamics will certainly be important for the significantly number of possible conformations and accessible folding pathways routes towards its minimum native state.

The hydrogen exchange rate of a certain amide group in a protein can also be used to study solvent accessibility and hydrogen bond strength. Because non-hydrogen bonded protons are in constant exchange with the solvent, their exchange rates depend on their protection level and bond strength. Protons participating in hydrogen bonds, for instance, will be more protected and thus have lower exchange rates than those who are solvent exposed.

In **Table 5** the published hydrogen exchange rates of Im7 are compared with the temperature and urea induced chemical shift perturbations. Only two residues with measurable exchange rates are among the ones perturbed by urea. This means that urea targets preferably residues that are accessible and solvent exposed. On the contrary, about 1/3 of the residues with significant temperature induced shifts have measurable exchange rates. This evidences that temperature shifts also reflect the strength of hydrogen bonds that are important for the stabilization of secondary structure elements.

All data points out for the importance of the hydrophobic collapse mechanism in which collapse of the chain precedes helix formation. Those phenomena should be strongly linked between helix propensity and though hydrogen bonds and hydrophobicity.

**Table 5: Comparison between the hydrogen exchange rates ($k_{ex}$) [*57*] and the chemical shift perturbation by thermal (T) and chemical (urea) denaturation (this work) of Im7 protein.** Chemical shifts variations larger than the corrected standard deviation to zero were considered as significant and marked with the "x".

| | Location | T | Urea | $k_{ex}$ ($h^{-1}$) |
|---|---|---|---|---|
| Asp9 | Random Coil | | | 0.3193 |
| Tyr10 | Random Coil | | | 0.0353 |
| Thr11 | Random Coil | | | 0.0333 |
| Val16 | Helix I | | | 0.0022 |
| Gln17 | Helix I | x | | 0.0075 |
| Leu18 | Helix I | x | | 0.0054 |
| Glu21 | Helix I | x | x | 0.0764 |
| Glu23 | Helix I | x | | 0.4505 |
| Leu37 | Helix II | | | 0.4266 |
| Leu38 | Helix II | | | 0.0888 |
| Phe41 | Helix II | | | 1.8620 |
| Val42 | Helix II | | | 0.0072 |
| Leu53 | Helix III | x | | 5.1000 |
| Ile54 | Helix III | | | 2.9000 |
| Tyr55 | Helix III | | | 3.8000 |
| Tyr56 | Helix III | | | 2.1000 |
| Gly67 | Helix IV | x | | 0.3875 |
| Val69 | Helix IV | x | | 0.0172 |
| Ile72 | Helix IV | | | 0.0159 |
| Lys73 | Helix IV | | | 0.0383 |
| Glu74 | Helix IV | | | 0.0488 |
| Trp75 | Helix IV | | | 0.0392 |
| Arg76 | Helix IV | | | 0.0452 |
| Ala77 | Helix IV | x | x | 0.0896 |
| Ala78 | Helix IV | x | | 0.3076 |
| Lys85 | Random Coil | | | 1.1098 |

## *5 – Conclusions*

In this work following the two approaches, Temperature and Urea, to promote the unfolding of Im7 and Im9, we were able to point out that urea targets preferably residues that are accessible and solvent exposed, while temperature results show that the residues most perturbed with increasing temperature are mostly located in loop regions between helices

The results shown that entire loop regions on both proteins may act as concerted units during the unfolding process, and contribute for favorable interactions that delimit and stabilize native-like structural features on the urea-unfolded state, allowing buried regions to be less solvent exposed. These regions are clearly very important due to their inherent dynamics that will certainly be important for the significantly number of possible conformations and accessible folding pathways routes towards its minimum native state.

With our experimental data, we have some insights of the importance of the hydrophobic collapse mechanism in which collapse of the chain precedes helix formation. Those phenomena should be strongly linked between helix propensity and though hydrogen bonds and hydrophobicity.

## 6 – References


1. Lin, M. M., and Zewail, A. H. (2012) Protein folding - simplicity in complexity, *Ann Phys-Berlin 524*, 379-391.
2. Dobson, C. M. (2003) Protein folding and misfolding, *Nature 426*, 884-890.
3. Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961) Kinetics of Formation of Native Ribonuclease during Oxidation of Reduced Polypeptide Chain, *P Natl Acad Sci USA 47*, 1309-1314.
4. Anfinsen, C. B. (1973) Principles That Govern Folding of Protein Chains, *Science 181*, 223-230.
5. Levintha.C. (1968) Are There Pathways for Protein Folding, *J Chim Phys Pcb 65*, 44-&.
6. Kim, P. S., and Baldwin, R. L. (1982) Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding, *Annu Rev Biochem 51*, 459-489.
7. Fersht, A. R. (1997) Nucleation mechanisms in protein folding, *Curr Opin Struc Biol 7*, 3-9.
8. Jackson, S. E., and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry 30*, 10428-10435.
9. Daggett, V., and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding?, *Trends Biochem Sci 28*, 18-25.
10. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2003) Unifying features in protein-folding mechanisms, *P Natl Acad Sci USA 100*, 13286-13291.
11. Gianni, S., Geierhaas, C. D., Calosci, N., Jemth, P., Vuister, G. W., Travaglini-Allocatelli, C., Vendruscolo, M., and Brunori, M. (2007) A PDZ domain recapitulates a unifying mechanism for protein folding, *Proc Natl Acad Sci U S A 104*, 128-133.
12. Baldwin, R. L. (1994) Protein-Folding - Matching Speed and Stability, *Nature 369*, 183-184.
13. Lazaridis, T., and Karplus, M. (1997) "New view" of protein folding reconciled with the old through multiple unfolding simulations, *Science 278*, 1928-1931.
14. Ghosh, K., Ozkan, S. B., and Dill, K. A. (2007) The ultimate speed limit to protein folding is conformational searching, *J Am Chem Soc 129*, 11920-11927.
15. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat Struct Biol 4*, 10-19.
16. Onuchic, J. N., LutheySchulten, Z., and Wolynes, P. G. (1997) Theory of protein folding: The energy landscape perspective, *Annu Rev Phys Chem 48*, 545-600.
17. Plaxco, K. W., and Gross, M. (2001) Unfolded, yes, but random? Never!, *Nat Struct Biol 8*, 659-660.
18. Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How Fast-Folding Proteins Fold, *Science 334*, 517-520.
19. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985) Hydrophobicity of Amino-Acid Residues in Globular-Proteins, *Science 229*, 834-838.
20. Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., Hendricks, M. M., Iimura, S., Gajiwala, K., Scholtz, J. M., and Grimsley, G. R. (2011) Contribution of hydrophobic interactions to protein stability, *J Mol Biol 408*, 514-528.

21. Rose, G. D., and Wolfenden, R. (1993) Hydrogen bonding, hydrophobicity, packing, and protein folding, *Annu Rev Biophys Biomol Struct 22*, 381-415.
22. Baldwin, R. L. (2007) Energetics of protein folding, *J Mol Biol 371*, 283-301.
23. Strickler, S. S., Gribenko, A. V., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T., Loladze, V. V., and Makhatadze, G. I. (2006) Protein stability and surface electrostatics: A charged relationship, *Biochemistry 45*, 2761-2766.
24. Gallivan, J. P., and Dougherty, D. A. (1999) Cation-pi interactions in structural biology, *Proc Natl Acad Sci U S A 96*, 9459-9464.
25. Crowley, P. B., and Golovin, A. (2005) Cation-pi interactions in protein-protein interfaces, *Proteins 59*, 231-239.
26. Pan, Y., Zhang, K., Qi, J., Yue, J., Springer, T. A., and Chen, J. (2010) Cation-pi interaction regulates ligand-binding affinity and signaling of integrin alpha4beta7, *Proc Natl Acad Sci U S A 107*, 21388-21393.
27. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) Disulfide bonds and protein folding, *Biochemistry 39*, 7032.
28. Royer, C. A. (2008) The nature of the transition state ensemble and the mechanisms of protein folding: a review, *Arch Biochem Biophys 469*, 34-45.
29. Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. R. (1989) Mapping the transition state and pathway of protein folding by protein engineering, *Nature 340*, 122-126.
30. Matouschek, A., Kellis, J. T., Serrano, L., and Fersht, A. R. (1989) Mapping the Transition-State and Pathway of Protein Folding by Protein Engineering, *Nature 340*, 122-126.
31. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding, *J Mol Biol 224*, 771-782.
32. Gianni, S., Ivarsson, Y., Jemth, P., Brunori, M., and Travaglini-Allocatelli, C. (2007) Identification and characterization of protein folding intermediates, *Biophys Chem 128*, 105-113.
33. Brockwell, D. J., and Radford, S. E. (2007) Intermediates: ubiquitous species on folding energy landscapes?, *Curr Opin Struc Biol 17*, 30-37.
34. Jackson, S. E. (1998) How do small single-domain proteins fold?, *Fold Des 3*, R81-R91.
35. Kramer, G., Boehringer, D., Ban, N., and Bukau, B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins, *Nat Struct Mol Biol 16*, 589-597.
36. Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I., Jr., and Bustamante, C. (2011) The ribosome modulates nascent protein folding, *Science 334*, 1723-1727.
37. Eichmann, C., Preissler, S., Riek, R., and Deuerling, E. (2010) Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy, *Proc Natl Acad Sci U S A 107*, 9111-9116.
38. Pace, C. N., Huyghues-Despointes, B. M. P., Fu, H. L., Takano, K., Scholtz, J. M., and Grimsley, G. R. (2010) Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins, *Protein Sci 19*, 929-943.
39. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea, *Science 293*, 487-489.

40. Tanford, C. (1964) Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions, *J Am Chem Soc 86*, 2050-&.
41. Bennion, B. J., and Daggett, V. (2003) The molecular basis for the chemical denaturation of proteins by urea, *Proc Natl Acad Sci U S A 100*, 5142-5147.
42. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res 25*, 3389-3402.
43. Gorski, S. A., Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2001) Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9, *J Mol Biol 312*, 849-863.
44. Wallis, R., Reilly, A., Rowe, A., Moore, G. R., James, R., and Kleanthous, C. (1992) Invivo and Invitro Characterization of Overproduced Colicin E9 Immunity Protein, *Eur J Biochem 207*, 687-695.
45. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) $^{1}H$, $^{13}C$ and $^{15}N$ Chemical-Shift Referencing in Biomolecular Nmr, *J Biomol Nmr 6*, 135-140.
46. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol Nmr 6*, 277-293.
47. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins 59*, 687-696.
48. Kawahara, K., and Tanford, C. (1966) Viscosity and Density of Aqueous Solutions of Urea and Guanidine Hydrochloride, *J Biol Chem 241*, 3228-3232.
49. Warren, J. R., and Gordon, J. A. (1966) On Refractive Indices of Aqueous Solutions of Urea, *J Phys Chem-Us 70*, 297-300.
50. Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation, *J Magn Reson B 108*, 94-98.
51. Schumann, F. H., Riepl, H., Maurer, T., Gronwald, W., Neidig, K. P., and Kalbitzer, H. R. (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions, *J Biomol Nmr 39*, 275-289.
52. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server, *Methods Mol Biol 112*, 531-552.
53. Bjellqvist, B., Basse, B., Olsen, E., and Celis, J. E. (1994) Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions, *Electrophoresis 15*, 529-539.
54. Wallis, R., Reilly, A., Barnes, K., Abell, C., Campbell, D. G., Moore, G. R., James, R., and Kleanthous, C. (1994) Tandem overproduction and characterisation of the nuclease domain of colicin E9 and its cognate inhibitor protein Im9, *Eur J Biochem 220*, 447-454.
55. Hong, J., Jing, Q., and Yao, L. (2013) The protein amide (1)H(N) chemical shift temperature coefficient reflects thermal expansion of the N-H...O=C hydrogen bond, *J Biomol Nmr 55*, 71-78.

56. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C., and Radford, S. E. (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9, *J Mol Biol 286*, 1597-1608.
57. Gorski, S. A., Le Duff, C. S., Capaldi, A. P., Kalverda, A. P., Beddard, G. S., Moore, G. R., and Radford, S. E. (2004) Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7, *J Mol Biol 337*, 183-193.

# 7 – Appendix

## 7.1 Appendix A

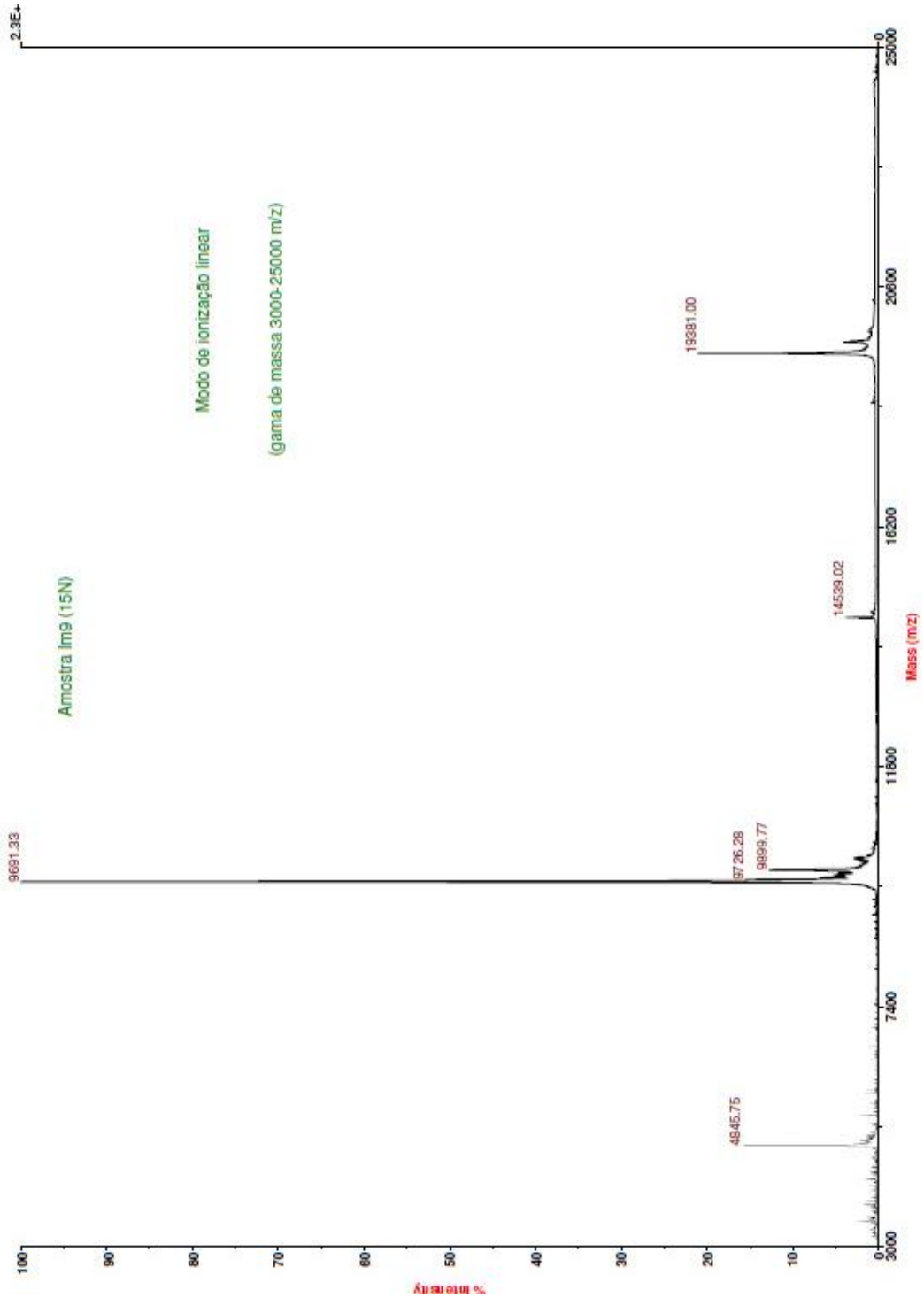



**Figure A1Time of Flight mass spectrum of the Im9 protein.** The experimentally determined molecular mass of 9691.33 Da is in accordance with the theoretical value of 9689.71 Da.

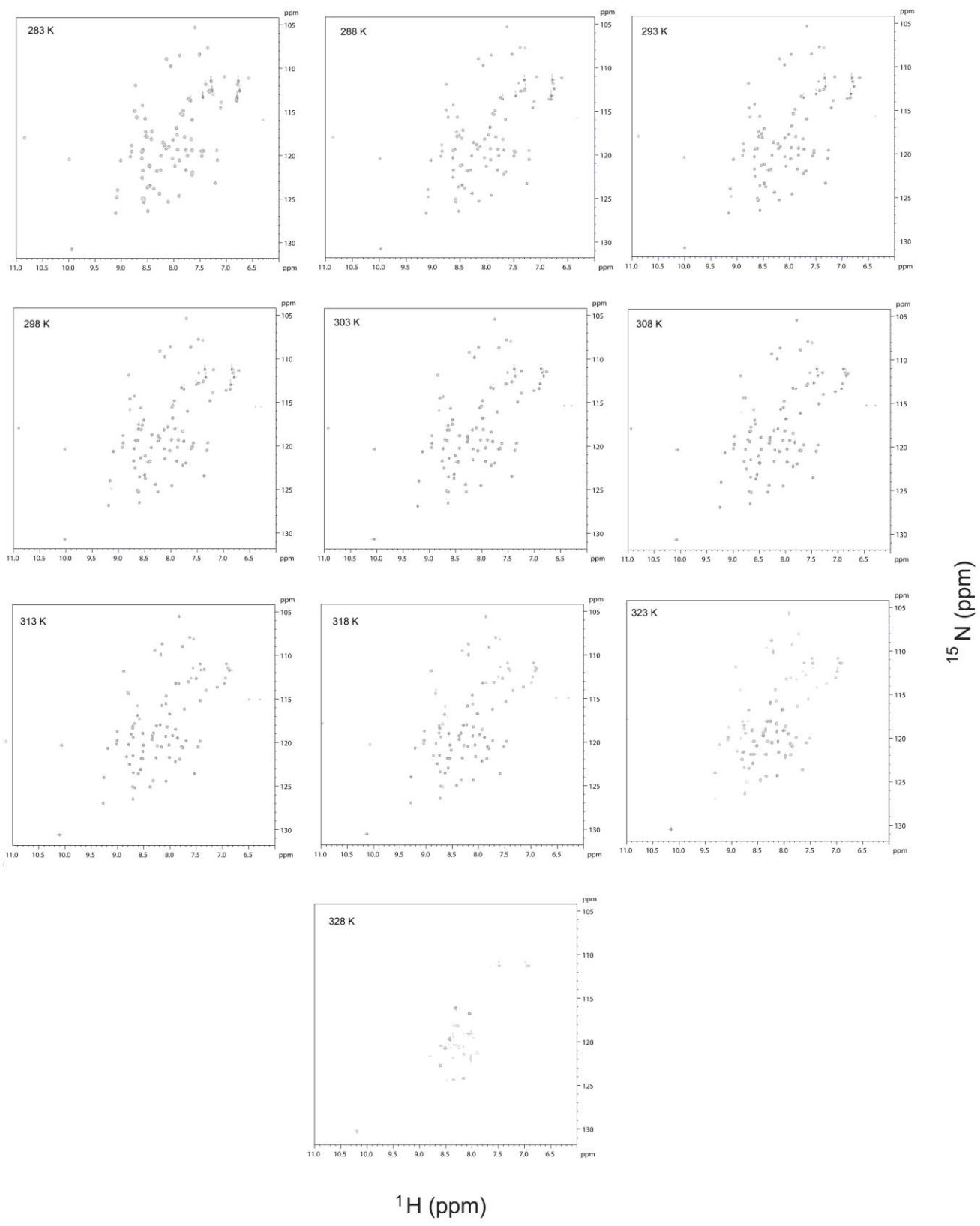



**Figure A2: $^{1}$H-$^{15}$N HSQC spectra of Im9 at different temperatures.** The temperatures used in each HSQC of are indicated inside each panel.

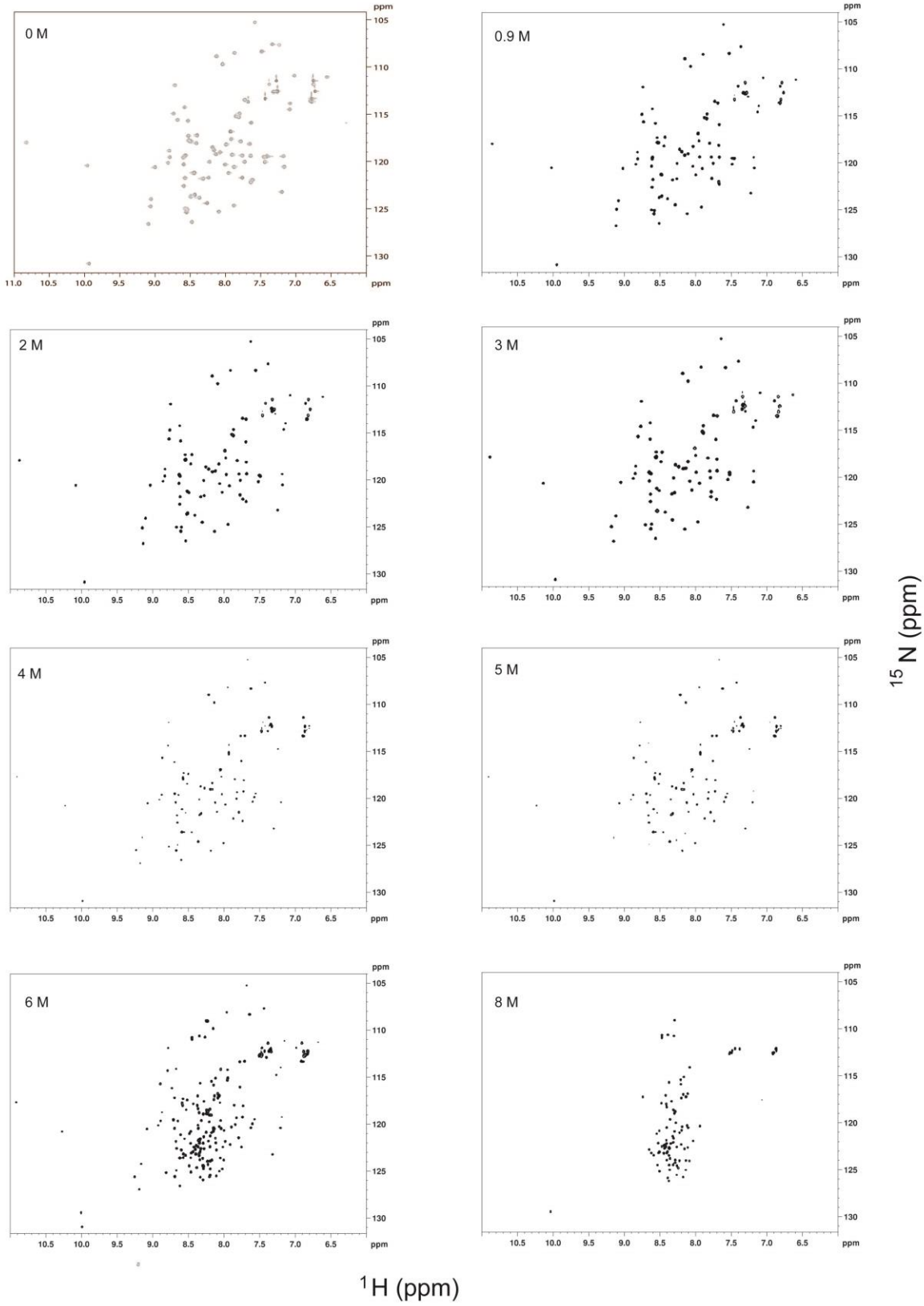



**Figure A3: $^{1}$H-$^{15}$N HSQC spectra of Im9 at different concentrations of urea.** The urea concentrations used in each HSQC of are indicated inside each panel.

