Exploring the limits of solution-state NMR in application to challenging proteins in drug discovery

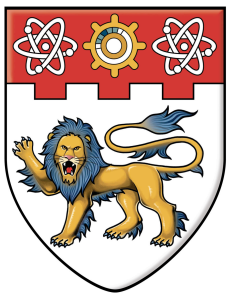
Russell Raj Pandian Senthamarai

2012

Russell Raj Pandian Senthamarai. (2012). Exploring the limits of solution-state NMR in application to challenging proteins in drug discovery. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/50874

https://doi.org/10.32657/10356/50874



NANYANG TECHNOLOGICAL UNIVERSITY

Exploring the limits of solution-state NMR in application to challenging proteins in drug discovery

RUSSELL RAJ PANDIAN SENTHAMARAI

SCHOOL OF BIOLOGICAL SCIENCE

2012

Exploring the limits of solution-state NMR in application to challenging proteins in drug discovery

Russell Raj Pandian Senthamarai

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2012

Acknowledgements

Life would not start without the blessings of the parents and my life in this world has also been the same. The blessings of my “Grand family” have always driven the wind of success towards me. Today this is the moment of truth that I can make my “Grand family” to be proud of me. I just can’t thank them but to make them proud in all the ways I can.

I am so thankful to my PhD supervisor, Prof. Konstantin, Pervushin. He has trusted on me to succeed in the path of this research work. His belief, knowledge, guidance, suggestions, advice and many more has groomed me throughout the course of my study. I wish to express all my deep appreciations and indebtedness to Prof. Konstanin, Pervushin for introducing me to the euphoric field of NMR.

I thank Prof. Gerhard Gruber, Associate Chair (Research), SBS-NTU, for his constant support and encouragement. I like to thank Prof. Julien, Lescar for his helpful discussions and suggestions and also providing me the facilities in his laboratory during my work with NS3 helicase protein. I would like to thank Prof. Yuguang, Mu for his constant encouragement throughout my Ph.D research and for the enlightening discussions, as a member of my Thesis Advisory Committee. I thank Dr. Ilya, Kuprov for his valuable comments and suggestions during the development of spin dynamics analysis tool. I also thank The Biomolecular Magnetic Resonance Center (BMRZ) of the Goethe University, Frankfurt, Germany for their generous allocation of the NMR time for the NS3 helicase work.

I thank all my lab mates, former and present, for their support and encouragement during the course of my research work.

It is the friends; no matter where they are in the world we seek to share us, when we need just a blind support. My dear friends Bakka, Senthil, Srini sir and lastly but clearly not the least, Senior, are those who supported constantly throughout my Ph. D research.

I am grateful to School of Biological Sciences, Nanyang Technological University, Singapore for providing the Ph. D studentship with Graduate Scholarship without which this work would not have been possible.

This piece of work is dedicated to my beloved family and teachers.

Acknowledgements i

Table of Contents v

List of Figures vii

List of Tables ix

Abbreviations xi

Abstract xiii

Table of Contents

Introduction 1
1 Introduction 3
1.1 NMR spectroscopy of high molecular weight proteins 3
1.1.1 Increasing sensitivity and spectral resolution –TROSY 6
1.1.2 Supra-molecular structures (MWt>200kDa) – CRIPT/CRINEPT 7
1.1.3 Multi-dimensional NMR experiments 8
1.1.4 Stable Isotope labeling 8
1.1.5 Paramagnetic labeling methods 8
1.2 Challenging systems studied by NMR 9
1.2.1 Malate synthase G 9
1.2.2 MurB enzyme 10
1.2.3 Maltose binding protein 11
1.2.4 VDAC – Voltage Dependent Anion Channel 12
1.2.5 Dengue virus NS2B-NS3 protease 13
1.3 Objective 22
1.3.1 Aims and scope of this thesis 22

Exploring the limits of NMR experiments: unified spin dynamics simulations, analysis and comparisons 23
2. Exploring the limits of NMR experiments: unified spin dynamics simulations, analysis and comparisons 25
2.1 Introduction 25
2.1.1 NMR spin-dynamics simulations prorgams 25
2.1.1.1 Analytical spin dynamics simulation programs 26
2.1.1.2 Numerical spin dynamics simulation programs 27
2.1.2 Spin dynamics 28
2.2 NMR-RDB: a relational database of protein pulse sequences 29
2.2.1 Functional architecture of NMR-DB and Spinach 29
2.3 Algorithms and implementation 31
2.3.1 The density operator propagation 32
2.3.2 Simulation of spin relaxation of the density operator 33
2.3.3 Magnetization transfer graphs and relaxation profiles 34
2.3.4 Performance metrics 35
2.4 Results and discussion 38
2.4.1 Construction of a generic spin system 38
2.4.2 Performance annotation and analysis of TROSY pulse sequences 39
2.4.3 Sensitivity in multidimensional experiments 44
2.4.4 WIKI based online NMR-DB for the setup, execution, sharing and analytics of complex NMR experiments 46

DENV4 NS3 helicase: backbone resonance assignment 49
3. DENV4 NS 3 helicase: backbone resonance assignment 51
3.1 Introduction 51
3.1.1 Dengue virus 52
3.1.1.1 NS3 protein 53
3.1.1.2 NS3 protease domain 53
3.1.1.3 NS3 helicase domain 54

3.2 Materials and Methods 57
3.2.1 Materials 57
3.2.1.1 Chemicals 57
3.2.1.2 Electrophoresis Reagents 57
3.2.1.3 Molecular Biology Materials 57
3.2.1.4 Chromatography 57
3.2.1.5 Instrumentation and accessories 58
3.2.1.6 Computer Software 58
3.3 Methods 59
3.3.1 Stable isotope labeling of DENV4 NS3 helicase 59
3.3.1.1 Uniform ^{2}H, ^{15}N and ^{2}H, ^{15}N, ^{13}C labeling of NS3 helicase 59
3.3.1.2 Selective 1-^{13}C-Leu, Ala, Val and Phe labeled NS3 helicase 60
3.3.2 DENV4 NS3 helicase expression 61
3.3.2.1 NS3 helicase protein expression protocol 61
3.3.3 Purification of DENV4 NS3 helicase 62
3.3.3.1 NS3 helicase purification protocol 62
3.4 Results and Discussion 64
3.4.1 Expression and purification of NS3 helicase 64
3.4.2 NMR spectroscopy of DENV4 NS3 helicase 65
3.4.2.1 Resonance assignment process 67
3.4.2.2 1-^{13}C-Leu, 1-^{13}C-Ala selective labeling in a ^{2}H, ^{15}N background 69
3.4.3. Preliminary PRE studies 70
3.4.4 Design, high-level expression and purification of fragments of DENV 4 NS3 helicase 72
3.4.4.1 Design of a truncated β hairpin in subdomain 2 72
3.4.4.2 High-level expression and purification of the designed fragments of DENV4 NS3 helicase 73

Conclusion 75
4. Conclusion and future perspectives 77

References 81

Appendix – A 95
Appendix – B 96
Appendix – C 97
Appendix – D 99
Appendix – E 100

Author's Publications 101

List of Figures

Introduction

Figure 1.1: Malate Synthase G 10
Figure 1.2: UDP-N-Acetylenolpyruvylglucosamine reductase (murB) 11
Figure 1.3: Maltose Binding Protein 12
Figure 1.4: Voltage Dependent Anion Channel-1 in LDAO miscelles 13
Figure 1.5: DENV2 NS2B-NS3 protease 14

Exploring the limits of NMR experiments: unified spin dynamics simulations, analysis and comparisons

Figure 2.1: Functional architecture of Spinach-NMR-RDB 30
Figure 2.2: Time line of HMQC pulse sequence 35
Figure 2.3: Spinach performance measures 37
Figure 2.4: Spin-system 39
Figure 2.5: Performance annotation of the TROSY pulse sequences 40
Figure 2.6: Magnetization flow graph of TROSY pulse sequences 42
Figure 2.7: Magnetization flow graph of TROSY pulse sequence 43
Figure 2.8: Predicted vs experimental S/N between a set of 2D $^{1}H,^{15}N$-correlation experiments 44
Figure 2.9: Sensitivity prediction across dimension 46
Figure 2.10: Screenshot of NMR-DB 47
Figure 2.11: Functional architecture of NMR-DB 48

DENV4 NS3 helicase: backbone resonance assignment

Figure 3.1: Dengue virus genome architecture 52
Figure 3.2: DENV4 NS3 protease-helicase 53
Figure 3.3: Flavivirus NS3 helicase – RNA – AMPPNP ternary complex 55
Figure 3.4: Purification analysis of DENV4 NS3 helicase 64
Figure 3.5: Amino acid sequence of DENV4 NS3 helicase 64
Figure 3.6: 1D ^{1}H NMR spectrum of DENV4 NS3 helicase 65
Figure 3.7: Conformation exchange induced line-broadening 66
Figure 3.8: Effect of non-hydrolisable ATP analogue 67
Figure 3.9: Sequence specific assignment 68
Figure 3.10: Sequential information from selectively 1-^{13}C amino acid labelled NS3 helicase 70
Figure 3.11: Paramagnetic relaxation enhancement study of NS3 helicase 71
Figure 3.12: [^{1}H, ^{15}N]-TROSY-HSQC spectrum of DENV4 NS3 helicase 71
Figure 3.13: Subdomain 1 and 2 of DENV4 NS3 helicase 73
Figure 3.14: Purification analysis of designed fragments 74

Conclusion

Figure 4.1: Statistical distribution of peak amplitudes for DENV4 NS3 helicase 77

List of tables

Introduction

Table 1.1: NMR experiments and their parameters used for Malate Synthase G 15
Table 1.2: NMR experiments and parameters used for UDP-N-acetylenolpyruvylglucosamine reductase (MurB) 17
Table 1.3: NMR experiments and parameters used for the backbone resonance assignment of Maltose Binding Protein
Table 1.4: NMR experiments and amino acid selective isotope labeled samples used for the structure of VDAC-1 Protein 18
Table 1.5 Summary of the strategies used and their percentile contributions to the total achieved backbone resonances assignment 20

Exploring the limits of NMR experiments: unified spin dynamics simulations, analysis and comparisons

Table 2.1a: Simulated performance of 2D ^{1}H-^{15}N correlation experiments for uniformly ^{13}C,^{15}N-labeled protein ($\alpha X\beta 2$ Integrin I domain) with t_c = 20 ns. 45
Table 2.1b: Simulated performance of 2D ^{1}H-^{15}N correlation experiments for uniformly ^{2}H,^{13}C,^{15}N-labeled protein (NS3 helicase) with t_c = 40 ns. 45

DENV4 NS3 helicase: backbone resonance assignment

Table 3.1a: M9 medium composition 60
Table 3.1b: M9 Medium supplements 60
Table 3.2: Growth conditions 61

Abbreviations

1D	One dimensional
2D	Two dimensional
3D	Three dimensional
4D	Four dimensional
ADP	Adenosine Di-Phosphate
ATP	Adenosine Tri-Phosphate
BMRB	Biological Magnetic Resonance Bank
CARA	Computer Aided Resonance Assignment
CPMG	Carr-Purcell-Meiboom-Gill
CRIPT	Cross-Relaxation Induced Polarization Transfer
CSA	Chemical Shift Anisotropy
CTE	Coherence Transfer Elements
DD	Dipole Dipole
DMSO	DiMethyl Sulfoxide
DNA	Deoxy ribo-Nucleic Acid
DNP	Dynamic Nuclear Polarization
DTT	DiThioThreitol
E. coli	Excherichia coli
GAMMA	General Approach to Magnetic resonance Mathematical Analysis
HMQC	Heternuclear Multiple Quantum Correlation
HSQC	Heteronuclear Single Quantum Correlation
HTS	High Throughput Screening
Hz	Hertz
INEPT	Insensitive Nuclei Enhanced Polarization Transfer
IPTG	Isopropyl-β-D-Thiogalactoside
JNES	J-Neighborhood of Every Spin
kDa	Kilo Dalton
L-TROSY	Longitudinal-TROSY
MWCO	Molecular Weight Cut-Off
NMR	Nuclear Magnetic Resonance
NMR-RDB	NMR Relational DataBase
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect SpectroscopY
NTA	Nitrilotriacetic acid
NTP	Nucleotide Tri-Phosphate
NS2B	Non-Structural protein 2B
NS3	Non-Structural protein 3
NS5	Non-Structural protein 5
PAGE	Polyacrylamide Gel Electrophoresis

