

Functional dynamics of the anti-HIV lectin OAA and NMR
methodology for the study of protein dynamics

Dissertation

for the award of the degree

“Doctor of Philosophy”

Division of Mathematics and Natural Sciences
of the Georg-August-Universität Göttingen

within the doctoral program *Molecular Biology*
of the Georg-August University School of Science (GAUSS)

submitted by

Marta Gião Carneiro

From Setúbal, Portugal

Göttingen 2015

Thesis Committee

Prof. Dr. Christian Griesinger

Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Jörg Enderlein

Research group for Single Molecule Spectroscopy and Imaging for Biophysics and Complex Systems, III. Institute of Physics, University of Göttingen, Göttingen, Germany

Dr. Tim Grüne

Paul Scherrer Institute, Villigen, Switzerland

Members of the Examination Board

Prof. Dr. Christian Griesinger

Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Jörg Enderlein

Research group for Single Molecule Spectroscopy and Imaging for Biophysics and Complex Systems, III. Institute of Physics, University of Göttingen, Göttingen, Germany

Dr. Tim Grüne

Paul Scherrer Institute, Villigen, Switzerland

Further members of the Examination Board

Prof. Dr. Kai Tittmann

Department of Bioanalytics, Albrecht-von-Haller-Institute for Plant Sciences, University of Göttingen, Göttingen, Germany

Prof. Dr. Markus Zweckstetter

Research group for Structure Determination of Proteins Using NMR, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Bert de Groot

Research group for Computational Biomolecular Dynamics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Date of oral examination: November 18th 2015

Affidavit

Herewith I declare, that I prepared this thesis titled “Functional dynamics of the anti-HIV lectin OAA and NMR methodology for the study of protein dynamics” on my own and with no other sources and aids than quoted.

Göttingen, September 29th 2015

ACKNOWLEDGEMENTS

This thesis would not be possible without the help and support of many people, to whom I am very grateful.

First and foremost, I would like to express my gratitude to Prof. Dr. Christian Griesinger and Dr. Donghan Lee for the opportunity to work on these interesting and stimulating projects, and especially for all the mentoring and guidance throughout the years.

I would like to thank my thesis committee members, Prof. Dr. Jörg Enderlein and Dr. Tim Grüne, for the constructive discussions and support. I further extend my appreciation to the members of the extended examination board for evaluating my dissertation.

I am also grateful to Prof. Dr. Angela Gronenborn and Dr. Leonardus Koharudin, for a very fruitful collaboration with regard to the OAA project.

I am thankful to Dr. David Band and Dr. Pablo Trigo-Mourino, from whom I have learned so much, to Dr. Michael T. Sabo, Dr. Adam Mazur and Dr. Jithender G. Reddy, with whom I worked closely.

To the “cubic creatures”, for creating such an exceptional working environment, and everyone in NMR II for the friendly atmosphere: Thank you!

I am indebted to our system administrator, Dr. Dirk Bockelmann, and our secretary, Petra Breiner, who made my life much easier on so many occasions.

An especially warm thanks to the friends that never let time and distance come between us, to my “Göttingen family”, and to Jan, who continuously help and support me in so many ways.

Finally, I am profoundly grateful to my family, who is always there for me.

RELATED PUBLICATIONS

The work presented in this thesis led to the following publications:

Carneiro, M. G., Reddy, J. G., Griesinger, C., Lee, D. (2015). Speeding-up exchange-mediated saturation transfer experiments by Fourier transform. *J. Biomol. NMR*, in press

Carneiro, M. G., Koharudin, L. M., Ban, D., Sabo, T. M., Trigo-Mourino, P., Mazur, A., Griesinger, C., Gronenborn, A. M., Lee, D. (2015). Sampling of glycan-bound conformers by the anti-HIV lectin *Oscillatoria agardhii* agglutinin in the absence of sugar, *Angew. Chem. Int. Ed.* *54*, 6462-6465

Carneiro, M. G., Koharudin, L. M., Griesinger, C., Gronenborn, A. M., Lee, D. (2015). ^1H , ^{13}C and ^{15}N resonance assignment of the anti-HIV lectin from *Oscillatoria agardhii*, *Biomol. NMR Assign.* *9*, 317-319

ABSTRACT

Specific recognition between biomolecules is at the basis of all biological processes, and a mechanistic description of molecular recognition is crucial for a detailed understanding of these processes. Two limiting models are currently discussed in the context of molecular recognition: (i) induced-fit, which postulates that conformational changes between free and bound states are the result of the binding interaction and (ii) conformational selection, according to which the binding interaction selects one of the multiple conformers that pre-exist in equilibrium in the absence of binding partner. A necessary (but not sufficient) condition to demonstrate a conformational selection scenario is the characterization of conformational substates in the absence of binding partner comparable to the conformations seen for the bound forms. NMR spectroscopy is a powerful technique for studying conformational heterogeneity in solution, given its sensitivity to a broad range of motional timescales in solution with atomic resolution.

The first part of this thesis is dedicated to the study of functional dynamics of the lectin OAA, aimed at the elucidation of the molecular recognition mechanism underlying its anti-HIV activity, which stems from binding to high-mannose glycans on the viral envelope glycoproteins. Previously determined X-ray crystallography structures identified a distinct conformational change between the free and sugar-bound protein. By using a variety of NMR methods we show that both sugar-free and sugar-bound conformations are conformational substates of the free protein. Further, our results indicate that the sugar-bound conformation is highly populated even in the absence of sugar, suggesting that recognition of high-mannose glycans by OAA proceeds by conformational selection within the ground state. These insights may guide further optimization and/or development of preventive anti-HIV therapeutics.

The second part of the thesis is concerned with the development of new strategies aimed at extending the efficacy and accuracy of two NMR methods frequently used to investigate lowly populated conformational states. We show that the demanding experimental time required by exchange-mediated saturation transfer experiments can be reduced by two-fold by making use of Fourier transform and linear prediction. We also demonstrate that the simultaneous analysis of data collected with at least two radio-frequency field strengths is necessary for extracting reliable exchange parameters from

these experiments. Additionally, we present a method for the identification of dynamic clusters based on model selection using the Akaike information criterion. The efficiency of the method is discussed in the context of synthetic CPMG relaxation dispersion data, but the principles outlined here can be easily applied on the analysis of a variety of experiments.

TABLE OF CONTENTS

Acknowledgements	v
Related publications	vii
Abstract	ix
Table of contents	1
1 Introduction	3
1.1 Molecular recognition mechanisms	3
1.2 Protein dynamics by NMR	5
1.3 The anti-HIV lectin <i>Oscillatoria agardhii</i> agglutinin	13
1.4 Thesis outline	16
2 Resonance assignment of OAA	17
2.1 Introduction	17
2.2 Results	18
2.3 Discussion	25
2.4 Materials and Methods	25
3 Functional dynamics of OAA	28
3.1 Introduction	28
3.2 Results	29
3.3 Discussion	41
3.4 Materials and methods	44
4 Improving NMR methods for the study of protein dynamics	50

4.1	Introduction	50
4.2	Results and discussion	51
4.3	Materials and Methods	68
5	Discussion and conclusion	72
6	References	75
7	Appendix	83
7.1	Exchange parameters from ^{15}N CPMG relaxation dispersion	83
7.2	Flow diagrams of the clustering algorithm	85
7.3	Synthetic datasets used for testing the clustering algorithm	87
7.4	Pulse Programs	90
	Curriculum Vitae	103

1 INTRODUCTION

1.1 MOLECULAR RECOGNITION MECHANISMS

Highly specific interactions between biomolecules are a key aspect in all biological processes. The characterization of the molecular recognition mechanisms governing such interactions is therefore crucial for understanding biology at the molecular level. (Kahn and Plaxco, 2010)

A first attempt to describe enzymatic specificity was proposed by Fischer in the end of the 19th century. (Fischer, 1894) Fischer's "lock-and-key" model described the binding site as a rigid template, whose structure is complementary to the structure of the ligand. The incompatibility of a rigid template model with several enzymatic reactions led to the formulation of a new hypothesis, which postulates that the substrate induces conformational changes at the binding site (Figure 1.1). (Koshland, 1958) Structural differences observed between a variety of intermolecular complexes and its free components (not only enzyme-substrate reactions), has given apparent support to the "induced fit" model as a general mechanism of molecular recognition. (Gerstein et al., 1994)

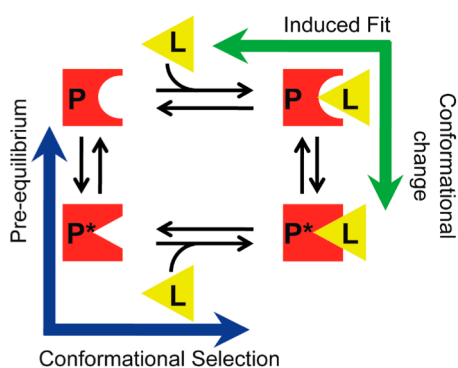


Figure 1.1 Schematic representation of molecular recognition processes via induced fit or conformational selection. The induced fit model (green arrow) postulates that conformational changes in the protein are induced by initial binding to the ligand. The conformational selection model (blue arrow) proposes that a conformation complementary to the ligand is part of an ensemble of conformations sampled by the free protein, which is selected for binding.

Although the induced fit model acknowledges some degree of plasticity in the binding site, it still assumes that the protein adopts a single, stable conformation under given experimental conditions. However, it has become clear that proteins sample a multitude of thermally accessible conformational substates that may be relevant for molecular recognition. (Boehr et al., 2009; Cooper, 1976; Frauenfelder et al., 1991; Yang et al., 2003) Building on this description of the energy landscape of proteins, and extending the funnel model of protein folding onto intermolecular interactions, Nussinov and co-workers generalized the alternative “conformational selection” model (Figure 1.1). (Ma et al., 1999; Tsai et al., 1999) The model shares some key aspects with the Monod-Wyman-Changeux model of allostery, (Monod et al., 1965) and postulates the pre-existence of a conformational equilibrium, with conformers complementary to the ligand being selected for binding, resulting in a population shift towards the favored conformation. The complementary conformation might correspond to the lowest energy conformation (ground state) of the free protein (a limit case which would correspond to the “lock-and-key” model), or to a higher energy conformation (excited state). According to this model, structural differences between free and bound states determined by X-ray crystallography or conventional NMR methods do not suffice to characterize the molecular recognition mechanism, as they most likely report on the lower-energy states, representing an incomplete picture of the structural heterogeneity experienced by the free protein. (Boehr et al., 2009) The characterization of conformational differences between the ground state and excited states (Boehr et al., 2006; Eisenmesser et al., 2005; Lange et al., 2008; Zhang et al., 2007) or within ground states (Carneiro et al., 2015c; Lange et al., 2008) comparable to conformational differences between free and bound conformations has provided substantial support to the conformational selection model. Moreover, it has been suggested that despite being thermodynamically reasonable, an induced fit mechanism would be too slow to be relevant biologically. (Bosshard, 2001)

Despite the core differences between the induced fit and the conformational selection models outlined above, the determination of the molecular recognition mechanism at play is not straightforward. A structural demonstration of the induced fit mechanism would require structural information sampled over the course of the binding event to show how specific interactions lead to conformational changes. Evidence for conformational selection should include the characterization of bound-like conformations in the free state complemented with studies demonstrating that the ligand interacts

preferentially with those conformations. (Boehr et al., 2009; Bosshard, 2001) Additionally, the two mechanisms can in principle be distinguished by monitoring the observed rate constant of binding as a function of varying ligand and protein concentrations, (Gianni et al., 2014; Hammes et al., 2009; Vogt and Di Cera, 2012) or as a function of mutations that affect the conformational equilibrium but not the binding. (Weikl and Boehr, 2012; Weikl and von Deuster, 2009) However, limitations in the experimental techniques (particularly in terms of detection of fast transitions between different, potentially lowly populated, conformational states), (Gianni et al., 2014; Vogt and Di Cera, 2012) and possible indirect effects of mutations on the binding free energies (Weikl and von Deuster, 2009) might render the determination of the binding mechanism difficult in practice.

Ultimately, it is likely that both induced fit and conformational selection mechanisms occur, with the relative concentrations of ligand and protein determining the dominant mechanism, (Greives and Zhou, 2014; Hammes et al., 2009) and that initial binding through a conformational selection mechanism is followed by further (induced) conformational adjustments in the complex. (Anthis et al., 2011; Bucher et al., 2011; Silva et al., 2011; Wlodarski and Zagrovic, 2009) Recently, Nussinov and co-workers proposed an “extended conformational selection model”, which unifies the lock-and-key, the induced fit and the original conformational selection model, by taking into account changes in the energy landscapes due to the encounter between the binding partners. (Csermely et al., 2010) The further characterization of the energy landscape and conformational ensembles might thus prove valuable for the better understanding of protein functionality and potentially guide drug design or protein engineering efforts.

1.2 PROTEIN DYNAMICS BY NMR

NMR spectroscopy has gained popularity in the study of protein (and molecules in general) dynamics due to its unique capability to investigate a broad range of motional timescales in solution (Figure 1.2) with atomic resolution. (Palmer III, 2004)

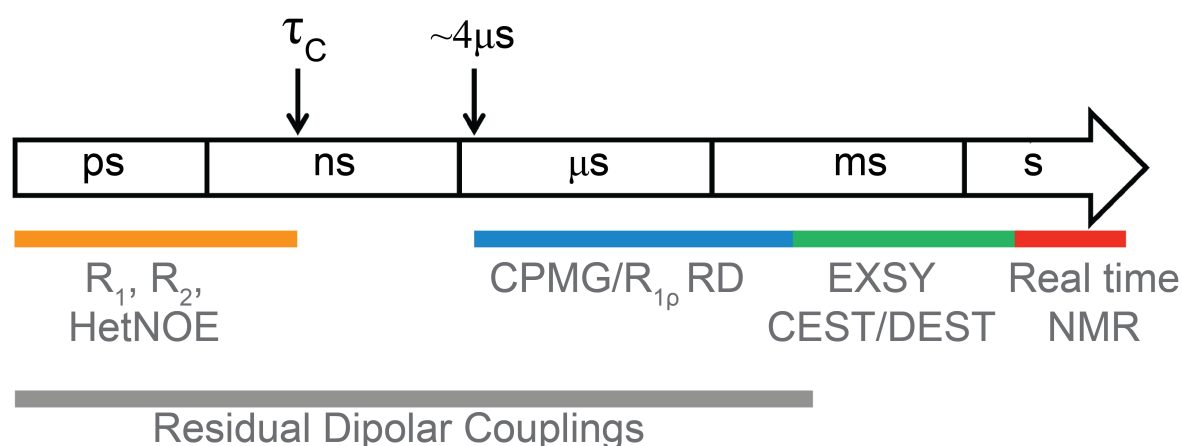


Figure 1.2 Timescales of motions probed by different NMR experiments.

Briefly, fast internal motions occurring on timescales up to the overall correlation time (τ_c , usually in the order of nanoseconds) of the molecule in solution influence the intrinsic nuclear spin relaxation resulting from the time-dependent stochastic modulation of dipole-dipole interactions and electronic environment surrounding the nuclei and can be probed with longitudinal relaxation rates (R_1), transverse relaxation rates (R_2) and heteronuclear NOE (hetNOE) measurements. (Kay et al., 1989) These methods will not be further discussed in this thesis but comprehensive reviews are available. (Cavanagh et al., 2007; Luginbühl and Wüthrich, 2002; Palmer III, 2004) Conformational fluctuations on the slower microsecond-millisecond timescale can be probed by experiments that explore the contribution of conformational exchange (*vide infra*) to the nuclear spin relaxation, namely relaxation dispersion (RD) experiments. (Carr and Purcell, 1954; Davis et al., 1994; Deverell et al., 1970; Meiboom and Gill, 1958) Technical limitations constrain the timescale of motions accessible with these experiments, and even though methodological advances have pushed the sensitivity of RD experiments up to one-digit microseconds (Ban et al., 2012; Smith et al., 2015), motions between this limit and the correlation time τ_c cannot be studied with the techniques mentioned so far. This window of timescales can however be probed with residual dipolar couplings (RDC) measurements. RDCs provide time and conformational-averaged information on the orientation of inter-nuclear vectors up to the millisecond timescale and thus report on the amplitude (but not kinetics) of motions within the nanosecond-millisecond timescale. (Ban et al., 2013b; Jensen et al., 2008; Lange et al., 2008; Nodet et al., 2009; Tolman et al., 1997) Finally, the effects of conformational exchange on a slower timescale (milliseconds-seconds) can be investigated

with exchange spectroscopy (EXSY (Farrow et al., 1994; Jeener et al., 1979)) and exchange-mediated saturation transfer experiments (CEST (Vallurupalli et al., 2012) and DEST (Fawzi et al., 2011)), and even slower motions (occurring within seconds and beyond) can be identified by monitoring changes in consecutively recorded spectra (real time NMR). (Schanda and Brutscher, 2005; Schanda et al., 2006) Methods that report on μ s-ms motions are of special interest to the study of functional dynamics of proteins, as many biological processes occur with time constants within this range, and have been extensively reviewed. (Ban et al., 2013b; Palmer III, 2014; Palmer III et al., 2001)

1.2.1 CONFORMATIONAL EXCHANGE IN NMR SPECTROSCOPY

Conformational exchange (or chemical exchange) corresponds to motions that alter the magnetic environment of a nucleus, resulting in a time-dependent modulation of its resonance frequency that contributes to the magnetization transfer between sites and to the transverse relaxation of the nuclear spin, and is theoretically described by the Bloch-McConnell equation (see also Chapter 4). (McConnell, 1958; Palmer III et al., 2001) Chemical exchange contribution to the transverse relaxation ($R_{2,ex}$) results in larger effective transverse relaxation rates ($R_{2,eff} = R_{2,0} + R_{2,ex}$, where $R_{2,0}$ corresponds to the transverse relaxation rate constant in the absence of chemical exchange) and therefore broadens the resonance line. Additionally the position of the observed resonances can also be affected, depending on the relative values of the resonance frequency difference and the rate of exchange between the different conformations. (Palmer III et al., 2001) Figure 1.3 demonstrates the effects of chemical exchange on the lineshape and position of observable resonances, for equally populated states and for unevenly populated states, assuming a two-state model described by Eq. (1.1):



where a and b refer to two distinct magnetic environments sampled by a nucleus, k_{ab} and k_{ba} are the forward and reverse rate constants, respectively, and the exchange rate $k_{ex} = k_{ab} + k_{ba}$. The equilibrium populations of state a and b correspond to $p_a = k_{ba} / k_{ex}$ and $p_b = 1 - p_a = k_{ab} / k_{ex}$.

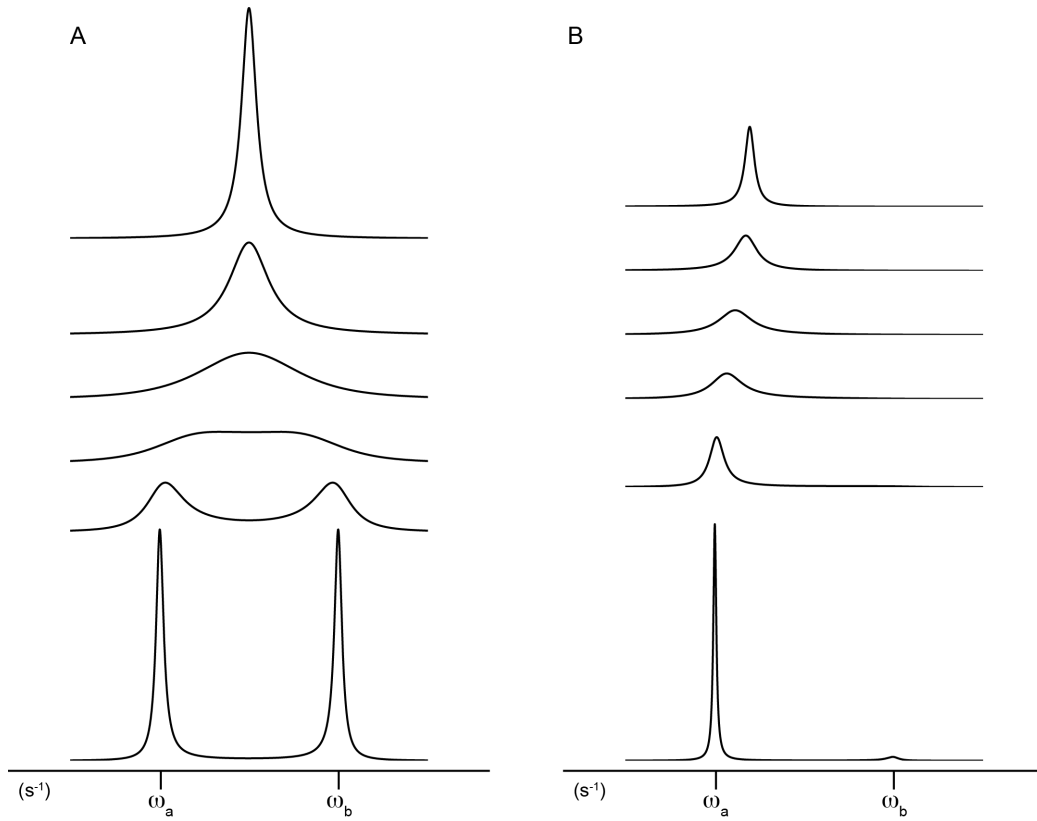


Figure 1.3 Effect of chemical exchange on NMR spectra. Chemical exchange for a two-site system with **(A)** equal populations ($p_a = p_b = 0.5$) and **(B)** skewed populations ($p_a = 0.8$, $p_b = 0.2$). The spectra were calculated using the Bloch-McConnell equation and values of k_{ex} of (from top to bottom in both A and B) 15000, 6000, 3000, 2000, 750 and 150 s^{-1} , $\Delta\omega = |\omega_b - \omega_a| = 3000 s^{-1}$ and $R_{2,0}^a = R_{2,0}^b = 0 s^{-1}$.

If the two states are equally populated (Figure 1.3 A) and the exchange rate slower than the chemical shift difference between the states ($k_{ex} < \Delta\omega$, where $\Delta\omega = |\omega_b - \omega_a|$) well resolved resonances are observed for each state, with the effective relaxation rate constants for each site corresponding to: (Palmer III et al., 2001)

$$\begin{aligned}
 R_{2,eff}^a &= R_{2,0}^a + p_b k_{ex} \\
 R_{2,eff}^b &= R_{2,0}^b + p_a k_{ex}
 \end{aligned}
 \tag{1.2}$$

As k_{ex} increases, the resonances become broader, up to the point where the linewidth is comparable to the chemical shift difference resulting in the coalescence of the two lines ($k_{ex} \approx \Delta\omega$). For $k_{ex} > \Delta\omega$ a single resonance is observed, at the population-averaged chemical shift ($\omega_{average} = p_a \omega_a + p_b \omega_b$), characterized by an also population-averaged transverse relaxation rate constant: (Palmer III et al., 2001)

$$R_{2,eff}^{ave} = p_a R_{2,0}^a + p_b R_{2,0}^b + p_a p_b \Delta\omega^2 / k_{ex} \quad (1.3)$$

From Eq. (1.3) it becomes apparent that for $k_{ex} > \Delta\omega$ the chemical shift information ($\Delta\omega$) and the relative populations are convoluted, and cannot be uniquely determined. Instead, a structural amplitude parameter $\Phi_{ex} = p_a p_b \Delta\omega^2$ is available. (Palmer III et al., 2001)

The relative values of k_{ex} and $\Delta\omega$ are thus used to define the chemical shift timescale: if $k_{ex} < \Delta\omega$ exchange is slow, whereas $k_{ex} > \Delta\omega$ corresponds to fast exchange. At the coalescence point, when $k_{ex} \approx \Delta\omega$, exchange is in the intermediate regime. (Cavanagh et al., 2007; Palmer III et al., 2001) With regard to the NMR experiments shown in Figure 1.2, relaxation dispersion experiments are suitable to study systems in all exchange regimes (provided that the exchange rate falls on the timescale probed by these experiments). EXSY on the other hand, requires well-resolved resonances to be observed for each state and is therefore applicable only to systems in slow exchange.

If the populations between the two states are highly skewed, the signal corresponding to the minor state (i.e., the least populated state) is both less intense (since it is proportional to the population) and broader (since $R_{2,ex}$ is proportional to $p_a k_{ex}$, where p_a is the population of the major state) and may become undetectable even in the slow exchange regime (Figure 1.3 B), rendering this type of systems difficult to study with EXSY. Relaxation dispersion and CEST/DEST experiments on the other hand, rely on changes observed in the major state signal and are applicable even if the minor state signal is not detectable. Such a scenario with highly skewed populations corresponds precisely to interconversions between ground and excited states that might be functionally relevant in the context of molecular recognition. Indeed, relaxation dispersion experiments and more recently exchange-mediated saturation transfer experiments (CEST/DEST) arose as powerful techniques to study conformational exchange solution as they provide kinetic (exchange rate), thermodynamic (relative populations) and structural (chemical shift differences) information, and are sensitive to the presence of lowly populated states. (Palmer III, 2014) In particular, Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (Carr and Purcell, 1954; Meiboom and Gill, 1958) and exchange-mediated saturation transfer (CEST/DEST) (Fawzi et al., 2011; Vallurupalli et al., 2012) experiments are an integral part of the work presented here and are briefly discussed below.

1.2.1.1 CPMG relaxation dispersion

CPMG relaxation dispersion relies on the application of 180° refocusing pulses to probe kinetic processes. In short, once the net magnetization is rotated into the transverse (xy) plane, magnetization corresponding to different sites (with different chemical shifts) will start precessing around the z -axis with different frequencies, resulting in the dephasing of the net transverse magnetization (and consequent line broadening). The degree of dephasing is directly proportional to the time T during which the magnetization rotates freely (i.e., no radiofrequency pulses are applied). In the absence of exchange, if a 180° pulse is applied at $T/2$, the direction of precession is inverted and, after another $T/2$ period, the magnetization is effectively refocused. In contrast, if the precession frequency of a magnetization vector varies stochastically (due to chemical exchange), the application of a 180° pulse in the middle of T will not result in the complete refocusing of the magnetization, leading to a broader linewidth. The refocusing will however improve if more 180° pulses are applied during T , ultimately resulting in complete refocusing if the rate of application of pulses (CPMG frequency $\nu_{\text{CPMG}} = 1/(2\tau)$, where τ is the delay between pulses) is faster than the exchange rate. (Neudecker et al., 2009; Palmer III, 2014) The CPMG relaxation dispersion experiments thus consist in monitoring the signal intensity, $I(\nu_{\text{CPMG}})$, of the major state signal as a function of ν_{CPMG} . In practice, $I(\nu_{\text{CPMG}})$ is converted into an effective transverse relaxation rate ($R_{2,\text{eff}}$, Eq. (3.6)). The relaxation dispersion profiles correspond then to the variation of $R_{2,\text{eff}}$ as a function of ν_{CPMG} , which depends on the exchange rate, relative populations and chemical shift differences between the interconverting states. Examples of typical relaxation dispersion curves for a site without and with exchange contributions to $R_{2,\text{eff}}$ are shown in Figure 1.4.

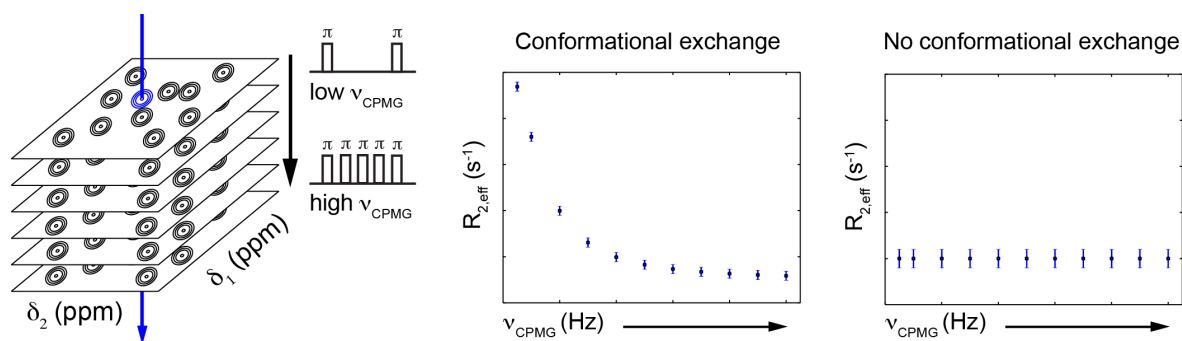


Figure 1.4 Schematic depiction of CPMG relaxation dispersion experiments. A series of 2D experiments is recorded with varying ν_{CPMG} (frequency of 180° pulses). The effective relaxation rate ($R_{2,\text{eff}}$, calculated from the signal intensity) is monitored as a function of ν_{CPMG} . Sites undergoing conformational exchange display a non-linear dependence of $R_{2,\text{eff}}$ on

ν_{CPMG} whereas sites without detectable exchange show no variation of $R_{2,\text{eff}}$. Fitting of the dispersion curves to the appropriate models (Chapter 3) provides kinetic, thermodynamic and structural information on the exchanging system.

1.2.1.2 Exchange-mediated saturation transfer

Conformational exchange with slow exchange rates and lowly populated states result in minute changes in $R_{2,\text{eff}}$ and are therefore difficult to detect and characterize using CPMG relaxation dispersion. (Vallurupalli et al., 2012) Two recently developed experiments — dark-state exchange-mediated saturation transfer (DEST; (Fawzi et al., 2011)) and chemical-exchange mediated saturation transfer (CEST; (Vallurupalli et al., 2012)) — are particularly suited to analyze such situations. Both experiments rely on the observation that, in an exchanging system, the perturbation of one of the sites by radio-frequency irradiation will be (partially) transferred to the other site. (Mayer and Meyer, 1999; Ward et al., 2000) Consequently, for a site undergoing two-state exchange as described by Eq. (1.1), perturbation of the minor-state magnetization (by applying a weak radio-frequency field at its resonance position) will be (partially) transferred to the major-state, causing a decrease in the intensity of the major-state signal. In contrast, radio-frequency irradiation off-resonance with either state will cause no disturbance in the spectrum. CEST/DEST experiments thus consist in monitoring the intensity of a signal of interest as a function of the frequency offset at which the weak radio-frequency field is applied. Typical CEST and DEST profiles are shown in Figure 1.5.

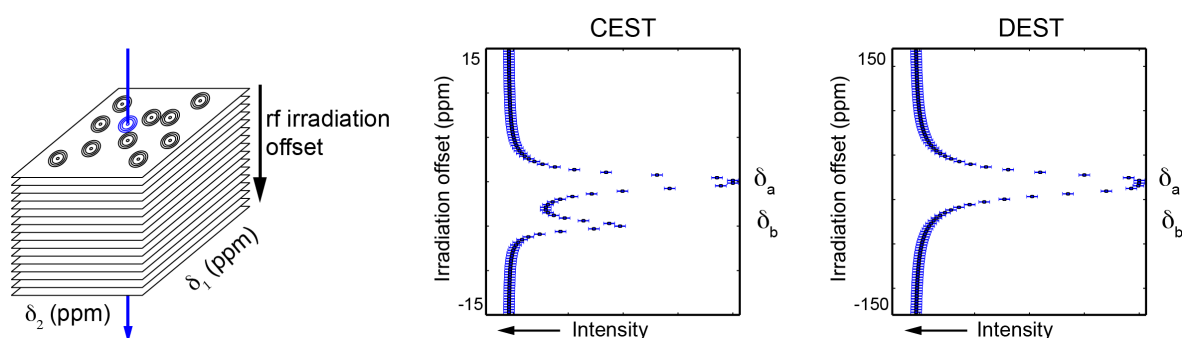


Figure 1.5 Schematic depiction of exchange-mediated saturation transfer experiments. A series of 2D spectra are recorded with a weak radio-frequency (rf) field being applied at varying offsets. The intensity of a signal of interest is monitored as a function of the irradiation offset (CEST/DEST profile). Selective irradiation of the minor state is accomplished by making use of the chemical shift difference (CEST) or R_2 difference (DEST) between the states. The perturbation of the minor state is transferred to the major (observable) state by chemical exchange. Fitting of the CEST/DEST profiles to the appropriate models (Chapter 4) provides kinetic, thermodynamic and structural information on the exchanging system.

CEST and DEST are essentially identical, differing only on the type of exchange targeted. CEST experiments, like relaxation dispersion experiments, target exchange between a highly and a lowly populated state, and rely on the chemical shift difference between the two states for the selective irradiation of the “invisible” state. (Palmer III, 2014; Vallurupalli et al., 2012) DEST experiments on the other hand, target exchange between a low and a high (> 1 MDa) molecular weight species. The high-molecular-weight species tumbles slowly in solution, and is therefore characterized by a large R_2 causing the signals from this species to be broaden beyond detection. In DEST experiments it is then the large difference in the line-width between the low and the high molecular weight species that is used to selectively irradiate the “invisible” state. (Fawzi et al., 2011; Fawzi et al., 2012; Palmer III, 2014) The set-up of the exchange-mediated saturation transfer experiment thus depends on the type of exchange targeted: CEST experiments require the use of very weak radio-frequency fields and small offset incremental steps to ensure that the resonance frequency of the minor state is selectively irradiated (Carneiro et al., 2015b; Vallurupalli et al., 2012); DEST experiments rely on the use of stronger radio-frequency fields irradiated at offsets far off-resonance from the major (visible) state to attain selective irradiation of the minor state. (Fawzi et al., 2011; Fawzi et al., 2012) Additionally, the use of at least two different radio-frequency field strengths is necessary to obtain reliable exchange parameters from both CEST and DEST experiments. (Carneiro et al., 2015b; Fawzi et al., 2012)

It should be noted that, although this section focused on relaxation-based experiments, experimental NMR parameters in general correspond to population-averaged values of individual conformations if the system undergoes conformational exchange in the fast regime in the chemical-shift timescale. (Jardetzky, 1980; Wüthrich, 1986) Thus, conformational dynamics can also be inferred from inconsistencies between experimentally derived structural parameters (such as interatomic distances derived from nuclear Overhauser effect (NOE) measurements, or dihedral angles derived from J coupling constants measurements) and a given (set) of structure(s) (e.g. (Carneiro et al., 2015c; Kessler et al., 1988); see also Chapter 3).