## 7.2 Appendix B

**Table B1**: $^{1}H$ and $^{15}N$ chemical shifts (δ) table for Im7 at different Temperatures

| | δ (ppm) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 283K | | 288K | | 298K | | 303K | | 308K | | 313K | |
| | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ |
| **E2** | 125.46 | 8.76 | - | - | - | - | - | - | - | - | - | - |
| **L3** | 125.69 | 8.54 | 125.70 | 8.55 | 125.72 | 8.58 | 125.70 | 8.59 | 125.69 | 8.61 | 125.66 | 8.62 |
| **K4** | 123.24 | 7.88 | 123.21 | 7.92 | 123.13 | 7.99 | 123.08 | 8.03 | 123.02 | 8.06 | 122.96 | 8.10 |
| **N5** | 115.76 | 8.51 | 115.77 | 8.54 | 115.81 | 8.60 | 115.85 | 8.63 | 115.91 | 8.66 | 116.02 | 8.69 |
| **S6** | 112.26 | 8.16 | 112.24 | 8.18 | 112.21 | 8.23 | 112.18 | 8.25 | 112.18 | 8.28 | 112.24 | 8.31 |
| **I7** | 125.87 | 9.38 | 125.95 | 9.42 | 126.11 | 9.49 | 126.17 | 9.52 | 126.21 | 9.55 | 126.21 | 9.57 |
| **S8** | 112.32 | 7.44 | 112.37 | 7.48 | 112.48 | 7.56 | 112.54 | 7.60 | 112.60 | 7.64 | 112.71 | 7.69 |
| **D9** | 118.09 | 7.79 | 118.13 | 7.83 | 118.24 | 7.92 | 118.30 | 7.96 | 118.36 | 8.00 | 118.48 | 8.04 |
| **Y10** | 120.33 | 8.20 | 120.36 | 8.25 | 120.43 | 8.33 | 120.44 | 8.37 | 120.46 | 8.42 | 120.47 | 8.45 |
| **T11** | 112.55 | 8.90 | 112.50 | 8.93 | 112.43 | 8.98 | 112.38 | 9.00 | 112.37 | 9.03 | 112.40 | 9.04 |
| **E12** | 122.01 | 9.18 | 122.13 | 9.22 | 122.35 | 9.30 | 122.44 | 9.34 | 122.54 | 9.37 | 122.63 | 9.39 |
| **A13** | 118.76 | 8.37 | 118.79 | 8.40 | 118.85 | 8.45 | 118.87 | 8.47 | 118.91 | 8.49 | 119.01 | 8.51 |
| **E14** | 119.56 | 8.06 | - | - | - | - | - | - | - | - | - | - |
| **F15** | 124.53 | 8.59 | 124.46 | 8.63 | 124.29 | 8.69 | 124.19 | 8.72 | 124.09 | 8.75 | 123.91 | 8.78 |
| **V16** | 119.93 | 8.65 | 119.91 | 8.68 | 119.83 | 8.74 | 119.78 | 8.77 | 119.68 | 8.79 | 119.71 | 8.82 |
| **Q17** | 118.05 | 7.57 | 118.11 | 7.62 | 118.22 | 7.71 | 118.26 | 7.75 | 118.32 | 7.79 | 118.42 | 7.84 |
| **L18** | 122.81 | 7.59 | 122.78 | 7.64 | 122.70 | 7.74 | 122.65 | 7.79 | 122.59 | 7.84 | 122.53 | 7.89 |
| **L19** | 119.27 | 7.78 | 119.28 | 7.82 | 119.29 | 7.89 | 119.28 | 7.93 | 119.27 | 7.96 | 119.20 | 8.00 |
| **K20** | 120.92 | 8.45 | 120.89 | 8.49 | 120.79 | 8.55 | 120.70 | 8.58 | 120.60 | 8.60 | 120.46 | 8.61 |
| **E21** | 123.54 | 8.16 | 123.45 | 8.20 | 123.24 | 8.28 | 123.11 | 8.32 | 122.92 | 8.35 | 122.68 | 8.38 |
| **I22** | 120.79 | 8.04 | 120.77 | 8.07 | - | - | - | - | - | - | - | - |
| **E23** | 118.83 | 7.79 | 118.90 | 7.84 | 119.06 | 7.95 | 119.14 | 7.99 | 119.26 | 8.04 | 119.43 | 8.09 |
| **K24** | 119.36 | 7.62 | 119.39 | 7.68 | 119.45 | 7.79 | 119.48 | 7.85 | 119.50 | 7.91 | 119.57 | 7.97 |
| **E25** | 119.14 | 8.30 | 119.19 | 8.34 | 119.28 | 8.42 | 119.30 | 8.45 | 119.34 | 8.48 | - | - |
| **N26** | 119.25 | 8.50 | 119.15 | 8.52 | 118.96 | 8.57 | 118.86 | 8.59 | 118.78 | 8.61 | 118.73 | 8.63 |
| **A28** | 124.74 | 7.49 | 124.71 | 7.58 | 124.65 | 7.77 | 124.63 | 7.86 | 124.64 | 7.95 | 124.72 | 8.04 |
| **A29** | 121.87 | 8.55 | 121.93 | 8.54 | 122.05 | 8.52 | 122.07 | 8.51 | 122.13 | 8.50 | 122.20 | 8.50 |
| **T30** | 109.14 | 8.04 | 109.58 | 8.12 | 110.43 | 8.27 | 110.79 | 8.33 | 111.15 | 8.39 | 111.46 | 8.43 |
| **D31** | 121.91 | 8.58 | 121.81 | 8.61 | 121.63 | 8.65 | 121.56 | 8.67 | 121.53 | 8.69 | 121.54 | 8.71 |
| **D32** | 118.97 | 8.34 | 119.00 | 8.38 | 119.17 | 8.45 | 119.21 | 8.48 | 119.30 | 8.50 | - | - |
| **V33** | 120.21 | 8.04 | 120.13 | 8.07 | 119.92 | 8.14 | 119.83 | 8.16 | 119.72 | 8.19 | 119.62 | 8.22 |
| **L34** | 122.39 | 8.70 | 122.39 | 8.71 | 122.40 | 8.73 | 122.41 | 8.73 | 122.47 | 8.72 | 122.56 | 8.71 |
| **D35** | 115.76 | 8.40 | 115.89 | 8.41 | 116.19 | 8.44 | 116.33 | 8.45 | 116.52 | 8.47 | 116.77 | 8.48 |
| **V36** | 120.45 | 7.25 | 120.37 | 7.33 | 120.24 | 7.47 | 120.14 | 7.54 | 120.06 | 7.61 | 119.95 | 7.68 |
| **L37** | 120.99 | 8.18 | 120.99 | 8.23 | 121.00 | 8.31 | 121.01 | 8.35 | 121.07 | 8.39 | - | - |
| **L38** | 119.00 | 9.08 | 119.05 | 9.11 | 119.13 | 9.14 | 119.17 | 9.16 | 119.21 | 9.17 | 119.30 | 9.17 |
| **E39** | 119.93 | 7.97 | 119.96 | 8.03 | - | - | - | - | - | - | - | - |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **H40** | 119.66 | 8.03 | - | - | - | - | - | - | - | - | - | - |
| **F41** | 118.58 | 8.87 | 118.56 | 8.91 | 118.52 | 8.98 | 118.47 | 9.00 | 118.45 | 9.03 | 118.45 | 9.04 |
| **V42** | 121.21 | 8.49 | 121.20 | 8.52 | 121.15 | 8.58 | 121.10 | 8.61 | 121.06 | 8.63 | 120.98 | 8.65 |
| **K43** | 120.94 | 8.11 | 120.98 | 8.17 | 121.02 | 8.27 | 121.06 | 8.32 | 121.10 | 8.37 | | |
| **T45** | 109.32 | 7.55 | 109.34 | 7.59 | 109.43 | 7.69 | 109.48 | 7.74 | 109.57 | 7.78 | 109.68 | 7.84 |
| **E46** | 112.41 | 7.29 | 112.85 | 7.37 | 113.85 | 7.53 | 114.35 | 7.61 | 114.86 | 7.69 | 115.44 | 7.77 |
| **H47** | 119.69 | 7.11 | 119.75 | 7.16 | 119.91 | 7.26 | 119.95 | 7.31 | 120.04 | 7.36 | 120.11 | 7.42 |
| **D49** | 121.43 | 10.77 | 121.45 | 10.80 | 121.48 | 10.86 | 121.47 | 10.87 | 121.46 | 10.89 | 121.43 | 10.90 |
| **G50** | 107.87 | 8.00 | 107.90 | 8.05 | 107.97 | 8.14 | 107.98 | 8.19 | 108.01 | 8.24 | 108.08 | 8.29 |
| **T51** | 114.71 | 9.28 | 114.70 | 9.29 | 114.68 | 9.30 | 114.64 | 9.30 | 114.61 | 9.29 | 114.54 | 9.28 |
| **D52** | 126.18 | 8.75 | 126.18 | 8.79 | 126.11 | 8.85 | 126.04 | 8.88 | 125.95 | 8.90 | 125.77 | 8.91 |
| **L53** | 115.63 | 7.27 | 115.68 | 7.32 | 115.87 | 7.42 | 115.98 | 7.46 | 116.13 | 7.51 | - | - |
| **I54** | 114.21 | 7.01 | 114.11 | 7.04 | 113.89 | 7.10 | 113.77 | 7.12 | 113.65 | 7.15 | 113.50 | 7.18 |
| **Y55** | 114.61 | 7.17 | 114.62 | 7.22 | 114.74 | 7.30 | 114.84 | 7.34 | 114.99 | 7.39 | - | - |
| **Y56** | 118.19 | 8.59 | 118.32 | 8.61 | 118.65 | 8.65 | 118.85 | 8.65 | - | - | - | - |
| **S58** | 116.96 | 8.67 | 116.93 | 8.66 | 116.84 | 8.64 | 116.77 | 8.63 | 116.71 | 8.62 | 116.61 | 8.61 |
| **D59** | 121.58 | 8.59 | 121.83 | 8.62 | 122.25 | 8.66 | 122.41 | 8.68 | 122.55 | 8.70 | - | - |
| **N60** | 115.63 | 8.32 | 115.66 | 8.37 | 115.76 | 8.47 | 115.82 | 8.51 | 115.93 | 8.55 | 116.10 | 8.59 |
| **R61** | 116.86 | 7.58 | 117.08 | 7.64 | 117.56 | 7.76 | 117.80 | 7.82 | 118.07 | 7.88 | 118.41 | 7.95 |
| **D62** | 121.15 | 8.60 | 121.24 | 8.62 | 121.44 | 8.63 | 121.53 | 8.64 | 121.62 | 8.65 | 121.70 | 8.65 |
| **D63** | 124.27 | 8.69 | 124.14 | 8.69 | 123.84 | 8.69 | 123.65 | 8.69 | 123.44 | 8.69 | 123.18 | 8.69 |
| **S64** | 117.47 | 7.92 | 117.41 | 7.99 | 117.31 | 8.14 | 117.25 | 8.21 | 117.19 | 8.28 | 117.14 | 8.35 |
| **E66** | 116.06 | 9.02 | 116.09 | 9.06 | 116.19 | 9.15 | 116.23 | 9.19 | 116.31 | 9.22 | 116.46 | 9.24 |
| **G67** | 112.23 | 8.53 | 112.12 | 8.55 | 111.84 | 8.58 | 111.69 | 8.59 | 111.51 | 8.61 | 111.29 | 8.62 |
| **I68** | 124.44 | 8.27 | 124.44 | 8.32 | 124.42 | 8.39 | 124.39 | 8.42 | 124.33 | 8.45 | 124.21 | 8.47 |
| **V69** | 118.77 | 8.30 | 118.86 | 8.37 | 119.07 | 8.51 | 119.16 | 8.57 | 119.26 | 8.63 | 119.36 | 8.68 |
| **K70** | 120.15 | 7.99 | 120.21 | 8.04 | 120.36 | 8.12 | - | - | - | - | - | - |
| **I72** | 119.20 | 8.59 | 119.22 | 8.62 | 119.20 | 8.69 | 119.20 | 8.72 | 119.24 | 8.74 | 119.33 | 8.76 |
| **K73** | 120.06 | 9.26 | 120.08 | 9.29 | 120.14 | 9.34 | 120.16 | 9.37 | 120.23 | 9.39 | 120.33 | 9.40 |
| **E74** | 119.29 | 8.57 | 119.36 | 8.62 | 119.50 | 8.71 | 119.58 | 8.75 | - | - | - | - |
| **W75** | 122.84 | 8.86 | 122.82 | 8.90 | 122.79 | 8.98 | 122.77 | 9.01 | 122.73 | 9.04 | 122.71 | 9.06 |
| **R76** | 118.74 | 9.21 | 118.72 | 9.23 | 118.68 | 9.26 | 118.65 | 9.28 | 118.66 | 9.29 | 118.73 | 9.29 |
| **A77** | 120.35 | 7.78 | 120.43 | 7.84 | 120.58 | 7.94 | 120.66 | 8.00 | 120.76 | 8.05 | 120.90 | 8.10 |
| **A78** | 120.71 | 7.95 | 120.76 | 8.00 | 120.85 | 8.11 | 120.88 | 8.16 | 120.92 | 8.21 | 120.99 | 8.26 |
| **N79** | 113.71 | 7.05 | 113.58 | 7.08 | 113.31 | 7.12 | 113.16 | 7.15 | 113.01 | 7.17 | 112.87 | 7.19 |
| **G80** | 108.30 | 7.55 | 108.36 | 7.60 | 108.48 | 7.71 | 108.52 | 7.76 | 108.57 | 7.82 | 108.62 | 7.88 |
| **K81** | 119.37 | 8.08 | 119.42 | 8.13 | 119.60 | 8.21 | - | - | - | - | - | - |
| **G83** | 110.77 | 8.44 | 110.80 | 8.46 | 110.86 | 8.50 | 110.86 | 8.52 | 110.87 | 8.54 | 110.84 | 8.56 |
| **F84** | 116.40 | 7.95 | 116.44 | 7.95 | 116.56 | 7.96 | 116.62 | 7.97 | 116.71 | 7.97 | 116.85 | 7.99 |
| **K85** | 124.28 | 8.60 | 124.20 | 8.64 | 124.05 | 8.72 | 123.97 | 8.76 | 123.89 | 8.79 | 123.83 | 8.82 |
| **Q86** | 125.54 | 8.72 | 125.52 | 8.75 | 125.47 | 8.79 | 125.42 | 8.81 | 125.36 | 8.82 | 125.25 | 8.83 |
| **G87** | 117.05 | 8.13 | 117.05 | 8.15 | 117.04 | 8.17 | 117.03 | 8.19 | 117.02 | 8.20 | 117.01 | 8.21 |