PCR	Polymerase Chain Reaction
PCS	Pseudo-Contact Shifts
PDB	Protein Data Bank
PEP	Preservation of Equivalent Pathways
PFG	Pulse Field Gradient
POMA	Product Operator formalism using Mathematica
ppm	parts per million
PRE	Paramagnetic Relaxation Enhancement
RF	Radio-Frequency
RNA	RiboNucleic Acid
SAXS	Small Angle X-ray Scattering
SDS	Sodium dodecyl sulphate
SEA	Solvent Exposed Amide
SIMPLTN	Simulation of PuLse and Two dimensional Nmr
SIMPSON	SImulation Package for SOlid state Nmr spectroscopy
SNR	Signal to Noise Ratio
ST2-PT	Single Transition to Single Transition Polarization Transfer
STD	Saturation Transfer Difference
TOCSY	Total Correlation SpectroscopY
TROSY	Transverse Relaxation Optimized Spectroscopy
URL	Uniform Resource Locator
VNMRS	Virtual NMR Spectrometer
XML	eXtensible Markup Language

Abstract

Solution state Nuclear Magnetic Resonance Spectroscopy, despite its emergence as an important technique has its own limits: rapid transverse relaxation proportional to the molecular size of the protein, slow longitudinal relaxation and resonance overlap with increasing molecular size. While the limitations have been addressed consistently the quantitative limits of solution NMR experiments remain unclear. In the present work, we reported a unified framework, which enables a uniform benchmarking of all the available solution NMR experiments in terms of their sensitivity and tolerance to transverse relaxation. Ability to predict the sensitivity per unit time across dimensions has also been demonstrated and validated. Subsequently, an online NMR-DataBase has been developed accumulating the most commonly used solution NMR experiments. A backbone resonance assignment of DENV4 NS3 helicase, a 51 kDa protein, has been discussed. Less than 10% of the resonance assignment has been achieved with triple resonance. With the introduction of selectively labeled amino acids, another 4 residues have been assigned, portraying the necessity of alternative approaches such as PRE/PCS.

Introduction

1 Introduction:

1.1 NMR spectroscopy of high molecular weight proteins

Since the introduction of Nuclear Magnetic Resonance Spectroscopy as a method for structural studies of biological macromolecules in the early 1980s, the method has been noted for its ability to offer structural and dynamical details of biomolecules and macromolecular complexes at the atomic resolution. NMR technology is currently being used in a variety of research and industrial applications including characterization of molecular structure and dynamics, metabolic profiling of living organisms, structure-based drug discovery, and quality control in production processes. A detailed timetable of the progress of NMR spectroscopy (Emsley *et al.,* 2007) showed an acceleration of the technical development in the field of biomolecular NMR enabling structure determination of proteins and nucleic acids and their complexes with ever increasing molecular weight. Today NMR spectroscopy is one of the established methods for the three-dimensional structure determination of biomolecules, along side with X-ray crystallography (Wuthrich 1998) and cryo-EM (Saibil 2000) . More than 9000 structures in the Protein Data Bank (Berman *et al.,* 2000) as of November 2011 have been determined by NMR spectroscopy. This remarkable achievement would have not been possible without the extensive development in the spectrometer hardware, multi-dimensional NMR techniques, isotope labeling strategies and software for analyzing the multi-dimensional spectra.

Despite a significant proportion of obtained structures of biomolecules deposited in PDB has been solved by NMR, upon examination of BMRB statistics (http://www.bmrb.wisc.edu/ref_info/), the following inference related to size distribution of the deposited structures can be made: 90% of the NMR derived structures are in the range of 2 kDa to 25 kDa with a maximum at 8 to 10 kDa with only a few structures with molecular weights exceeding 25 kDa (Guntert 2011). This paucity of NMR structures above 25 kDa reflects the increasing effort and cost involved in the structure determination. The prerequisite of a successful NMR structural study is availability of individual resonances-resolved nD NMR spectra recorded with sufficiently high signal-to-noise ratio (SNR). For biomolecules with molecular sizes exceeding 25 kDa both NMR resolution power and SNR are rapidly degrading (Wagner 1993; Kay *et al.,* 1997; Clore *et al.,* 1998). The limiting factors are lower solubility of high molecular weight proteins at concentrations required by current inherently low sensitivity of NMR signal-detection methods and resonance line

broadening due to fast transverse relaxation and extensive resonance overlap due to the high complexity of the spectra. Among this sensitivity is the single most critical factor in NMR performance and limits the feasibility and success of various applications. Compared to other analytical methods, NMR requires larger sample quantities and/or longer experiment times than related analytical methods. Over the past 20 years progress on this issue has been slow, with a sensitivity enhancement of about 20-fold achieved through modest improvements in NMR hardware and techniques, including the use of ultra high superconducting magnets (>20 T) and cryogenically cooled probes. However, at present, NMR requires more than 10 micromoles of a sample concentration in the volume of 150-500 μl, which is much larger than that required for state-of-the-art mass spectrometry which recently has been able to work in atto (10^{-18}) mole units.

Addressing these limitations a number of breakthroughs have been made in the past couple of decades, which can be broadly divided into hardware/signal-detection schemes improvements and methods utilizing better understanding of the underlying nuclear spin dynamics. The former developments include cryogenic probe technology involving a probe RF coil and/or built-in signal pre-amplifier that are cooled with a flow of cold He gas. The base noise of the NMR spectrum is proportional to $\sqrt{Rc(Tc+Ta)}$ where Rc is the resistance of the coil, Tc is the temperature of the RF coil, and Ta is the noise temperature of the pre-amplifier; note that we neglect the effect of sample noise. The signal-to-noise ratio (SNR) of the current cryogenic probe, in which the probe RF coils are cooled down to be 20-30K and the pre-amplifier to be 50K, is increased by about a factor of 4 compared with a conventional probe operating at room temperature. We shall note here, that although technologically challenging, further decrease of the operational temperatures holds a promise to double or triple existing levels of SNR. Another promise to improve SNR is stemming from the use of dynamic nuclear polarization (DNP). The origin of low sensitivity in NMR is due to the small magnetic moment of nuclear spins, which yields small Boltzmann polarization at thermal equilibrium condition. The sensitivity of NMR can be enhanced by hyperpolarization where the nuclear spin polarization is far beyond thermal equilibrium condition (Waldherr *et al.,* 2011). DNP is a hyperpolarization technique, which transfers a higher degree of order of electron spin polarization to nuclear spin polarization. Theoretically, it is capable of increasing the NMR sensitivity

by a factor of up to 10^4 under ideal conditions. Recently, DNP has been applied mainly to solid state NMR and has improved its sensitivity by ~10^2. However, it is still not practical for solution NMR because dramatic enhancement can be obtained only when the sample is polarized at very low temperature (1-4 K) and/or in the device optimized for polarization other than in the NMR magnet. Oxford Instruments had commercialized the DNP device for solution NMR. However, it can be applied only on a single scan one-dimensional experiment because it takes more than several hours to polarize the sample for acquiring one transient NMR data.

Attempts at construction of theoretically optimal NMR pulse sequences and NMR experiments utilizing favorable magnetization transfer pathways led to TROSY (Transverse Relaxation-Optimized Spectroscopy)-class of NMR experiments (Pervushin *et al.,* 1997a), CRINEPT (cross-correlated relaxation induced polarization transfer) and Methyl-TROSY (Tugarinov *et al.,* 2003a) in association with suitable isotope labeling schemes (Kay *et al.,* 1997), the effective size limit for the observation of NMR signals in solution has been extended several fold.

In the presence of spin relaxation in large proteins or in the crowding environment in living-cell (Sakakibara *et al.,* 2009), new quantum bounds imposing fundamental limitations on the sensitivity of NMR experiments and eventually their applicability to proteins of interest shall be developed promising further increase of sensitivity of biological NMR experiments by a factor of 1.2-2.0. Systematic benchmarking of multi-dimensional NMR experiments is a critical prerequisite for optimal allocation of NMR resources for structural analysis of challenging proteins, e.g. large proteins with limited solubility or proteins prone to aggregation. A set of fundamental benchmarks for all commonly used protein NMR experiments enabling accurate simulation and benchmarking of NMR experiments on large spin systems needs to be developed, an issue considered in subsequent chapters of this thesis. A key feature is the ability to use a single spin system to simulate the majority of solution state NMR experiments, thus providing the (hitherto unavailable) unified framework for pulse sequence evaluation. This development enables predicting relative sensitivity of different implementations of NMR experiments, thus providing a basis for comparison, optimization and, eventually, automation of NMR analysis. The fundamental studies on quantum bounds will result in optimal time-shared data acquisition schemes such as targeted acquisition (Jaravine *et al.,* 2006) , where several NMR experiments are acquired simultaneously with time-domain sampling schemes

adjusted to the selected acquisition target, for example, full assignment of resonances and/or the protein structure. This approach holds promises to further improve SNR per unit time by at least factor of 1.5. At the same time, rigorous (non-empirical) analytical framework for assessment of optimality for this type of new NMR experiments for rapid data collection is entirely missing.

In the following sections, a description and use of these new techniques in the observation of high-resolution NMR spectra of macromolecules and in the context of high molecular weight is given. A brief discussion on NMR resonance assignment of challenging large molecular weight proteins is also given towards the end of this chapter.

1.1.1 Increasing sensitivity and spectral resolution –TROSY

With the advent of high-field magnets biomolecular NMR reached intrinsic sensitivity and resonance lines separation, otherwise termed spectral resolution, sufficient for protein work (Wagner 1993; Wuthrich 1998). At higher magnetic fields (Optimum lies between 950MHz - 1050MHz), chemical shift anisotropy of ^{1}H, ^{15}N and ^{13}C nuclei contributes significantly to transverse relaxation, besides the relaxation due to dipole-dipole coupling. TROSY technique uses CSA relaxation at higher fields to cancel field-independent dipolar relaxation. In brief: in heteronuclear two-spin systems, the NMR signal of each resonance is split into two components by the scalar spin-spin coupling. Therefore, in a 2D correlation experiment one observes a four-line fine structure. However, it has been known that the individual multiplet components have different transverse relaxation times and, hence, different line widths, which are mixed by decoupling in conventional multidimensional NMR techniques (refer (Pervushin *et al.,* 1997a) for the literature review). On the other hand, using TROSY technique the multiplet structure is not decoupled, rather only the narrowest, most slowly relaxing line of each multiplet is retained.

Combination of a new polarization element that retains 50% (as opposed to 25%) of the initial proton polarization and by the use of steady-state heteronuclear magnetization, available due to the absence of decoupling during chemical shift encoding in the indirect dimension, partially compensates the sensitivity loss as a result of the use of only one out of four multiplet components (Pervushin *et al.,* 1998a; Pervushin *et al.,* 1998b). And thus, TROSY technique offers superior sensitivity than the conventional techniques when working with large molecular sizes and extends the

size limit amenable to NMR to ~150kDa. With the implementation of TROSY elements in triple resonance experiments, a fourfold increase in sensitivity can be obtained for proteins with high molecular range (Salzmann *et al.,* 1998). However, the prerequisite to apply TROSY technique to large molecular size macromolecule requires a proportion of non-exchangeable protons to be replaced with ^{2}H depending on the molecular size.

1.1.2 Supra-molecular structures (MWt>200kDa) – CRIPT/CRINEPT

For molecular sizes above 200 kDa, for which transverse relaxation during polarization transfers becomes a limiting factor, application of CRINEPT (cross-correlated relaxation-enhanced polarization transfer) yields a significant gain in sensitivity. While TROSY employs INEPTs (insensitive nuclei enhanced polarization transfer) for the transfer of magnetization between different types of nuclei via spin-spin coupling, CRINEPT combines CRIPT (cross-correlated relaxation induced polarization transfer) with INEPT. In this manner, CRINEPT effectively circumvents the rapid transverse relaxation: transfer efficiency proportional to the effective molecular size (Riek *et al.,* 1999).