1.3 THE ANTI-HIV LECTIN *OSCILLATORIA AGARDHII* AGGLUTININ

The initial steps in human immunodeficiency virus (HIV) infection, namely attachment to cellular receptors and entry of the virus into the host cells, are mediated by the envelope protein complex. (Chan and Kim, 1998; Freed and Martin, 1995) This complex, also known as the envelope spike, is a trimer of highly glycosylated gp120/gp41 heterodimers. (Freed and Martin, 1995; Ward and Wilson, 2015) Interestingly, although the viral proteins are glycosylated by host cell machinery, differences in the processing of these sites result in dense clusters of high-mannose type glycans that are not observed in mammalian glycoproteins. (Bonomelli et al., 2011; Pritchard et al., 2015) The intrinsic high-density of glycosylation sites combined with trimerization of gp120/gp41 heterodimers is thought to restrict the access to the endoplasmic reticulum and Golgi α -mannosidases, resulting in a predominance of high-mannose glycans on gp120 that is conserved for different viral clades. (Bonomelli et al., 2011; Pritchard et al., 2015) This dense high-mannose glycan coat plays a direct role in viral transmission, by binding to cellular receptors, and in evasion of the host's immune system, by shielding the underlying protein surface from potentially neutralizing antibodies. (Balzarini, 2007; Horiya et al., 2014)

Agents that directly bind glycans on the viral envelope may, thus, compromise viral entry. Interestingly, carbohydrate-binding agents do not prevent the initial interaction between the virus and the host cell, but block downstream events that would lead to the entry of the virus. (Balzarini, 2007) Moreover, viral evolution due to prolonged exposure to such carbohydrate-binding agents could lead to (partial) deletion of the glycan shield and exposure of potentially immunogenic epitopes in the envelope spike. (Balzarini, 2007) Indeed, up to date several broad neutralizing antibodies ((Horiya et al., 2014) and references therein) and many lectins from plant, algal and cyanobacterial origin (reviewed in (Akkouh et al., 2015)) were found to interfere with HIV infection by binding to the glycan moieties of gp120. Algal and cyanobacterial lectins are particularly promising for the development of topical microbicides given their potent and broad-spectrum anti-HIV activity. (Huskens and Schols, 2012)

The agglutinin from the cyanobacteria *Oscillatoria agardhii* (OAA) is one such lectin with anti-HIV activity. OAA binds exclusively to high-mannose N-glycans, and the

5-mannose branched core of high-mannose type glycans (Man α (1-3)[Man α (1-6)]Man α (1-6)[Man α (1-3)]Man; hereafter termed Man5) has been identified as the minimal high affinity ligand. (Sato et al., 2007) This carbohydrate-recognition epitope is unique to OAA-family lectins. (Koharudin and Gronenborn, 2011; Koharudin et al., 2012; Whitley et al., 2013) Likely due to its unique binding determinants, OAA also shows potent and broad-spectrum activity against HIV strains resistant to other carbohydrate-binding agents. (Férris et al., 2014)

The crystal structures of free and Man5-complexed OAA (Figure 1.6) revealed a β -barrel like structure with two symmetrically located binding sites. Notably, the crystal structures show a distinct two-fold symmetry that is also seen at the primary structure level.



Figure 1.6 Amino-acid sequence and crystal structures of OAA. (A) Amino acid sequence alignment of the two sequence repeats. Identical residues are colored black, and the residues comprising binding site 1 and 2 are marked by the orange and yellow boxes, respectively. (B) Ribbon representation of the crystal structure of free OAA (PDB code 3S5V). The first 67 residues are colored in cyan and the following 66 residues in gray to highlight the symmetric properties of the structure. (C) Ribbon representation of the crystal structure of Man5-bound OAA (PDB code 3S5X). One Man5 molecule per binding site is shown in stick representation.

In addition, binding between OAA and Man5 was investigated by NMR. The binding is in slow exchange on the chemical shift timescale, suggesting a relatively tight

binding, and similar intensity changes are seen for equivalent residues in binding site 1 and 2, indicating similar affinities of each binding site to Man5. Interestingly, not only the overall fold of OAA is identical in the free and Man5-bound forms, but also the binding site conformations seem relatively unperturbed by the presence of ligand (Figure 1.7).

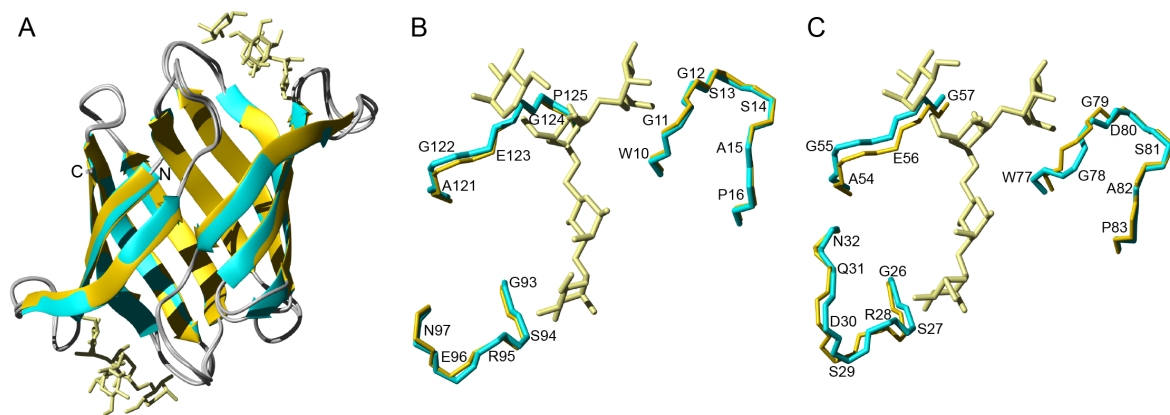


Figure 1.7 Comparison between free and Man5-bound OAA crystal structures. (A) Best fit superposition of the C α atoms for the crystal structures of free (cyan) and Man5-bound (yellow) OAA. OAA structures are shown in ribbon representation, and Man5 is shown in stick representation. (B) Conformational details of binding site 1. (C) Conformational details of binding site 2. For each binding site, the best fit superposition of the heavy backbone atoms for the free (cyan) and Man5-bound (yellow) crystal structures is shown.

In fact, virtually no difference is found in the conformation of binding site 1 (Figure 1.7 B) in the absence and presence of Man5. Binding site 2, on the other hand, shows a small but clear conformational change, namely the orientation of the peptide bond between W77 and G78 is flipped by $\sim 180^\circ$ (Figure 1.7 C), with the conformation seen in the Man5-bound structure corresponding to the one seen in the free and bound state of binding site 1. However, careful evaluation of the crystal packing indicates that the conformation of binding site 1 in the free crystal structure is associated with protein-protein contacts. Further NMR relaxation experiments indicate that the behavior of both binding sites in solution is similar for the free protein. (Koharudin and Gronenborn, 2011) The conformational changes between the free and bound forms identified for binding site 2 are thus expected to occur also in binding site 1.

While the crystal structures determined by Koharudin and Gronenborn provide invaluable information regarding the free and the bound conformations of OAA, they are insufficient to characterize the molecular recognition mechanism essential for OAA's anti-

HIV activity. The characterization of the dynamics of OAA in solution might prove crucial for its further development and optimization as anti-HIV therapeutics.

1.4 THESIS OUTLINE

The work conducive to this thesis focused primarily on the elucidation of the mechanism by which OAA recognizes Man5, the event leading to its anti-HIV activity, by making use of the unique advantages of NMR spectroscopy described above. Additionally, continuous efforts to extend the efficacy and accuracy of NMR methodology used for the characterization of protein dynamics resulted in the development of two new tools.

The results are presented and discussed in the subsequent four chapters. Chapter 2 describes the resonance assignment of OAA, an essential step for atomic-resolution NMR studies. Chapter 3 is focused on the characterization of functional dynamics of free OAA in solution, and relaxation dispersion studies, NOE and *J*-coupling constants analysis, and the solution structure of OAA are presented. In Chapter 4 the two new strategies for the analysis of protein dynamics by NMR developed in the course of my PhD work are presented. First, a strategy that enables a two-fold reduction in the experimental time and accurate extraction of exchange parameters from exchange-mediated saturation transfer experiments is presented. Second, a clustering algorithm, aimed at the identification of dynamic clusters within a protein, is demonstrated using relaxation dispersion data. Finally, a general discussion and conclusion of the results presented in Chapters 2–4 is provided in Chapter 5.

2 RESONANCE ASSIGNMENT OF OAA

2.1 INTRODUCTION

Assignment of each resonance in the NMR spectrum to an atom in the chemical structure is an essential step for detailed NMR studies. Common assignment approaches require isotopically enriched proteins and are based on a combination of different experiments that correlate resonances either via through-bond or via through-space interactions between nuclei (an overview of several experiments commonly used for assignment of proteins by NMR can be found in (Sattler et al., 1999)).

In the strategy used here, backbone sequential assignments were obtained using the 3D HNCA experiment, (Grzesiek and Bax, 1992) which correlates the backbone amide resonances, $^1\text{H}_\text{N}$ and ^{15}N , to the $^{13}\text{C}_\alpha$ resonances of their own and preceding amino acids, together with ^{15}N -resolved [$^1\text{H}, ^1\text{H}$]-NOESY. (Marion et al., 1989a; Marion et al., 1989b; Zuiderweg and Fesik, 1989) Aliphatic side-chain resonances were assigned based on the 3D HCCH-TOCSY experiment (Bax et al., 1990), which connects the aliphatic $^1\text{H}, ^{13}\text{C}$ resonances of a given amino acid. Finally, aromatic and amide side-chain resonances were assigned based on intra-residual cross-peaks observed in NOESY experiments. (Wüthrich, 1986) In contrast to the 3D HNCA and 3D HCCH-TOCSY experiments, in which the different nuclei are correlated on the basis of through-bond scalar couplings, NOESY experiments correlate ^1H nuclei through-space interaction, and are useful not only for resonance assignment but are also the main source of distance restraints used in structure determination protocols (see also Chapter 3). The combination of these different experiments yielded the nearly complete assignment of all the expected resonances. (Carneiro et al., 2015a)

2.2 RESULTS

2.2.1 BACKBONE ASSIGNMENT

Sequence-specific backbone ^{15}N , $^1\text{H}_\text{N}$, and $^{13}\text{C}_\alpha$ assignments at 298 K were primarily obtained using the 3D HNCA experiment (Figure 2.1), and are in good agreement with the previously reported ^{15}N and $^1\text{H}_\text{N}$ assignments. (Koharudin et al., 2011)

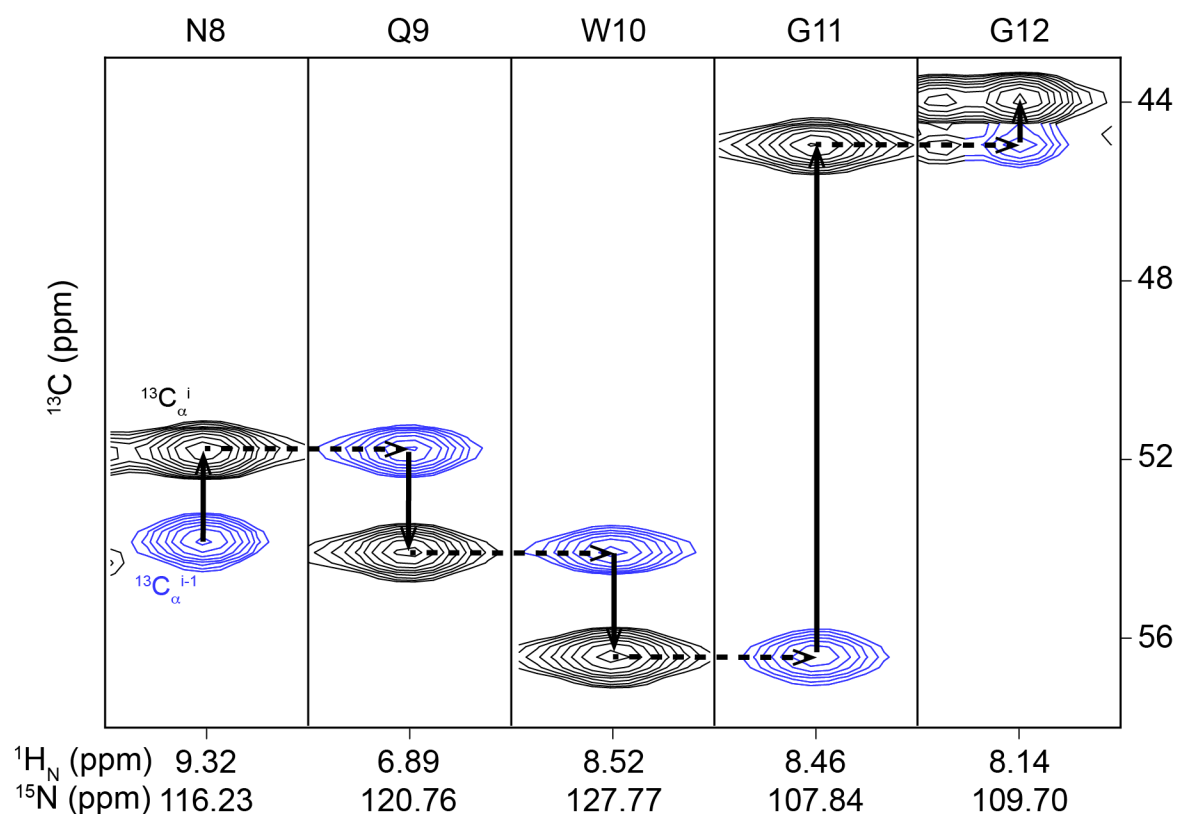


Figure 2.1 ^{13}C - ^1H strips from a 3D HNCA spectrum of ^{15}N , ^{13}C -labeled OAA at 298 K. The strips were taken at the ^{15}N chemical shifts of residues 8-12 and centered on the corresponding $^1\text{H}_\text{N}$ chemical shifts (indicated at the bottom of each strip). The sequential and intra-residual connectivities are indicated by solid and dashed lines, respectively. The sequence-specific assignment is indicated at the top of each strip by the one-letter amino acid code and the sequence number. $^{13}\text{C}_\alpha^i$ and $^{13}\text{C}_\alpha^{i-1}$ peaks are colored in black and blue, respectively.

Eventual ambiguities were resolved using a 3D ^{15}N -resolved [$^1\text{H}, ^1\text{H}$]-NOESY experiment. In particular, the ambiguity in the ^{15}N and $^1\text{H}_\text{N}$ assignments of S40/S107, G41/G108 and D42/D109 due to degeneracy in their C_α and C_β chemical shifts (Koharudin et al., 2011) could be resolved based on the observation of sequential NOE connectivities (Figure 2.2). (Wüthrich, 1986)

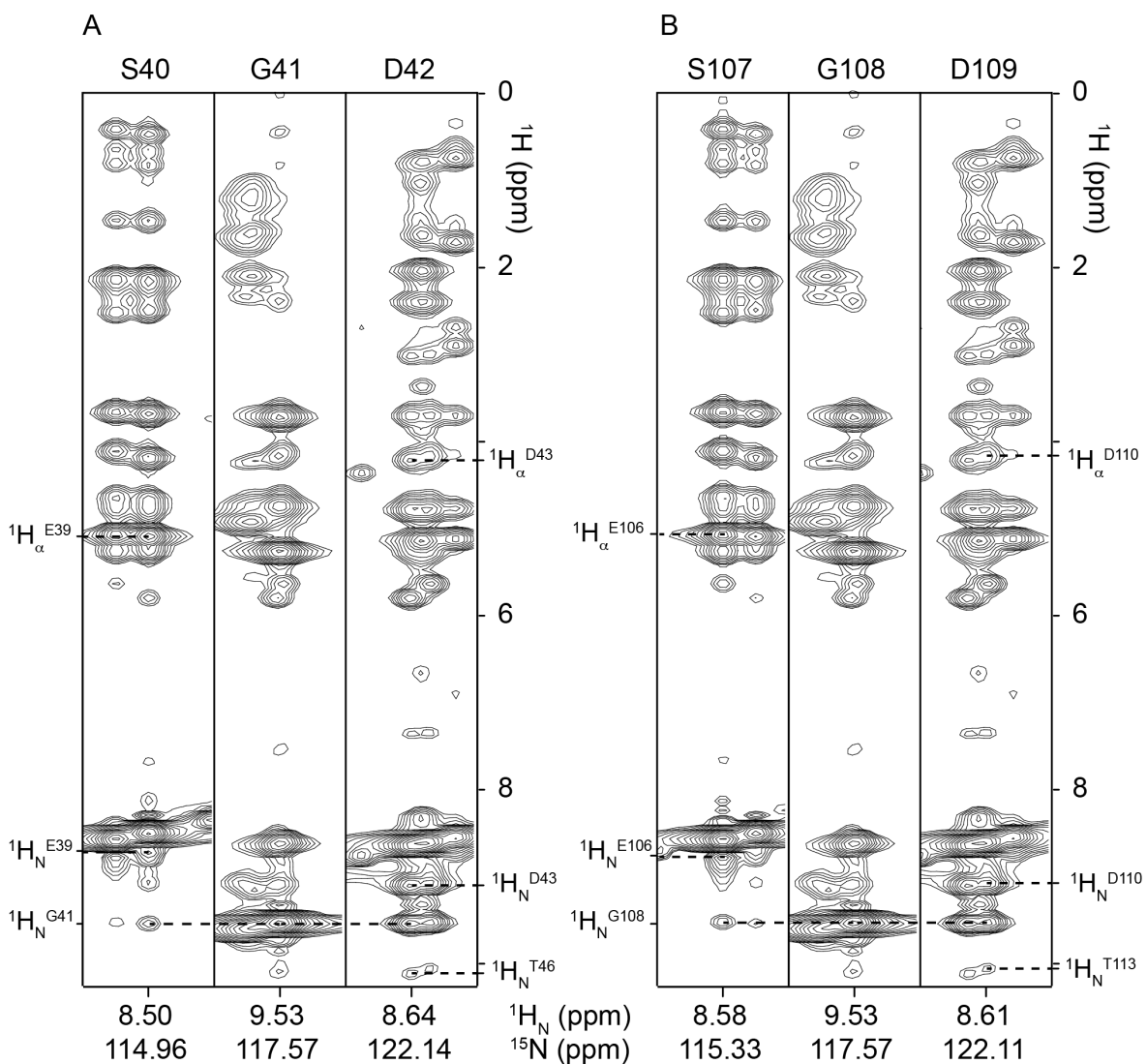


Figure 2.2 Assignment of backbone amide resonances of S40 - D42 and S107 - D109 based on NOE connectivities. (A) Section of the 3D ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY, taken at the ^{15}N chemical shifts of residues 40-42 and centered on the corresponding $^1\text{H}_\text{N}$ chemical shifts (indicated at the bottom of each strip). (B) Section of the 3D ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY, taken at the ^{15}N chemical shifts of residues 107-109 and centered on the corresponding $^1\text{H}_\text{N}$ chemical shifts (indicated at the bottom of each strip). The sequence-specific assignment is indicated at the top of each strip by the one-letter amino acid code and the sequence number. The inter-residual connectivities used for the unambiguous assignment of the backbone resonances of S40/S107, G41/G108 and D42/D109 are indicated by dashed lines and the corresponding assignments indicated on the side.

Additionally, the previously unassigned amide resonance of N69 could be assigned based on the combined analysis of the 3D HNCA and the ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY, and was found to be overlapped with the amide resonance of G116 (Figure 2.3).

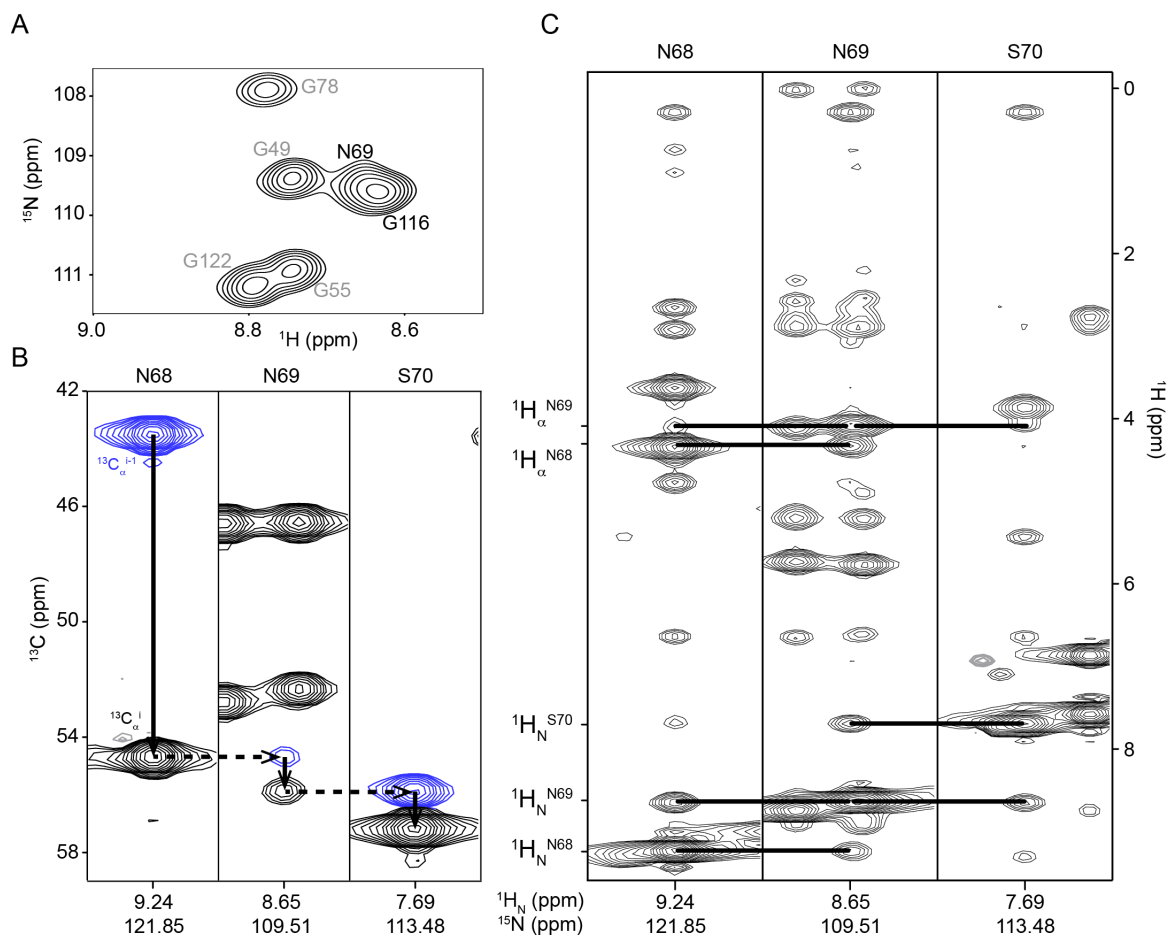


Figure 2.3 Assignment of the backbone amide resonance of N69. (A) Section of the 2D ^1H - ^{15}N HSQC spectrum of free OAA. (B) Section of the 3D HNCA spectrum. (C) Section of the 3D ^{15}N -resolved $[\text{H}, \text{H}]$ -NOESY. The strips of the 3D spectra were taken at the ^{15}N chemical shifts of residues 68-70 and centered on the corresponding $^1\text{H}_\text{N}$ chemical shifts (indicated at the bottom of each strip). The sequence-specific assignment is indicated at the top of each strip by the one-letter amino acid code and the sequence number. The sequential and intra-residual connectivities are indicated by solid and dashed lines, respectively.

Based on this approach, all the expected backbone resonances for ^{15}N , $^1\text{H}_\text{N}$, and $^{13}\text{C}_\alpha$ nuclei were assigned. Remaining carbonyl carbon assignments were achieved with a 3D HNCO, which correlates an amide group to its preceding carbonyl carbon. (Grzesiek and Bax, 1992; Kay et al., 1994; Schleucher et al., 1993) The complete resonance assignment of the backbone amide groups is shown in Figure 2.4.

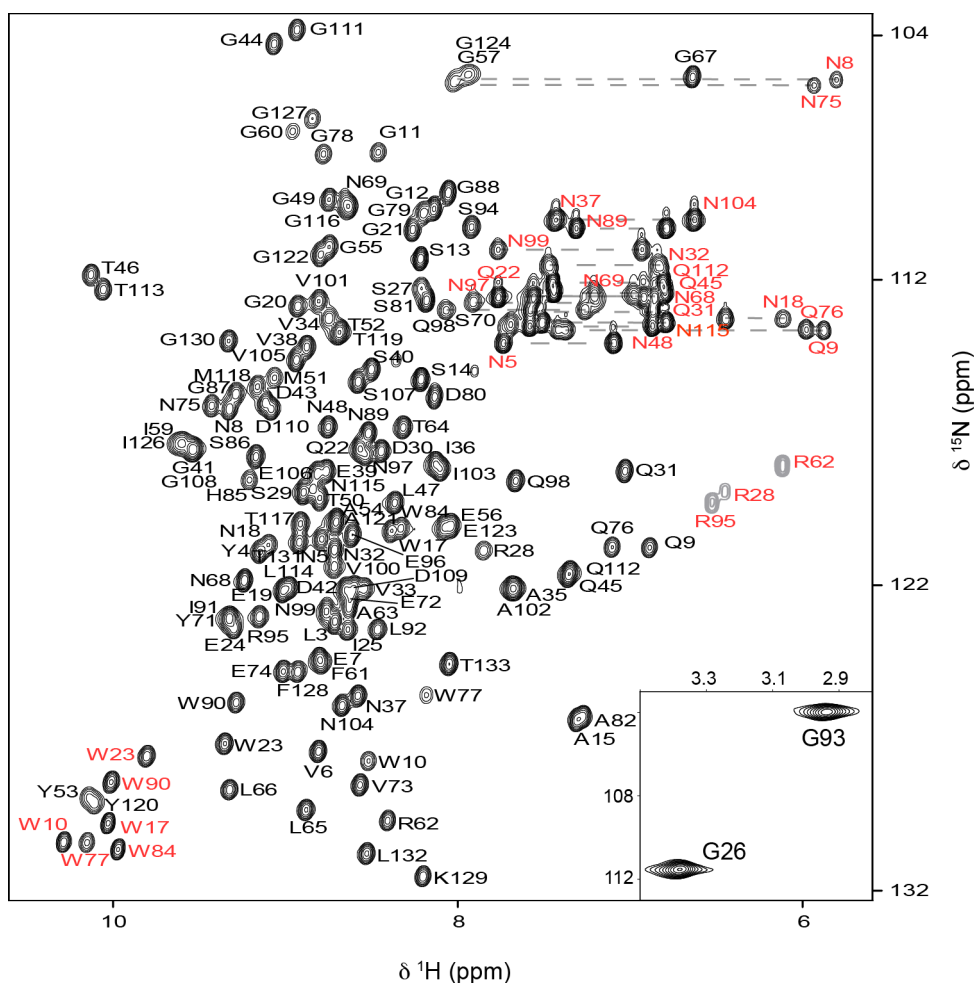


Figure 2.4 2D ^1H - ^{15}N HSQC spectrum of ^{15}N -labeled OAA at 298 K. Backbone resonance assignments are indicated by the one-letter amino acid code and the sequence number in black. The amide resonances of G26 and G93 exhibit unusual proton chemical shifts and are shown in the inset. Side-chain resonances are indicated by the one-letter amino acid code and the sequence number in red. Resonances corresponding to the amino groups of N and Q are connected with a dotted line. Arginine side-chain resonances are folded in the ^{15}N dimension and contoured in grey.

The good signal dispersion in the 2D ^1H - ^{15}N HSQC is indicative of a folded structure. Furthermore, the amide resonances of G26 and G93 exhibit unusual proton chemical shifts. The proximity of the amide group of these residues to the aromatic side chains of W90 and W23 (Figure 2.5), respectively, suggests that the upfield shift is due to ring current effects (Koharudin et al., 2011). Indeed, the deviation of the observed amide proton chemical shift of G26 and G93 from the average chemical shift (-4.95 and -5.39 ppm, respectively) is in agreement with the structure-based predicted deviation due to ring current effects (-4.87 and -4.80, for G26 and G93 respectively).

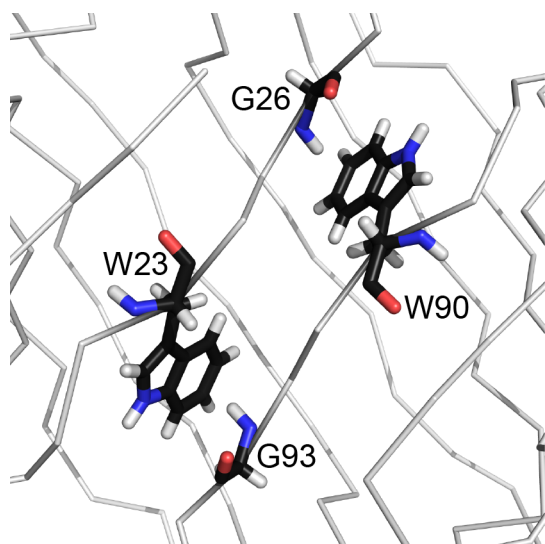


Figure 2.5 Structural detail of the chemical environment surrounding G26 and G93. The proximity of the amide proton of G26 to the indole ring of W90 leads to its large upfield shift. An equivalent situation is observed between the amide proton of G93 and the indole ring of W23. The backbone and side chain atoms of G26, W23, W90 and G93 are shown in stick representation. Hydrogen atoms were added to the original X-ray structure (PDB code 3S5V) using Amber99sb force-field (Hornak et al., 2006) implemented in Gromacs 4.5. (Hess et al., 2008)

Additionally, H_{α} resonances were tentatively assigned based on the ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY and validated during side-chain assignment (section 2.2.2).

2.2.2 SIDE-CHAIN ASSIGNMENT

The assignment of aliphatic ^{13}C resonances was achieved through the analysis of a 3D HCCH-TOCSY experiment, using the $^1\text{H}_{\alpha}$ and $^{13}\text{C}_{\alpha}$ chemical shifts as a starting point (example shown for V105 in Figure 2.6). The corresponding ^1H resonances were assigned based on intra-residual NOE cross-peaks observed in a $^{13}\text{C}_{\text{aliphatic}}$ -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY spectrum. (Wüthrich, 1986) The combined analysis of the 3D HCCH-TOCSY and $^{13}\text{C}_{\text{aliphatic}}$ -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY spectra yielded the complete resonance assignment of aliphatic ^{13}C and ^1H resonances.

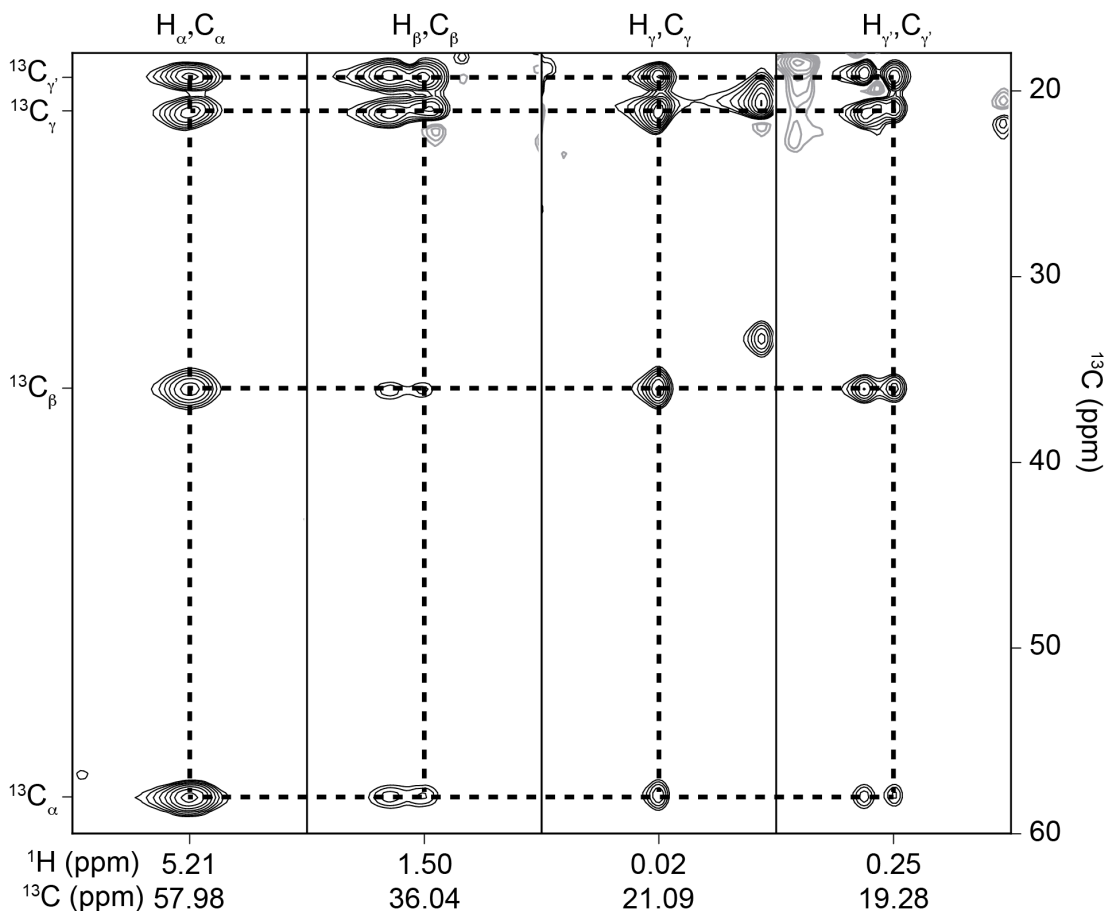


Figure 2.6 ^{13}C - ^1H strips from 3D HCCH-TOCSY spectrum for residue V105. The strips were taken at the ^{13}C chemical shifts and centered on the ^1H chemical shifts indicated at the bottom of each strip, which correspond to the assignments shown the top. The intra-residual connectivities are indicated by the dashed lines. The chemical shift assignment of the aliphatic carbons of V105 is shown on the left.

The assignment of aromatic side-chains was obtained via intra-residual NOE cross-peaks between aliphatic and aromatic protons observed in ^{13}C -resolved $[\text{H}, \text{H}]$ -NOESY spectra. (Wüthrich, 1986) An example of this approach is shown in Figure 2.7 for Y4. The corresponding ^{13}C resonances were tentatively assigned based on a 2D ^1H - ^{13}C HSQC and verified against the ^{13}C -resolved $[\text{H}, \text{H}]$ -NOESY spectra. Based on this approach, all ^1H and ^{13}C aromatic resonances of tyrosine, tryptophan, histidine and phenylalanine residues could be assigned, with the exception of the aromatic ring of Y71 due to spectral overlap. Of note, the $^1\text{H}_{\delta 1}/^1\text{H}_{\delta 2}$, $^{13}\text{C}_{\delta 1}/^{13}\text{C}_{\delta 2}$, $^1\text{H}_{\epsilon 1}/^1\text{H}_{\epsilon 2}$, $^{13}\text{C}_{\epsilon 1}/^{13}\text{C}_{\epsilon 2}$ resonances of each tyrosine and phenylalanine are degenerate.