**Table B2:** $^{1}H$ and $^{15}N$ chemical shifts (δ) table for Im9 at different Temperatures

| | δ (ppm) | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 283K | | 288K | | 293K | | 298K | | 303K | | 308K | | 313K | | 318K | | 323K | |
| | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ |
| **L3** | 124.68 | 8.46 | 124.68 | 8.47 | 124.68 | 8.48 | 124.68 | 8.49 | 124.67 | 8.50 | 124.66 | 8.51 | 124.64 | 8.52 | 124.62 | 8.52 | 124.60 | 8.53 |
| **K4** | 122.20 | 7.82 | 122.21 | 7.85 | 122.21 | 7.88 | 122.21 | 7.91 | 122.20 | 7.94 | 122.20 | 7.98 | 122.18 | 8.01 | 122.16 | 8.04 | 122.15 | 8.07 |
| **H5** | 115.96 | 8.73 | 115.93 | 8.75 | 115.92 | 8.76 | 115.93 | 8.77 | 115.95 | 8.79 | 115.98 | 8.80 | 116.04 | 8.82 | 116.12 | 8.83 | 116.25 | 8.84 |
| **S6** | 113.74 | 7.92 | 113.68 | 7.95 | 113.63 | 7.97 | 113.59 | 8.00 | 113.56 | 8.02 | 113.50 | 8.05 | 113.47 | 8.09 | 113.45 | 8.12 | 113.46 | 8.16 |
| **I7** | 126.88 | 9.29 | 126.97 | 9.32 | 127.04 | 9.36 | 127.09 | 9.39 | 127.15 | 9.42 | 127.19 | 9.44 | 127.22 | 9.47 | 127.24 | 9.49 | 127.22 | 9.51 |
| **S8** | 112.83 | 7.46 | 112.67 | 7.49 | 112.51 | 7.51 | 112.37 | 7.54 | 112.21 | 7.56 | 112.08 | 7.58 | 111.95 | 7.61 | 111.81 | 7.63 | 113.03 | 7.78 |
| **D9** | 118.42 | 7.85 | 118.47 | 7.89 | 118.54 | 7.93 | 118.59 | 7.97 | 118.64 | 8.01 | 118.70 | 8.05 | 118.75 | 8.09 | 118.82 | 8.13 | 118.91 | 8.17 |
| **Y10** | 120.59 | 8.20 | 120.59 | 8.24 | 120.59 | 8.28 | 120.59 | 8.32 | 120.60 | 8.35 | 120.60 | 8.40 | 120.64 | 8.44 | 120.64 | 8.47 | 120.64 | 8.50 |
| **T11** | 112.19 | 8.92 | 112.18 | 8.95 | 112.17 | 8.97 | 112.15 | 9.00 | 112.12 | 9.02 | 112.09 | 9.05 | 112.08 | 9.08 | 112.07 | 9.09 | 112.08 | 9.11 |
| **E12** | 124.25 | 9.25 | 124.26 | 9.29 | 124.26 | 9.33 | 124.27 | 9.36 | 124.27 | 9.39 | 124.28 | 9.42 | 124.28 | 9.46 | 124.27 | 9.48 | 124.24 | 9.51 |
| **A13** | 118.75 | 8.38 | 118.85 | 8.45 | 118.93 | 8.50 | 119.00 | 8.56 | 119.08 | 8.61 | 119.15 | 8.67 | 119.20 | 8.72 | 119.28 | 8.76 | 119.37 | 8.81 |
| **E14** | 119.63 | 7.96 | 119.62 | 7.99 | 119.60 | 8.03 | 119.57 | 8.06 | 119.53 | 8.09 | 119.51 | 8.12 | 119.48 | 8.15 | 119.43 | 8.18 | 119.40 | 8.21 |
| **F15** | 123.98 | 8.69 | 123.94 | 8.73 | 123.88 | 8.76 | 123.84 | 8.79 | 123.80 | 8.83 | 123.75 | 8.86 | 123.72 | 8.90 | 123.67 | 8.93 | 123.60 | 8.96 |
| **L16** | 119.81 | 8.99 | 119.84 | 9.03 | 119.87 | 9.06 | 119.92 | 9.10 | 119.95 | 9.14 | 119.98 | 9.17 | 120.03 | 9.20 | 120.06 | 9.23 | 120.13 | 9.26 |
| **Q17** | 122.46 | 7.84 | 122.50 | 7.88 | 122.50 | 7.93 | 122.51 | 7.97 | 122.50 | 8.01 | 122.49 | 8.05 | 122.46 | 8.10 | 122.43 | 8.13 | 122.35 | 8.18 |
| **L18** | 123.49 | 7.40 | 123.56 | 7.46 | 123.63 | 7.51 | 123.70 | 7.57 | 123.74 | 7.63 | 123.80 | 7.68 | 123.84 | 7.74 | 123.87 | 7.79 | 123.87 | 7.85 |
| **V19** | 117.89 | 8.12 | 117.97 | 8.17 | 118.02 | 8.21 | 118.09 | 8.26 | 118.13 | 8.30 | 118.19 | 8.34 | 118.22 | 8.38 | 118.27 | 8.42 | 118.32 | 8.45 |
| **T20** | 118.11 | 8.69 | 118.04 | 8.71 | 117.94 | 8.73 | 117.85 | 8.74 | 117.75 | 8.75 | 117.65 | 8.77 | 117.53 | 8.78 | 117.42 | 8.78 | | |
| **T21** | 119.55 | 8.11 | 119.54 | 8.13 | 119.54 | 8.16 | 119.54 | 8.19 | 119.55 | 8.22 | 119.55 | 8.25 | 119.56 | 8.28 | 119.55 | 8.31 | 119.52 | 8.33 |
| **I22** | 119.73 | 7.61 | 119.94 | 7.68 | 120.15 | 7.74 | 120.36 | 7.79 | 120.53 | 7.85 | 120.69 | 7.90 | 120.84 | 7.95 | 121.00 | 7.99 | 121.15 | 8.04 |
| **C23** | 119.85 | 8.80 | 119.78 | 8.83 | 119.70 | 8.85 | 119.61 | 8.87 | 119.51 | 8.90 | 119.41 | 8.92 | 119.31 | 8.93 | 119.21 | 8.95 | 119.14 | 8.96 |
| **N24** | 115.45 | 8.05 | 115.52 | 8.09 | 115.57 | 8.13 | 115.63 | 8.17 | 115.66 | 8.20 | 115.73 | 8.24 | 115.80 | 8.27 | 115.87 | 8.31 | 115.99 | 8.34 |
| **A25** | 124.92 | 8.08 | 124.90 | 8.12 | 124.88 | 8.15 | 124.83 | 8.18 | 124.80 | 8.22 | 124.74 | 8.24 | 124.68 | 8.27 | 124.61 | 8.30 | 124.54 | 8.33 |
| **D26** | 119.06 | 8.36 | 119.09 | 8.39 | 119.11 | 8.42 | 119.13 | 8.44 | 119.17 | 8.46 | 119.18 | 8.49 | 119.20 | 8.51 | 119.23 | 8.53 | 119.30 | 8.55 |
| **T27** | 108.62 | 7.68 | 108.71 | 7.73 | 108.80 | 7.78 | 108.89 | 7.82 | 109.00 | 7.87 | 109.11 | 7.91 | 109.24 | 7.96 | 109.36 | 7.99 | 109.52 | 8.04 |
| **S28** | 115.86 | 8.88 | 115.94 | 8.91 | 116.01 | 8.95 | 116.08 | 8.98 | 116.16 | 9.01 | 116.25 | 9.03 | 116.36 | 9.06 | - | - | - | - |
| **S29** | 112.91 | 7.52 | 112.94 | 7.57 | 113.00 | 7.62 | 113.05 | 7.68 | 113.13 | 7.74 | 113.19 | 7.79 | 113.28 | 7.85 | 113.39 | 7.90 | 113.55 | 7.96 |
| **E30** | 125.02 | 9.26 | 125.09 | 9.29 | 125.14 | 9.31 | 125.19 | 9.33 | 125.25 | 9.35 | 125.29 | 9.36 | - | - | - | - | - | - |
| **E31** | 119.63 | 8.77 | 119.64 | 8.79 | 119.64 | 8.81 | 119.64 | 8.83 | 119.65 | 8.85 | 119.66 | 8.87 | 119.67 | 8.90 | 119.68 | 8.92 | 119.71 | 8.93 |
| **E32** | 119.68 | 7.84 | 119.71 | 7.88 | 119.73 | 7.92 | 119.74 | 7.95 | 119.74 | 7.99 | 119.74 | 8.02 | 119.74 | 8.05 | 119.73 | 8.08 | 119.74 | 8.12 |
| **L33** | 120.82 | 7.36 | 120.80 | 7.41 | 120.79 | 7.46 | 120.77 | 7.51 | 120.76 | 7.56 | 120.75 | 7.61 | 120.74 | 7.67 | 120.75 | 7.72 | 120.79 | 7.77 |
| **K35** | 119.78 | 7.66 | 119.84 | 7.69 | 119.88 | 7.74 | 119.94 | 7.77 | 119.97 | 7.81 | 120.03 | 7.85 | 120.06 | 7.89 | 120.10 | 7.93 | 120.17 | 7.97 |
| **L36** | 120.85 | 8.08 | 120.83 | 8.13 | 120.82 | 8.18 | 120.81 | 8.23 | 120.80 | 8.27 | 120.79 | 8.32 | 120.77 | 8.36 | 120.75 | 8.40 | 120.78 | 8.43 |
| **T38** | 117.48 | 8.60 | 117.44 | 8.64 | 117.37 | 8.68 | 117.32 | 8.72 | 117.26 | 8.75 | 117.22 | 8.78 | 117.15 | 8.82 | 117.11 | 8.84 | 117.03 | 8.87 |
| **H39** | 121.47 | 8.16 | 121.58 | 8.21 | 121.67 | 8.27 | 121.77 | 8.32 | 121.85 | 8.37 | 121.92 | 8.41 | 122.00 | 8.46 | 122.08 | 8.50 | 122.11 | 8.54 |
| **F40** | 120.40 | 9.01 | 120.41 | 9.05 | 120.42 | 9.09 | 120.42 | 9.12 | 120.42 | 9.15 | 120.42 | 9.18 | 120.42 | 9.22 | 120.42 | 9.24 | 120.45 | 9.26 |
| **E41** | 122.01 | 8.77 | 122.01 | 8.82 | 122.00 | 8.86 | 121.99 | 8.91 | 121.97 | 8.95 | 121.92 | 8.98 | 121.89 | 9.02 | 121.84 | 9.06 | 121.74 | 9.09 |
| **E42** | 120.30 | 8.43 | 120.37 | 8.48 | 120.41 | 8.53 | 120.47 | 8.57 | 120.50 | 8.62 | 120.54 | 8.66 | 120.58 | 8.70 | 120.60 | 8.74 | 120.64 | 8.77 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **M43** | 114.53 | 8.78 | 114.54 | 8.82 | 114.55 | 8.86 | 114.57 | 8.90 | 114.59 | 8.93 | 114.62 | 8.96 | 114.64 | 8.99 | 114.68 | 9.01 | 114.75 | 9.03 |