Recently the development focal point was shifted towards targeted acquisition for real-time NMR spectroscopy (Jaravine *et al.,* 2006; Wong *et al.,* 2008). A target-oriented approach for the acquisition of information in biomolecular NMR spectroscopy combines concurrent data accumulation, processing, and monitoring of spectral quality. Real-time estimation of parameters allows acquisition to be stopped when results are complete and have a specified precision. An incremental nonuniform sampling scheme ensures the optimization of resolution sensitivity. In order to establish an optimal balance between completeness of resonance assignment and losses of cross-peaks due to dynamic processes/degradation of protein, assignment of backbone resonances is set as a stirring criterion for dynamically controlled targeted nonlinear NMR data acquisition. Combined these methods may result in hyperdimensional NMR spectroscopy often equipped with nonlinear sampling (Jaravine *et al.,* 2008). It is very clear that optimization and full utilization of magnetization available in the concurrently run NMR experiments dictated by fundamental limits of spin dynamics is critical for maximizing SNR and to this end the overall chance for success for this class of experiments.

1.1.3 Multi-dimensional NMR experiments

Often, spectra of large molecular weight proteins suffer from extensive overlap of resonances that renders unambiguous interpretation of cross-peaks virtually impossible, especially in the case of single polypeptide chain proteins. However, several exciting developments from over the past year promise to extend the NMR methodology from 2D in to 3D and 4D NMR (Clore *et al.,* 1991). While the conventional approach attempts to improve the resolution of 2D NMR spectra by increasing the digital resolution and by using strong resolution-enhancement, digital filtering functions at the expense of sensitivity, the alternative approach improves the digital resolution by increasing the dimensionality of the spectrum (Kay *et al.,* 1990). The application of 3D and 4D NMR to large proteins is limited as the coherence transfer efficiency drops during long magnetization transfers. Clore and co-workers have reviewed a number of 3D and 4D pulse sequences developed so far (Clore *et al.,* 1991).

1.1.4 Stable isotope labeling

Complexity of the NMR spectra increases with increasing size of proteins without inherent symmetry. This prevents the detailed analysis of the multidimensional spectra, despite the high resolution and sensitivity offered by aforementioned techniques. However, promising level of development in biochemical approaches, the spectral complexity can be brought down to manageable level, even for larger proteins without inherent symmetry, using different isotope labeling strategies (Yamazaki *et al.,* 1998; Xu *et al.,* 1999; Tugarinov *et al.,* 2006; Takeuchi *et al.,* 2007; Liu *et al.,* 2009; Gans *et al.,* 2010; Muona *et al.,* 2010; Ayala *et al.,* 2011).

1.1.5 Paramagnetic labeling methods

In the last decade paramagnetic labeling of proteins has gained importance as it offers powerful, long-range structural information for the analysis of structure and dynamics of proteins as well as protein-ligand complexes (Jia *et al.,* 2011). In addition, it offers widely used tools for sensitivity enhancement of NMR experiments, resonance assignment and studies on conformational exchange (Otting 2010). The paramagnetic effects include paramagnetic relaxation enhancement (PRE), paramagnetic shifts or pseudocontact shifts (PCS), paramagnetic molecular alignment and DD/CSA cross-correlation, the former pair of effects offering long-range

structural and principle vector-directional information for the structural analysis. Recently the full structure determination of a protein using paramagnetic shift data in conjunction with the protein structure prediction algorithm Rosetta (Christophe Schmitz *et al.,* 2009) was demonstrated and available under the name Rosetta-PCS.

Paramagnetic effects are induced either by site-specific incorporation of a paramagnetic agent or titration of a paramagnetic agent against the protein while chemical modification of the protein have also been reported (Pintacuda *et al.,* 2002; Man *et al.,* 2010; Yagi *et al.,* 2010; Graham *et al.,* 2011). Furthermore, a number of paramagnetic agents are reported to suit different requirements (Graham *et al.,* 2011; Yagi *et al.,* 2011).

With this arsenal of NMR techniques and advancements in isotope labeling methods, a few of large molecular weight targets have been attempted and deposited in PDB upon successful completion of resonance assignment and structure determination. In the following section, we will discuss the largest single polypeptide proteins or protein complexes successfully studied by NMR and relevant to our primary protein target presented in this thesis, NS3 helicase from Dengue virus.

1.2 Challenging systems studied by NMR

Continuing technical advancement coupled with methodological development, armed NMR spectroscopy to address challenging high molecular weight proteins. Since, this thesis discuss the attempts made for the backbone resonance assignment of DENV4 NS3 helicase – a single polypeptide chain protein with molecular weight of 51 kDa, here we restrict the discussion to single polypeptide chain proteins above 30 kDa. A brief description on the NMR study/characterization of the notable proteins fall under the category follows.

1.2.1 Malate synthase G

The latest and current "world-record" (at the time of writing) for the size limit amenable to NMR is, Malate synthase G enzyme, a protein of 723 amino acids, solution structure of which has been solved by NMR spectroscopy (Tugarinov *et al.,* 2005a) and later refined using small angle X-ray scattering (SAXS) (Grishaev *et al.,* 2008). MSG is a real challenge to NMR, in the sense that it is a monomeric globular multidomain protein with no inherent symmetry as shown in Figure 1.1. It shall be noted that as a protein Malate synthase G showed exceptionally high solubility and

conformational homogeneity under the conditions optimal for NMR measurements greatly facilitating its study by NMR.

Nearly complete backbone resonance assignment of MSG has been achieved with the introduction of 4D NMR techniques (Yang *et al.,* 1999b; Tugarinov *et al.,* 2002). In addition to introduction of methyl-TROSY experiments (Tugarinov *et al.,* 2003a) and new isotope labeling techniques (Tugarinov *et al.,* 2003b), the methyl side chain resonance assignment has also been achieved for MSG. Further refinements of the structure have been made using SAXS data (Grishaev *et al.,* 2008). The list of experiments and the corresponding parameters has been given in Table 1.1. All the cross-peaks stemming from the backbone amino groups are found in the TROSY spectra as shown in Figure 1.1 with only modest spectrum overlap despite the large size of the protein. For this protein a set of TROSY-class triple resonance experiments was enough to achieve almost complete assignment of the backbone resonances.

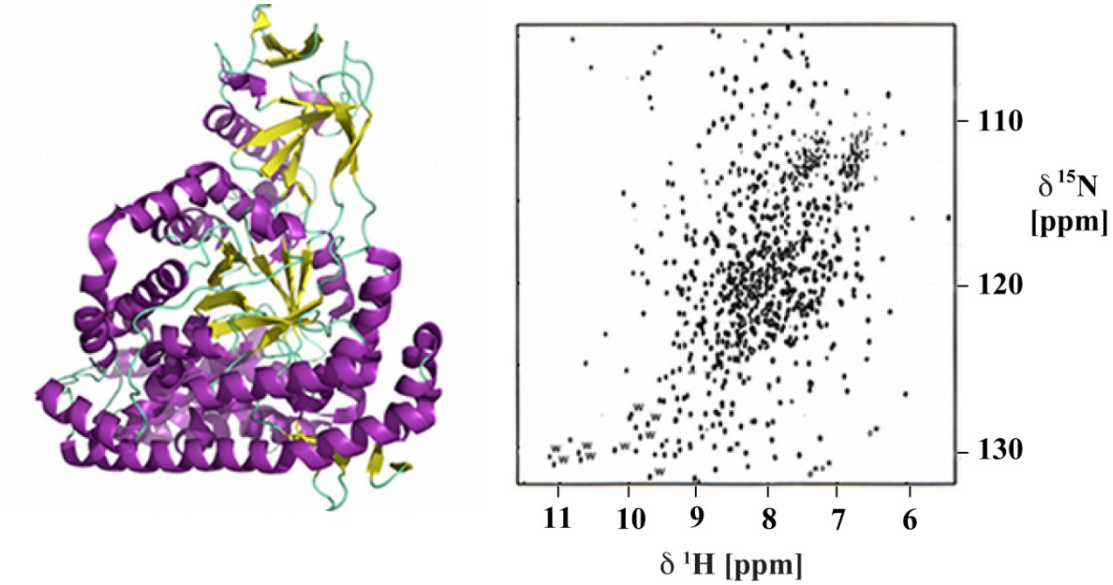


Figure 1.1: Malate Synthase G: Cartoon representation three dimensional structure of Malate Synthase G. (PDB: 1D8C.pdb) (Howard *et al.,* 2000) is shown with its fingerprint [^{1}H,^{15}N]-TROSY-HSQC spectrum (spectrum adopted from (Tugarinov *et al.,* 2002) and used for the illustrative purpose)

1.2.2 MurB enzyme

MurB, known as UDP-N-acetylenolpyruvylglucosamine reductase, is a monomeric, 347-residue (38.5 kDa) flavoenzyme essential for bacterial cell-wall biosynthesis (Pucci *et al.,* 1992). Cartoon representation three dimensional structure of MurB is shown in Figure 1.2. A nearly complete backbone resonance – 96% of the backbone atoms, $^{13}C\alpha$, $^{1}H_N$, ^{15}N and ^{13}CO for the substrate free and 90% of the

backbone atoms for the ligand bound MurB, has been achieved with the conventional triple resonance NMR experiments coupled with the uniform deuterium labeling of the protein (Constantine *et al.,* 1997). The experiments used and the parameters there on are listed in Table 1.2. As with MSG, exceptionally high structural homogeneity and solubility of the protein resulting in the presence of all expected cross-peaks in the HSQC spectrum as shown in Figure 1.2 helped to assign the protein backbone resonances without the use of PRE/PCS or site specific mutagenesis.

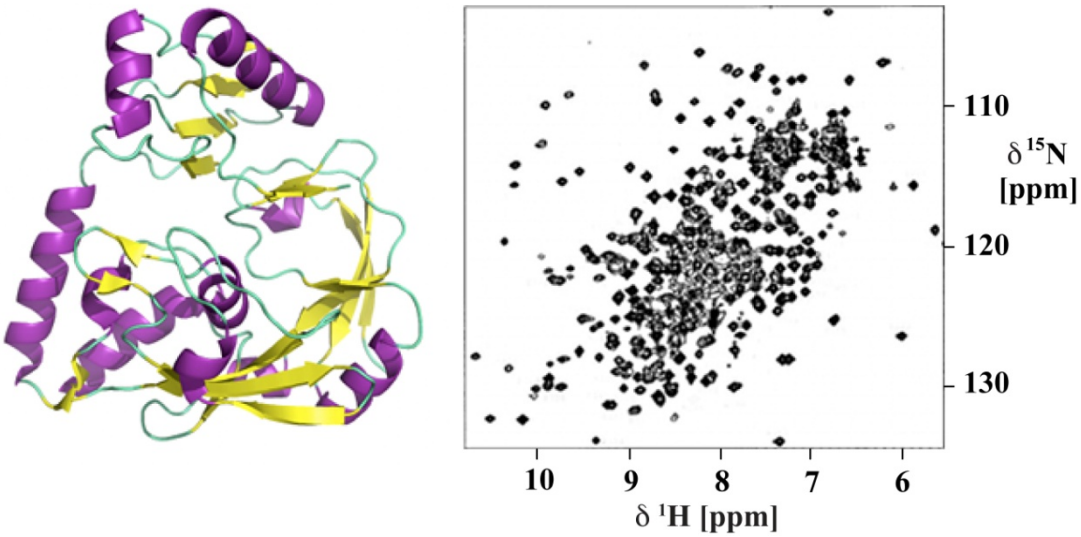


Figure 1.2: UDP-N-Acetylenolpyruvylglucosamine reductase (MurB): Cartoon representation three dimensional structure of UDP-N-Acetylenolpyruvylglucosamine reductase. (PDB: 2GQU) (Kim *et al.,* 2007) is shown with its fingerprint [^{1}H, ^{15}N]-HSQC spectrum (spectrum adopted from (Constantine *et al.,* 1997) and used for the illustrative purpose).

1.2.3 Maltose binding protein

Maltose binding protein is a part of the maltose/maltodextrin system of *Escherichia coli*. It is a 370 amino acid single polypeptide chain monomeric protein of molecular weight ~ 42 kDa. The three dimensional structure of maltose binding protein is shown in Figure 1.3. Over 95% of the backbone resonance assignment has been reported with the combination of triple resonance experiments and uniform deuterium labeling. Furthermore, the selective protonation of methyl groups of Val, Leu and Ile (δ1 methyl group only) was helpful in obtaining long range NOE based restraints greatly assisting the three dimensional structure determination (Gardner *et al.,* 1997). MBP, well known, as a solubility tag in protein expression systems, is one of the stable and soluble proteins under a wide range of experimental NMR conditions. High solubility of MBP enables the nearly complete resonance assignment using only

TROSY based triple resonance experiments in combination with the deuterium labeling with selective protonation of methyl groups. Figure 1.3 shows a high quality [^{1}H, ^{15}N]-TROSY-HSQC spectrum with well-dispersed homogeneous peaks. A detailed description of NMR experiments, corresponding parameters are given in Table 1.3.