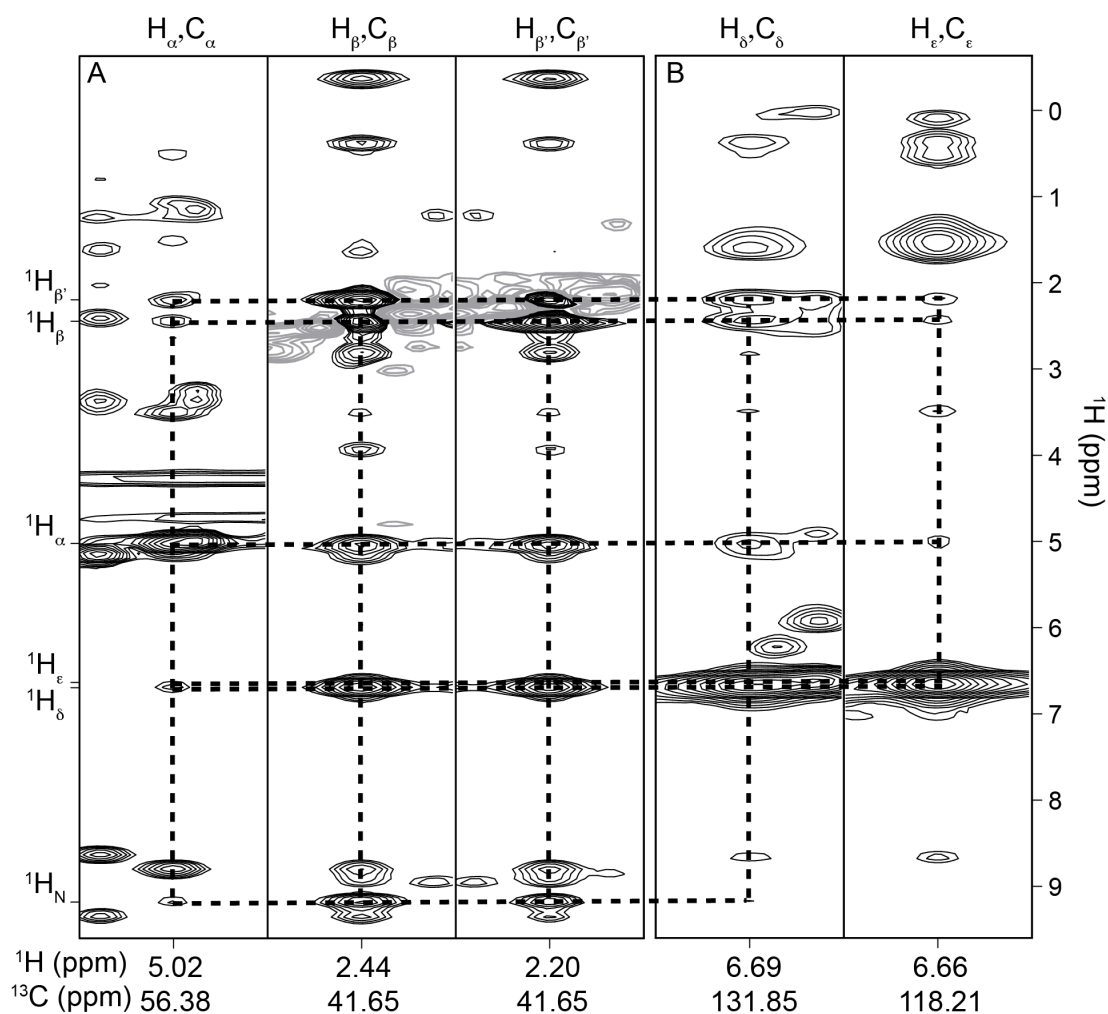


Figure 2.7 ^1H - ^1H strips from a 3D $^{13}\text{C}_{\text{aliphatic}}$ -resolved $[\text{H}, \text{H}]$ -NOESY (A) and a 3D $^{13}\text{C}_{\text{aromatic}}$ -resolved $[\text{H}, \text{H}]$ -NOESY (B) spectra for residue Y4. The strips were taken at the ^{13}C chemical shifts and centered on the ^1H chemical shifts indicated at the bottom of each strip which correspond to the assignments shown the top. The intra-residual connectivities are indicated by the dashed lines. The chemical shift assignment of the aliphatic and aromatic protons of Y4 is shown on the left.

Amide side-chain resonances of asparagine and glutamine residues were assigned via intra-residual NOE cross-peaks between $^1\text{H}_\beta$ and $^1\text{H}_\delta$ (asparagine) or between $^1\text{H}_\gamma$ and $^1\text{H}_\epsilon$ (glutamine) observed in the observed in the $^{13}\text{C}_{\text{aliphatic}}$ -resolved $[\text{H}, \text{H}]$ -NOESY and ^{15}N -resolved $[\text{H}, \text{H}]$ -NOESY spectra. (Wüthrich, 1986) In addition, $^1\text{H}_\epsilon/^{15}\text{N}_\epsilon$ resonances corresponding to arginine side-chains were assigned using the same approach. This strategy resulted in the complete assignment of amide side-chains of asparagine and glutamine, and $\text{N}_\epsilon\text{H}_\epsilon$ groups of arginine residues (Figure 2.4).

2.3 DISCUSSION

Resonance assignment is a crucial step to analyze any NMR spectrum. Following a strategy making use of 3D HNCA, 3D HCCH TOCSY, ^{13}C -resolved [^1H , ^1H]-NOESY and ^{15}N -resolved [^1H , ^1H]-NOESY spectra a nearly complete assignment of OAA was obtained. (Carneiro et al., 2015a)

All ^{15}N , ^1H and $^{13}\text{C}_\alpha$ backbone resonances were assigned, with exception of ^{15}N , $^1\text{H}_\text{N}$ of A2, the N-terminal residue of OAA, which is not observable due to exchange with the bulk water. Likewise, terminal amine side-chain groups of lysines, and the amine side-chain groups of histidines exchange with the bulk water and are not observable. Additionally, the η - NH_2 groups of all arginines, which are usually broadened at the temperature used (25 °C) due to the flipping of the guanidinium group (Henry and Sykes, 1995), were not observed. Apart from these exceptions, all amide side-chains were observed and assigned. All aliphatic and aromatic ^{13}C and ^1H resonances were also assigned, with the exception of the aromatic resonances of Y71 since spectral overlap prevented an unequivocal assignment.

The nearly complete assignment described here paved the way for further NMR studies on OAA, namely the characterization of OAA's functional dynamics described in Chapter 3.

2.4 MATERIALS AND METHODS

2.4.1 SAMPLE PREPARATION

NMR samples of free OAA consisted of 2 mM of ^{15}N -, or ^{15}N and ^{13}C -labeled OAA in 20 mM sodium acetate (pH 5.0), 20 mM sodium chloride, 3 mM sodium azide, and 90/10% $\text{H}_2\text{O}/\text{D}_2\text{O}$ or 100% D_2O . OAA was expressed and purified as described in (Koharudin et al., 2011) and kindly provided by L.M.I. Koharudin and A. M. Gronenborn (University of Pittsburgh, USA).

2.4.2 NMR SPECTROSCOPY

A summary of the experiments used to assign ^1H , ^{13}C , and ^{15}N resonances is given in Table 2.1.

Table 2.1 List of NMR experiments used to obtain resonance assignments

Experiment (mixing time)	Nuclei*	Data size (complex points)*	Acquisition time (ms)*
2D [^{15}N , ^1H]-HSQC	^{15}N , ^1H	256, 1024	92.7, 91.7
3D HNCA	^{15}N , ^{13}C , ^1H	104, 100, 1024	37.9, 16.6, 91.7
3D HNCO**	^{13}C , ^{15}N , ^1H	54, 64, 1536	13.2, 19.5, 106.5
3D ^{15}N -resolved [^1H , ^1H]-NOESY (60 ms)	^1H , ^{15}N , ^1H	218, 100, 1024	19.6, 36.2, 91.7
2D [^{13}C , ^1H]-HSQC***	^{13}C , ^1H	200, 1024	15.2, 91.7
3D HCCH-TOCSY (20 ms)	^{13}C , ^{13}C , ^1H	140, 140, 2048	14.0, 14.0, 243.3
3D ^{13}C -resolved [^1H , ^1H]-NOESY*** (80 ms)	^1H , ^{13}C , ^1H	256, 120, 1024	22.5, 15.6, 91.7
2D [^{13}C , ^1H]-HSQC****	^{13}C , ^1H	160, 1024	35.2, 121.6
3D ^{13}C -resolved [^1H , ^1H]-NOESY**** (80 ms)	^1H , ^{13}C , ^1H	200, 100, 1024	20.4, 17.8, 104.4

* The dimensions were collected in the listed order, with the acquisition dimension listed last

** The ^{13}C carrier frequency was placed around the center of the carbonyl region, at 177 ppm. The 3D HNCO spectrum was kindly provided by L.M.I. Koharudin and A.M. Gronenborn (University of Pittsburgh, USA)

*** The ^{13}C carrier frequency was placed around the center of the aliphatic region, at 40 ppm.

**** The ^{13}C carrier frequency was placed around the center of the aromatic region, at 125 ppm.

The [^{15}N , ^1H]-HSQC (Bodenhausen and Ruben, 1980; Mori et al., 1995), HNCA (Grzesiek and Bax, 1992), ^{15}N -resolved [^1H , ^1H]-NOESY (Marion et al., 1989a; Marion et al., 1989b; Zuiderweg and Fesik, 1989), [$^{13}\text{C}_{\text{aliphatic}}$, ^1H]-HSQC (Bodenhausen and Ruben, 1980; Mori et al., 1995) and $^{13}\text{C}_{\text{aliphatic}}$ -resolved [^1H , ^1H]-NOESY (Muhandiram et al., 1993) spectra were recorded on a 800 MHz AVANCE III spectrometer equipped with a 5 mm triple-resonance cryogenic probe. The HNCO spectrum (Grzesiek and Bax, 1992; Kay et al., 1994; Schleucher et al., 1993) was recorded on a 900 MHz AVANCE spectrometer equipped with a 5 mm triple-resonance cryogenic probe. The (H)CCH-TOCSY (Bax et al., 1990) and [$^{13}\text{C}_{\text{aromatic}}$, ^1H]-HSQC (Bodenhausen and Ruben, 1980; Mori et al., 1995) experiments were acquired on a 600 MHz AVANCE III spectrometer equipped with a 5 mm triple resonance room-temperature probe. The $^{13}\text{C}_{\text{aromatic}}$ -resolved [^1H , ^1H]-NOESY (Muhandiram et al., 1993) spectrum was recorded on a 700 MHz AVANCE III spectrometer equipped with a 5 mm triple resonance room-temperature probe. All experiments were measured at 298 K. ^1H chemical shifts were referenced to the water

resonance; ^{13}C and ^{15}N chemical shifts were referenced indirectly based on frequency ratios. (Wishart et al., 1995)

2.4.3 DATA ANALYSIS

All spectra were processed using NMRPipe (Delaglio et al., 1995). Manual assignment of backbone and side-chain resonances was performed using CARA. (Keller, 2004)

2.4.3.1 Ring current shifts prediction

Ring current shifts of the amide protons of G26 and G93 were predicted with MOLMOL (Koradi et al., 1996) using the Johnson-Bovey model. (Case, 1995; Johnson and Bovey, 1958) Predictions were based on the X-ray structure of free OAA (PDB code 3S5V), to which hydrogen atoms were added using Amber 99sb force-field (Hornak et al., 2006) implemented in Gromacs 4.5. (Hess et al., 2008)

The observed shift was calculated assuming an average chemical shift of 8.33 ppm for glycine amide protons. (Ulrich et al., 2008)

3 FUNCTIONAL DYNAMICS OF OAA

3.1 INTRODUCTION

The X-ray crystal structures of sugar-free and sugar-bound OAA revealed a distinct conformational change in binding site 2, namely a $\sim 180^\circ$ flip of the orientation of the peptide bond between W77 and G78. Interestingly, the equivalent peptide bond in binding site 1 is essentially unperturbed by the presence of ligand, and adopts the conformation seen in the bound binding site 2 both in the free and in the ligand-complexed form. However, the presence of protein-protein contacts in the crystal lattice of free OAA suggests that the observed bound conformation in binding site 1 of free OAA might be artifactual. (Koharudin and Gronenborn, 2011)

In terms of molecular recognition mechanisms, the presence of distinct conformations in the absence and presence of ligand suggests that binding of Man5 by OAA occurs by induced fit. However, the X-ray structures report on the lowest energy conformations (ground state), and a scenario in which the bound conformation exists as a higher-energy conformational substate (excited state) in the absence of ligand is also conceivable, corresponding to a conformational selection mechanism. (Boehr et al., 2009) Indeed, relaxation dispersion experiments have been widely used to identify and characterize excited states that are relevant for binding interactions. (Boehr et al., 2006; Eisenmesser et al., 2005; Korzhnev et al., 2009; Sugase et al., 2007) Additionally, conformational fluctuations within the ground state, which may be linked to the binding process, can lie on outside of the detectable limit of relaxation dispersion. (Ban et al., 2011; Bouvignies et al., 2011; Lange et al., 2008) In order to elucidate the general mechanism by which OAA recognizes high-mannose glycans, we investigated whether the X-ray bound conformation is sampled in solution in the absence of sugar, either as an excited or as a ground substate.

The nature of the conformational change seen in the X-ray structures of OAA (peptide bond flip) results in significant changes in the distance between adjacent backbone amide groups and in the backbone dihedral angles, which can be readily probed with different NMR observables. Variations in the distance between protons can be studied with nuclear Overhauser effect (NOE) measurements, due to the steep distance dependence

of this effect ($NOE \propto d^{-6}$, where d represents the distance between two atoms). (Wüthrich, 1986) Changes in the backbone dihedral angles, on the other hand, can be probed by measuring three-bond J coupling constants (3J), whose dependence on dihedral angles is given by the empirical Karplus equation. (Karplus, 1963) NOEs and 3J coupling constants are also some of the most common NMR parameters used for structural determinations. In fact, NOEs are essential in defining the tertiary structure of protein since the effect is propagated through space and only up to about 5 Å. (Wüthrich, 1986) 3J coupling constants, in contrast, provide only information on the local conformation of the peptide chain. (Güntert, 1998) Both NOEs and 3J coupling constants provide short-range restraints. In addition to the NOE and 3J coupling, residual dipolar coupling constants (RDCs) can serve as structural restraints. Unlike NOEs and 3J coupling constants, RDCs provide long-range restraints since they provide information on the orientation of internuclear bond vectors in a global alignment frame. However, RDCs are averaged out under isotropic conditions and their measurement requires the use of anisotropic alignment media. (Bax et al., 2001)

Our efforts to describe the conformational space of OAA, described in this chapter, started with the analysis of ${}^{15}\text{N}$ CPMG relaxation dispersion experiments. Curiously, an excited state was identified, but it does not resemble the sugar-bound state. Further analysis of proton-proton distances (in terms of NOEs) and backbone dihedral angles (in terms of 3J coupling constants), and structure determination by NMR, based on NOEs and RDCs restraints ensued. A comparison of the different NMR parameters with the structural X-ray data indicates that both sugar-free and sugar-bound conformations coexist in solution in the absence of ligand.

3.2 RESULTS

3.2.1 ${}^{15}\text{N}$ CPMG RELAXATION DISPERSION DETECTS EXCITED STATE

The possibility of sugar-bound conformers corresponding to high-energy conformers sampled by free OAA was addressed with ${}^{15}\text{N}$ CPMG relaxation dispersion experiments on sugar-free OAA. At 298 K, the backbone amides of 32 residues were

found to undergo chemical exchange. However at this temperature exchange is fast in the chemical shift timescale for most residues, impeding the extraction of chemical shift information (Table 7.1). By lowering the temperature to 277 K, a total of 60 backbone amide resonances display significant relaxation dispersion, indicative of conformational exchange (representative relaxation dispersion curves are shown in Figure 3.1 B). These correspond to residues spread throughout the entire protein (Figure 3.1).

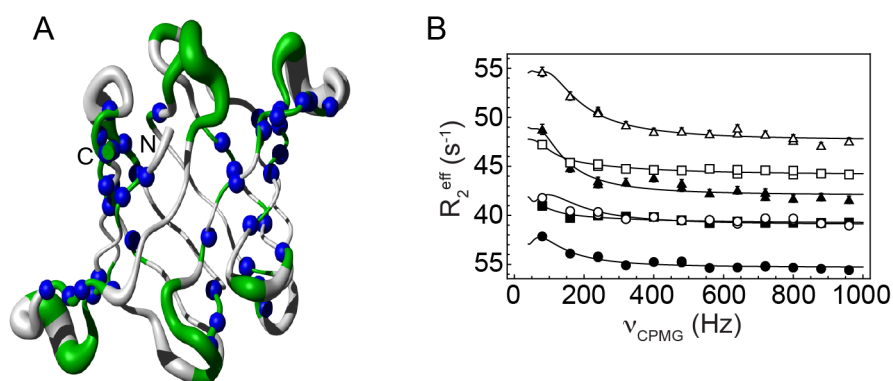


Figure 3.1 Structural mapping of residues undergoing conformational exchange in the absence of ligand. (A) Residues displaying significant ^{15}N relaxation dispersion are colored in green, with the blue spheres highlighting those outside of the sugar binding sites. (B) ^{15}N relaxation dispersion curves of M51 (triangles), W90 (circles), and T117 (squares) amide resonances. Relaxation dispersion data was collected at two fields (800 MHz, empty symbols; 600 MHz, full symbols). Experimental errors were estimated on the basis of duplicate measurements and are within the size of each symbol. Solid lines represent the best fits to two-state exchange model on a per-residue basis.

More importantly, analysis of the dispersion curves on a per-residue basis indicates that for 22 sites exchange is slow in the chemical shift timescale at this temperature, providing structural information about the excited state. Of note, the wide dispersion of k_{ex} obtained from the individual fits (from $16 \pm 13 \text{ s}^{-1}$ up to $5925 \pm 9280 \text{ s}^{-1}$) indicates that not all residues report on the same exchange process (Table 7.2). This observation is also corroborated by the high reduced- χ^2 obtained for the global fit of all residues ($\chi^2_{\text{red., global}} = 4.6$). Even residues within the same binding site (e.g. W10 and G11 in binding site 1, and W77 and G78 in binding site 2) show distinct exchange rates ($k_{\text{ex}}^{\text{W10}} = 2498 \pm 148 \text{ s}^{-1}$, $k_{\text{ex}}^{\text{G11}} = 1075 \pm 71 \text{ s}^{-1}$, $k_{\text{ex}}^{\text{W77}} = 198 \pm 500 \text{ s}^{-1}$, $k_{\text{ex}}^{\text{G78}} = 1622 \pm 109 \text{ s}^{-1}$), suggesting that the conformational exchange events probed by ^{15}N CPMG relaxation dispersion might not be related to binding.

In order to further assess whether the excited state corresponds to the sugar-bound state we compare the results from the ^{15}N CPMG relaxation dispersion experiments with the ones obtained from direct analysis of free and ligand-saturated OAA spectra. If the

conformational exchange detected by relaxation dispersion experiments corresponds to the sampling of the sugar-bound conformation we would expect (1) a similar spatial distribution of the amide resonances significantly perturbed upon addition of Man5 and of those undergoing chemical exchange, and (2) agreement between the chemical shift differences fitted from the relaxation dispersion profiles and those obtained from direct analysis of the peak positions in ^1H - ^{15}N HSQC spectra of free and ligand-saturated OAA. Figure 3.2 summarizes the comparison between the results obtain from these experiments. In contrast to the widespread distribution of residues undergoing conformational exchange, the residues whose amide resonances are significantly perturbed upon addition of Man-5 are restricted to the binding sites. Moreover, the chemical shift differences between the ground and the excited states do not agree, nor correlate (Pearson correlation coefficient of 0.36), with the chemical shift differences between free and bound states for residues within and outside the binding sites. Taken together, these observations indicate that the excited state does not correspond to the sugar-bound state.

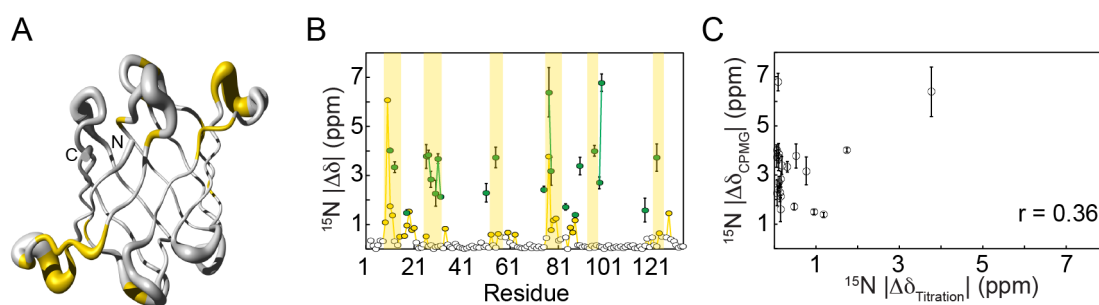


Figure 3.2 Comparison between chemical shift differences between ground and excited states and sugar-free and sugar-bound resonances. (A) Structural mapping of residues significantly perturbed upon addition of sugar. (B) The ^{15}N chemical shift difference ($|\Delta\delta|$) between the ground and excited state, extracted from CPMG relaxation dispersion experiments on free OAA, (green circles) do not agree with those measured directly from ^1H - ^{15}N HSQC spectra of sugar-free and sugar-saturated OAA (yellow circles). The white circles represent residues that do not present significant chemical shift perturbation upon addition of ligand. Areas associated with sugar binding are shaded yellow. (C) The chemical shift differences between ground and excited state ($|\Delta\delta_{\text{CPMG}}|$) also do not correlate with the chemical shift differences between sugar-free and sugar-bound OAA ($|\Delta\delta_{\text{Titration}}|$). The absence of correlation is highlighted by the Pearson correlation coefficient (r) of 0.36.

Alternatively, the excited state detected could correspond to a (partially) unfolded state. This hypothesis was tested, following the same strategy as before, by comparing the chemical shift differences between the ground and the excited states to the chemical shift differences between the native free state and predicted random coil values. The striking lack of correlation, evidenced by the Pearson correlation coefficient of -0.02 (Figure 3.3), indicates that this is also not the case.

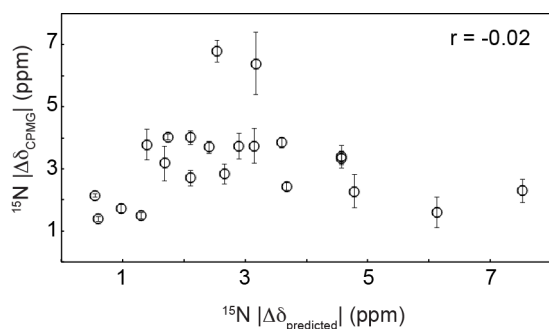


Figure 3.3 Comparison between chemical shift differences between ground and excited state and chemical shift differences between the native state and predicted random coil chemical shifts. The ^{15}N chemical shift differences derived from CPMG relaxation dispersion ($|\Delta\delta_{\text{CPMG}}|$) do not correlate with the chemical shift differences between the native state and predicted random coil ($|\Delta\delta_{\text{predicted}}| = |\delta_{1\text{H-}^{15}\text{N HSQC}} - \delta_{\text{random coil}}|$).

3.2.2 NOES INDICATE PRESENCE OF BOUND CONFORMER IN THE ABSENCE OF LIGAND

The free and bound conformations seen in X-ray structures in binding site 2 are characterized by distinct interatomic distances between the backbone amide protons (H_N) of W77 and G78, with the very short distance (2.2 Å) in the absence of sugar increasing to 4.4 Å when OAA forms the complex with Man5 (Figure 3.4 A and B), and between the H_N and each H_α of G78 (Figure 3.4 C and D).

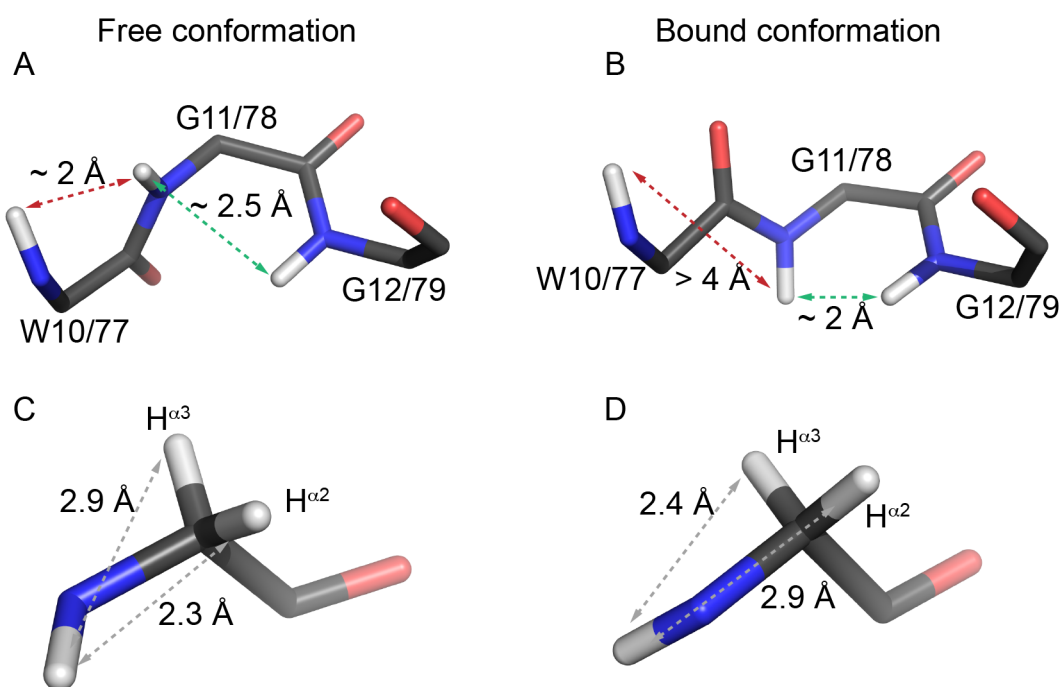


Figure 3.4 Free and bound conformations are characterized by distinct H-H distances. Stick representations of the backbone conformations characteristic of (A) sugar-free (PDB code 3S5V) and (B) sugar-bound (PDB code 3S5X) crystal structures. A detailed representation of G11/G78 in the sugar-free (C) and in the sugar-bound (D) is shown. The distance between the backbone amide protons (H_N) of W77 and G78 in the binding site 2 (dashed red arrows) is significantly shorter in the sugar-free conformation (2.2 Å) than in the sugar-bound conformation (4.4 Å). The distance between the H_N of G11/78 and G12/79 is considerably less affected and is shown in dashed green arrows for comparison. The two conformations also show distinct interatomic distances between the H_N and each H_α of the central glycine

(G11/78). Hydrogen atoms were added to the original X-ray structures using Amber99sb force-field (Hornak et al., 2006) implemented in Gromacs 4.5. (Hess et al., 2008)

Such differences in inter-proton distances can be analyzed by nuclear Overhauser effect (NOE) measurements, given its steep distance dependence ($1/d^6$, where d is the interatomic distance). Based on the distances measured from the X-ray structures, the free conformation ($d = 2.2 \text{ \AA}$) would give rise to a strong NOE cross-peak between the H_N of W77 and G78, whereas the bound conformation ($d = 4.4 \text{ \AA}$) would give rise to a weak cross-peak between those H_N . Similarly, a stronger cross-peak between H_N and $H_{\alpha 3}$ and a weaker cross-peak between H_N and $H_{\alpha 2}$ of G78 are expected for the bound conformation, whereas the opposite relation (i.e., a weaker cross-peak between H_N and $H_{\alpha 3}$ and a stronger cross-peak between H_N and $H_{\alpha 2}$ of G78) is expected for the free conformation. Furthermore, if the bound conformation seen in the binding site 1 in the sugar-free X-ray structure is indeed an artifact due to crystal packing effects, the same relations are expected between the H_N atoms of W10 and G11 and the H_N and H_{α} atoms of G11. Cross-sections of the NOESY spectrum of Man5-saturated OAA are shown in Figure 3.5. The very weak cross-peaks at the noise level between H_N of W77 and H_N of G78, and H_N of W10 and H_N of G11 are consistent with an interatomic distance larger than 4 \AA . Likewise, the relative intensity of the cross-peaks between H_N and each H_{α} of G78 and G11 are consistent with the inter-proton distances derived from the X-ray bound structure.

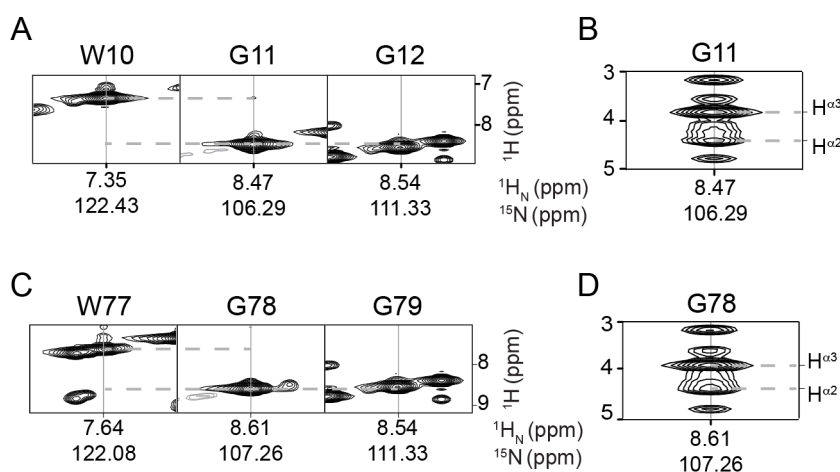


Figure 3.5 NOE cross-peaks of Man5-saturated OAA are in agreement with the Man5-bound X-ray structure. Cross sections of the 3D ^{15}N -resolved $[[^1\text{H}, ^1\text{H}]]$ -NOESY spectrum of Man5-saturated OAA for the backbone region and Gly side-chain region of binding site 1 (A and B, respectively) and binding site 2 (C and D, respectively). The very weak NOE cross-peaks between H_N W10 and 1H_N G11 (A) and H_N W77 and 1H_N G78 (C) are in agreement with the interatomic distance characteristic of the sugar-bound X-ray structure (4.4 \AA). The relative intensities of the NOE cross-peaks between H_N and each H_{α} of G11 (B) and G78 (D) are also in agreement with the interatomic distances in the sugar-bound conformation.

Conversely, the NOESY spectrum of free OAA (Figure 3.6) exhibits a sizable NOE cross-peak between H_N of W77 and H_N of G78. However, this cross-peak is weaker than the expected from the distance in the X-ray crystal structure when compared to the intensity of the NOE cross-peak between H_N of G78 and H_N of G79. Spectral overlap at both 298 and 277 K prevent a similar analysis for binding site 1. Also the relative intensities of the NOE cross-peaks between H_N and each H_α of G78, while clearly different from the ones in sugar saturating conditions, differ from the expected from the free crystal conformation. Notably, NOE cross-peaks between H_N and each H_α of G11 are also well resolved and its relative intensities are similar to the ones in binding site 2 (Figure 3.7). Since it has been shown that NOE intensities can be fully reproduced based on a known structure (Edwards et al., 2014), these results indicate that the X-ray structure of free OAA alone is insufficient to explain the experimental data.

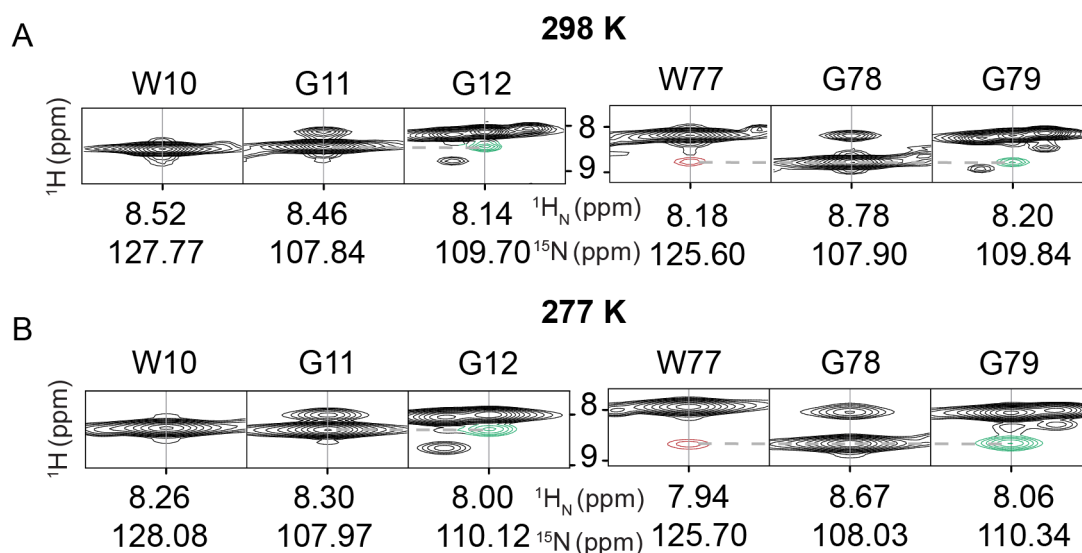


Figure 3.6 Backbone NOE cross-peaks of free OAA indicate conformational equilibrium between the sugar-free and the sugar-bound X-ray conformations. Cross sections of the 3D ^{15}N -resolved ^1H , ^1H -NOESY spectrum of free OAA for the backbone region of binding site 1 (W10-G12) and binding site 2 (W77-G79), at two different temperatures (298 K, **A**; 277 K, **B**). The NOE cross-peaks corresponding to the distances depicted in **Figure 3.4 A** between H_N W10/77 and H_N G11/78 (red) and between H_N G11/78 and H_N G12/79 (green). The relative intensity of these cross-peaks indicates that both the sugar-free and the sugar-bound X-ray conformations are present in the absence of ligand.

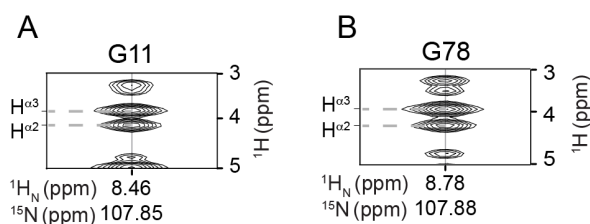


Figure 3.7 Side-chain NOE cross-peaks of free OAA indicate conformational equilibrium between the sugar-free and the sugar-bound X-ray conformations. Cross sections of the 3D ^{15}N -resolved ^1H , ^1H -NOESY spectrum of free OAA at 298 K for the side chain region of (A) G11 and (B) G78. The relative intensities of the NOE cross-peaks between H_N and each H_α of G11 (A) and G78 (B) do not match the interatomic distances in the X-ray sugar-free conformation and are best explained by the presence of both the sugar-free and the sugar-bound X-ray conformations.