| **T44** | 105.53 | 7.78 | 105.57 | 7.82 | 105.60 | 7.87 | 105.63 | 7.91 | 105.68 | 7.95 | 105.72 | 7.98 | 105.79 | 8.02 | 105.86 | 8.06 | 105.95 | 8.10 |
| **E45** | 111.19 | 7.22 | 111.30 | 7.28 | 111.41 | 7.33 | 111.53 | 7.39 | 111.64 | 7.44 | 111.75 | 7.49 | 111.88 | 7.55 | 112.01 | 7.60 | 112.15 | 7.65 |
| **H46** | 119.71 | 7.38 | 119.76 | 7.42 | 119.83 | 7.46 | 119.89 | 7.50 | 119.95 | 7.54 | 120.04 | 7.58 | 120.10 | 7.62 | 120.18 | 7.66 | 120.29 | 7.71 |
| **G49** | 108.77 | 8.07 | 108.82 | 8.12 | 108.84 | 8.17 | 108.87 | 8.21 | 108.91 | 8.26 | 108.93 | 8.30 | 108.95 | 8.35 | 108.98 | 8.39 | 109.05 | 8.43 |
| **S50** | 120.70 | 10.15 | 120.68 | 10.18 | 120.66 | 10.20 | 120.63 | 10.21 | 120.61 | 10.23 | 120.60 | 10.24 | 120.58 | 10.26 | 120.53 | 10.26 | 120.47 | 10.26 |
| **D51** | 126.67 | 8.67 | 126.72 | 8.72 | 126.75 | 8.76 | 126.76 | 8.80 | 126.77 | 8.84 | 126.77 | 8.87 | 126.75 | 8.90 | 126.71 | 8.93 | 126.61 | 8.95 |
| **L52** | 114.76 | 7.29 | 114.87 | 7.35 | 114.97 | 7.40 | 115.08 | 7.46 | 115.19 | 7.51 | 115.31 | 7.57 | 115.44 | 7.62 | 115.60 | 7.68 | 115.75 | 7.73 |
| **I53** | 111.32 | 6.76 | 111.43 | 6.81 | 111.53 | 6.86 | 111.62 | 6.91 | 111.73 | 6.96 | 111.84 | 7.01 | 111.95 | 7.07 | 112.07 | 7.12 | 112.23 | 7.18 |
| **Y54** | 114.12 | 7.28 | 114.12 | 7.32 | 114.14 | 7.36 | 114.15 | 7.40 | 114.19 | 7.44 | 114.22 | 7.47 | 114.28 | 7.51 | 114.35 | 7.54 | 114.45 | 7.57 |
| **Y55** | 117.57 | 8.72 | 117.65 | 8.75 | 117.71 | 8.78 | 117.79 | 8.81 | 117.88 | 8.83 | 117.97 | 8.85 | 118.06 | 8.87 | 118.18 | 8.88 | 118.33 | 8.88 |
| **K57** | 122.86 | 8.79 | 122.85 | 8.82 | 122.84 | 8.85 | 122.82 | 8.88 | 122.80 | 8.91 | 122.78 | 8.94 | 122.76 | 8.96 | 122.75 | 8.98 | 122.71 | 8.99 |
| **E58** | 123.80 | 8.63 | 123.73 | 8.66 | 123.67 | 8.68 | 123.60 | 8.70 | 123.52 | 8.72 | 123.44 | 8.74 | 123.36 | 8.76 | 123.26 | 8.77 | 123.15 | 8.79 |
| **G59** | 115.20 | 8.93 | 115.10 | 8.95 | 114.99 | 8.96 | 114.88 | 8.98 | 114.76 | 8.99 | 114.61 | 9.00 | 114.48 | 9.00 | 114.25 | 9.01 | 113.99 | 9.00 |
| **D60** | 121.99 | 7.96 | 122.00 | 8.00 | 122.00 | 8.04 | 122.01 | 8.08 | 122.00 | 8.13 | 121.99 | 8.17 | 121.98 | 8.21 | 121.92 | 8.25 | 121.86 | 8.30 |
| **D61** | 121.47 | 8.64 | 121.45 | 8.66 | 121.43 | 8.67 | 121.40 | 8.69 | 121.36 | 8.70 | 121.74 | 8.86 | 121.27 | 8.71 | 121.80 | 8.94 | 121.13 | 8.72 |
| **D62** | 121.47 | 8.64 | 121.55 | 8.69 | 121.59 | 8.73 | 121.65 | 8.78 | 121.69 | 8.82 | 121.33 | 8.71 | 121.77 | 8.90 | 121.22 | 8.72 | 121.86 | 8.97 |
| **S63** | 125.57 | 8.29 | 125.57 | 8.34 | 125.56 | 8.39 | 125.54 | 8.44 | 125.50 | 8.49 | 125.45 | 8.53 | 125.36 | 8.58 | 125.25 | 8.61 | 125.06 | 8.65 |
| **P64** | 115.56 | 8.01 | 115.62 | 8.07 | 115.71 | 8.12 | 115.74 | 8.18 | 115.86 | 8.24 | 115.94 | 8.29 | 116.02 | 8.35 | 116.10 | 8.40 | 116.21 | 8.46 |
| **S65** | 109.96 | 8.24 | 110.00 | 8.27 | 110.02 | 8.29 | 110.05 | 8.31 | 110.10 | 8.33 | 110.13 | 8.35 | 110.18 | 8.37 | 110.24 | 8.39 | 110.34 | 8.40 |
| **G66** | 115.20 | 8.00 | 115.17 | 8.05 | 115.12 | 8.09 | 115.09 | 8.14 | 115.06 | 8.18 | 115.00 | 8.23 | 114.94 | 8.27 | 114.85 | 8.32 | 114.72 | 8.36 |
| **I67** | 125.35 | 8.74 | 125.37 | 8.78 | 125.37 | 8.81 | 125.37 | 8.84 | 125.37 | 8.87 | 125.36 | 8.89 | 125.36 | 8.91 | 125.33 | 8.92 | 125.25 | 8.92 |
| **V68** | 119.50 | 8.33 | 119.58 | 8.37 | 119.64 | 8.40 | 119.72 | 8.43 | 119.78 | 8.47 | 119.84 | 8.50 | 119.90 | 8.54 | 119.95 | 8.56 | 119.97 | 8.59 |
| **N69** | 118.46 | 8.19 | 118.45 | 8.24 | 118.43 | 8.29 | 118.42 | 8.33 | 118.40 | 8.37 | 118.36 | 8.41 | 118.34 | 8.45 | 118.33 | 8.48 | 118.32 | 8.51 |
| **T70** | 120.58 | 8.79 | 120.59 | 8.83 | 120.59 | 8.87 | 120.58 | 8.90 | 120.56 | 8.93 | 120.52 | 8.96 | 120.48 | 8.98 | 120.41 | 9.00 | 120.33 | 9.02 |
| **V71** | 122.09 | 8.51 | 122.10 | 8.55 | 122.10 | 8.59 | 122.10 | 8.62 | 122.09 | 8.65 | 122.09 | 8.69 | 122.11 | 8.72 | 122.11 | 8.75 | 122.11 | 8.77 |
| **K72** | 120.88 | 9.20 | 120.89 | 9.23 | 120.90 | 9.26 | 120.91 | 9.30 | 120.91 | 9.33 | 120.92 | 9.35 | 120.93 | 9.38 | 120.95 | 9.40 | 121.01 | 9.43 |
| **Q73** | 118.04 | 8.72 | 118.14 | 8.75 | 118.24 | 8.79 | 118.33 | 8.82 | 118.43 | 8.85 | 118.52 | 8.88 | 118.59 | 8.91 | 118.66 | 8.93 | 118.77 | 8.96 |
| **W74** | 121.98 | 8.44 | 122.00 | 8.49 | 122.02 | 8.54 | 122.04 | 8.60 | 122.09 | 8.65 | 122.09 | 8.69 | 122.09 | 8.74 | 122.10 | 8.79 | 122.12 | 8.83 |
| **R75** | 119.13 | 8.99 | 119.11 | 9.03 | 119.10 | 9.07 | 119.08 | 9.10 | 119.06 | 9.14 | 119.04 | 9.17 | 119.01 | 9.20 | 118.99 | 9.23 | 119.00 | 9.25 |
| **A76** | 120.34 | 7.64 | 120.40 | 7.70 | 120.48 | 7.76 | 120.53 | 7.81 | 120.60 | 7.87 | 120.68 | 7.92 | 120.73 | 7.98 | 120.81 | 8.03 | 120.91 | 8.08 |
| **A77** | 120.30 | 7.92 | 120.36 | 7.98 | 120.39 | 8.03 | 120.43 | 8.08 | 120.47 | 8.12 | 120.50 | 8.17 | 120.54 | 8.22 | 120.59 | 8.27 | 120.66 | 8.32 |
| **N78** | 113.95 | 6.98 | 113.89 | 7.01 | 113.80 | 7.04 | 113.72 | 7.07 | 113.65 | 7.10 | 113.57 | 7.13 | 113.49 | 7.16 | 113.41 | 7.19 | 113.31 | 7.22 |
| **G79** | 107.86 | 7.54 | 107.93 | 7.58 | 107.98 | 7.63 | 108.04 | 7.68 | 108.08 | 7.73 | 108.14 | 7.77 | 108.19 | 7.82 | 108.25 | 7.87 | 108.30 | 7.92 |
| **K80** | 118.13 | 7.98 | 118.20 | 8.03 | 118.25 | 8.07 | 118.32 | 8.12 | 118.36 | 8.16 | 118.42 | 8.21 | 118.45 | 8.25 | 118.52 | 8.29 | 118.60 | 8.33 |
| **S81** | 118.33 | 8.61 | 118.36 | 8.66 | 118.38 | 8.72 | 118.40 | 8.76 | 118.42 | 8.81 | 118.43 | 8.86 | 118.43 | 8.91 | 118.42 | 8.95 | 118.34 | 8.98 |
| **G82** | 119.29 | 8.28 | 119.32 | 8.32 | 119.38 | 8.37 | 119.41 | 8.41 | 119.44 | 8.45 | 119.49 | 8.49 | 119.51 | 8.53 | 119.56 | 8.56 | 119.63 | 8.59 |
| **F83** | 116.16 | 7.84 | 116.20 | 7.85 | 116.25 | 7.87 | 116.29 | 7.88 | 116.34 | 7.90 | 116.38 | 7.91 | 116.43 | 7.92 | 116.49 | 7.93 | 116.59 | 7.94 |
| **K84** | 124.12 | 8.57 | 124.07 | 8.61 | 124.04 | 8.65 | 124.00 | 8.69 | 123.95 | 8.72 | 123.91 | 8.76 | 123.85 | 8.80 | 123.81 | 8.83 | 123.75 | 8.86 |
| **Q85** | 125.64 | 8.75 | 125.62 | 8.77 | 125.58 | 8.79 | 125.56 | 8.81 | 125.53 | 8.83 | 125.49 | 8.85 | 125.45 | 8.87 | 125.39 | 8.88 | 125.30 | 8.89 |
| **G86** | 117.11 | 8.13 | 117.10 | 8.14 | 117.07 | 8.16 | 117.06 | 8.17 | 117.04 | 8.18 | 117.03 | 8.19 | 117.02 | 8.21 | 117.01 | 8.22 | 117.00 | 8.23 |