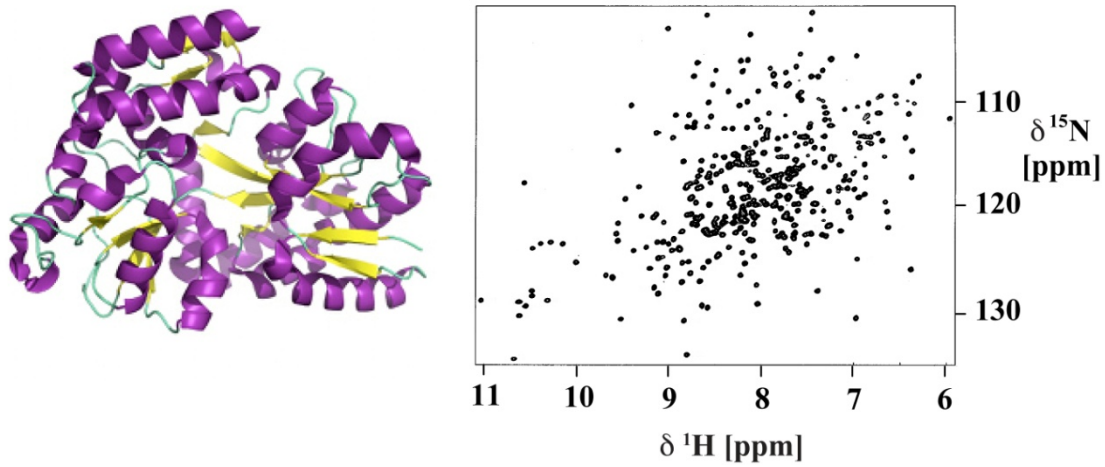


Figure 1.3: Maltose Binding Protein: Cartoon representation of Maltose binding protein from *Streptococcus pneumoniae*. (PDB: 2XD3) (Abbott *et al.,* 2010) shown with its fingerprint [^{1}H, ^{15}N]-TROSY-HSQC spectrum (spectrum adopted from (Gardner *et al.,* 1998) and used for illustrative purpose)

1.2.4 Voltage Dependent Anion Channel – 1 (VDAC – 1)

The voltage-dependent anion channel (VDAC), showin in Figure 1.4, mediates trafficking of small molecules and ions across the eukaryotic outer mitochondrial membrane (Hiller *et al.,* 2008; Villinger *et al.,* 2010). VDAC stands as a remarkable achievement with solution NMR spectroscopy as it is a membrane protein, besides its large molecular weight. The structure of VDAC has been solved recently by solution NMR spectroscopy (Hiller *et al.,* 2008; Villinger *et al.,* 2010). Challenges stems from the process of sample preparation, as it requires refolding procedure where a significant loss of protein was encountered. The molecular size of VDAC-micelles complex has been estimated to be around 90kDa (Hiller *et al.,* 2008), posing it as a larger protein than the Malate Synthase G, in addition to its unstable nature. While the selective isotope labeled amino acids helped partially for the resonance assignment, non-uniform sampling schemes with multi-domain decomposition and compression sensing techniques were essential for assignment. Apart from uniformly ^{15}N, ^{13}C and ^{2}H labeled VDAC-1 and ^{2}H labeling of LDAO, a sum of 23 selectively isotope-labeled amino acids incorporated samples was required in the structure determination.

Furthermore, paramagnetic labeling of detergents for mapping the residues lining the membrane interface was used as additional knowledge. In addition to the huge "wet-lab" efforts, the structure determination of VDAC-1 in LDAO miscelles was made possible with massive amount of ultra high field NMR time spent for acquiring sufficient quality triple resonance experiments.

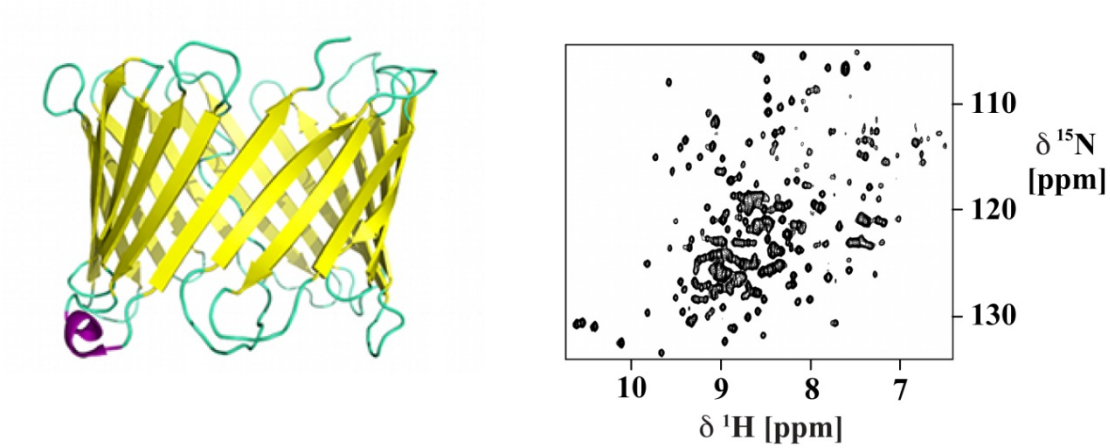


Figure-1.4: VDAC-1 in LDAO micelles: Cartoon representation three dimensional structure of Voltage Dependent Anion Channel – 1 as solved by the solution-state NMR (Hiller *et al.,* 2008) is shown with its fingerprint [^{1}H,^{15}N] – TROSY-HSQC spectrum. (spectrum adopted from (Hiller *et al.,* 2008) and used for the illustrative purpose)

1.2.5 Dengue virus NS2B-NS3 protease

Dengue virus NS3 protein plays a crucial role in viral replication (Matusan *et al.,* 2001). NS3 protein is a bifunctional enzyme, possessing protease activity at its N-terminal and helicase activity in its C-terminal end while NS2B cofactor is the requirement for the efficient activity of NS3 protease (Chambers *et al.,* 1991; Falgout *et al.,* 1991). Despite its relatively low molecular weight ~27 kDa, compared to the targets discussed above, NS2B-NS3pro complex, shown in Figure 1.5A manifests as a challenging target for NMR spectroscopy by its highly dynamic nature. The effect of conformational exchange was significant to an extent that in the absence of the inhibitor, the NMR spectrum of DENV NS3pro was unsuitable for resonance assignment, indicating the conformational exchange on the micro to millisecond scale as evidenced from the [^{1}H, ^{15}N]-HSQC spectrum shown in Figure 1.5C. However, complex formation with the inhibitor produced spectrum with better uniform peak intensities, as shown in Figure 1.5D besides better dispersion out of the crowded spectral region at the center, due to fact that C-terminal of NS2B wraps around NS3pro and overall adopts a closed conformation as depicted as cartoon representation in Figure 1.5B.

In addition, nearly 18 selective site-specific mutations were introduced, contributing ~20% to the overall backbone resonance assignment. Subsequently, two sets of combinatorially isotope labeled samples, first set with different combinations of ^{15}N labeled amino acids and the second set with different combinations of selected ^{15}N, ^{13}C amino acids and uniformly ^{15}N labeled remaining residues were used for the prupose of alleviating severe resonance overlap but with marginal success.

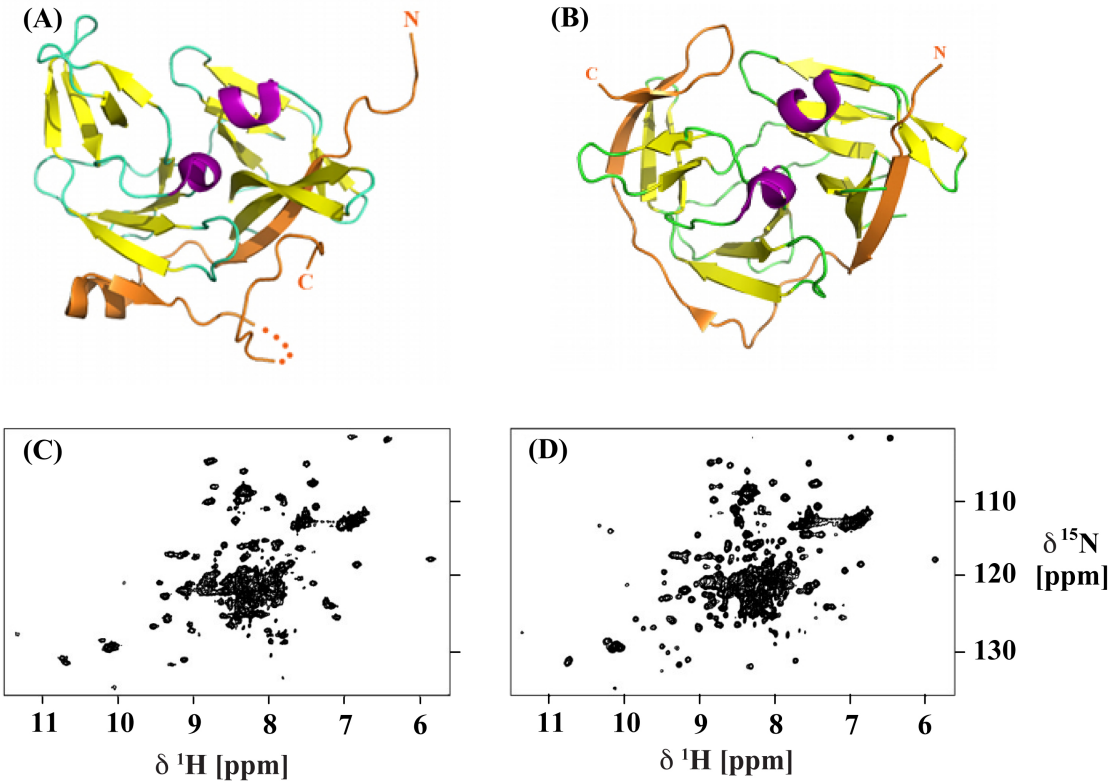


Figure 1.5: DENV NS2B-NS3pro complex: (A) Cartoon representation three-dimensional structure of open conformation NS3 protease-NS2B complex (PDB: 2FOM). NS2B is shown in orange with the tails annotated. The missing residues were indicated by dotted line. (B) Cartoon representation three-dimensional structure of closed conformation NS3 protease-NS2B complex (PDB: 2FP7). NS2B is shown in orange colour for clarity. (C) [^{1}H, ^{15}N]-HSQC spectrum of NS3 protease (without inhibitor). (D) [^{1}H, ^{15}N]-HSQC spectrum of NS3 protease with 5 fold excess of inhibitor. Spectra were adopted from (De La Cruz *et al.,* 2011).

In addition, the conformation of the NS2B-NS3pro has also been studied by attaching Lanthanide ion dipicholinic acid complex [$Ln(DPA)_3$] to the engineered binding site as reported by G. Otting and colleagues (Jia *et al.,* 2011). Despite the use of PCS for the resonance assignment and conformational analysis, PRE and PCS contributed to only about ~30% of the resonance assignment, emphasizing the complexity involved in the resonance assignment of proteins with highly dynamic nature. The scenario with NS3pro is much comparable with NS3 helicase, where the later case is further complicated by its challenging molecular size..