In order to test whether the experimental NOE data could be reconciled with the X-ray structural data, we have estimated the relative populations of sugar-free and sugar-bound conformation (assuming that only those two states exist) that would result in the observed experimental NOE intensities. Due to the nonlinear dependence of the NOE intensity on the interatomic distance ($NOE \propto d^{-6}$) relative populations cannot be estimated by directly comparing experimental intensities with population weighted average distances. Instead, the relative intensities of two cross-peaks are directly comparable to the ratio of the sixth power of the distances between the two sets of two atoms as: (Carneiro et al., 2015c)

$$\frac{d_{AB}^{-6}}{d_{CD}^{-6}} \approx \frac{I_{AB}^{Normalized}}{I_{CD}^{Normalized}} \quad (3.1)$$

where AB and CD refer to two sets of two atoms, and the NOE cross-peak intensities (I) are normalized to the corresponding NOE diagonal peak (i.e., $I_{ij}^{Normalized} = I_{ij}/I_{ii}$). The experimental intensity ratios can now be analyzed as the ratio of population weighted average distances, assuming a two-state model, as:

$$\frac{I_{AB}^{Normalized}}{I_{CD}^{Normalized}} \approx \frac{p_1 d_{(AB)_1}^{-6} + p_2 d_{(AB)_2}^{-6}}{p_1 d_{(CD)_1}^{-6} + p_2 d_{(CD)_2}^{-6}} \quad (3.2)$$

$$p_2 = 1 - p_1 \quad (3.3)$$

where the subscripted numbers 1 and 2 refer to two distinct conformations

Interestingly, the backbone NOE intensities of binding site 2 can be reconciled with the interatomic distances derived from the sugar-free and sugar-bound X-ray structures if relative populations of $\sim 50\%$ of the sugar-free and $\sim 50\%$ of the sugar-bound X-ray conformations are considered at 298 K. Using the NOE intensities of the H_α cross-peaks for G78 results in relative populations of $\sim 25\%$ of the sugar-free and $\sim 75\%$ of the sugar-bound X-ray conformations. NOE intensities and interatomic distances used for estimating populations are presented in Table 3.1. It should be noted that possible effects of spin diffusion and fast internal motions, which are known to affect the NOE rates, (Neuhaus and Williamson, 2000) are not taken into account here. Consequently, this analysis must be taken from a qualitative perspective, suggestive of the presence of a conformational

equilibrium between two highly populated states (not less than 25 and not more than 75%) represented by the sugar-free and sugar-bound X-ray conformations. Also at 277 K, a qualitative evaluation of the NOESY spectrum at that temperature (Figure 3.6 B) indicates that a conformational equilibrium between the same conformations with similar populations is present. Such analysis for binding site 1 is unattainable due to spectral overlap in the sugar-free NOESY spectra at both 298 and 277 K, and to the presence of protein-protein contacts in the sugar-free crystal lattice. Even so, the comparable NOE pattern seen for the H_{α} cross-peaks for G78 and G11 (Figure 3.7) suggests a similar behavior in both binding sites. In addition, populations estimated using the experimental G11 H_{α} cross-peaks and the distances derived from the X-ray structures for G78 (Table 3.1) are in excellent agreement with the ones estimated for binding site 2.

Table 3.1 Summary of interatomic distances and NOE intensities* used to estimate relative populations of sugar-free and sugar-bound conformations at 298 K in the absence of sugar.

X-ray structure**	Distance (Å)
Sugar-free (PDB code: 3S5V)	
H_N W77 - H_N G78	2.2
H_N G78 - H_N G79	2.9
H_N G78 - $H_{\alpha 2}$ G78	2.3
H_N G78 - $H_{\alpha 3}$ G78	2.9
Sugar-bound (PDB code: 3S5X)	
H_N W77 - H_N G78	4.4
H_N G78 - H_N G79	2.1
H_N G78 - $H_{\alpha 2}$ G78	2.9
H_N G78 - $H_{\alpha 3}$ G78	2.4
NOE cross-peak	Integral (a.u.)
H_N W77 - H_N W77	254146
H_N G78 - H_N G78	594756
H_N G79 - H_N G79	1406681
H_N W77 - H_N G78	8079
H_N G79 - H_N G78	69809
H_N G78 - $H_{\alpha 2}$ G78	51290
H_N G78 - $H_{\alpha 3}$ G78	77531
H_N G11 - $H_{\alpha 2}$ G11	51997
H_N G11 - $H_{\alpha 3}$ G11	80856

* A similar analysis for binding site 1 is precluded by spectral overlap in the sugar-free NOESY spectrum and the presence of protein-protein contacts in the sugar-free crystal.

** Hydrogen atoms were added to the original structures using Amber99sb force-field (Hornak et al., 2006) implemented in Gromacs 4.5. (Hess et al., 2008)

3.2.3 ${}^3J(H_N, H_\alpha)$ COUPLING CONSTANTS CORROBORATE PRESENCE OF SUGAR-BOUND CONFORMER IN THE ABSENCE OF SUGAR

Additional and complementary structural information regarding the ground state of OAA can be provided by three-bond J coupling constants (3J), given their dependence on dihedral angles. (Karplus, 1963) In particular, ${}^3J(H_N, H_\alpha)$, which depend on the ϕ dihedral angle, are remarkably sensitive to peptide bond flips. For glycines two different ${}^3J(H_N, H_\alpha)$ coupling constants can be measured, provided that the two H_α resonances are well resolved. These coupling constants depend on the angle θ : (Karplus, 1963)

$${}^3J(H_N, H_\alpha) = A \cos^2 \theta + B \cos \theta + C \quad (3.4)$$

where $\theta = \phi - 60^\circ$ for $H_{\alpha 3}$ (pro-S) and $\theta = \phi + 60^\circ$ for $H_{\alpha 2}$ (pro-R), and $A = (7.13 \pm 0.34)$, $B = -(1.31 \pm 0.13)$ and $C = (1.56 \pm 0.34)$ Hz. (Habeck et al., 2005)

The H_α resonances of G11 and G78 were stereospecifically assigned based on the comparison between the interatomic distances derived from the X-ray structures and the relative intensity of NOE cross-peaks. For example, as mentioned in section 3.2.2, the relative NOE intensities between the H_N - H_α groups of G11 and G78 in Man5 saturating conditions are in excellent agreement with the interatomic distances derived from the X-ray sugar-bound structure, if the stronger NOE cross-peak (up-field shifted) is assigned to $H_{\alpha 3}$ and the weaker NOE cross-peak (down-field shifted) is assigned to $H_{\alpha 2}$. Likewise, assigning the up-field shifted resonance to $H_{\alpha 3}$ and the down-field shifted resonance to $H_{\alpha 2}$ in the absence of ligand is in reasonable agreement with the relative populations of sugar-free and sugar-bound conformations estimated from the backbone NOE cross-peaks.

Following a similar strategy as employed for the analysis of NOE data, we compared the experimentally obtained ${}^3J(H_N, H_\alpha)$ coupling constants, to the ones predicted from the ϕ dihedral angles measured from the sugar-free and sugar-bound X-ray structures. Of note, an estimation of the H_α flip rates is required for an accurate quantitative analysis of ${}^3J(H_N, H_\alpha)$ coupling constants. (Vuister and Bax, 1993) Still, ${}^3J(H_N, H_\alpha)$ coupling constants can be qualitatively analyzed without this rate. Table 3.2 summarizes this comparison. The experimental ${}^3J(H_N, H_\alpha)$ coupling constants in Man5 saturating conditions for both G11 and G78 are consistent with those predicted from the sugar-complexed X-ray structure, with the larger coupling constant associated with $H_{\alpha 2}$ and the

smaller coupling constant associated with $H_{\alpha 3}$. In contrast, the relative magnitude of the experimental ${}^3J(H_N, H_{\alpha})$ coupling constants for G78 of sugar-free OAA is of opposite size to the one predicted from the free X-ray structure. Experimental ${}^3J(H_N, H_{\alpha})$ coupling constants can also be analyzed as a population averaged value of two conformations:

$${}^3J(H_N, H_{\alpha})_{exp} = p_1 {}^3J(H_N, H_{\alpha})_1 + p_2 {}^3J(H_N, H_{\alpha})_2 \quad (3.5)$$

where the subscripted numbers 1 and 2 refer to two distinct conformations.

Since for glycines two ${}^3J(H_N, H_{\alpha})$ coupling constants are available that report on the same structural parameter (ϕ angle), the relative populations were estimated using a least-squares approach taking into account both coupling constants simultaneously (see also section 3.4.3.4). This analysis results in relative populations of $\sim 60\%$ of the sugar-free and $\sim 40\%$ of the sugar bound X-ray conformations, thus corroborating the presence of a conformational equilibrium between these two conformations in the absence of ligand.

Table 3.2 Comparison between predicted* and experimental ${}^3J(H_N, H_{\alpha})$ coupling constants for G11 and G78.

G11 (binding site 1)	
Sugar-free	
ϕ ($^{\circ}$)	N/A**
Predicted ${}^3J(H_N, H_{\alpha 2})$ (Hz)	N/A**
Measured ${}^3J(H_N, H_{\alpha 2})$ (Hz)	Large (6.5 ± 0.5)
Predicted ${}^3J(H_N, H_{\alpha 3})$ (Hz)	N/A**
Measured ${}^3J(H_N, H_{\alpha 3})$ (Hz)	Small (4.7 ± 0.8)
Sugar-bound	
ϕ ($^{\circ}$)	100.9
Predicted ${}^3J(H_N, H_{\alpha 2})$ (Hz)	Large (9.2 ± 0.5)
Measured ${}^3J(H_N, H_{\alpha 2})$ (Hz)	Large (8.1 ± 0.2)
Predicted ${}^3J(H_N, H_{\alpha 3})$ (Hz)	Small (4.6 ± 0.4)
Measured ${}^3J(H_N, H_{\alpha 3})$ (Hz)	Small (3.4 ± 0.6)
G78 (binding site 2)	
Sugar-free	
ϕ ($^{\circ}$)	-89.7
Predicted ${}^3J(H_N, H_{\alpha 2})$ (Hz)	Small (5.8 ± 0.4)
Measured ${}^3J(H_N, H_{\alpha 2})$ (Hz)	Large (6.6 ± 0.5)
Predicted ${}^3J(H_N, H_{\alpha 3})$ (Hz)	Large (8.0 ± 0.4)
Measured ${}^3J(H_N, H_{\alpha 3})$ (Hz)	Small (4.9 ± 1)
Sugar-bound	
ϕ ($^{\circ}$)	111.4
Predicted ${}^3J(H_N, H_{\alpha 2})$ (Hz)	Large (9.8 ± 0.5)
Measured ${}^3J(H_N, H_{\alpha 2})$ (Hz)	Large (8.0 ± 0.2)
Predicted ${}^3J(H_N, H_{\alpha 3})$ (Hz)	Small (3.5 ± 0.4)
Measured ${}^3J(H_N, H_{\alpha 3})$ (Hz)	Small (3.7 ± 0.6)

* ${}^3J(H_N, H_{\alpha})$ predicted based on the dihedral angles ϕ extracted from the sugar-free and sugar-bound X-ray structures using the Karplus equation (Karplus, 1963) with the coefficients suggested by Habeck and coworkers. (Habeck et al., 2005)

** Analysis precluded by the presence of protein-protein contacts in the crystal lattice

A direct comparison between the $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants measured for G11 in the absence of sugar and the ones predicted from the X-ray sugar-free structure is prevented by the presence of protein-protein contacts within the crystal lattice involving binding site 1. Nonetheless, the striking similarity between the relative intensity of the H_N - H_α correlations observed for G11 and G78 (Figure 3.8), as well as between the extracted $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ (Table 3.2), indicates that the conclusions drawn for G78 can be extended to G11.

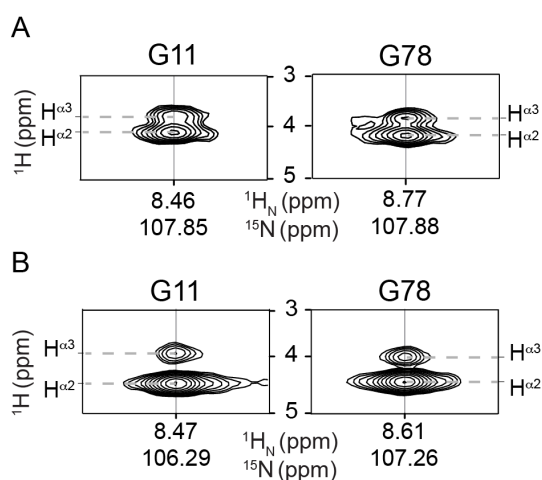


Figure 3.8 $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants suggest the presence of both the sugar-free and the sugar-bound conformations in the absence of Man5. Cross sections of the 3D-HNHA spectra of (A) Free OAA and (B) Man5-saturated OAA. The relative intensity of the two H_N - H_α cross-peaks depends on the dihedral angle ϕ .

3.2.4 NMR SOLUTION STRUCTURE OF FREE OAA

In order to confirm the presence of both sugar-free and sugar-bound conformations in solution, we have also determined the solution structure of free OAA. The structure was determined based on the complete resonance assignment (Chapter 2), and on a large number of experimental constraints (on average ~ 19 constraints per residue) derived from NOEs and RDCs. The experimental constraints used for structure determination as well as the structural statistics of the resulting ensemble are summarized in Table 3.3. The solution structure is well defined with atomic R.M.S.D. values relative to the mean coordinates of (0.7 ± 0.1) and (0.33 ± 0.05) Å for all backbone atoms and those in secondary structure elements, respectively. In addition, no violations of the experimental constraints are present in the final ensemble.

Table 3.3 Summary of experimental data, constraints and structural statistics

Constraints*	
Distance constraints	2419
Total NOE	2361
Intra-residue	175
Inter-residue	
Sequential ($ i-j = 1$)	512
Medium range ($ i-j < 5$)	278
Long range ($ i-j \geq 5$)	1396
H-bonds	58
Dipolar coupling constraints	
$^1D_{NH}$	89
Structural statistics*	
Violations	
Distance constraints ($> 0.5 \text{ \AA}$)	0
RDC constraints ($> 1.5 \text{ Hz}$)	0
R.M.S.D. from idealized geometry	
Bond lengths (\AA)	0.006
Bond angles ($^\circ$)	0.621
Improvers ($^\circ$)	1.003
Average pairwise R.M.S.D. (\AA)	
All backbone atoms	0.7 ± 0.1
Backbone atoms in secondary structures**	0.33 ± 0.05
Procheck Ramachandran statistics (%)**	
Most favored regions	98.15
Additionally allowed regions	1.69
Generously allowed regions	0.08
Disallowed regions	0.08

* Analyses performed with PSVS version 1.5. (Bhattacharya et al., 2007) Average distance violations were calculated using the sum over r^{-6} .

** For residues in regular secondary structures (3-9, 18-24, 33-40, 46-53, 58-65, 70-76, 84-91, 100-106, 113-120, 126-132), calculated using DSSP.

The solution structure of OAA displays a β -barrel fold composed of 10 antiparallel β -strands comprising residues 3-9, 18-24, 33-40, 46-53, 58-65, 70-76, 84-91, 100-106, 113-120, and 126-132 (Figure 3.9 A). Overall, the solution structure is very similar to the sugar-free and sugar-complexed X-ray structures, verified by the backbone R.M.S.D. values of 0.7 and 0.67 \AA , respectively. Details of the two previously identified binding sites (Koharudin et al., 2011; Koharudin and Gronenborn, 2011), comprising the loops between $\beta 1$ - $\beta 2$, $\beta 7$ - $\beta 8$ and $\beta 9$ - $\beta 10$ (binding site 1), and $\beta 2$ - $\beta 3$, $\beta 4$ - $\beta 5$ and $\beta 6$ - $\beta 7$ (binding site 2), are shown in Figure 3.9 A and B, respectively.

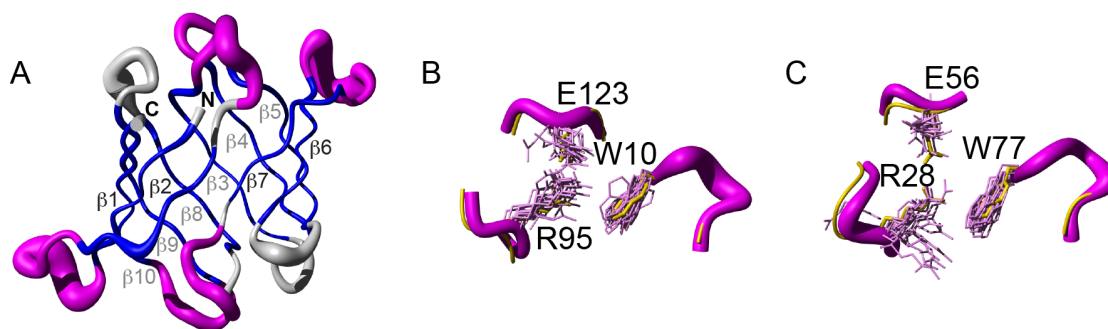


Figure 3.9 Solution structure of free OAA. (A) The overall fold of the solution structure of free OAA (PDB code 2MWH) comprises 10 antiparallel β -strands (blue) that form a β -barrel very similar to the sugar-free and sugar-bound X-ray structures (PDB code 3S5V and 3S5X, respectively). The two binding sites are highlighted in magenta (B) Details of binding site 1 conformation. (C) Details of binding site 2 conformation. Both binding sites resemble the sugar-bound conformation of the X-ray structure (yellow). Side-chains directly involved in Man5 binding (W10, R95 and E123 in binding site 1, and R28, E56 and W77 in binding site 2) are shown in stick representation. The mean position of the backbone C_{α} atoms is shown in tube representation, with the radius of the tube corresponding to the average deviation of the 20 lowest energy structures with respect to the mean structure.

As expected from the low backbone R.M.S.D. value, the overall structure of both binding sites in solution resembles the X-ray sugar-bound structure. Interestingly, this is also true for the side-chains which are directly involved in Man5 binding, namely W10, R95 and E123 in binding site 1, and R28, E56 and W77 in binding site 2 (Figure 3.9 B and C). Moreover, individual conformers of the solution structure exhibit the different backbone conformations, characterized by the different conformation of the peptide-bond between W10/77–G11/78, seen in the sugar-free and sugar-bound X-ray structures in both binding sites (Figure 3.10).

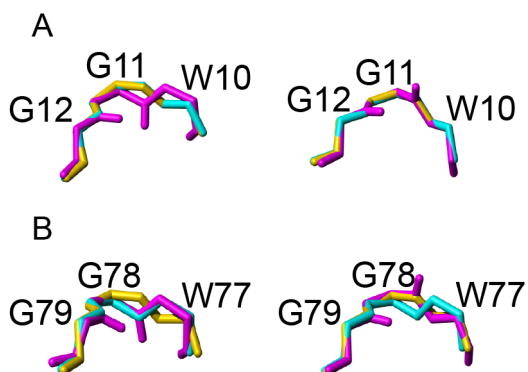


Figure 3.10 NMR solution structure of free OAA includes the sugar-free and the sugar-bound conformation in both binding sites. The different backbone conformations sampled by OAA in solution (magenta; PDB code 2MWH) are in excellent agreement with the sugar free (cyan; PDB code 3S5V) and the sugar-bound (yellow; PDB code 3S5X) conformations seen in the X-ray structures in both binding site 1 (A) and binding site 2 (B).

3.3 DISCUSSION

Given the popularity of the excited state model, and its apparent validity in several systems, (Boehr et al., 2006; Eisenmesser et al., 2005; Korzhnev et al., 2009; Sugase et al.,

2007) we first measured ^{15}N CPMG relaxation dispersion experiments to identify and characterize potential excited states sampled by free OAA. Indeed the presence of an excited state was established, however both the spatial distribution of residues undergoing conformational exchange between the ground and the excited state and the chemical shift difference between the two states does not correlate with differences observed between the free and sugar-saturated OAA. These results indicate that the excited state does not structurally resemble the bound state, suggesting it is not related to binding of Man5.

In addition, we explored the possibility of a conformational equilibrium between the free and the bound X-ray conformations within the ground state of free OAA. To this end, structural parameters in solution were analyzed and compared to the X-ray structures in the absence and presence of ligand. In particular, interproton distances and dihedral angles were queried through NOE and the $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants measurements. Additionally, the solution structure on OAA was determined and compared to the X-ray crystal structures.

The inter-proton distances and dihedral angles derived from NOE intensities and $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants measured on ligand-saturated OAA are in good agreement with the X-ray crystallography structure of Man5-bound OAA. These observations indicate that the conformation adopted by Man5-bound OAA in solution corresponds to the crystallographic structure. In contrast, both the NOE intensities and $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants measured on ligand-free OAA cannot be explained by the free crystal structure alone, but instead by a conformational equilibrium between the free and the bound crystallographic conformations. These results are corroborated by the solution structure of OAA, determined on the basis of NOE and RDC restraints: the free and the bound X-ray conformations are seen for both binding sites in different conformers of the NMR structural model. Of note, the $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants were not used in the structural calculation, and thus represent an independent assessment of the conformational space sampled by free OAA in solution. Additionally, the electron density map of sugar-free OAA suggests structural flexibility in binding site 2, and does not exclude the sugar-bound conformation (Figure 3.11). (Carneiro et al., 2015c)

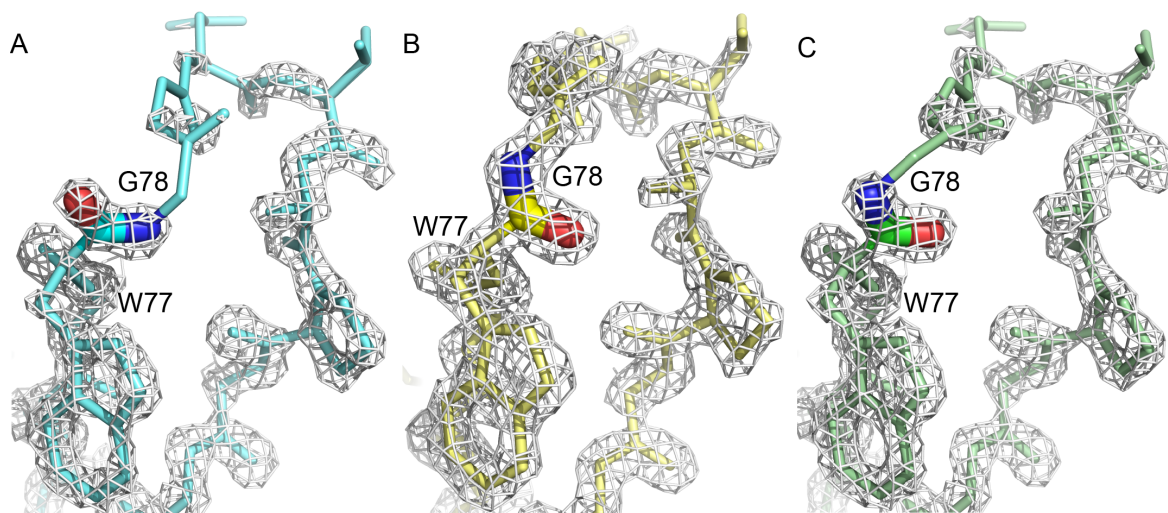


Figure 3.11 Electron density maps of OAA suggest structural flexibility in the free binding site 2. (A) Electron density map and X-ray structure of free OAA (PDB code 3S5V). The weak electron density in the backbone region around the peptide bond between W77 and G78 suggests structural flexibility. (B) Electron density map and X-ray structure of sugar-bound OAA (PDB code 3S5X; Man5 is not shown for clarity). The strong electron density throughout the loop suggests that in the presence of ligand the loop is in a predominant conformation. (C) Electron density map of free OAA (PDB code 3S5V) and a hybrid structural model, in which the peptide bond between W77 and G78 is flipped to mimic the bound conformation. The lack of defined electron density allows for a good fit of this model to the electron density map. The peptide bond between W77 and G78 is highlighted in all panels in stick representation, with the backbone carbonyl and amide groups colored in red and blue, respectively. All electron density maps are contoured at 2σ . The electron density maps were generated with Coot (Emsley et al., 2010) from PDB codes 3S5V and 3S5X, and the final images were generated in PyMOL. (Schrodinger, 2010)

Taken together, these results also suggest that the X-ray sugar-bound conformer is thermodynamically favorable (i.e. highly populated). Moreover, inspection of X-ray structures available for several OAA homologues, (Koharudin et al., 2012; Whitley et al., 2013) with high conservation of binding-site residues and Man5–protein interactions, revealed that the sugar-bound conformation is frequently seen in sugar-free structures. (Carneiro et al., 2015c) These observations indicate that these structural changes are associated with motions within the ground state, rather than between the ground and excited states.

3.4 MATERIALS AND METHODS

3.4.1 SAMPLE PREPARATION

NMR samples of free OAA consisted 2 mM of ^{15}N , or ^{15}N and ^{13}C labeled OAA in 20 mM sodium acetate (pH 5.0), 20 mM sodium chloride, 3 mM sodium azide, and 90/10% $\text{H}_2\text{O}/\text{D}_2\text{O}$ or 100% D_2O .

Sugar-saturated OAA sample consisted of 1 mM of ^{15}N labeled OAA and 12 mM of α 3, α 6-mannopentaose (Sigma-Aldrich), in 20 mM sodium acetate (pH 5.0), 20 mM sodium chloride, 3 mM sodium azide, and 90/10% $\text{H}_2\text{O}/\text{D}_2\text{O}$.

NMR sample used for RDCs measurement as prepared as described in (Rückert and Otting, 2000) and consisted of 500 μM of ^{15}N labeled OAA in a dilute liquid crystalline phase of N-octylpentaoxyethylene (C8E5) and octanol. The molar ratio of C8E5 to octanol of 0.87 and the C8E5 to water ratio was 5 % (w/w).

OAA was expressed and purified as described in (Koharudin et al., 2011) and kindly provided by A. M. Gronenborn (University of Pittsburgh).

3.4.2 NMR SPECTROSCOPY

3.4.2.1 ^{15}N -CPMG relaxation dispersion

Proton detected constant-time ^{15}N -CPMG relaxation dispersion experiments were recorded at 298 K on 800 MHz AVANCE III and 700 MHz AVANCE III spectrometers equipped with cryogenic and room-temperature triple resonance probes, respectively, and at 277 K on 800 MHz AVANCE III and 600 MHz AVANCE spectrometers equipped with cryogenic triple resonance probes, as described in (Ban et al., 2013a). A constant relaxation delay (T_{CP}) of 50 ms was used and CPMG frequencies from 80 to 960 Hz were employed. Duplicated experiments at three different CPMG frequencies were recorded for error estimation.

3.4.2.2 NOEs

^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY spectra were acquired for ligand-free ^{15}N labeled OAA at 298 and 277 K, and for sugar-saturated ^{15}N labeled OAA at 298 K. Experimental details regarding the spectrum recorded for ligand-free OAA at 298 K are provided in section 2.4.2. The ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY spectrum of ligand-free OAA at 277 K was acquired on a 800 MHz AVANCE III spectrometer equipped with a 5-mm triple-resonance cryogenic probe, with 178, 100, and 1024 complex points and acquisition times of 8.0, 18.1, and 45.9 ms in the ^1H (t_1), ^{15}N (t_2) and ^1H (t_3) dimensions, respectively, using 16 scans per t_1/t_2 increment. The ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY spectrum of sugar-saturated OAA was acquired at 298 K on a 900 MHz AVANCE spectrometer equipped with a 5-mm triple-resonance cryogenic probe, with 218, 108, and 1024 complex points and acquisition times of 8.6, 17.3, and 40.5 ms in the ^1H (t_1), ^{15}N (t_2) and ^1H (t_3) dimensions, respectively, using 16 scans per t_1/t_2 increment.

3.4.2.3 RDCs

For the determination of NH residual dipolar couplings, IPAP-HSQC experiments (Ottiger et al., 1998) were measured on ^{15}N labeled OAA samples in both isotropic and partially aligned conditions at 298 K, with the in-phase (IP) and anti-phase (AP) spectra recorded in an interleaved manner. The IPAP-HSQC experiment under isotropic conditions was recorded on a 600 MHz AVANCE III spectrometer equipped with a 5-mm triple-resonance room-temperature probe, with 712 and 2048 complex points per experiment and acquisition times of 172.3 ms and 121.6 ms in the ^{15}N (t_1) and ^1H (t_2) dimensions, respectively, using 32 scans per t_1 increment. The IPAP-HSQC spectrum of the partially aligned sample was acquired on a 700 MHz AVANCE spectrometer equipped with a 5-mm triple-resonance cryogenic probe, with 340 and 1024 complex points per experiment and acquisition times of 70.5 and 52.4 ms in the ^{15}N (t_1) and ^1H (t_2) dimensions, respectively, using 80 scans per t_1 increment.

3.4.2.4 $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants

For the determination of $^3J(\text{H}_\text{N}, \text{H}_\alpha)$ coupling constants 3D-HNHA spectra (Vuister and Bax, 1993) were acquired at 298 K on both free and sugar-saturated ^{15}N labeled OAA samples. The 3D-HNHA spectrum of free OAA was recorded on a 700 MHz AVANCE III

spectrometer equipped with a 5-mm triple-resonance room-temperature probe, with 106, 112 and 2048 complex points and acquisition times of 21.9, 5.7 and 104.4 ms in the ^{15}N (t_1), ^1H (t_2), and ^1H (t_3) dimensions, respectively, using 16 scans per t_1/t_2 increment. The 3D-HNHA spectrum of sugar-saturated OAA was recorded on a 800 MHz AVANCE III spectrometer equipped with a 5-mm triple-resonance cryogenic probe, with 112, 114 and 2048 complex points and acquisition times of 20.3, 5.0 and 91.7 ms in the ^{15}N (t_1), ^1H (t_2), and ^1H (t_3) dimensions, respectively, using 32 scans per t_1/t_2 increment.

3.4.3 DATA ANALYSIS

All spectra were processed using NMRPipe (Delaglio et al., 1995) and analyzed with CARA. (Keller, 2004)

3.4.3.1 ^{15}N CPMG relaxation dispersion

Peak intensities were quantified using the model-based linear equation system implemented in CARA. (Keller, 2004) The effective transverse relaxation rate (R_2^{eff}) at each CPMG frequency (ν_{CPMG}) was calculated as: (Korzhnev et al., 2004)

$$R_2^{\text{eff}}(\nu_{\text{CPMG}}) = -\frac{1}{T_{\text{CP}}} \ln \frac{I_i(\nu_{\text{CPMG}})}{I_0} \quad (3.6)$$

where $I_i(\nu_{\text{CPMG}})$ refers to the peak intensity at a given CPMG frequency, and I_0 to the peak intensity in the reference spectrum (for which $T_{\text{CP}} = 0$). Uncertainties associated with R_2^{eff} were calculated based on the root mean square deviation (R.M.S.D.) of the peak intensity estimated from repeat measurements as: (Korzhnev et al., 2004)

$$\Delta R_2^{\text{eff}}(\nu_{\text{CPMG}}) = \frac{1}{T_{\text{CP}}} \frac{\text{RMSD}}{I_i(\nu_{\text{CPMG}})} \quad (3.7)$$

Dispersion profiles that exhibit a difference in R_2^{eff} between the lowest and highest CPMG frequency larger than 2 s^{-1} were fitted on a per residue basis to a two-state model using the program *ShereKhan* (Mazur et al., 2013) to models assuming slow (Bloch-McConnell equation (McConnell, 1958)) and fast (Luz-Meiboom equation (Luz and Meiboom, 1963)) exchange regime. Model selection was based on the small-sample

corrected Akaike information criterion (AICc), calculated as: (Burnham and Anderson, 2002)

$$AICc = \chi^2 + 2m + \frac{2m(m+1)}{n-m-1} \quad (3.8)$$

where χ^2 is the target function value from a least-square minimization, m the number of parameters in a given model ($m = 5$ or 4 , for slow or fast exchange model, respectively) and n is the number of experimental R_2^{eff} values ($n = 24$). For each residue, the AICc difference between the two models ($\Delta AICc = AICc_{\text{fast}} - AICc_{\text{slow}}$) was evaluated, with residues for which $\Delta AICc > 10$ being assigned to slow exchange.

3.4.3.2 NOEs

NOE peak intensities were quantified using the model-based linear equation system implemented in CARA. (Keller, 2004) Estimation of populations based on NOE intensity ratios were derived as described in section 3.2.2.

3.4.3.3 RDCs

Backbone NH residual dipolar couplings ($D_{HN,N}$) were calculated as:

$$D_{HN,N} = T_{HN,N} - J_{HN,N} \quad (3.9)$$

where $T_{HN,N} = J_{HN,N} + D_{HN,N}$ corresponds to the total splitting between the downfield and upfield components of the IPAP-HSQC spectrum measured under anisotropic conditions, and $J_{HN,N}$ corresponds to the splitting between the downfield and upfield components of the IPAP-HSQC spectrum measured under isotropic conditions. Uncertainties (σ) in the measured splittings ($T_{HN,N}$ and $J_{HN,N}$) were estimated based on the uncertainties in the peak position (δ) of the up- and downfield components of the IPAP-HSQC spectra as:

$$\sigma_{\text{splitting}} = \sqrt{\left(\sigma_{\delta_{\text{upfield}}}^2 + \sigma_{\delta_{\text{downfield}}}^2\right)} \quad (3.10)$$

$$\sigma_{\delta} = LW/(S/N) \quad (3.11)$$

where LW corresponds to the line width at half height and S/N to the signal-to-noise ratio. (Bax et al., 2001) The total error associated with the measured RDCs was estimated as:

$$\sigma_{D_{HN,N}} = \sigma_{T_{HN,N}} + \sigma_{J_{HN,N}} \quad (3.12)$$

3.4.3.4 ${}^3J(H_N, H_\alpha)$ coupling constants

${}^3J(H_N, H_\alpha)$ coupling constants of both free and sugar-saturated OAA were calculated from the intensity ratio between cross and diagonal peaks in a 3D-HNHA spectra as: (Vuister and Bax, 1993)

$$\frac{I_{cross}}{I_{diagonal}} = -\tan^2(2\pi J(H_N, H_\alpha) \zeta) \quad (3.13)$$

where ζ refers to the scalar coupling labeling delay, and the peak intensities were quantified using the model-based linear equation system implemented in CARA. (Keller, 2004) The uncertainties associated with the calculated ${}^3J(H_N, H_\alpha)$ were estimated by Monte Carlo simulations with 1000 iterations and taking the signal-to-noise ratio of the measured peak intensities as experimental error.

${}^3J(H_N, H_\alpha)$ coupling constants were predicted from ϕ dihedral angles measured in the free and sugar-complexed OAA X-ray structures using Eq. (3.4). Uncertainties associated with these predicted ${}^3J(H_N, H_\alpha)$ were estimated by Monte Carlo simulations with 1000 iterations using the standard deviations of the Karplus coefficients from (Habeck et al., 2005).