**Table B3:** $^{1}H$ and $^{15}N$ chemical shifts (δ) table for Im7 at different Urea concentrations

| | δ (ppm) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0M | | 0.4M | | 0.9M | | 2M | | 3M | |
| | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ |
| **E2** | 125.44 | 8.79 | 125.44 | 8.79 | 125.47 | 8.81 | 125.55 | 8.85 | 125.56 | 8.90 |
| **L3** | 125.75 | 8.55 | 125.75 | 8.55 | 125.78 | 8.56 | 125.86 | 8.57 | 125.87 | 8.59 |
| **K4** | 123.25 | 7.89 | 123.25 | 7.89 | 123.22 | 7.90 | 123.19 | 7.92 | 123.19 | 7.95 |
| **N5** | 115.81 | 8.53 | 115.81 | 8.53 | 115.86 | 8.54 | 115.85 | 8.56 | 115.95 | 8.58 |
| **S6** | 112.26 | 8.17 | 112.26 | 8.17 | 112.22 | 8.17 | 112.18 | 8.18 | 112.17 | 8.19 |
| **I7** | 125.88 | 9.40 | 125.88 | 9.40 | 125.89 | 9.41 | 125.88 | 9.43 | 125.98 | 9.45 |
| **S8** | 112.38 | 7.45 | 112.38 | 7.45 | 112.39 | 7.47 | 112.48 | 7.49 | 112.50 | 7.51 |
| **D9** | 118.09 | 7.80 | 118.09 | 7.80 | 118.05 | 7.80 | 118.01 | 7.81 | 117.99 | 7.83 |
| **Y10** | 120.36 | 8.22 | 120.36 | 8.22 | 120.37 | 8.23 | 120.37 | 8.25 | 120.39 | 8.27 |
| **T11** | 112.54 | 8.91 | 112.54 | 8.91 | 112.54 | 8.91 | 112.54 | 8.91 | 112.51 | 8.93 |
| **E12** | 122.00 | 9.20 | 122.00 | 9.20 | 122.00 | 9.20 | 122.11 | 9.22 | 122.11 | 9.24 |
| **A13** | 118.77 | 8.39 | 118.77 | 8.39 | 118.76 | 8.40 | 118.77 | 8.42 | 118.86 | 8.44 |
| **E14** | 119.55 | 8.06 | 119.55 | 8.06 | 119.52 | 8.07 | 119.51 | 8.08 | 119.51 | 8.08 |
| **F15** | 124.47 | 8.60 | 124.47 | 8.60 | 124.47 | 8.61 | 124.37 | 8.63 | 124.37 | 8.64 |
| **V16** | 119.95 | 8.66 | 119.95 | 8.66 | 119.94 | 8.66 | 119.95 | 8.68 | 119.95 | 8.69 |
| **Q17** | 118.09 | 7.59 | 118.09 | 7.59 | 118.12 | 7.59 | 118.12 | 7.61 | 118.22 | 7.63 |
| **L18** | 122.80 | 7.60 | 122.80 | 7.60 | 122.77 | 7.60 | 122.75 | 7.61 | 122.73 | 7.63 |
| **L19** | 119.29 | 7.78 | 119.29 | 7.78 | 119.29 | 7.79 | 119.29 | 7.81 | 119.19 | 7.82 |
| **K20** | 120.90 | 8.45 | 120.90 | 8.45 | 120.91 | 8.46 | 120.82 | 8.46 | 120.80 | 8.48 |
| **E21** | 123.46 | 8.16 | 123.46 | 8.16 | 123.40 | 8.17 | 123.29 | 8.17 | 123.18 | 8.19 |
| **I22** | 120.76 | 8.05 | 120.76 | 8.05 | 120.71 | 8.06 | 120.70 | 8.07 | 120.68 | 8.08 |
| **E23** | 118.86 | 7.80 | 118.86 | 7.80 | 118.86 | 7.81 | 118.87 | 7.82 | 118.87 | 7.84 |
| **K24** | 119.42 | 7.62 | 119.42 | 7.62 | 119.41 | 7.63 | 119.51 | 7.63 | 119.51 | 7.64 |
| **E25** | 119.17 | 8.31 | 119.17 | 8.31 | 119.17 | 8.31 | 119.30 | 8.32 | 119.30 | 8.34 |
| **N26** | 119.20 | 8.51 | 119.20 | 8.51 | 119.17 | 8.52 | 119.08 | 8.53 | 118.98 | 8.55 |
| **V27** | 112.55 | 7.32 | 112.55 | 7.32 | 112.67 | 7.35 | 112.93 | 7.41 | 113.17 | 7.45 |
| **A28** | 124.83 | 7.53 | 124.83 | 7.53 | 124.92 | 7.57 | 125.11 | 7.64 | 125.22 | 7.70 |
| **A29** | 121.94 | 8.54 | 121.94 | 8.54 | 122.00 | 8.56 | 122.21 | 8.58 | 122.31 | 8.60 |
| **T30** | 109.05 | 8.05 | 109.05 | 8.05 | 109.04 | 8.05 | 108.94 | 8.05 | 108.84 | 8.06 |
| **D31** | 121.90 | 8.57 | 121.90 | 8.57 | 121.89 | 8.58 | 121.79 | 8.58 | 121.79 | 8.60 |
| **D32** | 119.04 | 8.36 | 119.04 | 8.36 | 119.08 | 8.37 | 119.08 | 8.40 | 119.19 | 8.42 |
| **V33** | 120.26 | 8.06 | 120.26 | 8.06 | 120.28 | 8.08 | 120.37 | 8.12 | 120.59 | 8.16 |
| **L34** | 122.44 | 8.71 | 122.44 | 8.71 | 122.47 | 8.71 | 122.54 | 8.72 | 122.65 | 8.73 |
| **D35** | 115.76 | 8.40 | 115.76 | 8.40 | 115.76 | 8.40 | 115.85 | 8.41 | 115.85 | 8.42 |
| **V36** | 120.42 | 7.27 | 120.42 | 7.27 | 120.38 | 7.29 | 120.28 | 7.32 | 120.26 | 7.35 |
| **L37** | 120.96 | 8.20 | 120.96 | 8.20 | 120.92 | 8.21 | 120.90 | 8.23 | 120.81 | 8.25 |
| **L38** | 119.01 | 9.09 | 119.01 | 9.09 | 119.03 | 9.10 | 119.08 | 9.11 | 119.08 | 9.12 |
| **E39** | 119.96 | 7.99 | 119.96 | 7.99 | 119.95 | 8.01 | 119.95 | 8.03 | 120.04 | 8.06 |
| **H40** | 119.63 | 8.05 | 119.63 | 8.05 | 119.63 | 8.06 | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **F41** | 118.56 | 8.88 | 118.56 | 8.88 | 118.55 | 8.89 | 118.54 | 8.90 | 118.54 | 8.92 |
| **V42** | 121.22 | 8.50 | 121.22 | 8.50 | 121.21 | 8.51 | 121.14 | 8.54 | 121.12 | 8.56 |
| **K43** | 120.94 | 8.13 | 120.94 | 8.13 | 120.92 | 8.14 | 120.92 | 8.16 | 121.00 | 8.18 |
| **T45** | 109.33 | 7.56 | 109.33 | 7.56 | 109.31 | 7.56 | 109.31 | 7.57 | 109.31 | 7.59 |
| **E46** | 112.49 | 7.30 | 112.49 | 7.30 | 112.50 | 7.31 | 112.49 | 7.32 | 112.58 | 7.34 |
| **H47** | 119.68 | 7.11 | 119.68 | 7.11 | 119.63 | 7.11 | 119.63 | 7.11 | 119.63 | 7.11 |
| **D49** | 107.88 | 8.01 | 107.88 | 8.01 | 107.87 | 8.02 | 107.87 | 8.03 | 107.90 | 8.05 |
| **G50** | 114.76 | 9.31 | 114.76 | 9.31 | 114.77 | 9.32 | 114.86 | 9.34 | 114.88 | 9.36 |
| **T51** | 126.20 | 8.78 | 126.20 | 8.78 | 126.20 | 8.80 | 126.20 | 8.84 | 126.20 | 8.87 |
| **D52** | 115.65 | 7.28 | 115.65 | 7.28 | 115.64 | 7.30 | 115.72 | 7.32 | 115.74 | 7.34 |
| **L53** | 114.21 | 7.02 | 114.21 | 7.02 | 114.12 | 7.03 | 114.11 | 7.05 | 114.01 | 7.07 |
| **I54** | 114.61 | 7.19 | 114.61 | 7.19 | 114.57 | 7.21 | 114.55 | 7.24 | 114.55 | 7.26 |
| **Y55** | 118.11 | 8.60 | 118.11 | 8.60 | 118.11 | 8.61 | 118.00 | 8.63 | 118.01 | 8.64 |
| **Y56** | 117.08 | 8.65 | 117.08 | 8.65 | 117.15 | 8.64 | 117.25 | 8.62 | 117.36 | 8.61 |
| **S58** | 121.57 | 8.58 | 121.57 | 8.58 | 121.57 | 8.57 | 121.53 | 8.54 | 121.45 | 8.51 |
| **D59** | 115.68 | 8.31 | 115.68 | 8.31 | 115.73 | 8.31 | 115.73 | 8.30 | 115.84 | 8.31 |
| **N60** | 116.72 | 7.58 | 116.72 | 7.58 | 116.60 | 7.58 | 116.39 | 7.59 | 116.28 | 7.60 |
| **R61** | 121.08 | 8.62 | 121.08 | 8.62 | 121.03 | 8.63 | 120.90 | 8.65 | 120.79 | 8.67 |
| **D62** | 124.32 | 8.71 | 124.32 | 8.71 | 124.36 | 8.72 | 124.37 | 8.74 | 124.39 | 8.76 |
| **D63** | 117.46 | 7.91 | 117.46 | 7.91 | 117.42 | 7.90 | 117.36 | 7.90 | 117.34 | 7.90 |
| **S64** | 116.13 | 9.03 | 116.13 | 9.03 | 116.17 | 9.04 | 116.28 | 9.06 | 116.38 | 9.09 |
| **E66** | 112.24 | 8.55 | 112.24 | 8.55 | 112.23 | 8.56 | 112.19 | 8.58 | 112.18 | 8.60 |
| **G67** | 124.43 | 8.29 | 124.43 | 8.29 | 124.38 | 8.30 | 124.37 | 8.31 | 124.37 | 8.33 |
| **I68** | 118.84 | 8.30 | 118.84 | 8.30 | 118.84 | 8.30 | 118.89 | 8.31 | 118.98 | 8.32 |
| **V69** | 120.14 | 8.01 | 120.14 | 8.01 | 120.16 | 8.03 | 120.16 | 8.05 | 120.17 | 8.08 |
| **K70** | 119.22 | 7.77 | 119.22 | 7.77 | 119.19 | 7.78 | 119.19 | 7.78 | 119.19 | 7.79 |
| **I72** | 119.19 | 8.59 | 119.19 | 8.59 | 119.19 | 8.59 | 119.09 | 8.60 | 119.07 | 8.61 |
| **K73** | 120.05 | 9.27 | 120.05 | 9.27 | 119.96 | 9.27 | 119.95 | 9.28 | 119.93 | 9.29 |
| **E74** | 119.30 | 8.58 | 119.30 | 8.58 | 119.30 | 8.59 | 119.29 | 8.61 | 119.29 | 8.63 |
| **W75** | 122.79 | 8.86 | 122.79 | 8.86 | 122.76 | 8.86 | 122.65 | 8.86 | 122.64 | 8.87 |
| **R76** | 118.70 | 9.21 | 118.70 | 9.21 | 118.66 | 9.21 | 118.65 | 9.20 | 118.62 | 9.21 |
| **A77** | 120.39 | 7.80 | 120.39 | 7.80 | 120.44 | 7.82 | 120.49 | 7.86 | 120.59 | 7.89 |
| **A78** | 120.70 | 7.96 | 120.70 | 7.96 | 120.71 | 7.96 | 120.70 | 7.97 | 120.70 | 7.98 |
| **N79** | 113.68 | 7.06 | 113.68 | 7.06 | 113.58 | 7.07 | 113.48 | 7.08 | 113.36 | 7.09 |
| **G80** | 108.30 | 7.56 | 108.30 | 7.56 | 108.30 | 7.57 | 108.40 | 7.59 | 108.40 | 7.61 |
| **K81** | 119.38 | 8.09 | 119.38 | 8.09 | 119.40 | 8.10 | 119.41 | 8.10 | 119.40 | 8.11 |
| **G83** | 110.77 | 8.44 | 110.77 | 8.44 | 110.77 | 8.45 | 110.77 | 8.46 | 110.77 | 8.47 |
| **F84** | 116.39 | 7.96 | 116.39 | 7.96 | 116.40 | 7.97 | 116.48 | 7.98 | 116.49 | 8.00 |
| **K85** | 124.26 | 8.61 | 124.26 | 8.61 | 124.25 | 8.62 | 124.16 | 8.63 | 124.07 | 8.64 |
| **Q86** | 125.56 | 8.74 | 125.56 | 8.74 | 125.58 | 8.75 | 125.67 | 8.77 | 125.67 | 8.79 |
| **G87** | 117.07 | 8.15 | 117.07 | 8.15 | 117.08 | 8.16 | 117.14 | 8.18 | 117.13 | 8.20 |