Table -1.1: NMR experiments and their parameters used for Malate Synthase G

Experiment	B_0 (MHz)	Dimension	Acquisition parameters									
			^{1}H			^{15}N			^{13}C			Time (hr)
			SW (Hz)	TD*	T_{acq}(ms)	SW (Hz)	TD*	T_{acq}(ms)	SW (Hz)	TD*	T_{acq}(ms)	
HNCACO	800	4D	12775.5	818	64	2268	40	17.2	CA: 3550	24	6.5	205
									CO: 2011	14	6.7	
HNCOCA	600	4D	9564.8	612	64	1700	40	22.9	CA: 2662.5	22	8.1	166
									CO: 1506	12	7.3	
$HNCO_{i-1}CA_i$	800	4D	12775.5	818	64	2268	40	17.2	CA: 3550	22	6.1	187
									CO: 2011	14	6.7	
[^{15}N,^{15}N] NOESY	800	4D	12775.5	818	64	^{15}N: 2268	32	13.9				183
						^{H}N: 3600	20	5.3				
						^{15}N: 2268	20	8.6				
HN(CO)CACB	600	3D	9000.9	576	64	1700	54	31.2	CACB:7380	35	4.7	109
HNCACB	800	3D	12775.5	818	64	2268	64	27.8	CACB:7803	40	5.1	157
HN(CA)CO	800	3D	12775.5	818	64	2268	64	27.8	CO: 2011	20	9.7	89
HNCO	600;800	3D	12775.5	818	64	2268	64	27.8	CO: 2011	28	13.7	26

Each dimension of every experiment, the following parameters are listed: type of nucleus evolved – ^{1}H, ^{15}N and ^{13}C, total number of acquired points - TD (* - complex points), sweep width in (**SW (Hz)**), and maximum acquisition time (**T_{acq}(ms)**). In addition, the spectrometer field (B0) and the total experimental acquisition time are listed for each experiment. 4D [^{1}H, ^{15}N]-TROSY-HNCACO, [^{1}H, ^{15}N]-TROSY-HNCOCA and [^{1}H, ^{15}N]-TROSY-$HNCO_{i-1}CA_i$ were recorded as

described by L. E. Kay and co-workers (Konrat *et al.,* 1999; Yang *et al.,* 1999b). The 4D [^{15}N, ^{15}N]-NOESY spectrum (mixing time of 90 ms) was recorded with the selection of the TROSY component in the ^{15}N and ^{1}H dimensions. Both 3D [^{1}H, ^{15}N]-TROSY-HN(CO)CACB and [^{1}H, ^{15}N]-TROSY-HNCACB experiments were recorded with non-constant-time ^{13}CB evolution periods. The ^{13}CA and ^{13}CB transfer delays were optimized for better sensitivity of ^{13}CB peaks and were set to 10 ms. For the 3D HN(CA)CO A ^{13}CA - ^{13}CO transfer period of 5.4 ms was used. 3D TROSY-HNCO experiment was recorded on both 600 and 800 MHz spectrometers. The parameters are listed for the 800 MHz experiment. Experiments HN(CO)CACB, HNCACB and HNCO were recorded with shared constant time evolution in ^{15}N dimension. Table adopted and modified from the publication by L. E. Kay and co-workers (Tugarinov *et al.,* 2002).

Table – 1.2: NMR experiments and parameters used for UDP-N-acetylenolpyruvylglucosamine reductase (MurB)

Experiment	B_0 (MHz)	Dimension	Acquisition parameters				Reference
			^{1}H	^{15}N	^{13}C	d1(s)	
			SW (Hz)	SW (Hz)	SW (Hz)		
HNCA	600	3D	11000	2000	CA: 5400	1.5	(Shan *et al.,* 1996)
HN(CA)CB	600	3D	11000	2000	CB: 9000	1.2	(Shan *et al.,* 1996)
HN(CO)CA	600	3D	11000	2000	CA: 5400	1.5	(Shan *et al.,* 1996)
HN(COCA)CB	600	3D	11000	2000	CB: 9000	1.4	(Shan *et al.,* 1996)
HNCO	600	3D	11000	2000	CO:1750	1.3	(Kay *et al.,* 1994)
HN(CA)CO	600	3D	11000	2000	CO:1750	1.4	(Matsuo *et al.,* 1996)
HBHA(CO)NH	600	3D	$^{1}H\alpha/\beta$: 4200 ^{1}H: 10000	1200	-	1.25	(Grzesiek *et al.,* 1993b)
^{15}N HSQC-NOESY	600	3D	$^{1}H^{N}$: 7700 ^{1}H: 11000	2000	-	2.4	(Farmer *et al.,* 1994)
HN(CACO)NH	600	4D	^{1}HN: 3850 ^{1}H: 11000	^{15}N: 1200 ^{15}N: 2000	CO: 2011	1.2	(Grzesiek *et al.,* 1993a)
^{15}N,^{15}N NOESY	600	4D	^{1}HN: 3850 ^{1}H: 11000	^{15}N: 1200 ^{15}N: 2000	-	1.4	(Farmer *et al.,* 1994)

For each dimension of every experiment, the following parameters are listed: type of nucleus evolved – ^{1}H, ^{15}N and ^{13}C. Sweep width and Longitudinal relaxation delay. Table generated with the data from the publication by K. L. Constantine and co-workers (Constantine *et al.,* 1997).

Table – 1.3: NMR experiments and parameters used for the backbone resonance assignment of Maltose Binding Protein

Experiment	B_0 (MHz)	Dimension	Acquisition Parameters									Time (hr)
			^{1}H			^{15}N			^{13}C			
			SW(Hz)	TD	T_{acq}(ms)	SW(Hz)	TD	T_{acq}(ms)	SW(Hz)	TD	T_{acq}(ms)	
CT-HNCA	600	3D	9000.9	576	64.0	1638.0	32	19.5	3921.6	92	23.5	51
CT-HN(CO)CA	600	3D	9000.9	576	64.0	1638.0	32	19.5	3921.6	92	23.5	58
CT-HN(CA)CB	600	3D	9000.9	576	64.0	1638.0	32	19.5	4199.9	92	21.9	115
CT-HN(COCACB)	600	3D	9000.9	576	64.0	1638.0	32	19.5	4199.9	92	21.9	52
HNCO	600	3D	9000.9	576	64.0	1638.0	32	19.5	1621.0	64	39.5	25
(HM)CMC(CM)HM	600	3D	9000.9	576	64.0	-	-	-	$^{13}C_{met}$: 3600.0	49	13.6	63
									$^{13}C_{att}$: 3600.0	89	24.7	
(H)C(CO)NH-TOCSY	600	3D	9000.9	576	64.0	1638.0	32	19.5	9178.5	64	7.0	71
(H)C(CA)NH-TOCSY	600	3D	9000.9	576	64.0	1638.0	32	19.5	8000.0	64	8.0	63
H(C)(CO)NH-TOCSY	600	3D	9000.9	576	64.0	1638.0	32	19.5	7500.5	64	8.5	71

Experimental parameters. For each dimension of every experiment, the following parameters are listed: total number of acquired points (*=complex points), sweep width (**SW(Hz)**), maximum acquisition time (**T_{acq}(ms)**). In addition, the total experimental acquisition time are listed for each experiment. Table adopted from the publication by L. E. Kay and co-workers (Gardner *et al.,* 1998).

Table – 1.4: NMR experiments and amino acid selective isotope labeled samples used for the structure of VDAC-1 Protein

Experiment	Dimension	Ref
TROSY	2D	(Pervushin *et al.,* 1997a)
TROSY-HNCA	3D	(Salzmann *et al.,* 1999)
TROSY-HNCO	3D	(Salzmann *et al.,* 1999)
TROSY-HNCOCA	3D	(Salzmann *et al.,* 1999)
TROSY-HNCACB	3D	(Salzmann *et al.,* 1999)
CT-TROSY-HNCA	3D	(Salzmann *et al.,* 1999)
[^{1}H,^{1}H]-NOESY-^{15}N-TROSY	3D	(Marion *et al.,* 1989; Zhu *et al.,* 2004)
[^{1}H,^{1}H]-NOESY-^{13}C-HMQC	3D	(Tugarinov *et al.,* 2004)
NUS-MDD-^{13}C-HMQC-[^{1}H,^{1}H]-NOESY-^{13}C-HMQC	4D	(Tugarinov *et al.,* 2005b)
NUS-MDD-^{15}N-HMQC-[^{1}H,^{1}H]-NOESY-^{13}C-HMQC	4D	(Hiller *et al.,* 2008)
[^{15}N,^{1}H]-TRACT	2D	(Lee *et al.,* 2006)
Samples	**Selectively isotope labeled amino acids**	
[U-99%- ^{2}H, ^{15}N]-VDAC-1 in LDAO		
[U-99%- ^{2}H, ^{13}C, ^{15}N]-VDAC-1		
[U-99%-^{2}H,^{13}C,^{15}N; 99%-^{1}Hδ-IL; 99%-^{1}Hγ-V]-VDAC-1		
[U-99%-^{2}H,^{15}N; 99%-^{1}Hδ,^{13}Cδ-IL; 99%-^{1}Hγ,^{13}Cγ-V]-VDAC-1		
[U-99%-^{2}H; 99%-^{15}N-Xxx]-VDAC-1	Xxx = Ala, Asn, Asp, Cys, Met, Gly + Ser, Ile, Leu, Lys, Phe, Thr, Tyr, Val	
[U-99%-^{2}H, ^{15}N; 99%-^{13}C - Yyy]-VDAC-1	Yyy = Ala, Ile, Leu, Lys, Phe, Tyr, Val, Pro	

All experiments were carried out at 30°C on Bruker spectrometers operating at field strengths of 500, 600, 750 and 900 MHz. All spectrometers were equipped with cryogenic probes. Two Non-Uniformly sampled spectra were recorded in 900MHz Spectrometer frequency. Table adopted from the publication by G. Wagner and co-workers (Hiller *et al.,* 2008).

Table 1.5: Summary of the strategies used and their percentile contributions to the total achieved backbone resonances assignment

Protein	Mol Wt (kDa)	Strategy				
		TRE[1]	Isotope Labelling[2]	PRE/PCS[3]	Mutagenesis[4]	Other experiments[5]
Malate Synthase G	81	~60%	~30%	-	-	Yes
UDP-N-Acetylenolpyruvylglucosamine reductase	38	~96%	-	-	-	-
Maltose Binding Protein	42	~95%	-	-	-	Yes
Voltage Dependent Anion Channel	32	~80%	Yes	Yes	-	Yes
DENV NS3 protease	27	~25%	Yes	~30%	Yes	-

[1] Triple Resonance Experiments with uniform [^{15}N, ^{13}C and ^{2}H] labeled samples
[2] Amino acid selective isotope labeling of protein samples
[3] Pramagnetic Relaxation Enhancement/PsuudoContact Shifts
[4] Site directed mutagenesis of selective amino acids
[5] Other experiments include Methyl-detected NMR, ^{13}C-detected NMR, various ligands binding, interactions with water, co-solvents (e.g., lipids) and various ligands including those used to increase thermodynamic stability of the sample.

A comprehensive summary of the challenging proteins and the strategies adopted/used for the resonance assignment and/or structure determination is listed in Table 1.5. Upon examination of the summary, it is evident that, for soluble proteins with high structural homogeneity, such as described in here Malate Synthase G, MurB and Maltose binding protein, TROSY based triple resonance NMR experiments in combination with uniform deuteration assisted with amino acid selective isotope labeling was sufficient for the resonance assignment and subsequent structure determination. On the contrary, proteins with inherently unstable nature due to their extensive intramolecular dynamics spanning several orders in the timescale manifested as conformational exchange-induced NMR line broadening. The sole use of TROSY-based triple resonance experiments using uniformly ^{2}H, ^{15}N and ^{13}C labeled proteins appear to be not sufficient for exhaustive backbone resonance assignment. Furthermore, such proteins require extensive number of selectively isotope labeled amino acids incorporated samples and paramagnetic spin labeling or the use of extensive mutagenesis.

DENV4 NS3 helicase, as discussed in Chapter-3, poses multiple of those challenges, including its low thermodynamic stability under conditions suitable for high resolution NMR that is further complicated by its high molecular weight. A combination of TROSY based triple resonance experiments with uniformly [^{2}H, ^{15}N and ^{13}C] labeled samples, selective amino acid-isotope labeling, paramagnetic relaxation enhancement and attempts to produce soluble and stable NS3 helicase fragments are discussed in Chapter-3.

1.3. Objective

1.3.1 Aims and scope of this thesis

It is evident from the discussion in the previous chapter that the solution NMR spectroscopy has been consistently evolving pushing its own limit with the introduction of novel NMR techniques and development of novel isotope labeling strategies. However, NMR studies of molecular structures beyond 50 kDa are still rare due to significant challenges encountered in NMR methodology, hardware requirements and financial costs. This calls for a careful resource allocation and reliable quantitative estimates of the expected outcome in terms of NMR spectral resolution and spectral signal-to-noise (SNR) ratios. We are confident that mathematical models capable of quantitative predicting the outcomes of planned experiments can make NMR structural work on high molecular weight entities and complexes a routine practice. This effort requires a complete and detailed analysis and evaluation of currently available NMR techniques for any given protein at hand. Such an analysis will enable not only NMR spectroscopists but also nonstructural biologists to carefully manage and allocate the available resources.