Relative populations from ${}^3J(H_N, H_\alpha)$ coupling constants were estimated by minimizing the target function:

$$\chi^2 = \sum_{i=1}^N \left(\frac{{}^3J(H_N, H_{\alpha,i})_{exp} - {}^3J(H_N, H_{\alpha,i})_{calc}}{\sigma_i} \right)^2 \quad (3.14)$$

where ${}^3J(H_N, H_{\alpha,i})_{exp}$ are the experimentally measured coupling constants, ${}^3J(H_N, H_{\alpha,i})_{calc}$ are the coupling constants calculated using Eq. (3.5) and σ_i is the sum of the experimental uncertainty and the uncertainties associated with the predicted coupling constants.

3.4.4 STRUCTURE DETERMINATION

NOE cross-peaks were automatically picked and assigned using the ATNOS/CANDID module of UNIO'10 (Herrmann et al., 2002a, b) and CYANA 3.96, (Güntert et al., 1997) using as input one chemical shift list, resulting from the sequence specific resonance assignment (Chapter 2), and three peak lists from the ^{15}N -, $^{13}\text{C}_{\text{aliphatic}}$ -, and $^{13}\text{C}_{\text{aromatic}}$ -resolved [$^1\text{H}, ^1\text{H}$]-NOESY spectra. The procedure resulted in 2419 NOE derived distance constraints. In addition, the presence of hydrogen bonds was evaluated based on initial structural calculations performed with CYANA 3.96 (Güntert et al., 1997). A total of 89 hydrogen bonds was observed in more than 50% of the 20 lowest energy conformers and included in subsequent runs as distance constraints of the distance between the acceptor (O) and the hydrogen atom of 1.8–2 Å, and 2.7–3.0 Å between the acceptor (O) and the donor (N). (Güntert, 1998)

The final structure calculations were performed with Xplor-NIH (Schwieters et al., 2003), and proceeded in several steps. An initial set of 200 structures was calculated and refined with standard simulated annealing protocols using only distance constraints. From those 200 structures, the ones that displayed (i) NOE violations larger than 0.5 Å, (ii) R.M.S.D. from idealized bond lengths larger than 0.01 Å, or (iii) R.M.S.D. from idealized bond angles larger than 2°, were excluded. The remainder 157 structures were used as input for a second simulated annealing step that included residual dipolar coupling constraints, in addition to the distance constraints. The 153 structures that met the acceptance criteria (i)–(iii) underwent a final energy minimization step with explicit solvent. The 20 lowest energy structures without any NOE violations were selected to represent the solution structure of OAA and were used to generate structural statistics using the PSVS 1.5 suite. (Bhattacharya et al., 2007) The program MOLMOL (Koradi et al., 1996) was used to analyze the structures and to prepare the structural representations in the figures. The NMR data used for the structural calculation was deposited in the Biological Magnetic Resonance Data Bank (accession code: 25324) and the solution structure of OAA was deposited in the Protein Data Bank (accession code: 2MWH).

4 IMPROVING NMR METHODS FOR THE STUDY OF PROTEIN DYNAMICS

4.1 INTRODUCTION

NMR has emerged as a powerful technique to study protein dynamics, providing information over a wide range of timescales with atomic resolution. Although the effects of conformational exchange on different NMR observables have been described several decades ago, new tools and techniques are constantly being developed to improve and extend the timescales probed and the parameters extracted that describe dynamic processes. ((Palmer III, 2014) and references therein)

In this chapter two new tools are described, which represent efforts to improve the study of protein dynamics by two NMR techniques, exchange-mediated saturation transfer and CPMG relaxation dispersion experiments.

Recently, exchange-mediated saturation transfer applications to protein NMR were developed, namely CEST (Vallurupalli et al., 2012) and DEST, (Fawzi et al., 2011) for the characterization of slow processes with very lowly populated conformations, and exchange between low molecular weight and high molecular weight assemblies, respectively. While exchange-mediated saturation transfer experiments provide a wealth of information on challenging systems, they require extensive measurement time and thus their applications to metastable systems might be hindered. In the first part of this chapter, a strategy based on Fourier transform and linear prediction is described, which enables a two-fold reduction in measurement time. In addition, we show that the simultaneous analysis of data collected with two radio-frequency field strengths is required for obtaining reliable exchange parameters. (Carneiro et al., 2015b)

The second part of this chapter is dedicated to a statistical procedure to identify dynamic clusters within a protein. A critical step in evaluating the site-specific data obtained from NMR experiments is to assess whether all residues are reporting on the same exchange process (in which case all residues display comparable kinetics and can be fitted to a global exchange rate and population) or if different residues report on different processes. In the limit, each residue reports on a different exchange process and must be

fitted individually. The selection of residues to be fitted together (i.e., reporting on the same dynamic process) is usually based on the user-dependent evaluation of the similarity between the fitting parameters obtained from individual fits, and validated with statistical tools such as F-test and AIC analysis. (Kovrigin and Loria, 2006) Alternatively, outliers of a global process are excluded by comparing the residuals obtained from individual fits ($\chi^2_{\text{individual}}$), to the per-site residuals obtained from a global fit (χ^2_{global}), with the sites experiencing conformational fluctuations distinct from the global process being identified by large values of $\chi^2_{\text{global}}/\chi^2_{\text{individual}}$. (Mulder et al., 2001) Here we propose an iterative statistical strategy to cluster residues undergoing chemical exchange based on model selection using the small-sample adjusted Akaike information criterion (AICc). (Hurvich and Tsai, 1989) This method is then independent of user interference and therefore at least potentially less biased. The method is demonstrated using CPMG relaxation dispersion data, but the same principle can be extended to other experiments (e.g., CEST/DEST, $R_{1\rho}$).

4.2 RESULTS AND DISCUSSION

4.2.1 SPEEDING-UP EXCHANGE-MEDIATED SATURATION TRANSFER EXPERIMENTS BY FOURIER TRANSFORM

CEST and DEST experiments are commonly described as saturation transfer experiments and analyzed with the numerical solution of the Bloch-McConnell equation. (McConnell, 1958) However, exchange-mediated saturation transfer and $R_{1\rho}$ experiments are essentially identical, the main difference being that in CEST/DEST experiments the magnetization is kept along the Z-axis during the relaxation delay, whereas in $R_{1\rho}$ experiments the magnetization is rotated (at least partially) into the transverse plane. (Palmer III, 2014; Zaiss and Bachert, 2013; Zhao et al., 2014) Further, approximated solutions of the $R_{1\rho}$ relaxation have been derived, based on the observation that the $R_{1\rho}$ relaxation rate corresponds to the largest eigenvalue of the Bloch-McConnell equation. Since this eigenvalue describes both transverse and longitudinal magnetization, the $R_{1\rho}$ analytical solution is also suitable to analyze CEST/DEST data. (Palmer III, 2014; Zaiss and Bachert, 2013) A comparison between the use of the numerical solution of the Bloch-McConnell equation and the $R_{1\rho}$ analytical solution is shown in Figure 4.1 and Table 4.1.

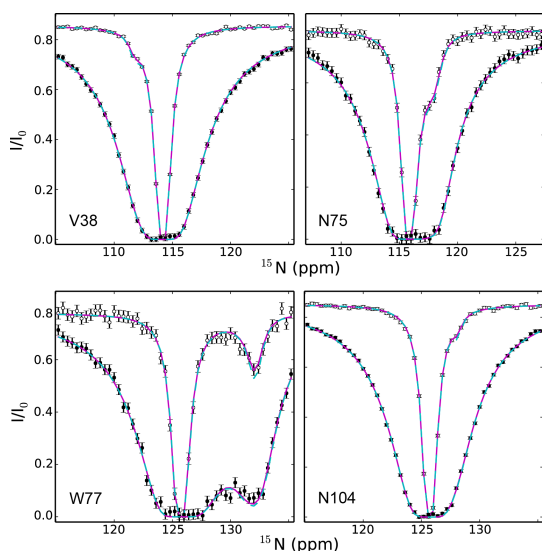


Figure 4.1 Both the analytical $R_{1\rho}$ solution and the numerical solution of the Bloch-McConnell equation are suitable to analyze exchange-mediated saturation transfer profiles. CEST profiles are shown for four representative residues of OAA (V38, N75, W77 and N104). For each residue, the data collected with two radio-frequency field strengths (15 and 75 Hz, open and filled circles, respectively) was simultaneously fitted to either the analytical approximation derived for $R_{1\rho}$ relaxation (magenta) or to the exact solution of the Bloch-McConnell equation (cyan). The curves provide essentially the same result and are therefore overlapped.

Table 4.1 Fitting parameters obtained from individual fit of CEST profiles of four residues of OAA with two radio-frequency field strengths (15 and 75 Hz) to either the analytical solution derived for $R_{1\rho}$ relaxation or to the Bloch-McConnell equation.

	Analytical	Bloch-McConnell
V38		
k_{ex} (s^{-1})	100 ± 28	94 ± 25
p_b (%) [*]	0.6 ± 0.2	0.6 ± 0.2
$\Delta\delta$ (ppm)	-2.43 ± 0.06	-2.44 ± 0.05
R_2^a (s^{-1})	25.6 ± 0.2	25.7 ± 0.2
R_2^b (s^{-1})	134 ± 29	109 ± 31
R_1 (s^{-1}) ^{**}	0.391 ± 0.002	0.390 ± 0.002
Reduced χ^2	1.14	1.13
N75		
k_{ex} (s^{-1})	342 ± 40	326 ± 39
p_b (%) [*]	2.6 ± 0.4	2.8 ± 0.4
$\Delta\delta$ (ppm)	2.28 ± 0.07	2.28 ± 0.07
R_2^a (s^{-1})	20.7 ± 0.7	20.8 ± 0.7
R_2^b (s^{-1})	26 ± 17	24 ± 14
R_1 (s^{-1}) ^{**}	0.556 ± 0.009	0.556 ± 0.009
Reduced χ^2	0.38	0.38
W77		
k_{ex} (s^{-1})	279 ± 33	258 ± 30
p_b (%) [*]	2.9 ± 0.4	3.2 ± 0.5
$\Delta\delta$ (ppm)	6.48 ± 0.06	6.48 ± 0.07
R_2^a (s^{-1})	26.5 ± 0.9	26.7 ± 0.8
R_2^b (s^{-1})	79 ± 27	70 ± 29
R_1 (s^{-1}) ^{**}	0.63 ± 0.01	0.63 ± 0.01
Reduced χ^2	0.69	0.68
N104		
k_{ex} (s^{-1})	112 ± 21	106 ± 21
p_b (%) [*]	0.5 ± 0.1	0.6 ± 0.1
$\Delta\delta$ (ppm)	2.40 ± 0.07	2.41 ± 0.07
R_2^a (s^{-1})	23.9 ± 0.2	24.0 ± 0.2

R_2^b (s ⁻¹)	102 ± 19	84 ± 19
R_1 (s ⁻¹)**	0.403 ± 0.003	0.403 ± 0.002
Reduced χ^2	0.98	0.98

*CEST profiles were fitted assuming a two-state model, such that $p_a + p_b = 1$.

**The longitudinal relaxation rate (R_1) is assumed to be the same for both states.

The fitting parameters and the quality of the fit obtained with the analytical solution are in excellent agreement with the ones obtained with the numerical solution of the Bloch-McConnell equation. This result indicates that both models are equally applicable to the analysis of CEST/DEST data. However, the utilization of the analytical solutions offers a few advantages compared to the Bloch-McConnell equation. First, the computation time is less demanding — using the synthetic data from Table 4.2, at least a three-fold reduction in computation time was obtained with the analytical solution. This aspect is especially beneficial when the uncertainty in the fitting parameters is estimated from Monte Carlo runs, for which typically more than 100 calculations have to be done. Furthermore, in the case that multiple residues are governed by a single dynamic process, and thus have to be fit simultaneously to obtain the kinetic parameters, the alleviation of computational time demands resulting from the usage of the analytical solution is significantly beneficial. Another advantage of the usage of the analytical solution is that one can infer the impact of the experimental parameters in the accuracy of fitted parameters. For this purpose, we examined the $R_{1\rho}$ analytical solution: (Baldwin and Kay, 2013; Trott and Palmer III, 2002)

$$R_{1\rho} = C_{R1}R_1 \cos^2 \theta + (C_{R2}R_2^a + R_{ex}) \sin^2 \theta \quad (4.1)$$

where $\cos^2 \theta = ((\delta_a - \delta_{RF}) + p_b \Delta\delta)^2 / (v_1^2 + ((\delta_a - \delta_{RF}) + p_b \Delta\delta)^2)$ and $\sin^2 \theta = v_1^2 / (v_1^2 + ((\delta_a - \delta_{RF}) + p_b \Delta\delta)^2)$; $\Delta\delta = \delta_b - \delta_a$; v_1 is the applied radio-frequency field strength, p_b is the population of the minor state, $\delta_{a/b}$ is the resonance frequency of the major (a) or minor (b) state and δ_{RF} is the resonance frequency at which the radio-frequency field v_1 is applied. The magnitude of the *sin* term is proportional to v_1 and, in the limit of very weak field strengths, the *sin* term is negligible and only the *cos* term contributes significantly to $R_{1\rho}$. The coefficient C_{R1} , Eq. (4.2), has explicit contributions from the intrinsic transverse relaxation rate of the major (R_2^a) and the minor (R_2^b) states, the exchange rate (k_{ex}), and the resonance frequencies and relative populations of both states:

$$C_{R1} = \frac{k_{ex}^2 + v_1^2 + (\delta_a - \delta_{RF})^2(\delta_b - \delta_{RF})^2 / ((\delta_a - \delta_{RF}) + p_b \Delta \delta)^2 + (c_1 + (c_2 - c_3)(R_2^b - R_2^a)) \tan^2 \theta}{k_{ex}^2 + (v_1^2 + (\delta_a - \delta_{RF})^2)(v_1^2 + (\delta_b - \delta_{RF})^2) / (v_1^2 + ((\delta_a - \delta_{RF}) + p_b \Delta \delta)^2)} + c_2 (R_2^b - R_2^a) \sin^2 \theta \quad (4.2)$$

where,

$$c_1 = p_a p_b \Delta \delta^2 \quad (4.3)$$

$$c_2 = 3p_b k_{ex} + \left(2p_b k_{ex} + \frac{v_1^2}{k_{ex}} + (R_2^b - R_2^a) + \frac{(R_2^b - R_2^a) p_b^2 k_{ex}^2}{(v_1^2 + (\delta_a - \delta_{RF})^2)} \right) \frac{(v_1^2 + (\delta_a - \delta_{RF})^2)}{v_1^2} \quad (4.4)$$

$$c_3 = 2k_{ex} + \frac{v_1^2}{k_{ex}} + (R_2^b - R_2^a) p_a \quad (4.5)$$

Again, the *sin* and *tan* terms in Eq. (4.2) are proportional to v_1 and, in the limit, are negligible. Thus, with the weak radio-frequency field strengths used in CEST experiments, often in the order of 10 to 20 Hz, the measured effect is nearly independent of R_2 . This observation suggests that with such weak field strengths the estimation of accurate transverse relaxation rates might be impaired. A larger contribution from R_2 can be found in the coefficients of the *sin* term in Eq.(4.1), C_{R2} and R_{ex} :

$$C_{R2} = \frac{c_4 + c_3(R_2^b - R_2^a)}{c_5 + c_3(R_2^b - R_2^a) \sin^2 \theta} \quad (4.6)$$

$$R_{ex} = \frac{c_1 k_{ex} + c_6(R_2^b - R_2^a)}{c_5 + c_3(R_2^b - R_2^a) \sin^2 \theta} \quad (4.7)$$

with,

$$c_4 = c_5 \csc^2 \theta - \left(k_{ex}^2 + v_1^2 + \frac{(\delta_a - \delta_{RF})^2(\delta_b - \delta_{RF})^2}{((\delta_a - \delta_{RF}) + p_b \Delta \delta)^2} \right) \cot^2 \theta - c_1 \quad (4.8)$$

$$c_5 = k_{ex}^2 + \frac{(v_1^2 + (\delta_a - \delta_{RF})^2)(v_1^2 + (\delta_b - \delta_{RF})^2)}{(v_1^2 + ((\delta_a - \delta_{RF}) + p_b \Delta \delta)^2)} \quad (4.9)$$

$$c_6 = p_b \left((v_1^2 + (\delta_a - \delta_{RF})^2) + k_{ex}^2 + (R_2^b - R_2^a) p_a k_{ex} \right) \quad (4.10)$$

The contribution of this term to the $R_{1\rho}$ can be augmented by using stronger radio-frequency field strengths, which should favor the determination of the R_2 rates. The relationship between the radio-frequency field strength and the fitted transverse relaxation rates was elucidated using simulated CEST data (Table 4.2 and Figure 4.2), for which all parameters are known *a priori*.

Table 4.2 Fitting parameters extracted from individual fit of CEST profiles with two different radio-frequency field strengths (15 or 75 Hz) and from simultaneous fit with the two radio-frequency field strengths (15 and 75 Hz), using the analytical solution.

	$\nu_1 = 15 \text{ Hz}$	$\nu_1 = 75 \text{ Hz}$	$\nu_1 = 15 \text{ and } 75 \text{ Hz}$
Simulated*			
$k_{\text{ex}} (150 \text{ s}^{-1})$	174 ± 43	72 ± 83	143 ± 7
$p_b (1\%)^{**}$	0.9 ± 0.3	2 ± 2	1.02 ± 0.06
$\Delta\delta (5 \text{ ppm})$	5.0 ± 0.4	4.95 ± 0.09	4.99 ± 0.02
$R_2^a (20 \text{ s}^{-1})$	19.8 ± 0.4	20.9 ± 0.9	20.2 ± 0.2
$R_2^b (100 \text{ s}^{-1})$	32 ± 39	65 ± 29	103 ± 8
$R_1 (0.5 \text{ s}^{-1})^{***}$	0.505 ± 0.004	0.50 ± 0.01	0.500 ± 0.003
Reduced χ^2	0.81	1.01	0.93
V38			
$k_{\text{ex}} (\text{s}^{-1})$	209 ± 65	903 ± 3535	101 ± 28
$p_b (\%)^{**}$	0.5 ± 0.2	2 ± 378	0.6 ± 0.2
$\Delta\delta (\text{ppm})$	-2.52 ± 0.07	-1 ± 1	-2.43 ± 0.06
$R_2^a (\text{s}^{-1})$	24.9 ± 0.5	28 ± 37	25.6 ± 0.2
$R_2^b (\text{s}^{-1})$	35 ± 62	0 ± 19	134 ± 29
$R_1 (\text{s}^{-1})^{***}$	0.392 ± 0.002	0.39 ± 0.02	0.391 ± 0.002
Reduced χ^2	1.94	0.43	1.14
W77			
$k_{\text{ex}} (\text{s}^{-1})$	340 ± 99	531 ± 116	279 ± 33
$p_b (\%)^{**}$	3 ± 1	2.1 ± 0.6	2.9 ± 0.4
$\Delta\delta (\text{ppm})$	6.5 ± 0.1	6.7 ± 0.2	6.48 ± 0.06
$R_2^a (\text{s}^{-1})$	24 ± 4	24 ± 4	26.5 ± 0.9
$R_2^b (\text{s}^{-1})$	77 ± 129	0 ± 40	79 ± 27
$R_1 (\text{s}^{-1})^{***}$	0.62 ± 0.02	0.64 ± 0.05	0.63 ± 0.01
Reduced χ^2	0.48	0.78	0.69

*Parameters used to simulate the CEST profiles are shown in parenthesis.

**CEST profiles were fitted assuming a two-state model, such that $p_a + p_b = 1$.

***The longitudinal relaxation rate (R_1) is assumed to be the same for both states.

As expected from the theoretical analysis above, the fitted R_2 , in particular R_2^b , obtained at a weak field strength (15 Hz) deviates largely from the target value, and the use of stronger fields (75 Hz) improves its determination. Interestingly, the fitted k_{ex} at 75 Hz, but not at 15 Hz, differs largely from the target value. That is, using data with a single radio-frequency impairs the estimation of either k_{ex} or R_2^b , despite the overall good quality of the fit (reduced χ^2 of 0.8 and 1.0 for 15 and 75 Hz, respectively; $\chi^2_{\text{reduced}} = \chi^2/(v-p-1)$, where χ^2 corresponds to Eq. (4.14), v corresponds to the number of data points and p to the number of fitting parameters), indicating that these parameters are also interdependent. Conversely,

the simultaneous fit of data at both 15 and 75 Hz results in accurate estimation of all parameters. The same trend was seen using experimental data, and examples are shown for V38 and W77 of OAA (Table 4.2 and Figure 4.2). Even though the quality of the three fits is comparable, the fitted values of R_2^b , k_{ex} and p_b are significantly different between the three analysis.

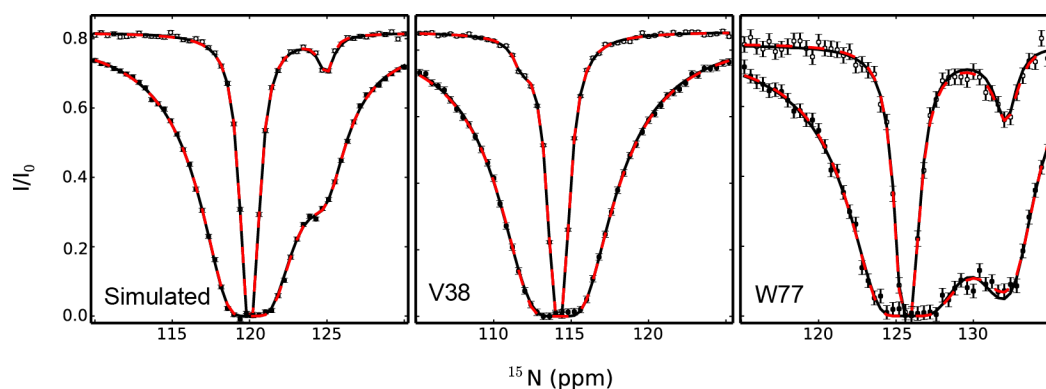


Figure 4.2 Simultaneous fit of two field strengths is required for obtaining reliable exchange parameters. CEST profiles with two radio-frequency field strengths (15 and 75 Hz, open and filled circles, respectively) were simulated using k_{ex} , p_b , δ_a , δ_b , R_1 , R_2^a , and R_2^b of 150 s⁻¹, 1%, 120 ppm, 125 ppm, 0.5 s⁻¹, 20.0 s⁻¹, and 100.0 s⁻¹, respectively. Experimental CEST profiles for the same radio-frequency field strengths are also shown for V38 and W77 of OAA. The best fit of the CEST profiles on a per residue and per field strength basis to the analytical solution is represented by the red dashed line. The best fit of the CEST profiles on a per residue basis (i.e., simultaneous fit of both radio-frequency field strengths) is represented by the black dashed line. Despite the overall equivalent quality of the two fits, some fitting parameters vary significantly (see also Table 4.2).

The requirement of recording data with more than one radio-frequency field strength along with the numerous frequency offsets measured for exchange-mediated saturation transfer profiles results in extensive measurement time. The frequency offset at which the radio-frequency field is applied is usually incremented in steps of 15 to 30 Hz. In a first attempt to reduce the experimental time required for CEST experiments, we analyzed the effect of under-sampling the frequency offsets. Table 4.3 and Figure 4.3 compare the CEST profiles and the fitted parameters for four representative residues of OAA, using a conventional incremental step (28 Hz per frequency offset; hereafter termed conventional dataset) or a twice-larger incremental step (56 Hz per frequency offset; hereafter termed reduced dataset). While $\Delta\delta$, p_b , R_2^a , and R_1 , are comparable for both data sets, k_{ex} and R_2^b can differ significantly between the two data sets, with residues undergoing conformational exchange with very slow exchange rates, small chemical shift differences and highly skewed populations (e.g., V38 and N75) being particularly affected.

In order to retain the advantages of reduced measurement time without compromising the estimation of the exchange parameters, we developed a strategy to extend a reduced profile using Fourier transform (FT). CEST and DEST profiles are in essence frequency domain spectra, and can be converted into a time-domain signal using inverse FT. (Hoch and Stern, 1996) In the time domain, the signal can be extrapolated using linear prediction (LP) algorithms commonly used to process NMR data. Given the small number of input data points in these cases, the LP extrapolation is limited to a factor of two (Hoch and Stern, 1996) The extrapolated time domain signal is reconverted into a frequency domain spectrum using FT.

Table 4.3 Fitting parameters derived from FT-CEST are comparable to the ones derived from conventional CEST. Fitting parameters obtained from the analysis of CEST profiles for V38, N75, W77 and N104 of OAA using reduced, FT-extended and conventional datasets.

	Reduced	FT	Conventional
V38			
k_{ex} (s ⁻¹)	55 ± 43	93 ± 46	100 ± 28
p_b (%) [*]	0.8 ± 0.8	0.6 ± 0.4	0.6 ± 0.2
$\Delta\delta$ (ppm)	-2.32 ± 0.07	-2.4 ± 0.1	-2.43 ± 0.06
R_2^a (s ⁻¹)	25.5 ± 0.6	25.6 ± 0.5	25.6 ± 0.2
R_2^b (s ⁻¹)	97 ± 47	106 ± 40	134 ± 29
R_1 (s ⁻¹) ^{**}	0.392 ± 0.003	0.394 ± 0.006	0.391 ± 0.002
Reduced χ^2	1.32	0.72	1.14
N75			
k_{ex} (s ⁻¹)	360 ± 57	347 ± 41	342 ± 40
p_b (%) [*]	2.6 ± 0.6	2.6 ± 0.5	2.6 ± 0.4
$\Delta\delta$ (ppm)	2.28 ± 0.09	2.29 ± 0.07	2.28 ± 0.07
R_2^a (s ⁻¹)	20 ± 1	20.5 ± 0.7	20.7 ± 0.7
R_2^b (s ⁻¹)	29 ± 26	29 ± 18	26 ± 17
R_1 (s ⁻¹) ^{**}	0.55 ± 0.01	0.55 ± 0.01	0.556 ± 0.009
Reduced χ^2	0.33	0.44	0.38
W77			
k_{ex} (s ⁻¹)	286 ± 43	262 ± 28	279 ± 33
p_b (%) [*]	2.8 ± 0.5	2.9 ± 0.4	2.9 ± 0.4
$\Delta\delta$ (ppm)	6.57 ± 0.09	6.50 ± 0.07	6.48 ± 0.06
R_2^a (s ⁻¹)	26 ± 1	26.5 ± 0.9	26.5 ± 0.9
R_2^b (s ⁻¹)	56 ± 39	83 ± 28	79 ± 27
R_1 (s ⁻¹) ^{**}	0.63 ± 0.02	0.63 ± 0.01	0.63 ± 0.01
Reduced χ^2	0.82	0.93	0.69
N104			
k_{ex} (s ⁻¹)	81 ± 38	91 ± 35	112 ± 21
p_b (%) [*]	0.5 ± 0.3	0.5 ± 0.2	0.5 ± 0.1
$\Delta\delta$ (ppm)	2.6 ± 0.1	2.5 ± 0.4	2.40 ± 0.07
R_2^a (s ⁻¹)	24.2 ± 0.2	24.0 ± 0.3	23.9 ± 0.2
R_2^b (s ⁻¹)	59 ± 25	112 ± 44	102 ± 19
R_1 (s ⁻¹) ^{**}	0.402 ± 0.004	0.397 ± 0.006	0.403 ± 0.003
Reduced χ^2	0.95	1.11	0.98

^{*}CEST profiles were fitted assuming a two-state model, such that $p_a + p_b = 1$.

^{**}The longitudinal relaxation rate (R_1) is assumed to be the same for both states.

The method was applied to the reduced datasets discussed above. The resulting FT and LP processed profiles (hereafter termed FT-CEST) have the same incremental step as the conventional datasets and are shown in Figure 4.3, with the corresponding fitted parameters summarized in Table 4.3. For all four residues, the FT-CEST profiles are comparable to the conventional ones. Moreover, the fitting parameters obtained from the analysis of the FT-CEST profiles are in good agreement with the ones obtained from the conventional datasets. In particular, the obtained k_{ex} and R_2^b values, which were the most affected by the use of reduced profiles, are in good agreement with the values obtained from the analysis of the conventional datasets. This type of processing can therefore overcome the limitations noted above for the reduced CEST datasets, while retaining the two-fold reduction in measurement time.

The strategy presented here combines the simultaneous analysis of data collected with two radio-frequency field strengths with Fourier transform and linear prediction processing (FT-CEST) to improve exchange-mediated saturation transfer experiments. The advantages compared to the conventional method are two fold: (i) the use of two radio-frequency field strengths ensures the accurate estimation of all fitting parameters, especially the exchange rate and the transverse relaxation rate of the lowly populated species; (ii) the FT-CEST approach reduces the required measurement time. (Carneiro et al., 2015c)

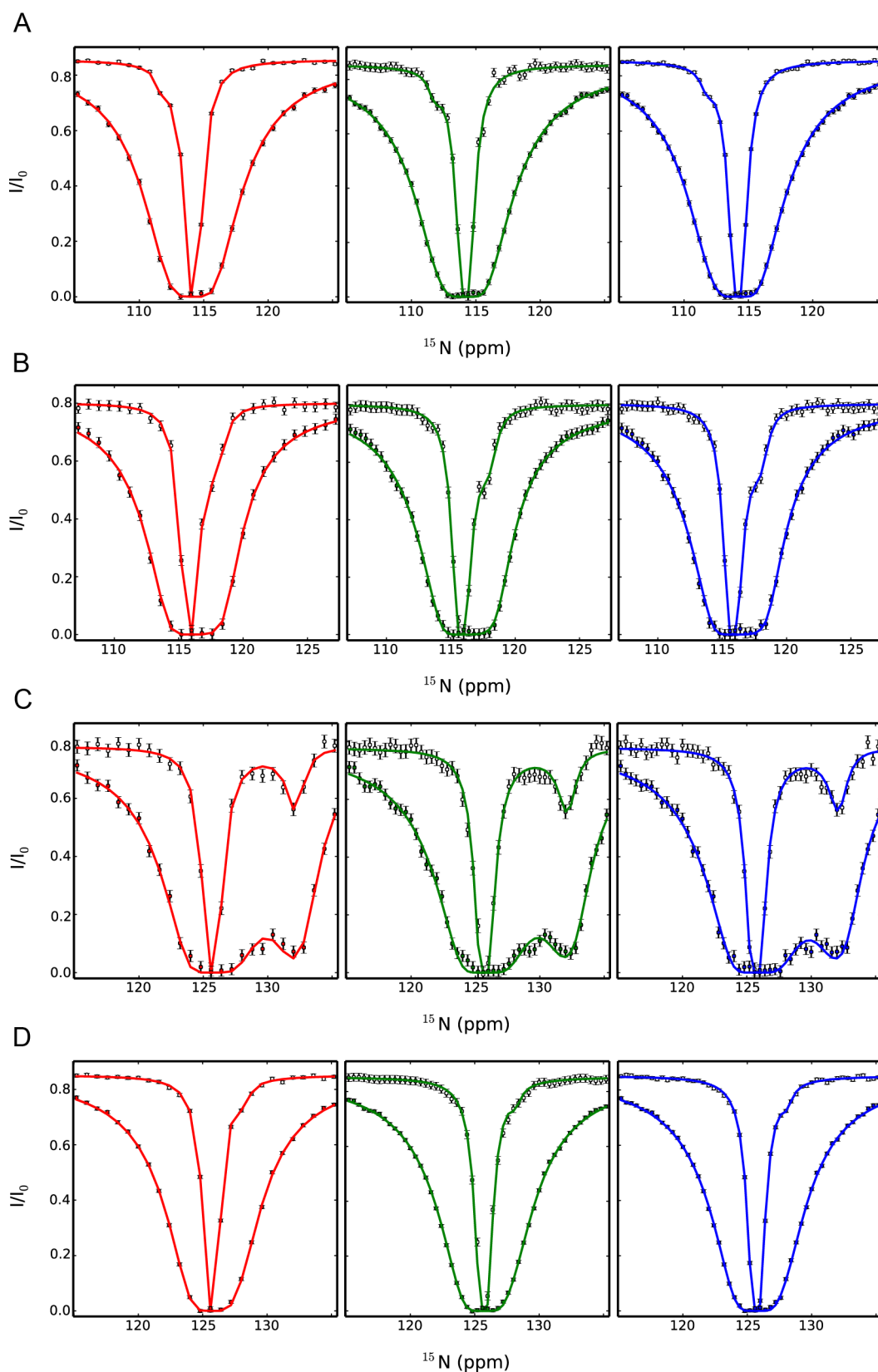


Figure 4.3 The experimental time required for exchange-mediated saturation transfer experiments can be reduced using Fourier transform and linear prediction. Reduced CEST profiles (56 Hz per frequency offset increment; red), FT-CEST profiles, obtained by FT and LP processing of the reduced datasets (28 Hz per frequency offset increment; green) and conventional CEST profiles (28 Hz per frequency offset increment; blue) for V38 (A), N75 (B), W77 (C) and N104 (D) of OAA. The data collected with two radio-frequency field strengths (15 and 75 Hz, open and filled circles, respectively). Solid lines correspond to the best simultaneous fit of the profiles on a per residue basis.

4.2.2 A STATISTICAL METHOD TO IDENTIFY DYNAMIC CLUSTERS

Initial analyses of experiments reporting on dynamic processes, such as relaxation dispersion or exchange-mediated saturation transfer, are based on fits to exchange models on a per-residue basis. Model fitting of CPMG relaxation dispersion data, assuming a 2-state model described by Eq. (1.1), enables the extraction of chemical shift differences ($\Delta\delta$), intrinsic transverse relaxation rates ($R_{2,0}$), exchange rates (k_{ex}) and relative populations (p_b), where $\Delta\delta$ and $R_{2,0}$ are residue-specific parameters and k_{ex} and p_b are potential global parameters (i.e., residues governed by the same dynamic process report the same k_{ex} and p_b but can have different $\Delta\delta$ and $R_{2,0}$). Of note, as described in section 1.2.1, the unambiguous determination of $\Delta\delta$ and p_b is only possible for systems in the slow exchange regime. For systems in fast exchange regime, a residue-specific structural amplitude parameter that includes contributions from both $\Delta\delta$ and p_b is obtained ($\Phi_{\text{ex}} = p_a p_b \Delta\delta^2$). A critical step in evaluating the site-specific data obtained from NMR experiments is then to assess which residues are reporting on the same exchange process. A common strategy relies on the user-based evaluation of the fitting parameters obtained from individual fits, with residues reporting on similar exchange rate and relative populations being clustered into different groups. Here we propose a user-independent, statistical method based on model selection for clustering residues reporting on different exchange processes. The method is described in terms of CPMG relaxation dispersion data, but the general concept is applicable to other types of experiments as well.