**Table B4:** $^{1}H$ and $^{15}N$ chemical shifts (δ) table for Im9 at different Urea concentrations

| | δ (ppm) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0M | | 0.4M | | 0.9M | | 2M | | 3M | | 4M | | 5M | |
| | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ | $^{15}N$ | $^{1}H$ |
| **L3** | 124.68 | 8.46 | 124.68 | 8.47 | 124.72 | 8.48 | 124.77 | 8.50 | 124.79 | 8.52 | 124.87 | 8.56 | 124.87 | 8.56 |
| **K4** | 122.21 | 7.81 | 122.25 | 7.84 | 122.33 | 7.87 | 122.31 | 7.93 | 122.34 | 7.98 | 122.40 | 8.06 | 122.41 | 8.07 |
| **H5** | 115.96 | 8.73 | 116.02 | 8.74 | 116.05 | 8.76 | 116.10 | 8.80 | 116.20 | 8.83 | 116.39 | 8.89 | 116.39 | 8.89 |
| **S6** | 113.72 | 7.92 | 113.72 | 7.92 | 113.71 | 7.93 | 113.70 | 7.93 | 113.68 | 7.94 | 113.63 | 7.96 | 113.63 | 7.96 |
| **I7** | 126.88 | 9.29 | 126.93 | 9.30 | 126.95 | 9.31 | 127.03 | 9.33 | 127.08 | 9.34 | 127.15 | 9.37 | 127.15 | 9.37 |
| **S8** | 112.83 | 7.46 | 112.82 | 7.47 | 112.80 | 7.47 | 112.73 | 7.49 | 112.69 | 7.50 | 112.59 | 7.52 | 112.59 | 7.52 |
| **D9** | 118.42 | 7.85 | 118.41 | 7.86 | 118.39 | 7.87 | 118.36 | 7.88 | 118.35 | 7.89 | 118.33 | 7.91 | 118.32 | 7.92 |
| **Y10** | 120.59 | 8.20 | 120.61 | 8.22 | 120.62 | 8.23 | 120.65 | 8.25 | 120.68 | 8.27 | 120.71 | 8.32 | 120.71 | 8.32 |
| **T11** | 112.19 | 8.92 | 112.19 | 8.92 | 112.19 | 8.93 | 112.18 | 8.94 | 112.18 | 8.95 | 112.16 | 8.97 | 112.16 | 8.97 |
| **E12** | 124.25 | 9.25 | 124.26 | 9.27 | 124.28 | 9.28 | 124.34 | 9.29 | 124.38 | 9.31 | 124.45 | 9.34 | 124.45 | 9.34 |
| **A13** | 118.75 | 8.38 | 118.78 | 8.40 | 118.82 | 8.42 | 118.87 | 8.45 | 118.91 | 8.48 | 119.00 | 8.52 | 119.00 | 8.52 |
| **E14** | 119.63 | 7.96 | 119.64 | 7.97 | 119.63 | 7.97 | 119.62 | 7.98 | 119.62 | 7.99 | 119.60 | 8.01 | 119.60 | 8.01 |
| **F15** | 123.98 | 8.69 | 123.97 | 8.70 | 123.95 | 8.71 | - | - | - | - | - | - | - | - |
| **L16** | 119.81 | 8.99 | 119.81 | 9.00 | 119.82 | 9.01 | 119.86 | 9.02 | 119.87 | 9.04 | 119.90 | 9.07 | 119.90 | 9.07 |
| **Q17** | 122.46 | 7.84 | 122.47 | 7.85 | 122.46 | 7.86 | 122.57 | 7.88 | 122.60 | 7.90 | 122.68 | 7.93 | 122.68 | 7.93 |
| **L18** | 123.49 | 7.40 | 123.48 | 7.41 | 123.47 | 7.42 | 123.47 | 7.44 | 123.47 | 7.46 | 123.48 | 7.50 | 123.48 | 7.50 |
| **V19** | 117.89 | 8.12 | 117.91 | 8.13 | 117.91 | 8.15 | 117.93 | 8.17 | 117.95 | 8.19 | 117.97 | 8.23 | 117.97 | 8.23 |
| **T20** | 118.33 | 8.61 | 118.39 | 8.62 | 118.46 | 8.64 | 118.55 | 8.66 | 118.61 | 8.68 | 118.71 | 8.71 | 118.71 | 8.71 |
| **T21** | 119.55 | 8.11 | 119.60 | 8.12 | 119.62 | 8.14 | 119.68 | 8.16 | 119.74 | 8.19 | 119.83 | 8.22 | 119.83 | 8.23 |
| **I22** | 119.73 | 7.61 | 119.75 | 7.63 | 119.78 | 7.65 | - | - | - | - | - | - | - | - |
| **C23** | 119.85 | 8.80 | 119.86 | 8.80 | 119.88 | 8.80 | 119.84 | 8.81 | 119.88 | 8.82 | 119.91 | 8.83 | 119.92 | 8.83 |
| **N24** | 115.45 | 8.05 | 115.45 | 8.06 | 115.44 | 8.07 | 115.41 | 8.08 | 115.39 | 8.10 | - | - | - | - |
| **A25** | 124.92 | 8.08 | 124.94 | 8.10 | 124.96 | 8.11 | 125.00 | 8.14 | 125.02 | 8.16 | 125.04 | 8.20 | 125.04 | 8.20 |
| **D26** | 119.06 | 8.36 | 119.08 | 8.37 | 119.09 | 8.39 | 119.11 | 8.41 | 119.13 | 8.43 | 119.18 | 8.47 | 119.18 | 8.47 |
| **T27** | 108.62 | 7.68 | 108.61 | 7.70 | 108.60 | 7.72 | 108.59 | 7.75 | 108.59 | 7.77 | 108.58 | 7.81 | 108.58 | 7.81 |
| **S28** | 115.86 | 8.88 | 115.89 | 8.90 | 115.89 | 8.92 | 115.91 | 8.97 | 115.92 | 9.00 | 115.95 | 9.06 | 115.95 | 9.06 |
| **S29** | 112.91 | 7.52 | 112.85 | 7.52 | 112.80 | 7.52 | - | - | - | - | - | - | - | - |
| **E30** | 125.02 | 9.26 | 125.11 | 9.28 | 125.19 | 9.30 | 125.38 | 9.34 | 125.52 | 9.37 | 125.76 | 9.43 | 125.77 | 9.43 |
| **E31** | 119.63 | 8.77 | 119.65 | 8.78 | 119.65 | 8.80 | 119.73 | 8.82 | 119.74 | 8.85 | 119.77 | 8.89 | 119.77 | 8.89 |
| **E32** | 119.68 | 7.84 | 119.66 | 7.85 | 119.64 | 7.86 | 119.61 | 7.88 | 119.58 | 7.89 | 119.53 | 7.92 | 119.52 | 7.92 |
| **L33** | 120.82 | 7.36 | 120.81 | 7.37 | 120.80 | 7.37 | 120.78 | 7.37 | 120.76 | 7.38 | 120.69 | 7.39 | 120.69 | 7.39 |
| **L36** | 120.85 | 8.08 | 120.86 | 8.09 | 120.87 | 8.10 | 120.89 | 8.12 | 120.90 | 8.14 | 120.93 | 8.17 | 120.92 | 8.18 |
| **T38** | 117.48 | 8.60 | 117.50 | 8.61 | 117.53 | 8.62 | 117.57 | 8.64 | 117.59 | 8.66 | 117.66 | 8.70 | 117.66 | 8.70 |
| **H39** | 121.47 | 8.16 | 121.51 | 8.18 | 121.53 | 8.19 | 121.58 | 8.22 | 121.63 | 8.25 | 121.70 | 8.29 | 121.71 | 8.29 |
| **F40** | 120.40 | 9.01 | 120.40 | 9.02 | 120.40 | 9.03 | 120.40 | 9.05 | 120.39 | 9.07 | 120.39 | 9.10 | 120.39 | 9.10 |
| **E41** | 122.01 | 8.77 | 122.03 | 8.78 | 122.04 | 8.79 | 122.06 | 8.81 | 122.07 | 8.83 | 122.08 | 8.85 | 122.09 | 8.86 |
| **E42** | 120.30 | 8.43 | 120.31 | 8.44 | 120.31 | 8.45 | 120.32 | 8.48 | 120.33 | 8.50 | 120.36 | 8.54 | 120.36 | 8.54 |