This thesis describes such an attempt to explore the limits and subsequently benchmark all the available solution NMR experiments on the unified basis. The information obtained thereon will enable us to predict the time and resources needed for any problem in hand. As the first step in exploring the performance of NMR experiments, a theoretical/analytical tool has been developed. Chapter-2 demonstrates the development and validation of a theoretical tool and formulation of unified framework for the uniform and systematic analysis and benchmarking of commonly available solution NMR experiments. A web-space for the deposition, analysis, sharing and automatic setup of these benchmarked NMR experiments in any spectrometer connected to the Internet has also been developed. Chapter-3 deals the challenges encountered and the strategies adopted for the resonance assignment of high molecular weight DENV4 NS3 helicase protein and the percentile of residues assigned. Finally, the Chapter-4 summarizes the results achieved in this work, and discusses the potential avenues for the future work.

Exploring the limits of NMR experiments: unified spin dynamics simulations, analysis and comparisons

2. Exploring the limits of NMR experiments: unified spin dynamics simulations, analysis and comparisons

2.1 Introduction

Historically, the NMR spectroscopy of large biological macromolecules is complicated by its low sensitivity, rapid transverse relaxation and spectral overlap. While the latter complications were addressed by experimental methods such as TROSY (Pervushin *et al.,* 1997a), CRIPT (Riek *et al.,* 1999) and hyper-dimensional NMR (Kay *et al.,* 1990) the sensitivity improvements are still continuing, since the molecular size remains to be a fundamental limitation: transverse relaxation time grows with the increasing effective molecular size. The sensitivity loss due to rapid transverse relaxation depends on the details of the trajectory of the density operator describing evolution of the entire coupled nuclear spin system stirred by the experimental radio frequency pulse sequence constituting an NMR experiment. Thus, a theoretical/computational tool that can efficiently describe the trajectory of the density operator under any pulse sequence would be of great application in studying, analyzing and benchmarking the NMR experiments. Such a tool has been developed and its abilities have been demonstrated in this section. Subsequently, a WIKI based online database has been created, depositing the commonly employed solution NMR experiments with their performance annotation.

2.1.1 NMR spin-dynamics simulation programs

Over the past, a number of computational approaches have evolved for the spin dynamics simulation, driven by the needs in the experimental NMR research (Bak *et al.,* 2000; Veshtort *et al.,* 2006). While the very first NMR program LAOCOON dated to as early as 1961 (Clark *et al.,* 1990; Gong *et al.,* 1995) a majority of programs developed so far are specialized in simulating NMR spectra (Ellison 1983; Clark *et al.,* 1990; Gong *et al.,* 1995) or a particular class of experiments (Schenker *et al.,* 1985; Kay *et al.,* 1986; Bock *et al.,* 1987). For the purpose of discussion the existing spin dynamics simulation programs can be broadly categorized as:

Analytical spin dynamics simulation programs

Numerical spin dynamics simulation programs

2.1.1.1 Analytical spin dynamics simulation programs

The first reported analytical spin dynamics tool is a FORTRAN-77 program: implementation of product operator simulations on a digital computer (Nakashima *et al.,* 1986). Following, a number of programs have been released addressing a specific need and/or experiments. Implementation of product operator rules based on product operator formalism (Richard R. Ernst Sep 1990) is the first available program in a modern computer: A *Mathematica notebook*[§] (John W 1991). The program performs product operator calculations and requires interactive typing of commands syntactically while lacking phase-cycling, pulsed field gradients and non-selective pulses.

The first complete and flexible implementation of computer manipulation of product operator expressions for the description of the spin dynamics under pulses and during chemical-shift and scalar coupling evolutions is a set of Maple procedures[‡] (Kanters *et al.,* 1993) closely followed by POMA (Guntert *et al.,* 1993) implementation of product operator rules for spin-½ nuclei in *Mathematica* (*www.wolfram.com*). While both of those programs provide analytical solutions for time evolution of weakly coupled spin-systems under the influence of rf-pulse, free precession delays and scalar coupling interactions, the former was equipped with spectra display options and the later incorporates the phase cycling future. While helpful in understanding spin dynamics under most pulse sequences, it is limited to coherent evolution of density operator of weakly coupled spins and very small systems.

MathNMR (Jerschow 2005) , the recent member under the analytical programs series, extending the limitations of POMA performs calculation of Redfield matrix elements, coherence filtering etc. It allows the calculations to be performed for any spin-system, containing spin-½ and quadrupolar spins. In a way, it remains as a complete analytical tool for the description and numerical spin dynamics simulation tool for the most pulse sequences in both solid-state and solution state NMR But, like its predecessors limited to very small spin-systems typically comprising less than 5 spins, besides failing to offer the numerical details.

[§] *Mathematica notebook* is a structured interactive document that can contain text, graphics, sound, calculations, typeset expressions, and user interface elements

[‡] *Maple procedure* is a portion of a code within a larger program that performs a specific task and is relatively independent of the remaining code.

In summary, analytical spin dynamics simulation packages, while helpful in understanding of most pulse sequences, limited to very small spin-systems and simple waveforms and thus become not suitable for performance evaluation and benchmarking of real complex solution NMR pulse sequences.

2.1.1.2 Numerical spin dynamics simulation programs

ANTIOPE (De Bouregas *et al.,* 1992), the first known general purpose numerical simulation program, focuses on the flexibility of experimentation and ease of use. While requiring spin-system and pulse sequence to be defined by the user notably, it requires no programming or theoretical skills, offering a consistent user interface and validation of user inputs. Results of the simulations could be plotted on screen, by inbuilt data manipulation and plotting modules. Recently, it acquired a user interface and undergone a significant changes in its implementation But, it simply computes the time evolution of single example of the spin system, opposing conventional experiments which superpose an enormous number of the same spin system each with its own orientation and sequence of random events. Thus, it is not efficient in simulating random processes and is not suitable for analysis and benchmarking of NMR experiments as proposed here.

GAMMA (Smith *et al.,* 1994); an object-oriented magnetic resonance tool kit, a strict super-set of C++ language encompassing abstract data types such as operators, superoperators and tensors, typically used in magnetic resonance simulations. Being a C++ library of classes and methods, it offers flexibility by allowing construction of various problem specific programs. The downside is, as with every other API type package, it requires significant programming skills from the user side. As a close analogue to GAMMA, BlochLib (Blanton 2003) extends to investigating the response of spin systems to complex interactions like radiation damping and field gradients. Relying heavily on optimized numerical libraries for its core calculations BlochLib has superior efficiency than GAMMA. But its improved efficiency comes with the cost of extensive programming knowledge from the user side. Despite these futures GAMMA and BlochLib consistently fails to simulate large spin systems, which can reasonably represent a globular protein encountered in practice.

SIMPLTN (Allman *et al.,* 1996) is designed to resemble a spectrometer, as much as possible, deliberately hiding the theoretical details of the spin dynamics and programming details. In contrast to GAMMA and BlochLib, it has its own compiler

that compiles the input pulse program inserting the correct values for pulses and delays and generates a sequence list for actual simulation. When most of the liquid-state experiments could be simulated by SIMPLTN, it doesn't account for dipolar relaxation and lacks selective pulses in its implementation, which is very common in modern day NMR experiments and thus, it becomes unsuitable for the objective proposed here.

Programs such as VNMRS (Nicholas *et al.,* 2000) and QSim (Helgstrand *et al.,* 2004) are capable of performing simulations employing pulse-field gradients, aside offering a user-friendly and intuitive graphical interface. Consistently, they fail to simulate spin-system of more than 5 spin ½.

Addressing the limitations and shortcomings posed by the above-mentioned packages, SPINEVOLUTION (Veshtort *et al.,* 2006) has been developed as a general spin dynamics simulation program utilizing the recent computational techniques and methodological approaches. Though it is tailor made for the simulations of solid-state MAS experiments, it could be employed for much of the solution-state experiments as well. Until recently, SPINEVOLUTION was the only program that can simulate more than ten spin-½ containing systems. While 10-spin ½ nuclei are enough for the description of a spin-system that reasonably represents the peptide backbone network, the complex effects of nuclear spin relaxation could not be realistically accounted.

2.1.2 Spin dynamics

Despite offering tailor made solution to specific problems and unique features, all the above-discussed spin dynamics simulation programs are consistently limited by the maximum number of spin ½ nuclei they can accommodate in their simulation. This is a consequence of very fact that the size of the density matrix scales exponentially with the increase in the spin ½ nuclei number in the spin-system under investigation.

Recently, the classical "exponential scaling wall" has been by-passed by the *adaptive state-space restriction* algorithm in association with zero track elimination and Krylov sub-space techniques (Kuprov *et al.,* 2007; Kuprov 2008). This enables the simulation of upto 200 spin ½ containing spin system to be simulated within the reach of a desktop-computer computational power (2.4GHz dual Core Intel CPU, 2GB DDR2 RAM).

2.2 NMR-RDB: a relational database of protein pulse sequences

First step for the analysis and subsequent benchmarking of the NMR experiments was a relational database, NMR-RDB, a XML-based portable database of NMR experiments and associated auxiliary data maintained outside NMR spectrometer was developed and interfaced with a novel software Spinach (Senthamarai *et al.,* 2010; Hogben *et al.,* 2011). The developed NMR-RDB interface to spinach algebraic core opens an avenue to simulate a variety of multi-channel, multi-dimensional NMR experiments employing phase cycles, simultaneous shaped pulses and composite pulse decoupling using a large generic spin system representing a fragment of protein. Full relaxation superoperator (Richard R. Ernst Sep 1990) including all cross-correlation terms is implemented in Spinach (Kumar *et al.,* 1996b) (Kumar *et al.,* 1996a). This development lays the foundation for creating a unified framework for performance annotation of a large number of the commonly employed NMR experiments with proteins in solution.

2.2.1 Functional architecture of NMR-DB and Spinach

The relational database is used to generate setup procedures for an NMR experiment or prepare input for Spinach. The block schematic is shown in Figure 2.1.

Relational Database of NMR experiments (RDB) is an XML formatted lightweight (stored as a single file) database grouping individual components of NMR experiments such as the pulse sequence, decoupling sequences, RF waveforms, calibration tables, sample-specific information and algorithms needed to set up and process NMR experiments using common attributes found in the data set. For example, all the files needed to setup a TROSY-HNCA experiment can be brought together under the corresponding experiment entry. Alternatively, these records can be annotated and grouped by attributes such as data type, contributing author, date of creation, URL reference, etc.

Relational database management system is implemented in the Lua application extension language (www.lua.org) as an extension script for the popular spectrum analysis program CARA (www.nmr.ch) and is itself stored in RDB making it self-contained. Since CARA is a single-file monolithic application (not requiring installation procedure) available for all common platforms, the RDB and management system are not constrained by the choice of platform. The RDB management system exports and imports all components of NMR experiments from and to spectrometer

data structure of TopSpin (Bruker Biospin AG), online database (www.trosy.com) or standalone files. For the selected spectrometer and sample, it generates macro performing experimental setup procedures at the spectrometer or semi-automatically translates the Bruker pulse programs and parameters to the Spinach representations for the numeric analysis. Due to access to the spectrometer calibration data, a semantic interpretation of the effects of RF pulses in terms of induced rotations as well as direct analysis of the performance of the experiments is easily available.