The key aspect of this approach is to formulate the different fits as models that are applied to the entire dataset (i.e., each model must describe all the residues displaying relaxation dispersion), such that the number of data points remains constant and only the number of fitting parameters differs between models. Each model is described by $2n$ residue-specific parameters ($\Delta\delta$ and $R_{2,0}$ for slow exchange regime, Φ_{ex} and $R_{2,0}$ for fast exchange regime) and $2m$ cluster-specific parameters (k_{ex} and p_b) for slow exchange regime or m cluster-specific parameters (k_{ex}) for fast exchange regime, where n and m correspond to the number of residues and number of clusters, respectively. Note that for the sake of generality, individual fits are considered here as single-residue clusters. The limiting cases of global (i.e. one cluster with n residues) and individual fits (i.e. n single-residue clusters) would then correspond to models with $2n + 2$ ($2n + 1$) and $4n$ ($3n$) parameters, respectively, for slow (fast) exchange regime. Within this framework we can

now conceive and compare any model with any number of single-residue and/or multiple-residues clusters, with the relative quality of each model being evaluated based on the small-sample adjusted Akaike information criterion (AICc): (Hurvich and Tsai, 1989)

$$AICc = \chi^2 + 2p + \frac{2p(p+1)}{v-p-1} \quad (4.11)$$

where v is the number of data points, p is the number of parameters of a given model, and χ^2 is the target function value from a least-square minimization defined as:

$$\chi^2 = \sum_{i=1}^N \left(\frac{exp_i - calc_i}{\sigma_i^{exp}} \right)^2 \quad (4.12)$$

where exp corresponds to experimental data, $calc$ corresponds to values calculated data using appropriated models, and σ_{exp} corresponds to the experimental uncertainty associated with exp .

It should be noted that AICc does not provide an absolute measure of how good a model is, rather it provides a relative measure of how good a model is compared with another (i.e., it identifies that model x is better than model y but that does not necessarily mean that model x is good). (Burnham and Anderson, 2002)

An overview of the clustering algorithm is shown in Figure 4.4 and the detailed flow diagrams of the model selection and iterative clustering processes are provided in the appendix (Figure 7.1 and Figure 7.2).

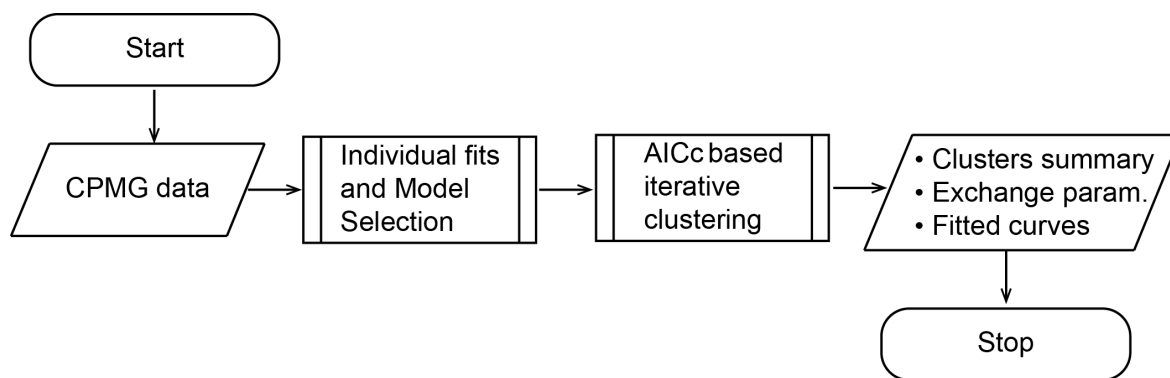


Figure 4.4 Flow diagram of the clustering procedure for CPMG relaxation dispersion data. Model selection between different clustering models is achieved based on AICc differences. Detailed flow diagrams for the individual fits and model selection, and AICc based iterative clustering procedures are given the appendix (**Figure 7.1** and **Figure 7.2**).

In the first stage of the algorithm each residue is fitted individually to both fast and slow exchange models described by the Luz-Meiboom (Luz and Meiboom, 1963), and the Bloch-McConnell (McConnell, 1958) equations, respectively. The AICc difference between the two models ($\Delta\text{AICc} = \text{AICc}_{\text{fast}} - \text{AICc}_{\text{slow}}$) is evaluated and only residues for which $\Delta\text{AICc} > 10$ (meaning that the fast exchange model is essentially not supported by the data (Burnham and Anderson, 2002)) are classified as undergoing exchange in the slow regime. The χ^2 value, the number of parameters, and the selected model are stored for further use during the clustering stage. The clustering step ensues and follows a “top-down” approach, that is, all residues initially belong to the same cluster (global fit), from which residues are iteratively excluded. The exclusion of a residue from the original cluster represents a new model composed of a globally fitted cluster with $n-1$ residues and an individual fitted residue (single-residue cluster). For global fits of more than one residue the exchange is assumed to be slow if at least one residue was previously classified as undergoing slow exchange. The first recursion of the residue-exclusion step thus requires the analysis of n models. The AICc value of each model is calculated based on the total χ^2 ($\chi^2_{\text{total}} = \chi^2_{\text{cluster}} + \chi^2_{\text{individual}}$) and number of parameters associated with it. The best model from the set of n models generated in the iterative exclusion process is compared to the previous best model (originally, the global fit model), with all model selection being based on AICc differences. If the best model (i.e., the model with the lowest AICc within a user-defined threshold) corresponds to the new model (one cluster of $n-1$ residues plus a single-residue cluster), the iterative exclusion step is repeated in the new “ $n-1$ ” cluster. The second recursion of the residue-exclusion step thus requires the analysis of $n-1$ models, each composed of a cluster of $n-2$ residues and two single-residue clusters. The residue

exclusion process is repeated until none of the newly generated models is better than the previous model. At this point, the residues that remained in the cluster are identified as the first dynamic cluster, and all residues that were excluded form a new cluster — that is, the model is composed of the first dynamic cluster, with j residues, and a second cluster of $n-j$ residues. This model is used as the starting point of a second recursion of the clustering step, during which residues will be excluded from the second cluster. The whole process is repeated until all residues are assigned into dynamic clusters, or all excluded residues are better fitted individually rather than simultaneously.

The “top-down” clustering approach was chosen over the alternative “bottom-up” approach (for which single-residue clusters are iteratively merged into larger clusters) since the former requires a smaller number of fitting calculations if multiple-residue clusters are indeed present in the dataset, and if the dataset is composed of more than 3 residues. For example, in the case of a global process involving four residues (a, b, c and d) the “top-down” approach requires 9 fitting calculations (clusters of a+b+c+d, a+b+c, a+b+d, a+c+d, b+c+d and individual fits of a, b, c and d), since the exclusion of any residue from the original cluster would not result in a better model than the original (correct) global fit. The “bottom-up” approach, on the other hand, requires 15 calculations (clusters of a+b, a+c, a+d, b+c, b+d and c+d in addition to the 9 cases mentioned above), since all pair-wise combinations of single-residues have to be initially evaluated.

The accuracy of the clustering algorithm was tested using numerous synthetic datasets, simulating ^{15}N CPMG relaxation dispersion data for residues undergoing exchange between 2 states. Each dataset consisted of 2 clusters of 3 residues each, corresponding to differences in k_{ex} , p_b or both k_{ex} and p_b (Table 7.3). In addition, a control dataset (dataset A) was generated for which only local parameters (i.e., R_2 and $\Delta\delta$) vary between the different residues. In order to assess the impact of the different exchange regimes in the clustering algorithm, three variants of each dataset were generated: (i) all residues in the slow exchange regime, (ii) all residues in the fast exchange regime, (iii) 3 residues in the slow exchange regime and 3 residues in the fast exchange regime (with at least one residue per cluster in each exchange regime). The same k_{ex} , p_b , and R_2 values were used to generate the different variants of each dataset, and the exchange regime was modulated by manipulating the residue-specific chemical shift differences ($\Delta\delta$) (Table 7.3). For each residue, effective R_2 values were calculated for 11 CPMG frequencies at two

different static magnetic field strengths (60.12 and 90.23 MHz) using the Bloch-McConnell equation and assuming a 2% random error.

Since relaxation dispersion curves depend on several parameters (namely, k_{ex} , p_{b} , and $\Delta\delta$), it is not straightforward to determine a universal parameter based only on k_{ex} and p_{b} that correlates with the accuracy of the clustering algorithm. Instead, the presence of various dynamic processes within a dataset can be potentially recognized from the reduced χ^2 value obtained from a global fit ($\chi^2_{\text{red., global}} = \chi^2 / (v - p - 1)$, where v corresponds to the number of data points and p to the number of fitting parameters): large values indicate that the fitted parameters do not adequately reproduce the experimental data, suggesting that not all sites report on the same exchange event (i.e., not all sites can be described with the same k_{ex} and p_{b}). Conversely, $\chi^2_{\text{red., global}} \approx 1$ indicates that the globally fitted k_{ex} and p_{b} describe the data properly, within the experimental error, suggesting that all sites report on the same exchange event. Since the clustering algorithm presented here is purely statistical, and AICc is based on χ^2 , it is expected that a single cluster is found in datasets for which $\chi^2_{\text{red., global}} \leq 1$, whereas different clusters should be identified in datasets for which $\chi^2_{\text{red., global}} > 1$.

Datasets comprising residues in the slow exchange regime indeed show a marked correlation between the accuracy of the clustering algorithm and the $\chi^2_{\text{red., global}}$ (Figure 4.5 A). From this correlation it becomes apparent that if the $\chi^2_{\text{red., global}}$ value is higher than 1.6 the algorithm correctly identifies the different clusters present in the dataset. For datasets yielding $\chi^2_{\text{red., global}}$ values between 1.0 and 1.6, a model better than the global fit (i.e., with lower AICc) is still found (Table 4.4), but the expected clusters are only partially identified and the results should be evaluated carefully. Datasets with $\chi^2_{\text{red., global}} \leq 1$ are always better fitted to a single global process.

Table 4.4 Clustering results for the synthetic datasets in slow exchange. The datasets are ordered by increasingly $\chi^2_{\text{red., global}}$, as presented in Figure 4.5.

Dataset	$\chi^2_{\text{red., global}}$	Nr. real clusters (residues)*	Nr. found clusters (residues)*	Δ AICc**
H	0.9	2 (123, 456)	1 (123456)	0.0
G	1.0	2 (123, 456)	1 (123456)	0.0
A	1.1	1 (123456)	1 (123456)	0.0
N	1.3	2 (123, 456)	2 (12345, 6)	24.8
F	1.3	2 (123, 456)	1 (123456)	0.0
M	1.3	2 (123, 456)	2 (12345, 6)	9.2
C	1.3	2 (123, 456)	1 (123456)	0.0
D	1.4	2 (123, 456)	2 (12345, 6)	15.1
K	1.5	2 (123, 456)	2 (1456, 23)	47.6
J	1.6	2 (123, 456)	2 (1456, 23)	37.3

I	1.7	2 (123, 456)	2 (123, 456)	54.4
Q	1.7	2 (123, 456)	2 (123, 456)	72.0
L	2.0	2 (123, 456)	2 (123, 456)	92.1
E	2.2	2 (123, 456)	2 (123, 456)	132.6
O	2.5	2 (123, 456)	2 (123, 456)	168.2
S	3.2	2 (123, 456)	2 (123, 456)	256.8
R	3.6	2 (123, 456)	2 (123, 456)	296.9
U	5.3	2 (123, 456)	2 (123, 456)	483.7
B	5.7	2 (123, 456)	2 (123, 456)	503.8
P	6.0	2 (123, 456)	2 (123, 456)	547.9
T	7.3	2 (123, 456)	2 (123, 456)	719.7
Y	7.4	2 (123, 456)	2 (123, 456)	731.0
X	11.4	2 (123, 456)	2 (123, 456)	1149.0
V	15.0	2 (123, 456)	2 (123, 456)	1577.9

* The residues belonging to each cluster are shown in parenthesis.

** $\Delta\text{AICc} = \text{AICc}_{\text{global fit}} - \text{AICc}_{\text{final model}}$. $\Delta\text{AICc} > 10$ indicate essentially no empirical support of the global model. (Burnham and Anderson, 2002)

The same trends can be seen for datasets composed of residues in the fast exchange regime (Figure 4.5 B). Notably, the $\chi^2_{\text{red., global}}$ values for most datasets is lower when all residues are in fast exchange, compared to when all residues are in slow exchange. This observation is not surprising, considering that in the fast exchange regime k_{ex} is the only global parameter, since p_b is convoluted in the residue-specific Φ_{ex} parameter. Therefore, a larger number of statistically robust solutions can be found that compensate mismatches in k_{ex} by over- or under-estimating the individual Φ_{ex} . Nonetheless, the same empirical threshold of $\chi^2_{\text{red., global}} > 1.6$ as observed for the slow-exchange datasets is observed for the fast-exchange datasets.

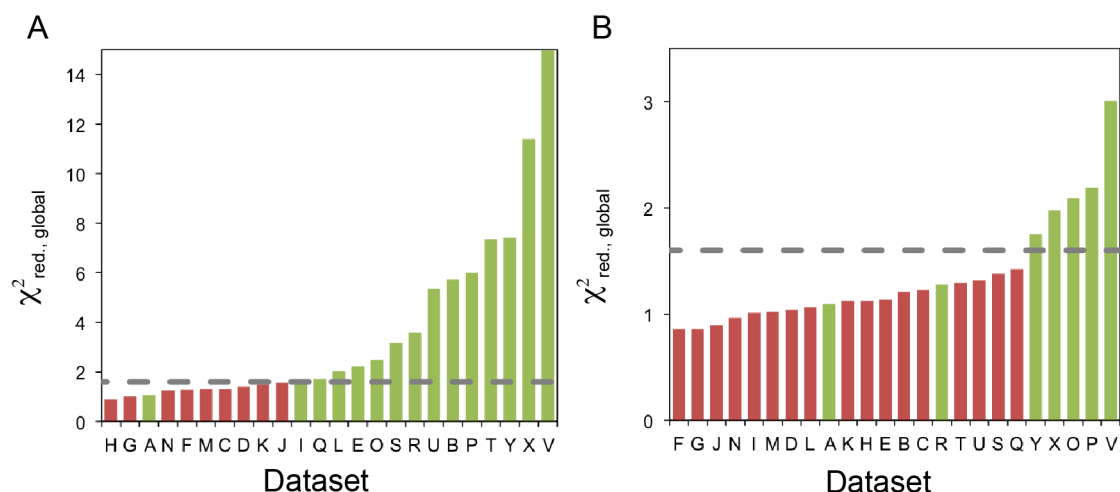


Figure 4.5 Reduced χ^2 of global fit correlates with clustering accuracy. The reduced χ^2 value obtained for the global fits ($\chi^2_{\text{red., global}}$) correlates with the accuracy of the clustering algorithm for (A) datasets in slow exchange and (B) datasets in fast exchange. Datasets for which the clusters were correctly identified by the algorithm are colored in green, and datasets for which the correct clusters were not found are shown in red. The dashed gray line represents $\chi^2_{\text{red., global}} = 1.6$. Dataset A corresponds to a control dataset for which only local parameters (i.e. $R_{2,0}$ and $\Delta\delta$) vary between the different residues.

In contrast to the datasets where all residues display the same exchange regime, a clear correlation between the accuracy of the algorithm and the $\chi^2_{\text{red., global}}$ was not observed for datasets composed of a mixture of residues in fast and slow exchange (Figure 4.6 A). In an attempt to find an empirical metric that predicts the accuracy of the algorithm, we evaluated the χ^2_{red} obtained when only residues in the same exchange regime are fitted globally. Two groups are thus independently fit, one comprising only residues in slow exchange, and one comprising only residues in fast exchange. Interestingly, the accuracy of the clustering algorithm for datasets with mixed exchange regimes seems to depend on the reduced χ^2 values obtained from both groups ($\chi^2_{\text{red., slow}}$, obtained from the global fit of residues in slow exchange, and $\chi^2_{\text{red., fast}}$, obtained from the global fit of residues in fast exchange). If both $\chi^2_{\text{red., slow}}$ and $\chi^2_{\text{red., fast}}$ are larger than the threshold value noted above (1.6), the algorithm indeed finds the correct clusters. The accuracy of the fitting algorithm for datasets with mixed exchange regimes thus correlates with the smaller of the two χ^2_{red} (Figure 4.6 B). Importantly, the independent global fit of residues in different exchange regimes only provides an empirical metric for predicting the accuracy of the algorithm; the clustering procedure starts from the global fit of all residues in the dataset as explained before.

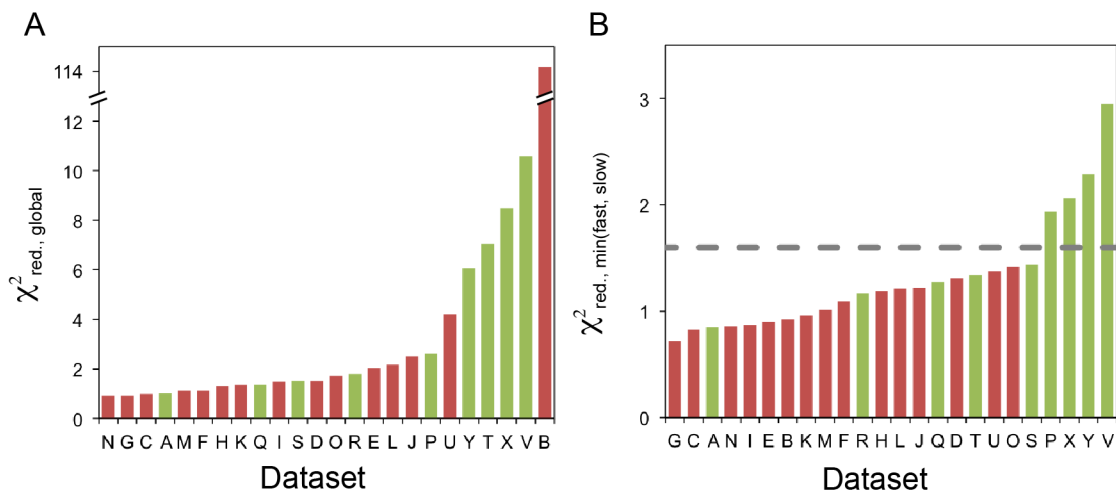


Figure 4.6 Clustering accuracy of datasets with both exchange regimes correlates with the minimum reduced χ^2 between the global fit of slow and fast residues. (A) For datasets with residues in both (fast and slow) exchange regimes, the reduced χ^2 value obtained for the global fits does not correlate with the accuracy of the clustering algorithm. (B) Instead, a correlation is found between the clustering accuracy and the minimum reduced χ^2 value obtained from independent global fits with only the residues in slow or in fast exchange ($\chi^2_{\text{red., min(fast, slow)}} = \min(\chi^2_{\text{red., fast}}, \chi^2_{\text{red., slow}})$). Datasets for which the clusters were correctly identified by the algorithm are colored in green, and datasets for which the correct clusters were not found are shown in red. The dashed gray line represents $\chi^2_{\text{red.,}} = 1.6$. Dataset A corresponds to a control dataset for which only local parameters (i.e. $R_{2,0}$ and $\Delta\delta$) vary between the different residues.

As mentioned above, evaluation of the different models throughout the clustering procedure is based on AICc differences ($\Delta\text{AICc} = \text{AICc}_{\text{model } i} - \text{AICc}_{\text{model } i+1}$), with a model with more parameters (model $i+1$, corresponding to the exclusion of a residue) only being accepted if the ΔAICc is larger than a user-defined threshold. Following the recommendations of Burnham and Anderson (Burnham and Anderson, 2002), we have used here an ΔAICc threshold of 7, indicative of considerably less empirical support of the model with less parameters (model i). Using this threshold we find no false-positives (i.e., identification of non-existing clusters) or false-negatives (i.e., non-identification of existing clusters) above the empirical χ^2_{red} based threshold discussed previously (Figure 4.5 and Figure 4.6). The use of lower ΔAICc thresholds might lead to over-interpretation of the variability in the data, and variations due to experimental error, rather than different dynamic processes, might lead to the selection of more complex models resulting in false-positives (Table 4.5).

Table 4.5 Clustering results for dataset P in fast exchange using two different ΔAICc thresholds (7 and 3).

Dataset	$\chi^2_{\text{red., global}}$	Nr. real clusters (residues)*	ΔAICc threshold = 7		ΔAICc threshold = 3	
			Nr. found clusters (residues)*	ΔAICc^{**}	Nr. found clusters (residues)*	ΔAICc^{**}
P	2.2	2 (123, 456)	2 (123, 456)	101.0	3 (123, 56, 4)	104.7

* The residues belonging to each cluster are shown in parenthesis.

** $\Delta\text{AICc} = \text{AICc}_{\text{global fit}} - \text{AICc}_{\text{final model}}$. $\Delta\text{AICc} > 10$ indicates essentially no empirical support of the global model. (Burnham and Anderson, 2002)

In summary, the results using synthetic data indicate that the algorithm works accurately for all exchange regimes, and applications of the algorithm to real systems will follow. Such a procedure can help identifying different exchange processes that might be relevant for characterizing the biological functions of biomolecules.

Additionally, the algorithm is based on the fitting software ShereKhan (Mazur et al., 2013) and benefits from all features of this package, namely the simultaneous fit of data collected at multiple static magnetic fields and the possibility of combining CPMG relaxation dispersion with HEROINE data (Ban et al., 2013a). Although the current implementation of the algorithm deals with CPMG relaxation data it can be easily applied to other types of site-specific data, such as CEST/DEST, $R_{1\rho}$ relaxation dispersion, titration data, among others.

4.3 MATERIALS AND METHODS

4.3.1 SAMPLE PREPARATION

NMR samples consisted of 2 mM ^{15}N labeled OAA in 20 mM sodium acetate (pH 5.0), 20 mM sodium chloride, 3 mM sodium azide and 90/10% $\text{H}_2\text{O}/\text{D}_2\text{O}$. OAA was expressed and purified as described in (Koharudin et al., 2011) and kindly provided by L.M.I. Koharudin and A. M. Gronenborn (University of Pittsburgh, USA).

4.3.2 NMR SPECTROSCOPY

^{15}N -CEST experiments were recorded at 277 K on a 700 MHz AVANCE III spectrometer equipped with a cryogenic triple resonance probe, as describe by Vallurupalli and coworkers. (Vallurupalli et al., 2012) ^{15}N radio-frequency field strengths of 15 and 75 Hz and a ^1H decoupling field strength of 3.5 kHz were applied during a relaxation delay (T_{ex}) of 400 ms. CEST data consisted of a series of 2D spectra, acquired in an interleaved fashion, corresponding to ^{15}N irradiation offsets incremented in steps of 0.4 ppm (28 Hz). Duplicated experiments at five different ^{15}N irradiation offsets were recorded for error estimation. Additionally, a reference experiment, for which $T_{\text{ex}} = 0$, was recorded.

4.3.3 DATA ANALYSIS

4.3.3.1 FT-CEST

All spectra were processed using NMRPipe. (Delaglio et al., 1995) Peak intensities were quantified using the model-based linear equation system implemented in CARA. (Keller, 2004) Peak intensities at each irradiation offset were normalized to the intensity of the reference experiment (I/I_0). Uncertainties associated with I/I_0 were estimated as the R.M.S.D. between I/I_0 from repeat measurements.

Reduced CEST profiles correspond to ^{15}N irradiation offsets incremented in steps of 0.8 ppm (56 Hz) and were generated from the conventional CEST profiles (irradiation offsets incremented in steps of 0.4 ppm).

FT-CEST profiles were obtained by further processing of the reduced profiles as follows. Each reduced profile was first inverted by subtracting the largest I/I_0 from all I/I_0 values and subsequently converted into a time-domain signal using inverse Fourier transform. The time domain signal was extrapolated using linear prediction (LP) to double the number of points. An LP prediction order of 2 was used and the LP filter was calculated using a singular value decomposition (SVD) method. (Hoch and Stern, 1996) The LP-extended time-domain signal was reconverted into a frequency domain spectrum using Fourier transform, which was inverted by subtracting the largest experimental I/I_0 from all I/I_0 values. The uncertainty in I/I_0 after the LP processing was estimated as:

$$\sigma_{I/I_0} = \sqrt{\sigma_{exp}^2 + \sigma_{LP}^2} \quad (4.13)$$

where σ_{exp} is the experimental uncertainty associated with I/I_0 estimated as described above and σ_{LP} is the uncertainty derived from the LP processing, estimated as the R.M.S.D. between the experimentally derived I/I_0 and the ones obtained after processing. The FT-CEST profiles were processed using an in-house written python script, using linear prediction and Fourier transform functions from the nmrglue module. (Helmus and Jaroniec, 2013)

All CEST profiles were analyzed with an in-house written python script that minimizes the target function:

$$\chi^2 = \sum_{i=1}^N \left(\frac{I_i^{exp} - I_i^{calc}}{\sigma_i^{exp}} \right)^2 \quad (4.14)$$

where I_{exp} corresponds to the experimental intensities, and I_{calc} to the intensities calculated using either the Bloch-McConnell equation (McConnell, 1958) or the analytical $R_{1\rho}$ expression. (Baldwin and Kay, 2013; Palmer III, 2014; Trott and Palmer III, 2002) Based on the Bloch-McConnell equation, intensities were calculated as: (Vallurupalli et al., 2012)

$$I^{calc} = M_T(I_z^a) / M_0(I_z^a) \quad (4.15)$$

$$M_T = \exp(A T_{ex}) M_0 \quad (4.16)$$

where I_z^a is the z-component of the angular momentum for the major state, M_0 is a column matrix with the populations of the major (p_a) and minor (p_b) states, and A is the system of the Bloch-McConnell equation for a single spin-1/2 system exchanging between two states in the homogenous form: (Helgstrand et al., 2000; McConnell, 1958)

$$\frac{d}{dt} \begin{pmatrix} E/2 \\ I_x^a \\ I_y^a \\ I_z^a \\ I_x^b \\ I_y^b \\ I_z^b \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -R_2^a - k_{ab} & -\delta_a & \omega_1 & k_{ba} & 0 & 0 \\ 0 & \delta_a & -R_2^a - k_{ab} & 0 & 0 & k_{ba} & 0 \\ 2R_1^a p_a & -\omega_1 & 0 & -R_1^a - k_{ab} & 0 & 0 & k_{ba} \\ 0 & k_{ab} & 0 & 0 & -R_2^b - k_{ba} & -\delta_b & \omega_1 \\ 0 & 0 & k_{ab} & 0 & \delta_b & -R_2^b - k_{ba} & 0 \\ 2R_1^b p_b & 0 & 0 & k_{ab} & -\omega_1 & 0 & -R_1^b - k_{ba} \end{pmatrix} \begin{pmatrix} E/2 \\ I_x^a \\ I_y^a \\ I_z^a \\ I_x^b \\ I_y^b \\ I_z^b \end{pmatrix} \quad (4.17)$$

where E is the identity operator and $I_j^{a/b}$ is the j -component of the angular momentum for the major or the minor state.

Based on the analytical $R_{1\rho}$ expression, intensities were calculated as: (Palmer III, 2014)

$$I^{calc} = \cos^2 \theta \exp(-T_{ex} R_{1\rho}) \quad (4.18)$$

where $R_{1\rho}$ is calculated using Eq. (4.1).

Regardless of the model used, the inhomogeneity in the radio-frequency field strength (ν_1) was taken into account by performing 10 calculations using evenly spaced ν_1 values between $\pm 2\sigma$ ($\sigma = 10\%$) around the mean value (15 or 75 Hz). The final calculated intensities correspond to a weighted average of the 10 calculations, assuming a Gaussian distribution. (Vallurupalli et al., 2012) All the fitted rates (k_{ab} , k_{ba} , R_2^a , R_2^b , and R_1) were constrained to be positive during the minimization, where $k_{ab} = p_b k_{ex}$ and $k_{ba} = (1-p_b) k_{ex}$, and we assume $R_1^a = R_1^b = R_1$. Initial values for the fitting parameters are defined as follows: R_2^a , R_2^b , and chemical shift difference between the major and the minor states ($\Delta\delta$) are user defined; R_1 and the chemical shift of the major state are automatically estimated from the data (the former as $R_1 = (-1/T_{ex}) \ln(I/I_0)$, where I is one of the experimental intensities measured off-resonance for both the major and the minor states and I_0 is the intensity for $T_{ex} = 0$, and the latter as the ν_1 irradiation offset for which the intensity ratio is minimal); k_{ab} and k_{ba} are estimated based on a grid search. Errors in the

fitted parameters were estimated by Monte Carlo simulations with 100 iterations and the experimental errors as the standard deviation for a given intensity ratio.

4.3.3.2 ^{15}N CPMG relaxation dispersion

The synthetic datasets were analyzed using the clustering algorithm described in section 4.2.2, using a ΔAICc threshold of 7. Each fitting step was performed using the program ShereKhan (Mazur et al., 2013), using either the Bloch-McConnell equation (McConnell, 1958) or the Luz-Meiboom equation (Luz and Meiboom, 1963).

4.3.4 SIMULATIONS

4.3.4.1 ^{15}N CEST

Synthetic data sets were created for a site with k_{ex} , p_b , δ_a , δ_b , R_1 , R_2^a , and R_2^b of 150 s^{-1} , 1%, 120 ppm, 125 ppm, 0.5 s^{-1} , 20.0 s^{-1} , and 100.0 s^{-1} respectively. CEST profiles were generated using the analytical $R_{1\rho}$ expression Eq. (4.1) for a relaxation delay of 400 ms, radio-frequency field strengths of 15 and 75 Hz, and a 70.12 MHz static magnetic field. Intensity ratios were calculated for 50 different irradiation offsets ranging from 110 to 130 ppm, assuming a Gaussian distribution centered on the theoretical value with a standard deviation of 0.006.

4.3.4.2 ^{15}N CPMG relaxation dispersion

Synthetic datasets composed of 6 residues each were generated using the Bloch-McConnell equation for a relaxation delay of 40 ms, CPMG frequencies of 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1000 Hz, and 60.12 and 90.23 MHz static magnetic fields. A summary of the k_{ex} , p_b , $\Delta\delta$ and R_2 used in the different datasets are summarized in Table 7.3. The effective R_2 rates were calculated assuming a Gaussian distribution centered on the theoretical value with a standard deviation of 2%.

5 DISCUSSION AND CONCLUSION

The main focus of this thesis was the characterization of the functional dynamics of the lectin OAA, aimed at the elucidation of the molecular recognition mechanism underlying its anti-HIV activity (Chapters 2 and 3). Despite the two-fold symmetry exhibited by OAA at the sequence and structural level, well-dispersed NMR spectra were acquired and the almost complete resonance assignments could be achieved (Chapter 2). Based on this assignment, detailed NMR studies were conducted for the characterization of the conformational properties sampled by free OAA in solution, as a means to elucidate the general mechanism underlying glycan recognition (Chapter 3).

Relaxation dispersion experiments, in particular CPMG relaxation dispersion, have lead to the identification of functionally relevant excited states in several systems. (Boehr et al., 2006; Eisenmesser et al., 2005; Korzhnev et al., 2009; Sugase et al., 2007) We thus investigated the possibility of the X-ray bound conformation being sampled by free OAA as a conformational excited state by measuring ^{15}N -CPMG relaxation dispersion. The experiments revealed that several residues undergo conformational exchange between the ground state, and lowly populated states ($0.1 \pm 0.2 \% > p_b < 18 \pm 16 \%$ at 277 K; Table 7.2). However, based on the comparison of the chemical shift differences and spatial distribution of the residues undergoing conformational exchange in free OAA and the ones significantly affected by the binding of Man5, we concluded that the excited state does not correspond to the bound conformation as seen in the X-ray structure. Moreover, the widespread of exchange rates and populations obtained from the residue-wise analysis of the ^{15}N -CPMG relaxation dispersion data (Table 7.1 and Table 7.2) suggests that not all residues report on the same exchange event. The conformational exchange detected with these experiments also does not resemble an unfolded state and may correspond to transient conformations not related to binding of Man5. (Carneiro et al., 2015c) We proceeded with the structural characterization of the ground state of OAA by analyzing inter-proton distances and backbone dihedral angles that are expected to be particularly sensitive to the conformational changes seen in the X-ray structures. Both parameters indicated that both the sugar-free and the sugar-bound conformations observed in the X-ray structures are sampled by OAA in the absence of sugar, and that the sugar-bound conformation is thermodynamically favorable (i.e. it is highly populated). These results

were corroborated by the solution structure of OAA, in which both free and bound X-ray conformations are seen in the final ensemble. (Carneiro et al., 2015c) Besides, the sugar-bound conformation is frequently seen in the sugar-free X-ray structures of several OAA homologues. (Carneiro et al., 2015c; Koharudin et al., 2012; Whitley et al., 2013) Taken together, the results presented in Chapter 3 indicate that the recognition of Man5 by OAA could occur by conformational selection within a ground-state ensemble. Further, these results suggest that a large population of the protein pre-exists in a “boundable” conformation, implying that the changes in the conformational entropy of the protein backbone upon binding are minimal. This however does not necessarily translate into minimal changes in the overall binding entropy, since backbone conformational changes could also affect the conformational entropy of the side-chains, (Smith et al., 2015) which has been shown to correlate with overall binding entropy, (Frederick et al., 2007) an aspect not explored in this study. Nonetheless, the insights provided here into the molecular recognition mechanism essential for the anti-HIV activity of OAA may help guide the development of this lectin as a potential microbicide.

The second part of this thesis (Chapter 4) describes two newly developed tools that extend the efficacy and accuracy of two NMR methods frequently used for the characterization of protein dynamics — exchange-mediated saturation transfer and CPMG relaxation dispersion.

Exchange-mediated saturation transfer type of experiments are particularly useful for the characterization of processes with slow kinetics and low populations, and between low molecular weight and high molecular weight species. (Fawzi et al., 2011; Vallurupalli et al., 2012) These experiments require the scanning of large spectral windows, with a series of 2D spectra being recorded as a weak radio frequency field is applied at varying offsets. Additionally, we show in Chapter 4 that measurements with two different field strengths are necessary for obtaining reliable exchange parameters. The resulting extensive measurement time may limit use of this type of experiments on metastable systems. In order to alleviate this drawback, we have developed a strategy based on Fourier transform to decrease the experimental time required by CEST/DEST experiments. The strategy is based in the observation that CEST profiles are essentially frequency spectra, analogous to a typical NMR spectrum. Thus, the Fourier transform of a CEST profile yields a time-domain signal that can be processed with techniques commonly used for improving the quality of NMR spectra. As a proof of principle we use linear prediction and obtained a

two-fold reduction in measurement time. The combination of the use of two radio-frequency field strengths and the FT-CEST processing results in more reliable estimation of exchange parameters and shorter acquisition times, aspects particularly important for the study of metastable systems.

Both CEST/DEST and CPMG relaxation dispersion provide site-specific information, and a critical step in evaluating the resulting data is to assess whether different (groups of) residues report on a single (or various) exchange process(es). For example, the analysis of the ^{15}N CPMG relaxation dispersion data obtained for OAA (Table 7.1 and Table 7.2) indicate that not all residues report on the same exchange process. Nonetheless, multiple residues might be governed by a single dynamic process, and should be fitted together. For this purpose, we have developed an iterative statistical method based on model selection using AICc. The algorithm was tested on a variety of synthetic data, simulating CPMG relaxation data in both slow and fast exchange regimes. In agreement with its statistical nature, the algorithm successfully identifies dynamic sub-clusters within a dataset, as long as the dataset cannot be properly described by a global set of parameters (i.e., the quality of the global fit of all residues is poor), and applications to real systems will follow. It should be noted that, although the algorithm was developed for CPMG relaxation dispersion data, the same principles could be applied to the analysis of other experiments.