| **M43** | 114.53 | 8.78 | 114.51 | 8.79 | 114.50 | 8.80 | 114.49 | 8.82 | 114.47 | 8.83 | 114.41 | 8.85 | 114.42 | 8.86 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **T44** | 105.53 | 7.78 | 105.53 | 7.79 | 105.53 | 7.80 | 105.53 | 7.82 | 105.52 | 7.83 | 105.52 | 7.86 | 105.52 | 7.86 |
| **E45** | 111.19 | 7.22 | 111.19 | 7.23 | 111.20 | 7.24 | 111.24 | 7.27 | 111.28 | 7.29 | 111.34 | 7.33 | 111.33 | 7.33 |
| **H46** | 119.71 | 7.38 | 119.69 | 7.38 | 119.67 | 7.38 | 119.63 | 7.38 | 119.61 | 7.38 | 119.52 | 7.38 | 119.52 | 7.38 |
| **G49** | 108.77 | 8.07 | 108.73 | 8.08 | 108.70 | 8.09 | 108.60 | 8.10 | 108.54 | 8.12 | 108.43 | 8.14 | 108.43 | 8.14 |
| **S50** | 120.74 | 10.19 | 120.74 | 10.19 | 120.78 | 10.22 | 120.83 | 10.28 | 120.91 | 10.33 | 121.02 | 10.42 | 121.03 | 10.43 |
| **D51** | 126.67 | 8.67 | 126.69 | 8.69 | 126.70 | 8.70 | 126.73 | 8.73 | 126.78 | 8.76 | 126.82 | 8.79 | 126.83 | 8.80 |
| **L52** | 114.76 | 7.29 | 114.79 | 7.31 | 114.84 | 7.32 | 114.88 | 7.36 | 114.93 | 7.39 | 115.00 | 7.44 | 115.00 | 7.44 |
| **I53** | 111.32 | 6.76 | 111.37 | 6.77 | 111.40 | 6.78 | 111.43 | 6.81 | 111.47 | 6.83 | 111.54 | 6.86 | 111.54 | 6.86 |
| **Y54** | 114.12 | 7.28 | 114.15 | 7.29 | 114.20 | 7.31 | 114.22 | 7.33 | 114.24 | 7.35 | 114.23 | 7.38 | 114.24 | 7.38 |
| **Y55** | 117.57 | 8.72 | 117.58 | 8.73 | 117.58 | 8.73 | 117.58 | 8.74 | 117.58 | 8.74 | 117.56 | 8.76 | 117.56 | 8.76 |
| **K57** | 122.86 | 8.79 | 122.86 | 8.80 | 122.85 | 8.80 | 122.85 | 8.81 | 122.85 | 8.83 | 122.84 | 8.85 | 122.84 | 8.85 |
| **E58** | 123.80 | 8.63 | 123.78 | 8.65 | 123.78 | 8.67 | - | - | - | - | - | - | - | - |
| **G59** | 115.20 | 8.93 | 115.15 | 8.94 | 115.09 | 8.95 | 114.96 | 8.95 | 114.85 | 8.96 | 114.66 | 8.98 | 114.65 | 8.98 |
| **D60** | 121.99 | 7.96 | 121.95 | 7.96 | 121.92 | 7.96 | 121.84 | 7.97 | 121.81 | 7.97 | 121.74 | 7.98 | 121.74 | 7.98 |
| **D61** | 121.47 | 8.64 | 121.49 | 8.66 | 121.50 | 8.67 | - | - | - | - | - | - | - | - |
| **D62** | 125.57 | 8.29 | 125.63 | 8.30 | 125.67 | 8.31 | 125.74 | 8.33 | 125.79 | 8.34 | 125.84 | 8.38 | 125.84 | 8.38 |
| **S63** | 115.56 | 8.01 | 115.55 | 8.02 | 115.54 | 8.04 | 115.53 | 8.06 | 115.52 | 8.09 | - | - | - | - |
| **S65** | 109.96 | 8.24 | 109.98 | 8.25 | 109.99 | 8.26 | 110.01 | 8.28 | 110.02 | 8.30 | 110.07 | 8.33 | 110.05 | 8.33 |
| **G66** | 115.20 | 8.00 | 115.13 | 8.02 | 115.04 | 8.03 | 114.89 | 8.06 | 114.77 | 8.09 | 114.57 | 8.13 | 114.56 | 8.13 |
| **I67** | 125.35 | 8.74 | 125.32 | 8.76 | 125.31 | 8.77 | 125.28 | 8.79 | 125.27 | 8.81 | 125.20 | 8.85 | 125.20 | 8.85 |
| **V68** | 119.29 | 8.28 | 119.28 | 8.29 | 119.28 | 8.30 | 119.29 | 8.31 | 119.29 | 8.33 | 119.30 | 8.36 | 119.30 | 8.36 |
| **N69** | 118.46 | 8.19 | 118.49 | 8.21 | 118.51 | 8.23 | 118.57 | 8.27 | 118.60 | 8.30 | 118.63 | 8.36 | 118.64 | 8.36 |
| **T70** | 120.58 | 8.79 | 120.59 | 8.80 | 120.61 | 8.81 | 120.64 | 8.83 | 120.66 | 8.84 | 120.69 | 8.87 | 120.69 | 8.87 |
| **V71** | 122.09 | 8.51 | 122.08 | 8.52 | 122.07 | 8.52 | 122.06 | 8.52 | 122.06 | 8.53 | 122.03 | 8.54 | 122.02 | 8.54 |
| **K72** | 120.88 | 9.20 | 120.87 | 9.21 | 120.85 | 9.21 | 120.82 | 9.23 | 120.80 | 9.24 | 120.79 | 9.27 | 120.79 | 9.27 |
| **Q73** | 118.03 | 8.72 | 118.07 | 8.72 | 118.09 | 8.73 | - | - | - | - | - | - | - | - |
| **W74** | 121.98 | 8.44 | 121.97 | 8.45 | 121.95 | 8.46 | 121.94 | 8.47 | 121.91 | 8.49 | 121.88 | 8.51 | 121.90 | 8.52 |
| **R75** | 119.13 | 8.99 | 119.12 | 9.00 | 119.11 | 9.01 | 119.10 | 9.02 | 119.08 | 9.03 | 119.04 | 9.05 | 119.03 | 9.06 |
| **A76** | 120.34 | 7.64 | 120.37 | 7.66 | 120.39 | 7.68 | 120.45 | 7.71 | 120.51 | 7.74 | 120.61 | 7.79 | 120.62 | 7.79 |
| **A77** | 120.30 | 7.92 | 120.30 | 7.93 | 120.30 | 7.95 | 120.30 | 7.97 | 120.30 | 7.98 | 120.30 | 8.02 | 120.30 | 8.02 |
| **N78** | 113.95 | 6.98 | 113.92 | 6.99 | 113.87 | 7.01 | 113.80 | 7.03 | 113.73 | 7.05 | 113.62 | 7.09 | 113.61 | 7.09 |
| **G79** | 107.86 | 7.54 | 107.87 | 7.55 | 107.89 | 7.56 | 107.91 | 7.58 | 107.93 | 7.59 | 107.94 | 7.62 | 107.95 | 7.62 |
| **K80** | 118.13 | 7.98 | 118.15 | 7.99 | 118.18 | 8.00 | 118.20 | 8.01 | 118.22 | 8.02 | 118.24 | 8.04 | 118.24 | 8.04 |
| **S81** | 118.11 | 8.69 | 118.11 | 8.70 | 118.10 | 8.71 | - | - | - | - | - | - | - | - |
| **G82** | 119.50 | 8.33 | 119.48 | 8.34 | 119.45 | 8.35 | 119.39 | 8.36 | 119.36 | 8.37 | 119.30 | 8.39 | 119.30 | 8.39 |
| **F83** | 116.16 | 7.84 | 116.17 | 7.85 | 116.18 | 7.86 | 116.21 | 7.89 | 116.25 | 7.91 | 116.29 | 7.95 | 116.30 | 7.95 |
| **K84** | 124.12 | 8.57 | 124.10 | 8.58 | 124.08 | 8.59 | 124.02 | 8.60 | 123.98 | 8.62 | 123.88 | 8.65 | 123.88 | 8.65 |
| **Q85** | 125.64 | 8.75 | 125.66 | 8.77 | 125.68 | 8.78 | 125.72 | 8.80 | 125.77 | 8.82 | 125.81 | 8.86 | 125.81 | 8.86 |
| **G86** | 117.11 | 8.13 | 117.11 | 8.15 | 117.14 | 8.16 | 117.16 | 8.18 | 117.18 | 8.21 | 117.21 | 8.24 | 117.21 | 8.25 |