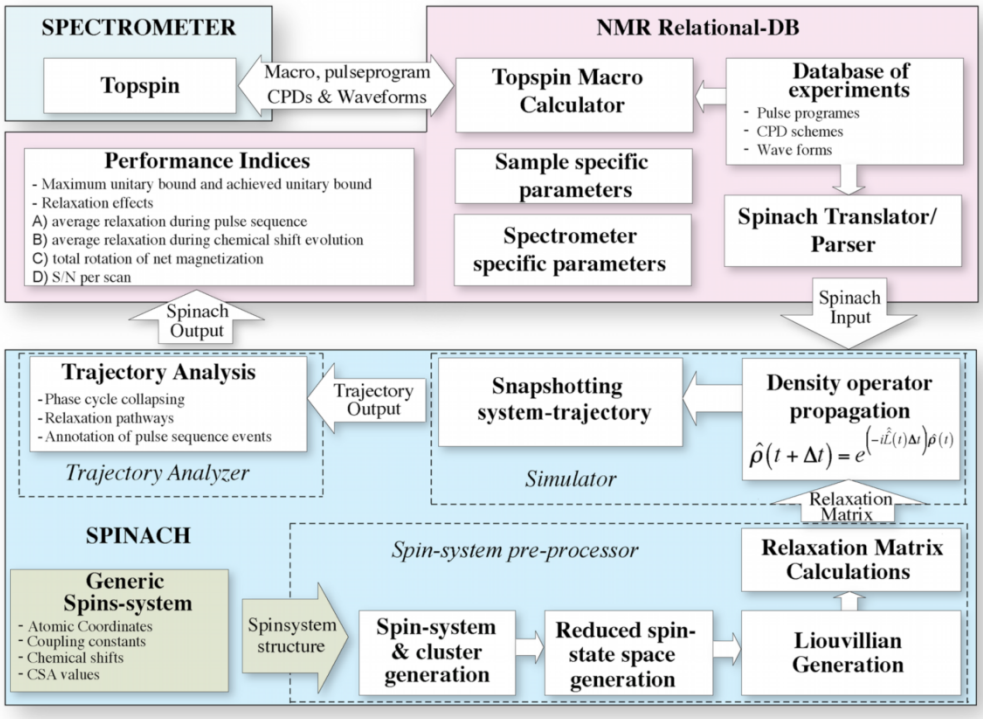


Figure 2.1: Functional architecture of Spinach-NMR-RDB: The flow of events involved in the setup and analysis of experiments deposited in the NMR-RDB. The arrows extend from the boxes, continuously, indicates, the procedure is automated and the arrows connected with the boxes indicates the need of manual intervention, despite the complete procedure could be automated with an external Matlab™ (www.mathworks.com) script. Adopted from K. Pervushin and co-workers (Senthamarai *et al*., 2010)

Spinach is functionally divided into three modules, Spin-system pre-processor, Simulator and Trajectory analyzer.

Spin system generator receives the description of coupled spin system as the input and generates the spin system data structure. This module builds the restricted basis set, subsequently generates the Liouvillian and the relaxation superoperator. Detailed procedures and algorithms of these steps are described by I. Kuprov (Kuprov 2008).

Simulator receives translated pulse sequence instructions in the form of spinach function calls and carries out those actions using the spin system structure obtained from the generator. At this level, the events such as delays and pulses can be quantized to trace evolution of the density operator with the desired time resolution. Snapshots of the density operator along its trajectory as well as auxiliary information such global time scale binding of the events, events durations, chemical shift evolution incremental cycles, phase cycles, quadrature schemes etc for the pulse sequence are stored for the subsequent analysis. In order to achieve this, a number of functions on top of the Spinach algebraic engine are implemented. Together with the snap-shotting of density operator, these functions enable monitoring the spin evolution/relaxation during the pulse sequence execution (Senthamarai *et al*., 2010).

Trajectory analyzer reads in the density operator trajectory and produces magnetization transfer graph upon collapsing the phase cycle. Pulse sequence events are tabulated and the norm of the density matrix and apparent relaxation rates along the productive pathways is outputted in conjunction with each pulse sequence events. In addition, a measure of deviation of the density operator from its relaxation-free trajectory is produced. In essence, this module semantically annotates the pulse sequence events and generates performance measures. The detailed description of algorithms used in the implementation of these modules has been discussed in the following section.

2.3 Algorithms and implementation

The state space restriction approximation adopted in Spinach is described here, in brief.

It has been noted that evolution of the density operator is confined in the restricted space of spin variables enabling computing evolution/relaxation superoperators for very large spin systems by means of adaptive state-space restriction[†] (Kuprov *et al.,* 2007; Richard R. Ernst Sep 1990). This is a typical situation in the analysis of solution NMR experiments with isotropically tumbling biomolecules. Unlike super-space, which is spanned by complete set of basis operators, the restricted spin-state space contains only the basis operators required to

[†] *A large number of states in big spin systems (very high multi spin-orders and orders linking the spins that are remote on the interaction graph), otherwise called as higher order coherences, are not essential in magnetic resonance simulations and can be safely dropped from the state-space*

describe evolution of coupled spin systems (Kuprov *et al.,* 2007; Kuprov 2008). Those basis operators comprising the restricted state-space are lower-order coherences and entanglements not higher than the number of J-couplings of individual spins. Thus, reduction in the size of the Liouvillian, by adopting adaptive state-space restriction, enables simulation of larger spin system, bypassing exponential scaling. Since, system trajectory, even in restricted state-space, at any particular time period confined to small Krylov sub-spaces (Kuprov 2008), the effective Liouvillian might still contain unpopulated dimensions, which are further contracted by Zero Track Elimination (ZTE) and/or Lanczos pruning procedures.

2.3.1 The density operator propagation

For the study of dynamics of any quantum mechanical system, the master evolution relaxation equation, of the following form, must be solved (Abragam 1961).

$$\frac{d}{dt}\hat{\rho} = -i[\mathrm{H},\hat{\rho}] - \hat{\hat{\Gamma}}\left(\hat{\rho} - \hat{\rho}_{eq}\right) \qquad 2.1$$

where H is spin Hamiltonian, $\hat{\hat{\Gamma}}$ is the relaxation superoperator and ρ_{eq} is the density operator at equilibrium. The inhomogeneous form of the Liouville-von Neumann equation can easily be rewritten in a homogeneous form when the unitary operator is included in the set of basis operators (Allard *et al.,* 1998). Thus Equation 2.1 becomes,

$$\frac{d}{dt}\hat{\rho}(t+\Delta t) = -\hat{\hat{L}}(t)\hat{\rho}(t) \qquad 2.2$$

where $\hat{\hat{L}}$ is the total homogeneous Liouvillian. Solving this equation, we obtain,

$$\hat{\rho}(t+\Delta t) = \exp\left(-i\hat{\hat{L}}(t)\Delta t\right)\hat{\rho}(t) \qquad 2.3$$

where $\rho(t)$ is the density matrix (including, if necessary, spatial degrees of freedom), $\hat{\hat{L}}(t)$is the Liouvillian superoperator (including relaxation and chemical kinetics contribution, if present) and Δt is kept sufficiently small. In most cases, temporal sequence of the NMR controls (rf-pulses, delays etc) can be compiled in to a sequence of matrices and the simulation can be carried forward by the repeated application of Equation 2.3 followed by projection onto the observable operator of interest (*A*).

$$A(t) = Tr\left(\hat{A}^{\dagger}\hat{\rho}(t)\right) \qquad 2.4$$

2.3.2 Simulation of spin relaxation of the density operator

Two implementations of relaxation superoperator are currently available – a local one, where relaxation superoperators are computed exactly in every spin cluster and then re-coupled (Kuprov *et al.,* 2007), and global, where the relaxation term between every pair of states in the restricted basis is calculated explicitly. The latter is based on the analysis of the direct product structure of individual states as described below.

The internal basis representation in Spinach is based on irreducible spherical tensor indexing. For every state, only the indices of its $\hat{T}_{l,m}$ components are stored. This provides a compact state description and permits the usage of a formidable body of analytical knowledge about the properties of such direct products[¶]. In the case of Bloch-Redfield-Wangsness relaxation superoperator (Goldman 2001) , it is known that $\hat{\hat{R}}$ has a very sparse structure – basis vectors differing in the state of more than three spins do not cross-relax under dipole and CSA interactions considered, and specific relaxation mechanisms have well-defined cross-relaxation paths. Thus, for each bilinear interaction Hamiltonian $\hat{H}_{ij}(t)$ connecting operators $\hat{T}_{l_i m_i}$ and $\hat{T}_{l_j m_j}$ a set of spin states may be identified with the members differing only in the operators $\hat{T}_{l_i m_i}$ and $\hat{T}_{l_j m_j}$ enabling compact tabulation of all relevant auto- and cross-relaxation rates for a generic pair of interacting spins computed. This table is used to rapidly calculate the relaxation superoperator (currently in the isotropic tumbling approximation) in the restricted state space. For each individual pair of spins, dipole-dipole interactions and the (axial) CSAs of each spin are included, thus enabling the simulation of the TROSY effect. Further computations of the cross-correlated cross-relaxation terms to DD and CSA interactions are restricted only to directly *J*-coupled spins. This significantly reduces the computation time by limiting the number of unique off-diagonal terms in the relaxation matrix. The computation is completed via iteration through all available DD and CSA interactions.

[¶] *Many routines in Spinach operate semi-analytically to take advantage of this*

2.3.3 Magnetization transfer graphs and relaxation profiles

In order to visualize the magnetization flow through the experiment, a pulse sequence timeline is built in Figure 2.2, with the nodes at each time point corresponding to the largest coefficients $c_i(t)$ in the expansion of the density operator in a chosen orthonormal basis $\{B_i(t)\}$ as shown,

$$\hat{\rho}(t) = \sum c_i(t) B_i \qquad 2.5$$

(time stepping is chosen to coincide with the major pulse sequence events). The node size is set proportional to $|c_i(t)|$. The coherence flux S_{ijk} between nodes *i* and *j* at the sequence state *k* is computed as,

$$S_{ijk} = \left(\left[\hat{\hat{U}}_k \hat{B}_i \right] \hat{B}_j^{\dagger} \right) \qquad 2.6$$

where $\hat{\hat{U}}_k$ is the total propagator for the sequence stage *k*. The display threshold for the absolute value of S_{ijk} was set to 0.03, which was found to provide a good balance between pathway importance and timeline clutter.

The resulting timelines have proven to be useful for visual and automatic identification of productive magnetization transfer pathways connecting the initial and target density operators and non-productive pathways resulting into the loss of magnetization. This separation of the density operator into productive and nonproductive parts at every stage in the pulse sequence enables construction of the coherence loss profile of NMR pulse sequences defined as an ordered list of the effective coherence loss rates r_k defined as,

$$r_k = \frac{\left|\left|\hat{\rho}_k^{prod}\right|\right| - \left|\left|\hat{\rho}_{k+1}^{prod}\right|\right|}{\Delta t_k} \qquad 2.7$$

where Δt_k is the duration of the *k*-th stage of the pulse sequence. With sufficiently detailed time quantization of the sequence, Equation 2.7 would approximate the time dependent rate of useful coherence loss $r(t) = \frac{d\left|\left|\hat{\rho}^{prod}(t)\right|\right|}{dt}$, so that the efficiency profile of the experiment can be constructed.

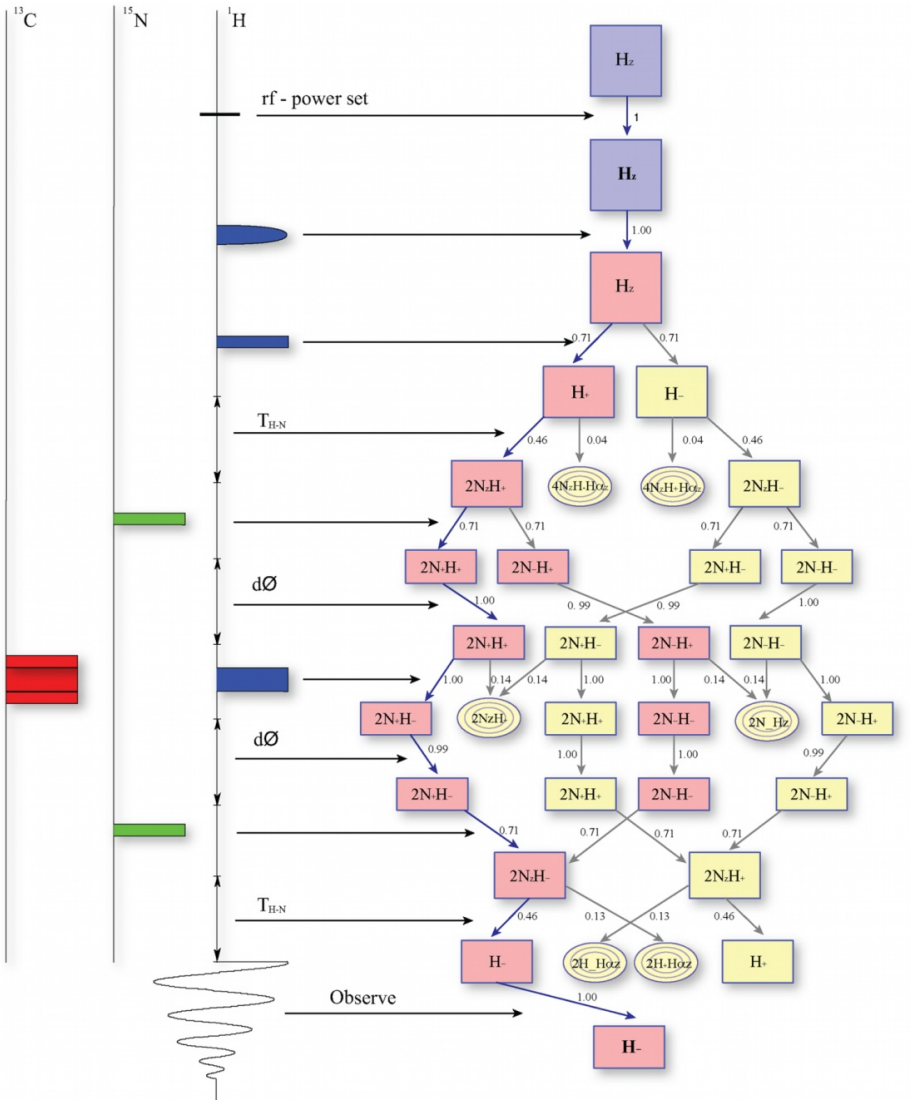


Figure 2.2: Time line of HMQC pulse sequence: Product operator annotation of HMQC pulse sequence is given. Operators carrying the magnetization to the final detectable operator are shown in magenta, while the operators shown in yellow do not end up in detection. Adopted from K. Pervushin and co-workers (Senthamarai *et al.,* 2010)