Both methodologies presented in Chapter 4 can aid in the identification and characterization of dynamic processes and lowly populated states sampled by biomolecules in solution, which ultimately will deepen our understanding of biologically relevant processes.

6 REFERENCES

- Akkouh, O., Ng, T., Singh, S., Yin, C., Dan, X., Chan, Y., Pan, W., and Cheung, R. (2015). Lectins with Anti-HIV Activity: A Review. *Molecules* *20*, 648.
- Anthis, N.J., Doucleff, M., and Clore, G.M. (2011). Transient, Sparsely Populated Compact States of Apo and Calcium-Loaded Calmodulin Probed by Paramagnetic Relaxation Enhancement: Interplay of Conformational Selection and Induced Fit. *J Am Chem Soc* *133*, 18966-18974.
- Baldwin, A., and Kay, L. (2013). An R1 ρ expression for a spin in chemical exchange between two sites with unequal transverse relaxation rates. *J Biomol NMR* *55*, 211-218.
- Balzarini, J. (2007). Targeting the glycans of glycoproteins: a novel paradigm for antiviral therapy. *Nat Rev Microbiol* *5*, 583-597.
- Ban, D., Funk, M., Gulich, R., Egger, D., Sabo, T.M., Walter, K.F.A., Fenwick, R.B., Giller, K., Pichierri, F., de Groot, B.L., *et al.* (2011). Kinetics of Conformational Sampling in Ubiquitin. *Angw Chem, Int Ed Engl* *50*, 11437-11440.
- Ban, D., Gossert, A.D., Giller, K., Becker, S., Griesinger, C., and Lee, D. (2012). Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *J Magn Reson* *221*, 1-4.
- Ban, D., Mazur, A., G. Carneiro, M., Sabo, T.M., Giller, K., Koharudin, L.I., Becker, S., Gronenborn, A., Griesinger, C., and Lee, D. (2013a). Enhanced accuracy of kinetic information from CT-CPMG experiments by transverse rotating-frame spectroscopy. *J Biomol NMR* *57*, 73-82.
- Ban, D., Sabo, T., Griesinger, C., and Lee, D. (2013b). Measuring Dynamic and Kinetic Information in the Previously Inaccessible Supra- μ s Window of Nanoseconds to Microseconds by Solution NMR Spectroscopy. *Molecules* *18*, 11904-11937.
- Bax, A., Clore, G.M., and Gronenborn, A.M. (1990). ^1H - ^1H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^1H and ^{13}C spectra of ^{13}C -enriched proteins. *J Magn Reson* *88*, 425-431.
- Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. *Methods Enzymol* *339*, 127-174.
- Bhattacharya, A., Tejero, R., and Montelione, G.T. (2007). Evaluating protein structures determined by structural genomics consortia. *Proteins: Struct, Funct, and Bioinf* *66*, 778-795.
- Bodenhausen, G., and Ruben, D.J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* *69*, 185-189.
- Boehr, D.D., McElheny, D., Dyson, H.J., and Wright, P.E. (2006). The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* *313*, 1638-1642.
- Boehr, D.D., Nussinov, R., and Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* *5*, 789-796.
- Bonomelli, C., Doores, K.J., Dunlop, D.C., Thaney, V., Dwek, R.A., Burton, D.R., Crispin, M., and Scanlan, C.N. (2011). The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS One* *6*, 1-7.
- Bosshard, H.R. (2001). Molecular Recognition by Induced Fit: How Fit is the Concept? *Physiology* *16*, 171-173.
- Bouvignies, G., Vallurupalli, P., Hansen, D.F., Correia, B.E., Lange, O., Bah, A., Vernon, R.M., Dahlquist, F.W., Baker, D., and Kay, L.E. (2011). Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* *477*, 111-114.

- Bucher, D., Grant, B.J., and McCammon, J.A. (2011). Induced Fit or Conformational Selection? The Role of the Semi-closed State in the Maltose Binding Protein. *Biochemistry* *50*, 10530-10539.
- Burnham, K., and Anderson, D. (2002). Model selection and multimodel inference: a practical information-theoretic approach (Springer).
- Carneiro, M., Koharudin, L.I., Griesinger, C., Gronenborn, A., and Lee, D. (2015a). ¹H, ¹³C and ¹⁵N resonance assignment of the anti-HIV lectin from *Oscillatoria agardhii*. *Biomol NMR Assign* *9*, 317-319.
- Carneiro, M., Reddy, J., Griesinger, C., and Lee, D. (2015b). Speeding-up exchange-mediated saturation transfer experiments by Fourier transform. *J Biomol NMR*, in press.
- Carneiro, M.G., Koharudin, L.M.I., Ban, D., Sabo, T.M., Trigo-Mourino, P., Mazur, A., Griesinger, C., Gronenborn, A.M., and Lee, D. (2015c). Sampling of Glycan-Bound Conformers by the Anti-HIV Lectin *Oscillatoria agardhii* agglutinin in the Absence of Sugar. *Angw Chem, Int Ed Engl* *54*, 6462-6465.
- Carr, H.Y., and Purcell, E.M. (1954). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys Rev* *94*, 630-638.
- Case, D. (1995). Calibration of ring-current effects in proteins and nucleic acids. *J Biomol NMR* *6*, 341-346.
- Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M., and Skelton, N.J. (2007). Protein NMR spectroscopy - Principles and practice, 2nd edn (Academic Press).
- Chan, D.C., and Kim, P.S. (1998). HIV entry and its inhibition. *Cell* *93*, 681-684.
- Cooper, A. (1976). Thermodynamic fluctuations in protein molecules. *Proceedings of the National Academy of Sciences of the United States of America* *73*, 2740-2741.
- Csermely, P., Palotai, R., and Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* *35*, 539-546.
- Davis, D.G., Perlman, M.E., and London, R.E. (1994). Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1-rho and T2 (CPMG) methods. *Journal of Magnetic Resonance, Series B* *104*, 266-275.
- 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.
- Deverell, C., Morgan, R.E., and Strange, J.H. (1970). Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame. *Molecular Physics* *18*, 553-559.
- Edwards, L.J., Savostyanov, D.V., Welderufael, Z.T., Lee, D., and Kuprov, I. (2014). Quantum mechanical NMR simulation algorithm for protein-size spin systems. *J Magn Reson* *243*, 107-113.
- Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E., and Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* *438*, 117-121.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D* *66*, 486-501.
- Farrow, N., Zhang, O., Forman-Kay, J., and Kay, L. (1994). A heteronuclear correlation experiment for simultaneous determination of ¹⁵N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR* *4*, 727-734.
- Fawzi, N.L., Ying, J., Ghirlando, R., Torchia, D.A., and Clore, G.M. (2011). Atomic-resolution dynamics on the surface of amyloid-beta protofibrils probed by solution NMR. *Nature* *480*, 268-272.

- Fawzi, N.L., Ying, J., Torchia, D.A., and Clore, G.M. (2012). Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nat Protocols* 7, 1523-1533.
- Féris, G., Huskens, D., Noppen, S., Koharudin, L.M.I., Gronenborn, A.M., and Schols, D. (2014). Broad anti-HIV activity of the *Oscillatoria agardhii* agglutinin homologue lectin family. *J Antimicrob Chemother*.
- Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* 27, 2985-2993.
- Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991). The energy landscapes and motions of proteins. *Science* 254, 1598-1603.
- Frederick, K.K., Marlow, M.S., Valentine, K.G., and Wand, A.J. (2007). Conformational entropy in molecular recognition by proteins. *Nature* 448, 325-329.
- Freed, E.O., and Martin, M.A. (1995). The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. *The Journal of Biological Chemistry* 270, 23883-23886.
- Gerstein, M., Lesk, A.M., and Chothia, C. (1994). Structural Mechanisms for Domain Movements in Proteins. *Biochemistry* 33, 6739-6749.
- Gianni, S., Dogan, J., and Jemth, P. (2014). Distinguishing induced fit from conformational selection. *Biophysical Chemistry* 189, 33-39.
- Greives, N., and Zhou, H.-X. (2014). Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. *Proceedings of the National Academy of Sciences* 111, 10197-10202.
- Grzesiek, S., and Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* 96, 432-440.
- Güntert, P. (1998). Structure calculation of biological macromolecules from NMR data. *Quarterly Reviews of Biophysics* 31, 145-237.
- Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program Dyana. *J Mol Biol* 273, 283-298.
- Habeck, M., Rieping, W., and Nilges, M. (2005). Bayesian estimation of Karplus parameters and torsion angles from three-bond scalar couplings constants. *J Magn Reson* 177, 160-165.
- Hammes, G.G., Chang, Y.-C., and Oas, T.G. (2009). Conformational selection or induced fit: A flux description of reaction mechanism. *Proceedings of the National Academy of Sciences* 106, 13737-13741.
- Helgstrand, M., Härd, T., and Allard, P. (2000). Simulations of NMR pulse sequences during equilibrium and non-equilibrium chemical exchange. *J Biomol NMR* 18, 49-63.
- Helmus, J., and Jaroniec, C. (2013). NmrGlue: an open source Python package for the analysis of multidimensional NMR data. *J Biomol NMR* 55, 355-367.
- Henry, G., and Sykes, B. (1995). Determination of the rotational dynamics and pH dependence of the hydrogen exchange rates of the arginine guanidino group using NMR spectroscopy. *J Biomol NMR* 6, 59-66.
- Herrmann, T., Güntert, P., and Wüthrich, K. (2002a). Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J Mol Biol* 319, 209-227.
- Herrmann, T., Güntert, P., and Wüthrich, K. (2002b). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24, 171-189.

- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* 4, 435-447.
- Hoch, J.C., and Stern, A.S. (1996). *NMR Data Processing* (Wiley-Liss).
- Horiya, S., MacPherson, I.S., and Krauss, I.J. (2014). Recent strategies targeting HIV glycans in vaccine design. *Nat Chem Biol* 10, 990-999.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct, Funct, and Bioinf* 65, 712-725.
- Hurvich, C.M., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297-307.
- Huskens, D., and Schols, D. (2012). Algal Lectins as Potential HIV Microbicide Candidates. *Mar Drugs* 10, 1476-1497.
- Jardetzky, O. (1980). On the nature of molecular conformations inferred from high-resolution NMR. *Biochimica et Biophysica Acta (BBA) - Protein Structure* 621, 227-232.
- Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1979). Investigation of exchange processes by two - dimensional NMR spectroscopy. *The Journal of Chemical Physics* 71, 4546-4553.
- Jensen, M.R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R.W.H., and Blackledge, M. (2008). Quantitative Conformational Analysis of Partially Folded Proteins from Residual Dipolar Couplings: Application to the Molecular Recognition Element of Sendai Virus Nucleoprotein. *J Am Chem Soc* 130, 8055-8061.
- Johnson, C.E., and Bovey, F.A. (1958). Calculation of Nuclear Magnetic Resonance Spectra of Aromatic Hydrocarbons. *The Journal of Chemical Physics* 29, 1012-1014.
- Kahn, K., and Plaxco, K. (2010). Principles of Biomolecular Recognition. In *Recognition Receptors in Biosensors*, M. Zourob, ed. (Springer New York), pp. 3-45.
- Karplus, M. (1963). Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J Am Chem Soc* 85, 2870-2871.
- Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28, 8972-8979.
- Kay, L.E., Xu, G.Y., and Yamazaki, T. (1994). Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H₂O Saturation. *J MagnReson A* 109, 129-133.
- Keller, R.L.J. (2004). *The computer aided resonance assignment tutorial* (CANTINA Verlag).
- Kessler, H., Griesinger, C., Lautz, J., Mueller, A., Van Gunsteren, W.F., and Berendsen, H.J.C. (1988). Conformational dynamics detected by nuclear magnetic resonance NOE values and J coupling constants. *J Am Chem Soc* 110, 3393-3396.
- Koharudin, L.M.I., Furrey, W., and Gronenborn, A.M. (2011). Novel fold and carbohydrate specificity of the potent anti-HIV cyanobacterial lectin from *Oscillatoria agardhii*. *J Biol Chem* 286, 1588-1597.
- Koharudin, L.M.I., and Gronenborn, A.M. (2011). Structural basis of the anti-HIV activity of the cyanobacterial *Oscillatoria agardhii* agglutinin. *Structure* 19, 1170-1181.
- Koharudin, L.M.I., Kollipara, S., Aiken, C., and Gronenborn, A.M. (2012). Structural Insights into the Anti-HIV Activity of the *Oscillatoria agardhii* Agglutinin Homolog Lectin Family. *J Biol Chem* 287, 33796-33811.
- Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics* 14, 51-55.

- Korzhev, D.M., Bezsonova, I., Lee, S., Chalikian, T.V., and Kay, L.E. (2009). Alternate Binding Modes for a Ubiquitin–SH3 Domain Interaction Studied by NMR Spectroscopy. *J Mol Biol* 386, 391-405.
- Korzhev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M., and Kay, L.E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430, 586-590.
- Koshland, D.E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences* 44, 98-104.
- Kovrigin, E.L., and Loria, J.P. (2006). Enzyme Dynamics along the Reaction Coordinate: Critical Role of a Conserved Residue. *Biochemistry* 45, 2636-2647.
- Lange, O.F., Lakomek, N.-A., Farès, C., Schröder, G.F., Walter, K.F.A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* 320, 1471-1475.
- Luginbühl, P., and Wüthrich, K. (2002). Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules. *Prog Nucl Magn Reson Spectrosc* 40, 199-247.
- Luz, Z., and Meiboom, S. (1963). Nuclear magnetic resonance study of the protolysis of trimethylammonium ion in aqueous solution - Order of the reaction with respect to solvent. *J Chem Phys* 39, 366-370.
- Ma, B., Kumar, S., Tsai, C.-J., and Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Engineering* 12, 713-720.
- Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M., and Clore, G.M. (1989a). Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.β. *Biochemistry* 28, 6150-6156.
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A., and Bax, A. (1989b). Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J Am Chem Soc* 111, 1515-1517.
- Mayer, M., and Meyer, B. (1999). Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angw Chem, Int Ed Engl* 38, 1784-1788.
- Mazur, A., Hammesfahr, B., Griesinger, C., Lee, D., and Kollmar, M. (2013). ShereKhan—calculating exchange parameters in relaxation dispersion data from CPMG experiments. *Bioinformatics* 29, 1819-1820.
- McConnell, H.M. (1958). Reaction rates by nuclear magnetic resonance. *J Chem Phys* 28, 430-431.
- Meiboom, S., and Gill, D. (1958). Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev Sci Instrum* 29, 688-691.
- Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12, 88-118.
- Mori, S., Abeygunawardana, C., Johnson, M.O., and Vanzijl, P.C.M. (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. *Journal of Magnetic Resonance, Series B* 108, 94-98.
- Muhandiram, D.R., Farrow, N.A., Xu, G.Y., Smallcombe, S.H., and Kay, L.E. (1993). A Gradient 13C NOESY-HSQC Experiment for Recording NOESY Spectra of 13C-Labeled Proteins Dissolved in H2O. *Journal of Magnetic Resonance, Series B* 102, 317-321.
- Mulder, F.A.A., Mittermaier, A., Hon, B., Dahlquist, F.W., and Kay, L.E. (2001). Studying excited states of proteins by NMR spectroscopy. *Nat Struct Mol Biol* 8, 932-935.

- Neudecker, P., Lundström, P., and Kay, L.E. (2009). Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding. *Biophysical Journal* *96*, 2045-2054.
- Neuhaus, D., and Williamson, M.P. (2000). *The Nuclear Overhauser Effect in structural and conformational analysis*, 2nd edn (Wiley-VCH).
- Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M.R., and Blackledge, M. (2009). Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *J Am Chem Soc* *131*, 17908-17918.
- Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J Magn Reson* *131*, 373-378.
- Palmer III, A.G. (2004). NMR characterization of the dynamics of biomacromolecules. *Chem Rev* *104*, 3623-3640.
- Palmer III, A.G. (2014). Chemical exchange in biomacromolecules: Past, present, and future. *J Magn Reson* *241*, 3-17.
- Palmer III, A.G., Kroenke, C.D., and Loria, J.P. (2001). Nuclear Magnetic Resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* *339*, 204-239.
- Pritchard, L.K., Spencer, D.I.R., Royle, L., Bonomelli, C., Seabright, G.E., Behrens, A.-J., Kulp, D.W., Menis, S., Krumm, S.A., Dunlop, D.C., *et al.* (2015). Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat Commun* *6*.
- Rückert, M., and Otting, G. (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J Am Chem Soc* *122*, 7793-7797.
- Sato, Y., Okuyama, S., and Hori, K. (2007). Primary Structure and Carbohydrate Binding Specificity of a Potent Anti-HIV Lectin Isolated from the Filamentous Cyanobacterium *Oscillatoria agardhii*. *J Biol Chem* *282*, 11021-11029.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Magn Reson Spectrosc* *34*, 93-158.
- Schanda, P., and Brutscher, B. (2005). Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J Am Chem Soc* *127*, 8014-8015.
- Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006). Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J Am Chem Soc* *128*, 9042-9043.
- Schleucher, J., Sattler, M., and Griesinger, C. (1993). Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNC0 Experiment. *Angw Chem, Int Ed Engl* *32*, 1489-1491.
- Schrodinger, LLC (2010). *The PyMOL Molecular Graphics System*, Version 1.3r1.
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Marius Clore, G. (2003). The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* *160*, 65-73.
- Silva, D.-A., Bowman, G.R., Sosa-Peinado, A., and Huang, X. (2011). A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PLoS Comput Biol* *7*, e1002054.
- Smith, C.A., Ban, D., Pratihari, S., Giller, K., Schwiegk, C., de Groot, B.L., Becker, S., Griesinger, C., and Lee, D. (2015). Population Shuffling of Protein Conformations. *Angw Chem, Int Ed Engl* *54*, 207-210.

- Sugase, K., Dyson, H.J., and Wright, P.E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* *447*, 1021-1025.
- Tolman, J.R., Flanagan, J.M., Kennedy, M.A., and Prestegard, J.H. (1997). NMR evidence for slow collective motions in cyanometmyoglobin. *Nat Struct Mol Biol* *4*, 292-297.
- Trott, O., and Palmer III, A.G. (2002). R1 ρ Relaxation outside of the Fast-Exchange Limit. *J Magn Reson* *154*, 157-160.
- Tsai, C.J., Kumar, S., Ma, B., and Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Science : A Publication of the Protein Society* *8*, 1181-1190.
- Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., *et al.* (2008). BioMagResBank. *Nucleic Acids Research* *36*, D402-D408.
- Vallurupalli, P., Bouvignies, G., and Kay, L.E. (2012). Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation. *J Am Chem Soc* *134*, 8148-8161.
- Vogt, A.D., and Di Cera, E. (2012). Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry* *51*, 5894-5902.
- Vuister, G.W., and Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HN,Ha) coupling constants in ¹⁵N-enriched proteins. *J Am Chem Soc* *115*, 7772-7777.
- Ward, A.B., and Wilson, I.A. (2015). Insights into the trimeric HIV-1 envelope glycoprotein structure. *Trends Biochem Sci* *40*, 101-107.
- Ward, K.M., Aletras, A.H., and Balaban, R.S. (2000). A New Class of Contrast Agents for MRI Based on Proton Chemical Exchange Dependent Saturation Transfer (CEST). *J Magn Reson* *143*, 79-87.
- Weickl, T.R., and Boehr, D.D. (2012). Conformational selection and induced changes along the catalytic cycle of Escherichia coli dihydrofolate reductase. *Proteins: Struct, Funct, and Bioinf* *80*, 2369-2383.
- Weickl, T.R., and von Deuster, C. (2009). Selected-fit versus induced-fit protein binding: Kinetic differences and mutational analysis. *Proteins: Struct, Funct, and Bioinf* *75*, 104-110.
- Whitley, M.J., Furey, W., Kollipara, S., and Gronenborn, A.M. (2013). Burkholderia oklahomensis agglutinin is a canonical two-domain OAA-family lectin: structures, carbohydrate binding and anti-HIV activity. *FEBS J* *280*, 2056-2067.
- Wishart, D., Bigam, C., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J., and Sykes, B. (1995). ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J Biomol NMR* *6*, 135-140.
- Wlodarski, T., and Zagrovic, B. (2009). Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proceedings of the National Academy of Sciences* *106*, 19346-19351.
- Wüthrich, K. (1986). *NMR of proteins and nucleic acids* (Wiley).
- Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T.-M., Rech, I., Cova, S., Xun, L., and Xie, X.S. (2003). Protein Conformational Dynamics Probed by Single-Molecule Electron Transfer. *Science* *302*, 262-266.
- Zaiss, M., and Bachert, P. (2013). Exchange-dependent relaxation in the rotating frame for slow and intermediate exchange – modeling off-resonant spin-lock and chemical exchange saturation transfer. *NMR in Biomedicine* *26*, 507-518.
- Zhang, Q., Stelzer, A.C., Fisher, C.K., and Al-Hashimi, H.M. (2007). Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* *450*, 1263-1267.

Zhao, B., Hansen, A.L., and Zhang, Q. (2014). Characterizing Slow Chemical Exchange in Nucleic Acids by Carbon CEST and Low Spin-Lock Field R1 ρ NMR Spectroscopy. *J Am Chem Soc* 136, 20-23.

Zuiderweg, E.R.P., and Fesik, S.W. (1989). Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* 28, 2387-2391.

7 APPENDIX

7.1 EXCHANGE PARAMETERS FROM ^{15}N CPMG RELAXATION DISPERSION

Table 7.1 Exchange parameters from individual fits of ^{15}N CPMG relaxation dispersion curves at 298 K. Model selection between slow and fast exchange models was accomplished as described in Chapter 3.

Residue	k_{ex} (s^{-1})	p_b (%)	$\Delta\delta$ (ppm)	ϕ_{ex} (ppm^2) ^{**}
Y4	2078 ± 165	--	--	0.025 ± 0.002
Q9	2822 ± 153	0.6 ± 0.2	2.5 ± 0.4	--
W10	3346 ± 157	--	--	0.120 ± 0.007
G11	3636 ± 130	--	--	0.221 ± 0.01
N18	2334 ± 47	0.7 ± 0.1	2.1 ± 0.1	--
G26	2644 ± 165	--	--	0.127 ± 0.008
S27	2724 ± 112	--	--	0.173 ± 0.007
R28	2442 ± 157	--	--	0.115 ± 0.007
S29	3094 ± 87	--	--	0.084 ± 0.003
Q31	3003 ± 44	--	--	0.179 ± 0.003
N32	2519 ± 101	--	--	0.072 ± 0.003
V34	3082 ± 90	--	--	0.166 ± 0.006
M51	2591 ± 321	--	--	0.046 ± 0.006
G55	2493 ± 122	--	--	0.095 ± 0.005
E56	3195 ± 199	--	--	0.095 ± 0.007
G60	3081 ± 264	--	--	0.140 ± 0.014
N68	2619 ± 140	--	--	0.035 ± 0.002
N75	2801 ± 79	--	--	0.12 ± 0.004
Q76	2523 ± 136	--	--	0.060 ± 0.003
W77	4310 ± 1121	--	--	0.9 ± 0.3
G78	3054 ± 84	--	--	0.173 ± 0.005
D80	3020 ± 47	--	--	0.259 ± 0.004
W84	2446 ± 150	--	--	0.064 ± 0.004
H85	3494 ± 487	--	--	0.05 ± 0.01
S94	56 ± 38	4.4 ± 3.1	4.7 ± 0.3	--
E96	3319 ± 112	--	--	0.134 ± 0.006
N97	1770 ± 189	0.4 ± 0.1	4.1 ± 0.2	--
Q98	3048 ± 265	--	--	0.063 ± 0.007
N99	2812 ± 249	1.5 ± 0.3	5.5 ± 0.2	--
V101	2950 ± 165	--	--	0.069 ± 0.005
G122	792 ± 288	0.6 ± 0.2	4.1 ± 0.2	--
E123	3454 ± 68	--	--	0.167 ± 0.004

** $\phi_{\text{ex}} = p_a p_b \Delta\omega^2$

Table 7.2 Exchange parameters from individual fits of ^{15}N CPMG relaxation dispersion curves at 277 K. Model selection between slow and fast exchange models was accomplished as described in Chapter 3.

Residue	k_{ex} (s^{-1})	p_{b} (%)	$\Delta\delta$ (ppm)	ϕ_{ex} (ppm^2)
Y4	673 ± 68	--	--	0.03 ± 0.001
N5	560 ± 87	--	--	0.023 ± 0.001
N8	1898 ± 205	--	--	0.034 ± 0.003
Q9	816 ± 95	--	--	0.039 ± 0.002
W10	2498 ± 148	--	--	0.206 ± 0.012
G11	1075 ± 71	1.1 ± 0.1	4.02 ± 0.09	--
S13	1094 ± 173	0.5 ± 0.1	3.3 ± 0.2	--
W17	468 ± 101	--	--	0.013 ± 0.001
N18	963 ± 29	1.6 ± 0.3	1.5 ± 0.1	--
E19	1821 ± 174	--	--	0.024 ± 0.002
G26	40 ± 35	18.5 ± 16.1	3.7 ± 0.5	--
S27	72 ± 25	12.2 ± 4.3	3.8 ± 0.2	--
R28	349 ± 259	2.7 ± 2.1	2.8 ± 0.3	--
S29	1764 ± 118	--	--	0.038 ± 0.002
D30	1185 ± 245	0.1 ± 0.2	2.3 ± 0.5	--
Q31	109 ± 56	6.3 ± 3.2	3.7 ± 0.2	--
N32	314 ± 41	2.7 ± 0.4	2.13 ± 0.05	--
V34	1868 ± 641	--	--	0.125 ± 0.034
S40	5925 ± 9280	0.2 ± 0.5	10 ± 11	--
G44	1795 ± 315	--	--	0.041 ± 0.006
N48	488 ± 109	--	--	0.01 ± 0.001
T50	1032 ± 133	--	--	0.019 ± 0.001
M51	166 ± 265	4.4 ± 7.1	2.3 ± 0.4	--
G55	91 ± 102	6.9 ± 7.7	3.7 ± 0.4	--
G60	1357 ± 172	--	--	0.055 ± 0.005
L66	2534 ± 946	--	--	0.034 ± 0.014
N68	1883 ± 98	--	--	0.071 ± 0.003
Y71	1498 ± 357	--	--	0.014 ± 0.003
V73	1520 ± 200	--	--	0.017 ± 0.002
E74	1287 ± 244	--	--	0.014 ± 0.002
N75	378 ± 126	2.1 ± 0.7	2.4 ± 0.1	--
Q76	1169 ± 110	--	--	0.045 ± 0.003
W77	198 ± 500	6.2 ± 15.8	6 ± 1	--
G78	1622 ± 109	0.6 ± 0.3	3.2 ± 0.6	--
W84	471 ± 55	2.9 ± 0.6	1.7 ± 0.1	--
H85	1731 ± 312	--	--	0.032 ± 0.005
S86	2533 ± 127	--	--	0.048 ± 0.003
G87	2774 ± 336	--	--	0.032 ± 0.005
G88	339 ± 4	3.2 ± 0.6	1.38 ± 0.09	--
W90	16 ± 13	16.8 ± 14.2	3.4 ± 0.4	--
L92	1368 ± 115	--	--	0.028 ± 0.002
S94	1214 ± 61	--	--	0.038 ± 0.001
R95	2271 ± 560	--	--	0.029 ± 0.007
E96	1235 ± 207	0.8 ± 0.2	4.0 ± 0.2	--
N97	1847 ± 145	--	--	0.054 ± 0.004
Q98	1223 ± 163	1 ± 0.3	2.7 ± 0.2	--
N99	1196 ± 339	1 ± 0.3	6.8 ± 0.4	--
V100	1445 ± 195	--	--	0.036 ± 0.004
E106	1532 ± 303	--	--	0.07 ± 0.014
S107	4444 ± 591	--	--	0.07 ± 0.014
Q112	3166 ± 588	--	--	0.087 ± 0.02

T113	3667 ± 1209	--	--	0.076 ± 0.034
N115	1295 ± 232	--	--	0.015 ± 0.002
T117	1174 ± 447	02 ± 0.3	1.4 ± 0.5	--
G122	88 ± 108	8.2 ± 10	3.7 ± 0.6	--
G127	2297 ± 421	--	--	0.065 ± 0.012
G130	414 ± 98	--	--	0.015 ± 0.001
T131	891 ± 101	--	--	0.019 ± 0.001
L132	237 ± 143	--	--	0.034 ± 0.014
T133	1022 ± 76	--	--	0.018 ± 0.001

$$*\phi_{\text{ex}} = p_a p_b \Delta \omega^2$$

7.2 FLOW DIAGRAMS OF THE CLUSTERING ALGORITHM

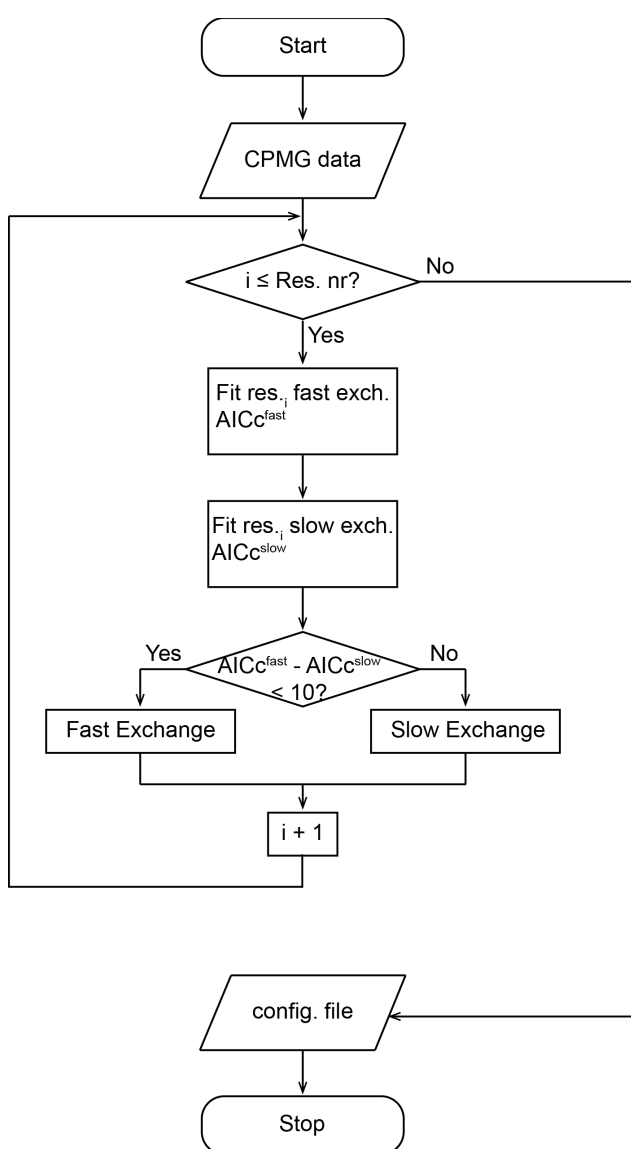


Figure 7.1 Flow diagram of the individual fits and model selection procedure of the clustering algorithm described in Chapter 4.

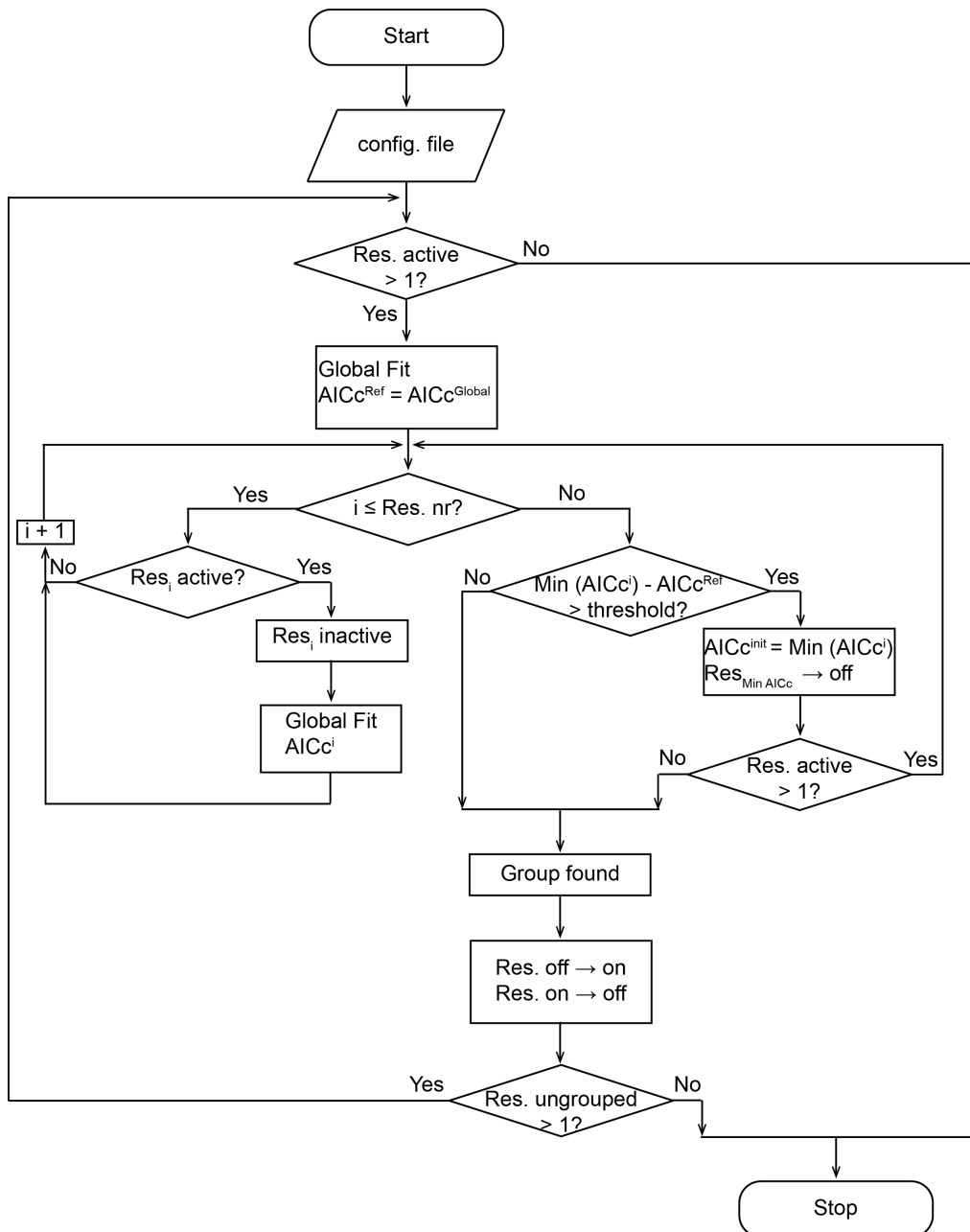


Figure 7.2 Flow diagram of the AICc based iterative clustering procedure described in Chapter 4.

7.3 SYNTHETIC DATASETS USED FOR TESTING THE CLUSTERING ALGORITHM

Table 7.3 Summary of the synthetic datasets used for testing the clustering algorithm described in Chapter 4. Each dataset was generated for different exchange regimes (slow, fast or both) by manipulating the chemical shift difference ($\Delta\delta$) of each residue.

A						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.1	1.5	4.5	21.0	1020	2
2	4.0	1.0	3.0	20.0	1020	2
3	3.7	1.4	2.8	19.5	1020	2
4	3.2	1.2	1.2	20.5	1020	2
5	3.0	0.8	0.8	19.0	1020	2
6	2.7	0.5	0.5	21.5	1020	2
B						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	0.8	4.5	20.0	1020	2
2	3.2	0.8	4.5	20.0	1020	2
3	3.2	0.8	0.8	20.0	1020	2
4	3.2	0.8	4.5	20.0	990	2
5	3.2	0.8	0.8	20.0	990	2
6	3.2	0.8	0.8	20.0	990	2
C						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	0.8	4.5	20.0	1000	4.5
2	3.2	0.8	4.5	20.0	1000	4.5
3	3.2	0.8	0.8	20.0	1000	4.5
4	3.2	0.8	4.5	20.0	1000	5
5	3.2	0.8	0.8	20.0	1000	5
6	3.2	0.8	0.8	20.0	1000	5
D						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	0.8	4.5	20.0	1000	8
2	3.2	0.8	4.5	20.0	1000	8
3	3.2	0.8	0.8	20.0	1000	8
4	3.2	0.8	4.5	20.0	1000	7
5	3.2	0.8	0.8	20.0	1000	7
6	3.2	0.8	0.8	20.0	1000	7
E						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	0.8	4.5	20.0	1000	8
2	3.2	0.8	4.5	20.0	1000	8
3	3.2	0.8	0.8	20.0	1000	8
4	3.2	0.8	4.5	20.0	1000	10
5	3.2	0.8	0.8	20.0	1000	10
6	3.2	0.8	0.8	20.0	1000	10
F						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	4.5	0.8	4.5	20.0	1500	19
2	4.5	0.8	4.5	20.0	1500	19
3	4.5	0.8	0.8	20.0	1500	19
4	4.5	0.8	4.5	20.0	1500	20
5	4.5	0.8	0.8	20.0	1500	20
6	4.5	0.8	0.8	20.0	1500	20
G						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)

1	3.2	0.8	4.5	20.0	920	2
2	3.2	0.8	4.5	20.0	920	2
3	3.2	0.8	0.8	20.0	920	2
4	3.2	0.8	4.5	20.0	1020	2
5	3.2	0.8	0.8	20.0	1020	2
6	3.2	0.8	0.8	20.0	1020	2

H

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	0.8	4.5	20.0	820	2
2	3.2	0.8	4.5	20.0	820	2
3	3.2	0.8	0.8	20.0	820	2
4	3.2	0.8	4.5	20.0	1020	2
5	3.2	0.8	0.8	20.0	1020	2
6	3.2	0.8	0.8	20.0	1020	2

I

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	0.8	4.5	20.0	620	1.9
2	3.2	0.8	4.5	20.0	620	1.9
3	3.2	0.8	0.8	20.0	620	1.9
4	3.2	0.8	4.5	20.0	1020	2
5	3.2	0.8	0.8	20.0	1020	2
6	3.2	0.8	0.8	20.0	1020	2

J

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	0.8	4.5	20.0	520	1.9
2	3.2	0.8	4.5	20.0	520	1.9
3	3.2	0.8	0.8	20.0	520	1.9
4	3.2	0.8	4.5	20.0	1020	2
5	3.2	0.8	0.8	20.0	1020	2
6	3.2	0.8	0.8	20.0	1020	2

K

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	1.0	4.0	20.0	820	10
2	3.2	1.0	4.0	20.0	820	10
3	3.2	1.0	1.0	20.0	820	10
4	3.2	1.0	4.0	20.0	1020	10
5	3.2	1.0	1.0	20.0	1020	10
6	3.2	1.0	1.0	20.0	1020	10

L

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	1.0	4.0	20.0	1000	20
2	3.2	1.0	4.0	20.0	1000	20
3	3.2	1.0	1.0	20.0	1000	20
4	3.2	1.0	4.0	20.0	900	18
5	3.2	1.0	1.0	20.0	900	18
6	3.2	1.0	1.0	20.0	900	18

M

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	3.2	1.0	4.0	20.0	1000	2
2	3.2	1.0	4.0	20.0	1000	2
3	3.2	1.0	1.0	20.0	1000	2
4	3.2	1.0	4.0	20.0	880	1.6
5	3.2	1.0	1.0	20.0	880	1.6
6	3.2	1.0	1.0	20.0	880	1.6

N

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.1	1.5	4.5	21.0	1000	2
2	2.7	1.0	3.0	20.0	1000	2

3	3.7	1.4	1.2	19.5	1000	2
4	3.2	1.2	2.8	20.5	900	1.4
5	3.0	0.8	0.8	19.0	900	1.4
6	4.0	0.5	0.5	21.5	900	1.4

O

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.1	1.5	4.5	21.0	1000	16
2	2.7	1.0	3.0	20.0	1000	16
3	3.7	1.4	1.2	19.5	1000	16
4	3.2	1.2	2.8	20.5	800	25
5	3.0	0.8	0.8	19.0	800	25
6	4.0	0.5	0.5	21.5	800	25

P

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.0	1.0	4.0	20.0	1000	20
2	5.0	1.0	4.0	20.0	1000	20
3	5.0	1.0	1.0	20.0	1000	20
4	5.0	1.0	4.0	20.0	1500	20
5	5.0	1.0	1.0	20.0	1500	20
6	5.0	1.0	1.0	20.0	1500	20

Q

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.0	1.0	4.0	20.0	1800	20
2	5.0	1.0	4.0	20.0	1800	20
3	5.0	1.0	1.0	20.0	1800	20
4	5.0	1.0	4.0	20.0	1500	20
5	5.0	1.0	1.0	20.0	1500	20
6	5.0	1.0	1.0	20.0	1500	20

R

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.0	1.0	4.0	20.0	1000	10
2	5.0	1.0	4.0	20.0	1000	10
3	5.0	1.0	1.0	20.0	1000	10
4	5.0	1.0	4.0	20.0	1500	10
5	5.0	1.0	1.0	20.0	1500	10
6	5.0	1.0	1.0	20.0	1500	10

S

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.0	1.0	4.0	20.0	1000	10
2	5.0	1.0	4.0	19.5	1000	10
3	5.0	1.0	1.0	21.0	1000	10
4	5.0	1.0	4.0	20.5	1500	10
5	5.0	1.0	1.0	19.0	1500	10
6	5.0	1.0	1.0	21.5	1500	10

T

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	5.0	1.0	6.5	20.0	1500	5
2	5.0	1.0	6.5	20.0	1500	5
3	5.0	1.0	1.0	20.0	1500	5
4	5.0	1.0	6.5	20.0	675	5
5	5.0	1.0	1.0	20.0	675	5
6	5.0	1.0	1.0	20.0	675	5

U

Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	6.0	1.0	6.5	20.0	2000	5
2	6.0	1.0	6.5	20.0	2000	5
3	6.0	1.0	1.0	20.0	2000	5
4	6.0	1.0	6.5	20.0	1000	5

5	6.0	1.0	1.0	20.0	1000	5
6	6.0	1.0	1.0	20.0	1000	5
V						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	6.0	1.0	6.5	20.0	2000	20
2	6.0	1.0	6.5	20.0	2000	20
3	6.0	1.0	1.0	20.0	2000	20
4	6.0	1.0	6.5	20.0	1000	20
5	6.0	1.0	1.0	20.0	1000	20
6	6.0	1.0	1.0	20.0	1000	20
X						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	6.0	1.0	6.5	20.0	2000	20
2	6.0	1.0	6.5	20.0	2000	20
3	6.0	1.0	1.0	20.0	2000	20
4	6.0	1.0	6.5	20.0	1100	20
5	6.0	1.0	1.0	20.0	1100	20
6	6.0	1.0	1.0	20.0	1100	20
Y						
Residue	$\Delta\delta$ (ppm) slow exch.	$\Delta\delta$ (ppm) fast exch.	$\Delta\delta$ (ppm) slow/fast exch.	R_2 (s ⁻¹)*	k_{ex} (s ⁻¹)	p_b (%)
1	6.0	1.0	6.5	20.0	2000	20
2	6.0	1.0	6.5	20.0	2000	20
3	6.0	1.0	1.0	20.0	2000	20
4	6.0	1.0	6.5	20.0	1200	20
5	6.0	1.0	1.0	20.0	1200	20
6	6.0	1.0	1.0	20.0	1200	20

* R_2 values used for generating data for a static field strength of 60.12 MHz. For data generated for a static field strength of 90.23 MHz, 2 s⁻¹ were added to the R_2 values reported on the table.

7.4 PULSE PROGRAMS

The pulse programs used to record ¹⁵N CPMG relaxation dispersion (Chapter 3) and exchange-mediated saturation transfer (Chapter 4) experiments are provided below in Bruker programming language.

7.4.1 ¹⁵N CPMG RELAXATION DISPERSION EXPERIMENT

Implemented by David Ban and Donghan Lee

```
;optimization of water flip back:
;- optimize watergate (o1,sp2,ph26)
;- optimize water flip back (sp1,ph16,ph18,ph17,ph19)

;K. Pervushin et al, PNAS, 94, 12366 (1997)

;p11 : power for 1H
;p12 : power for 13C
;p13 : power for 15N
```

```

;sp1 : water flipback power
;sp2 : water flipback power in watergate
;spnam1: gauss128_5
;spnam2: gauss128_5

;p1 : 90 degree hard pulse 1H
;p3 : 90 degree hard pulse 13C
;p4 : 180 degree hard pulse 13C (225d for 5/600)
;p5 : 90 degree hard pulse 15N
;p11 : water flipback pulse (1.5m)

;p20 : 1m (Cleaning Gradient )
;p21 : 500u (Gradient in first INEPT)
;p22 : 500u (Gradient in second INEPT)
;p23 : 900u (Gradient in watergate)
;p24 : 500u (Gradient in Z-filter)
;gpz0 : 80%
;gpz1 : 19%
;gpz2 : 15%
;gpz3 : 32%
;gpz4 : 60%

;d1 : relaxation delay
;d2 : INEPT delay (~2.7ms)
;d8 : length of ct-CPMG block
;in0 : 1/(2 SW) (Hz)

;l3 : vCPMG loop counter
;l11 : l3_max + 1

define delay INEPT1
define delay INEPT2
define delay INEPT3
define delay U1

#include <Avance_dl.incl>

#define GRADIENT0 10u p20:gp0 200u
#define GRADIENT1 10u p21:gp1 200u
#define GRADIENT2 10u p22:gp2 200u
#define GRADIENT3 10u p23:gp3 200u
#define GRADIENT4 10u p24:gp4 200u
#define GRADIENT5 10u p25:gp5 200u
#define GRADIENT6 10u p26:gp6 200u

"p2=2*p1"
"p6=2*p5"
"in0=infl/2"
"d0=in0/2-(p3*2 + 1.5u)"
"INEPT1=d2-(p21+p11+210u)-14u"
"INEPT2=d2-(p22+p11+210u)-14u"
"INEPT3=d2-(p23+p11+210u)-14u"

"U1=d2-p11-10u-p26-210u"

"l2 = 1"
"l4=l11-l3"

"d13=d1-d8"
"d14=d8/(16*l3)-p5"
"d16=d8/(16*l4)-p5"

1 10u ze
2 1m
; d1

```

```

;   d13
    20u p11:f1
    20u p12:f2
    20u p13:f3
    d13*0.5
    10u

4   d16
    (p6 ph21):f3
    d16
    d16
    (p6 ph21):f3
    d16
    d16
    (p6 ph20):f3
    d16
    d16
    (p6 ph22):f3
    d16
    d16
    (p6 ph20):f3
    d16
    d16
    (p6 ph20):f3
    d16
    d16
    (p6 ph21):f3
    d16
    d16
    (p6 ph23):f3
    d16
lo  to 4 times l4
    d13*0.5
    20u p11:f1
    20u p12:f2
    20u p13:f3

    20u LOCKH_ON
;-----first INEPT
5   (p5 ph20):f3
    GRADIENT0
    1m
    (p1 ph20):f1
    10u
    (p11:sp1 ph18:r):f1
    GRADIENT1
    INEPT1 p11:f1
    (center(p2 ph21):f1 (p6 ph20):f3)
    GRADIENT1
    INEPT1
    (p11:sp1 ph19:r):f1
    10u p11:f1
    (p1 ph23):f1
    GRADIENT4

    (p5 ph4):f3
;-----CPMG
11  d14
    (p6 ph21):f3
    d14
    d14
    (p6 ph21):f3
    d14
    d14
    (p6 ph20):f3
    d14
    d14
    (p6 ph22):f3
    d14

```

```

lo to 11 times 13
;-----U element
GRADIANT6
U1
(p11:sp3 ph15:r):f1
5u
5u p11:f1
(center(p2 ph20):f1 (p6 ph20):f3)
10u
(p11:sp3 ph15:r):f1
GRADIANT6
U1 p11:f1
;-----CPMG
21 d14
(p6 ph20):f3
d14
d14
(p6 ph20):f3
d14
d14
(p6 ph21):f3
d14
d14
(p6 ph23):f3
d14
lo to 21 times 13
(p5 ph5):f3
GRADIANT5
;-----15N evolution
if "12 %2 == 1" goto 31
(p5 ph2):f3
goto 32
31 (p5 ph1):f3
32 d0
(p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2
d0
;-----second INEPT
(p1 ph10):f1
10u
(p11:sp1 ph17:r):f1
GRADIANT2
INEPT2 p11:f1
(center(p2 ph20):f1 (p6 ph20):f3)
GRADIANT2
INEPT2
(p11:sp1 ph16:r):f1
10u p11:f1
(center(p1 ph20):f1 (p5 ph12):f3)
;-----WATERGATE
GRADIANT3
INEPT3
(p11:sp2 ph26:r):f1
10u p11:f1
(center(p2 ph20):f1 (p6 ph20):f3)
10u
(p11:sp2 ph26:r):f1
GRADIANT3
INEPT3 LOCKH_OFF
(p5 ph11):f3
;-----acquisition
go=2 ph31
lm mc #0 to 2 F1EA(ip10*2 & ip12*2 & ip17*2 & iu2,id0)
10u do:f1
10u do:f2
10u do:f3
10u LOCKH_OFF
exit

ph1 =1 3 2 0
ph2 =1 3 0 2

```

```
ph4 =0 0 0 0 2 2 2 2
ph5 =1
ph31=1 3 2 0 3 1 0 2
```

```
ph10=3 3 3 3
ph11=0 0 0 0
ph12=3 3 3 3
```

```
ph15=2
ph16=0
ph17=1
ph18=2
ph19=3
```

```
ph20=0
ph21=1
ph22=2
ph23=3
```

```
ph26=2
```

7.4.2 ^{15}N CPMG RELAXATION DISPERSION REFERENCE EXPERIMENT

Implemented by David Ban and Donghan Lee

```
;optimization of water flip back:
;- optimize watergate (o1,sp2,ph26)
;- optimize water flip back (sp1,ph16,ph18,ph17,ph19)

;K. Pervushin et al, PNAS, 94, 12366 (1997)

;p11 : power for 1H
;p12 : power for 13C
;p13 : power for 15N

;sp1 : water flipback power
;sp2 : water flipback power in watergate
;spnam1: gauss128_5
;spnam2: gauss128_5

;p1 : 90 degree hard pulse 1H
;p3 : 90 degree hard pulse 13C
;p4 : 180 degree hard pulse 13C (225d for 5/600)
;p5 : 90 degree hard pulse 15N
;p11 : water flipback pulse (1.5m)

;p20 : 1m (Cleaning Gradient )
;p21 : 500u (Gradient in first INEPT)
;p22 : 500u (Gradient in second INEPT)
;p23 : 900u (Gradient in watergate)
;p24 : 500u (Gradient in Z-filter)
;gpz0 : 80%
;gpz1 : 19%
;gpz2 : 15%
;gpz3 : 32%
;gpz4 : 60%

;d1 : relaxation delay
;d2 : INEPT delay (~2.7ms)
;d8 : length of ct-CPMG block
;in0 : 1/(2 SW) (Hz)

;l11 : l3_max + 1
```



```

define delay INEPT1
define delay INEPT2
define delay INEPT3
define delay U1

#include <Avance_dl.incl>

#define GRADIENT0    10u p20:gp0 200u
#define GRADIENT1    10u p21:gp1 200u
#define GRADIENT2    10u p22:gp2 200u
#define GRADIENT3    10u p23:gp3 200u
#define GRADIENT4    10u p24:gp4 200u
#define GRADIENT5    10u p25:gp5 200u
#define GRADIENT6    10u p26:gp6 200u

"p2=2*p1"
"p6=2*p5"
"in0=in0/2"
"d0=in0/2-(p3*2 + 1.5u)"
"INEPT1=d2-(p21+p11+210u)-14u"
"INEPT2=d2-(p22+p11+210u)-14u"
"INEPT3=d2-(p23+p11+210u)-14u"

"U1=d2-p11-10u-p26-210u"

"l2 = 1"
"l4=l11-l3"

"d13=d1-d8"
"d16=d8/(16*l4)-p5"

1 10u ze
2 1m
20u p11:f1
    20u p12:f2
    20u p13:f3
    d13*0.5
    10u

4 d16
    (p6 ph21):f3
    d16
    d16
    (p6 ph21):f3
    d16
    d16
    (p6 ph20):f3
    d16
    d16
    (p6 ph22):f3
    d16
    d16
    (p6 ph20):f3
    d16
    d16
    (p6 ph20):f3
    d16
    d16
    (p6 ph21):f3
    d16
    d16
    (p6 ph23):f3
    d16

10 to 4 times l4
    d13*0.5

```

```

20u p11:f1
20u p12:f2
20u p13:f3

20u LOCKH_ON
;-----first INEPT
5 (p5 ph20):f3
GRADIEN0
1m
(p1 ph20):f1
10u
(p11:sp1 ph18:r):f1
GRADIEN1
INEPT1 p11:f1
(center(p2 ph21):f1 (p6 ph20):f3)
GRADIEN1
INEPT1
(p11:sp1 ph19:r):f1
10u p11:f1
(p1 ph23):f1
GRADIEN4
(p5 ph4):f3
;-----U element
GRADIEN6
U1
(p11:sp3 ph15:r):f1
5u
5u p11:f1
(center(p2 ph20):f1 (p6 ph20):f3)
10u
(p11:sp3 ph15:r):f1
GRADIEN6
U1 p11:f1

(p5 ph5):f3
GRADIEN5
;-----15N evolution
if "12 %2 == 1" goto 31
(p5 ph2):f3
goto 32
31 (p5 ph1):f3
32 d0
(p3 ph23 1.5u p4 ph20 1.5u p3 ph23):f2
d0
;-----second INEPT
(p1 ph10):f1
10u
(p11:sp1 ph17:r):f1
GRADIEN2
INEPT2 p11:f1
(center(p2 ph20):f1 (p6 ph20):f3)
GRADIEN2
INEPT2
(p11:sp1 ph16:r):f1
10u p11:f1
(center(p1 ph20):f1 (p5 ph12):f3)
;-----WATERGATE
GRADIEN3
INEPT3
(p11:sp2 ph26:r):f1
10u p11:f1
(center(p2 ph20):f1 (p6 ph20):f3)
10u
(p11:sp2 ph26:r):f1
GRADIEN3
INEPT3 LOCKH_OFF
(p5 ph11):f3
;-----acquisition
go=2 ph31
1m mc #0 to 2 F1EA(ip10*2 & ip12*2 & ip17*2 & iu2,id0)

```

```

10u do:f1
10u do:f2
10u do:f3
10u LOCKH_OFF
exit

ph1 =1 3 2 0
ph2 =1 3 0 2
ph4 =0 0 0 0 2 2 2 2
ph5 =1
ph31=1 3 2 0 3 1 0 2

ph10=3 3 3 3
ph11=0 0 0 0
ph12=3 3 3 3

ph15=2
ph16=0
ph17=1
ph18=2
ph19=3

ph20=0
ph21=1
ph22=2
ph23=3

ph26=2

```

7.4.3 EXCHANGE-MEDIATED SATURATION TRANSFER EXPERIMENT

Implemented by Jithender G. Reddy and Donghan Lee

```

;15N-1H HSQC correlations without water saturation
;The delay for 3-9-19 watergate (d5) should be matched
;with 1/d;d=distance of next null point (in Hz).
;irradiation frequencies to be given as frequency list

;S. Mori et al, JMR B108, 94-98 (1995)

;p11 : power for 1H
;p12 : power for 13C
;p13 : power for 15N
;p113 : power for 15N waltz16 decoupling

;p1 : 90 degree hard pulse 1H
;p3 : 90 degree hard pulse 13C
;p4 : 180 degree hard 13C pulse (225d for 5/600)
;p5 : 90 degree hard pulse 15N
;pcpd3 : 90 deg cpd-pulse15N(waltz16,160u)

;d1 : relaxation delay
;d2 : INEPT delay (~2.7m)
;d5 : delay for 3-9-19=1/(Hz between nulls)

;in0 : 1/(2 SW) (Hz)

;p8 : CEST B1 length
;p20 : 1m (Gradient in first INEPT)
;p21 : 500u (Gradient in first INEPT)
;p22 : 500u (Gradient for z-filter)
;p23 : 500u (Gradient in first INEPT)
;p24 : 500u (Gradient for z-filter)
;p25 : 500u (Gradient in first INEPT)
;p26 : 500u (Gradient for z-filter)

```

```

;p27 : 500u (Gradient for z-filter)
;p28 : 500u (Gradient for z-filter)
;p29 : 1m (Gradient for second INEPT)
;gpz1 : 19%
;gpz2 : 30%
;gpz9 : 65%

#include <Avance_dl.incl>

define delay INEPT_W
define delay INEPT_1
define delay INEPT_2
define delay INEPT_3

#define GRADIENT0 10u p20:gp0 200u
#define GRADIENT1 10u p21:gp1 200u
#define GRADIENT2 10u p22:gp2 200u
#define GRADIENT3 10u p23:gp3 200u
#define GRADIENT4 10u p24:gp4 200u
#define GRADIENT5 10u p25:gp5 200u
#define GRADIENT6 10u p26:gp6 200u
#define GRADIENT7 10u p27:gp7 200u
#define GRADIENT8 10u p28:gp8 200u
#define GRADIENT9 10u p29:gp9 200u

define list<frequency>ONRES=<$FQ1LIST>

"p2=2*p1"
"p6=2*p5"

"in0=inf1/2"

"d0=in0/2-p5*2/3.14159-p1"
"d3=d5/2-p5"

"INEPT_1=d2-p21-210u"
"INEPT_2=d2-p23-210u"
"INEPT_3=d2-p27-210u"
"INEPT_W=d2-(p29+210u+p1*2.3846+d5*2.5)"

1 10u ze
2 1m do:f3
d1
10u p11:f1
10u p12:f2
20u p13:f3
20u LOCKH_ON

(p5 ph20):f3
GRADIENT0
10m
;-----first INEPT
(p1 ph20):f1
GRADIENT1
INEPT_1
(center(p2 ph21):f1 (p6 ph20):f3)
GRADIENT1
INEPT_1
(p1 ph21):f1
GRADIENT2
;-----
(p5 ph2):f3
GRADIENT3
INEPT_2
(center(p2 ph21):f1 (p6 ph20):f3)
GRADIENT3
INEPT_2

```

```

(p5 ph23):f3
;-----15N evolution and EST block
2u fq=cnst14 (bf ppm):f1
2u pl4:f3
2u fq=ONRES:f3
2u pl11:f1
GRADIEN4

2u cpds1:f1
(p8 ph20):f3
2u do:f1

GRADIEN5
2u pl3:f3
2u pl1:f1
2u fq=cnst15 (bf ppm):f1
2u fq=cnst17 (bf ppm):f3
(p5 ph1):f3
(d0 p2 ph23 d0):f1

(p5 ph20):f3
GRADIEN6
(p5 ph21):f3
GRADIEN7
INEPT_3
(center(p2 ph21):f1 (p6 ph20):f3)
GRADIEN7
INEPT_3
(p5 ph20):f3
GRADIEN8
;-----second INEPT
(p1 ph22):f1
GRADIEN9
INEPT_W
(p1*0.2308 ph21 d5 p1*0.6923 ph21 d5 p1*1.4615 ph21):f1
(d3 p6 ph20 d3):f3
(p1*1.4615 ph23 d5 p1*0.6923 ph23 d5 p1*0.2308 ph23):f1
GRADIEN9
INEPT_W pl13:f3 LOCKH_OFF
;-----acquisition
go=2 ph31 cpd3:f3
lm do:f3 mc #0 to 2
  F1I(ONRES.inc, 130)
  F1PH(ip1,id0)

10u do:f1
10u do:f2
10u do:f3
10u LOCKH_OFF
exit

ph2 =0 0 2 2
ph1 =0 2
ph31=2 0 0 2

ph20=0
ph21=1
ph22=2
ph23=3

```

7.4.4 EXCHANGE-MEDIATED SATURATION TRANSFER REFERENCE EXPERIMENT

Implemented by Jithender G. Reddy and Donghan Lee

```
;15N-1H HSQC correlations without water saturation
```

;The delay for 3-9-19 watergate (d5) should be matched
;with 1/d;d=distance of next null point (in Hz).

;S. Mori et al, JMR B108, 94-98 (1995)

```
;p11 : power for 1H
;p12 : power for 13C
;p13 : power for 15N
;p113 : power for 15N waltz16 decoupling

;p1 : 90 degree hard pulse 1H
;p3 : 90 degree hard pulse 13C
;p4 : 180 degree hard 13C pulse (225d for 5/600)
;p5 : 90 degree hard pulse 15N
;pcpd3 : 90 deg cpd-pulse15N(waltz16,160u)

;d1 : relaxation delay
;d2 : INEPT delay (~2.7m)
;d5 : delay for 3-9-19=1/(Hz between nulls)

;in0 : 1/(2 SW) (Hz)

;p8 : CEST B1 length
;p20 : 1m (Gradient in first INEPT)
;p21 : 500u (Gradient in first INEPT)
;p22 : 500u (Gradient for z-filter)
;p23 : 500u (Gradient in first INEPT)
;p24 : 500u (Gradient for z-filter)
;p25 : 500u (Gradient in first INEPT)
;p26 : 500u (Gradient for z-filter)
;p27 : 500u (Gradient for z-filter)
;p28 : 500u (Gradient for z-filter)
;p29 : 1m (Gradient for second INEPT)
;gpz1 : 19%
;gpz2 : 30%
;gpz9 : 65%
```

```
#include <Avance_dl.incl>
```

```
define delay INEPT_W
define delay INEPT_1
define delay INEPT_2
define delay INEPT_3
```

```
#define GRADIENT0 10u p20:gp0 200u
#define GRADIENT1 10u p21:gp1 200u
#define GRADIENT2 10u p22:gp2 200u
#define GRADIENT3 10u p23:gp3 200u
#define GRADIENT4 10u p24:gp4 200u
#define GRADIENT5 10u p25:gp5 200u
#define GRADIENT6 10u p26:gp6 200u
#define GRADIENT7 10u p27:gp7 200u
#define GRADIENT8 10u p28:gp8 200u
#define GRADIENT9 10u p29:gp9 200u
```

```
define list<frequency>ONRES=< $FQ1LIST >
```

```
"p2=2*p1"
"p6=2*p5"
```

```
"in0=inf1/2"
```

```
"d0=in0/2-p5*2/3.14159-p1"
"d3=d5/2-p5"
```

```
"INEPT_1=d2-p21-210u"
```

```
"INEPT_2=d2-p23-210u"
```

```
"INEPT_3=d2-p27-210u"
```

```
"INEPT_W=d2-(p29+210u+p1*2.3846+d5*2.5)"
```

```

1 10u ze
2 1m do:f3
  d1
  10u p11:f1
  10u p12:f2
  20u p13:f3
  20u LOCKH_ON

  (p5 ph20):f3
  GRADIENT0
  10m

;-----first INEPT
  (p1 ph20):f1
  GRADIENT1
  INEPT_1
  (center(p2 ph21):f1 (p6 ph20):f3)
  GRADIENT1
  INEPT_1
  (p1 ph21):f1
  GRADIENT2

;-----
  (p5 ph2):f3
  GRADIENT3
  INEPT_2
  (center(p2 ph21):f1 (p6 ph20):f3)
  GRADIENT3
  INEPT_2
  (p5 ph23):f3

;-----15N evolution
  2u fq=cnst14 (bf ppm):f1
  2u p14:f3
  2u fq=ONRES:f3
  2u p11:f1
  GRADIENT4

  GRADIENT5
  2u p13:f3
  2u p11:f1
  2u fq=cnst15 (bf ppm):f1
  2u fq=cnst17 (bf ppm):f3
  (p5 ph1):f3
  (d0 p2 ph23 d0):f1

  (p5 ph20):f3
  GRADIENT6
  (p5 ph21):f3
  GRADIENT7
  INEPT_3
  (center(p2 ph21):f1 (p6 ph20):f3)
  GRADIENT7
  INEPT_3
  (p5 ph20):f3
  GRADIENT8

;-----second INEPT
  (p1 ph22):f1
  GRADIENT9
  INEPT_W
  (p1*0.2308 ph21 d5 p1*0.6923 ph21 d5 p1*1.4615 ph21):f1
  (d3 p6 ph20 d3):f3
  (p1*1.4615 ph23 d5 p1*0.6923 ph23 d5 p1*0.2308 ph23):f1
  GRADIENT9
  INEPT_W p113:f3 LOCKH_OFF

;-----acquisition
  go=2 ph31 cpd3:f3
  1m do:f3 mc #0 to 2
    F1PH(ip1,id0)

```

```
;F1QF(iu3)
10u do:f1
10u do:f2
10u do:f3
10u LOCKH_OFF
exit
```

```
ph2 =0 0 2 2
ph1 =0 2
ph31=2 0 0 2
```

```
ph20=0
ph21=1
ph22=2
ph23=3
```


CURRICULUM VITAE

Name Marta Gião Carneiro
Date of birth 04 August 1987
Place of birth Setubal, Portugal

EDUCATION

2012 - Present **Ph.D. Student**, Molecular Biology, Max-Planck Institute for Biophysical Chemistry, Goettingen (Germany)
Advisors: Prof. Dr. Christian Griesinger, Dr. Donghan Lee

2010 - 2012 **M.Sc.**, Molecular Biology, Max-Planck Institute for Biophysical Chemistry, Goettingen (Germany)
Advisors: Prof. Dr. Christian Griesinger, Dr. Donghan Lee
Global Classification: Excellent (A)

2009 - 2010 **Post-graduation**, Structural and Functional Biochemistry, Universidade Nova de Lisboa, Caparica (Portugal)
Global National Classification: Very Good (17 out of 20)

2005 - 2009 **B.Sc.**, Molecular and Cellular Biology, Universidade Nova de Lisboa, Caparica (Portugal)
Global National Classification: Very Good (16 out of 20)

AWARDS AND STIPENDS

2014 Student travel stipend for the 55th ENC
2012 ISOTEC Student Sponsorship for the 2012 ICMRBS conference
2010 – 2011 Scholarship, Excellence Foundation for the Promotion of the Max-Planck Society
2009 Scholarship, Socrates/Erasmus program

SCIENTIFIC COMMUNICATIONS

Oral communications

2015 *Speeding-up exchange-mediated saturation transfer experiments by Fourier transform*, GRC Computational Aspects — Biomolecular NMR, Il Ciocco (Italy)

2014 *Anti-HIV lectin OAA samples glycan-bound conformers in solution in the absence of ligand*, VII Biennial GERMN, Madrid (Spain)

Poster Presentations

2014 *Solution ensemble of OAA reveals bound-like conformation in the absence of*

- ligand*, Marta G. Carneiro, Leonardus Koharudin, David Ban, T. Michael Sabo, Adam Mazur, Christian Griesinger, Angela M. Gronenborn, Donghan Lee, 55th ENC, Boston (U.S.A.)
- 2013 *Solution ensemble of OAA reveals bound-like conformation in the absence of ligand*, Marta G. Carneiro, Leonardus Koharudin, David Ban, T. Michael Sabo, Adam Mazur, Christian Griesinger, Angela M. Gronenborn, Donghan Lee, EUROMAR 2013, Crete (Greece)
- 2012 *N-glycan Recognition by the Cyanobacterial *Oscillatoria agardii* lectin*, Marta G. Carneiro, Leonardus Koharudin, David Ban, T. Michael Sabo, Adam Mazur, Christian Griesinger, Angela M. Gronenborn, Donghan Lee, XXVth ICMRBS, Lyon (France)
- 2008 *¹H, ¹⁵N and ¹³C assignment and NMR solution structure of rubredoxin*, Marta G. Carneiro, Carla S. Ferreira, R.M. Almeida, J.J.G. Moura, E.J. Cabrita, S.R. Pauleta, IV Bienal GERMN, Seville (Spain)

PUBLICATIONS

Carneiro, M. G., Reddy, J. G., Griesinger, C., Lee, D., *Speeding-up exchange-mediated saturation transfer experiments by Fourier transform*, J. Biomol. NMR. *in press* (2015)

Carneiro, M.G., Koharudin, L. M., Ban, D., Sabo, T. M., Trigo-Mourino, P., Mazur, A., Griesinger, C., Gronenborn, A. M., Lee, D., *Sampling of glycan-bound conformers by the anti-HIV lectin *Oscillatoria agardii* agglutinin in the absence of sugar*, Angew. Chem. Int. Ed. Engl. 54:6462 (2015)

Carneiro M.G., Koharudin, L. M., Griesinger, C., Gronenbron, A. M., Lee, D., *¹H, ¹³C and ¹⁵N resonance assignment of the anti-HIV lectin from *Oscillatoria agardii**, Biomol. NMR Assign. 9:317 (2015)

Bibow, S., Carneiro, M. G., Sabo, T. M., Schwiegk, C., Becker, S., Riek, R., Lee, D., *Measuring membrane protein bond orientations in nanodiscs via residual dipolar couplings*, Protein Sci. 23:851 (2014)

Ban, D., Mazur, A., Carneiro, M. G., Sabo, T.M., Giller, K., Koharudin, L. M., Becker, S., Gronenborn, A. M., Griesinger, C., Lee, D., *Enhanced accuracy of kinetic information from CT-CPMG experiments by transverse rotating-frame spectroscopy*, J. Biomol. NMR 57:73 (2013)