2.3.4 Performance metrics

The experiments deposited in NMR-RDB are performance annotated using the following parameters as the numerical quantum bounds (Glaser *et al.,* 1998; Pervushin *et al.,* 2005) in the absence $b_{\max}$ and in the presence $b^r_{\max}$ of relaxation.

$$b^r_{\max} = Tr\left(\hat{O}^{\dagger}\rho_{acq}\right) \qquad 2.8$$

where $\hat{O}$ is normalized observable operator and $\hat{\rho}_{acq}$ is the density operator at the beginning of the signal acquisition. To account for the quadrature detection scheme, these calculations can be performed twice for both the quadrature components.

One of the central performance measures computed with NMR-RDB/Spinach is the signal-to-noise ratio S/N in *n*D spectra in relation to S/N^{1D} in the reference 1D ^{1}H-detected spectrum measured with the same free induction decay (FID) acquisition parameters as the directly detected dimension of the *n*D spectrum. For simplicity, the following assumption has been made: all spectra are measured with one scan. If more scans are employed, e.g. *ns* > 1, then the resulting S/N are scaled up with $\sqrt{ns}$. In the first step, S/N^{2D} in two-dimensional orthogonal planes of the nD experiment are computed by time-domain sampling along the corresponding indirect dimension while keeping all other incremented delays to their minimum. Equation 2.9 provides the reconstructed one-dimensional slice along the indirect dimension of the 2D plane normalized to the amplitude of the reference 1D ^{1}H spectrum,

$$S(\omega) = F_+\left[\left(f_{tr}(t)e^{-R_{eff}t} + n(t)\right)f_{apod}(t)\right] \qquad 2.9$$

where F_+ is a discrete Fourier transform, R_{eff} is the effective relaxation rate defined above, f_{apod} is the apodization function and $n(t)$ is Gaussian noise with standard deviation typically set to 10% of $f(t)$ (see below), and *t* is the incremented time-delay in the indirectly sampled dimension. The normalized transfer function is then defined as

$$f_{tr}(t) = \left|\frac{Tr\left(\hat{O}^{\dagger}\hat{\rho}(t)\right)}{\max\left[\left|Tr\left(\hat{O}^{\dagger}\hat{\rho}(t)\right)\right|\right]}\right| \qquad 2.10$$

The simulations are performed for each component of the Cartesian orthogonal or echo-anti-echo quadrature-sampling scheme thus resulting in the complex, phase-sensitive spectrum. To compute S/N^{2D}_{H-X} 10% of Gaussian noise $n(t)$ is added to the time domain signal prior to apodization and Fourier transform, which corresponds to the signal-to-noise ratio in the reference 1D spectrum of 10:1. The parameters S/N^{2D}_{H-X} and $\Delta\nu_{1/2}$ are directly estimated from the simulated spectrum as shown in Figure 2.3. Due to the added random noise, the simulations are repeated 200 times to report the mean and RMSD values.

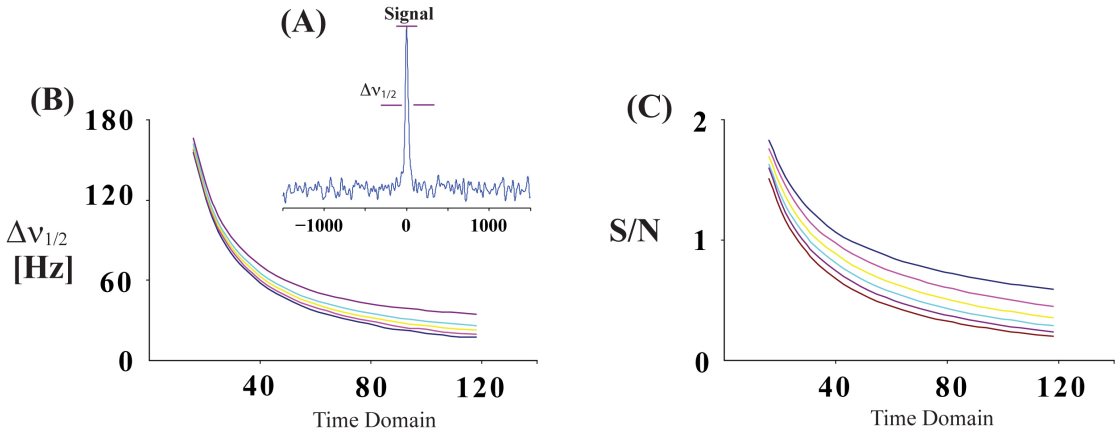


Figure 2.3: Spinach performance measures: (A) Figure shows simulated signal to noise simulation in TROSY experiment in the t1 indirectly detected dimension. (B) Linewidth (Full Width at Half Maxima) as a function of TD points for a range of relaxation rates. (C) S/N per unit time as a function of TD points for a range of B0. The differently colored lines correspond to simulations with the effective relaxation rate of chemical shift encoding coherence set to 10 s^{-1} till 60 s^{-1} with the step of 10 s^{-1}.

The sensitivity in nD spectrum is computed according to Equation 2.11

$$\left(S/N\right)^{nD} = b_{\max}^{r} \prod_{i=1}^{n-1} S/N_{H-X_i}^{2D} \qquad 2.11$$

where $b_{\max}^{r}$ is the maximum magnetization transfer bound in the presence of relaxation and the product is performed over the number of indirectly detected dimensions. For example, S/N^{3D} in the 3D HNCA experiment will be a product of S/N_{H-N}^{2D} and $S/N_{H-C\alpha}^{2D}$ spectral planes. Finally, an assumption was made that the same steady state level of initial spin polarization is generated for all experiments, thus avoiding the need for explicit modeling of the return of the spin polarization back to its thermal equilibrium.

2.4 Results and discussion

2.4.1 Construction of a generic spin system

The key development in the unified benchmarking framework described above is the ability to use a single generic spin system to adequately simulate the majority of solution state NMR experiments deposited in RDB. The use of the single spin system provides a uniform framework for performance annotation across majority of known pulse sequences. For the construction of such a system, a fragment from NS3 helicase (PDB ID: 2JLQ.pdb) comprising 27 *J*-coupled spins from residues Gln-384 to Ser-386 as well as five spatially proximate protons belonging to other residues, was adopted. Atomic coordinates and chemical shielding tensor values of each atom constitutes the generic spin-system can be found in Appendix-A. The positions of these protons were slightly adjusted to maintain the dipolar density at the backbone spins typically encountered in compact globular proteins Appendix-B reports the adjusted dipolar densities in the generated spin system and averaged ^{1}H-^{1}H dipolar densities found in NS3 helicase (PDB: 2JLQ.pdb). For perdeuterated proteins, all aliphatic ^{1}H are replaced with ^{2}H isotopes. At the time of writing, a typical calculation takes about 1 to 2 hours on an average workstation (2.4GHz dual core CPU, 2GB DDR2 RAM).

The values of coupling constants are adopted from NMRGuide 3.5 (Bruker Biospin AG). In cases of conformation-dependent coupling constants (V.F 1976), e.g. $^{3}J_{\mathrm{HNH}\alpha}$, a single value corresponding to the middle of the range is set. Chemical shielding tensor values of Cα, Cβ, carbonyl C, backbone N and amide protons were adopted from E. Oldfield and co-workers and G. Bodenhausen and co-workers (Sun *et al.,* 2002; Loth *et al.,* 2005) and generic chemical shielding values were used for the rest of the spins. The chemical shifts of all spins were set to the averaged values reported for the restricted set of diamagnetic proteins in the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu). Figure 2.4A shows a flattened and 3D representation of the constructed spin system with the coupling constants shown in red.

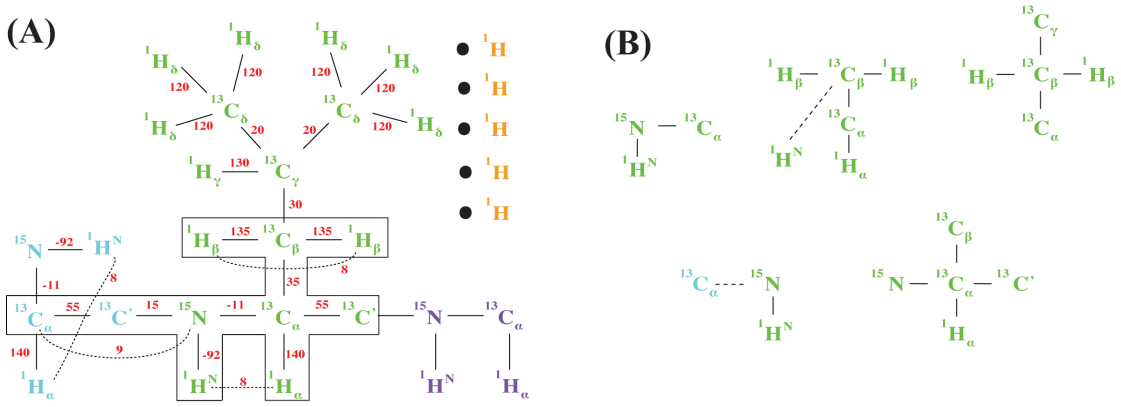


Figure 2.4: Spin-system: (A) Flattened 2D picture of the generic protein spin system used for sequence benchmarking. Atomic coordinates were taken from NS3 helicase (2JLQ.pdb). Typical coupling constants (in Hz) are shown in red. Dotted lines indicates couplings between indirectly connected spins (B) Selected clusters for C_α spin constructed by inspection of the magnetization transfer across C_α spin as described in Methods. Dotted line represents indirect connectivity between C_α and $^1H^N$. Adopted from (Senthamarai *et al*., 2010)

In order to construct the restricted spin state space, the coupled spins are grouped together to form clusters comprising spins entangled into highest spin orders excited in pulse sequences. Each cluster size is kept to minimum and for computational efficiency does not exceed 6. The main cluster set is automatically generated by the JNES algorithm (Kuprov *et al*., 2007) and is manually extended after inspection of propagation of highest order coherences encountered in triple resonance experiments. It is important to note that clusterization is a *basis set generation strategy* -– the spin system is always simulated in its entirety in the resulting connectivity-adapted basis set (Kuprov *et al*., 2007). Figure 2.4B depicts the clusters of importance in magnetization transfer via C_α as a hub.

2.4.2 Performance annotation and analysis of TROSY pulse sequences

Two experimental schemes of TROSY experiments from NMR-RDB employing ST2-PT (Pervushin *et al.,* 1998b) and improved coherence transfer elements (CTE) (Yang *et al.,* 1999a) (Nietlispach 2005) have been simulated using NMR-RDB/Spinach framework with the generic spin system shown in Figure 2.4A. The density operator trajectory analysis including schematic representation of the pulse sequences are shown in Figure 2.5 (note that ST2-PT and improved CTE differ by pulses marked with asterisks).

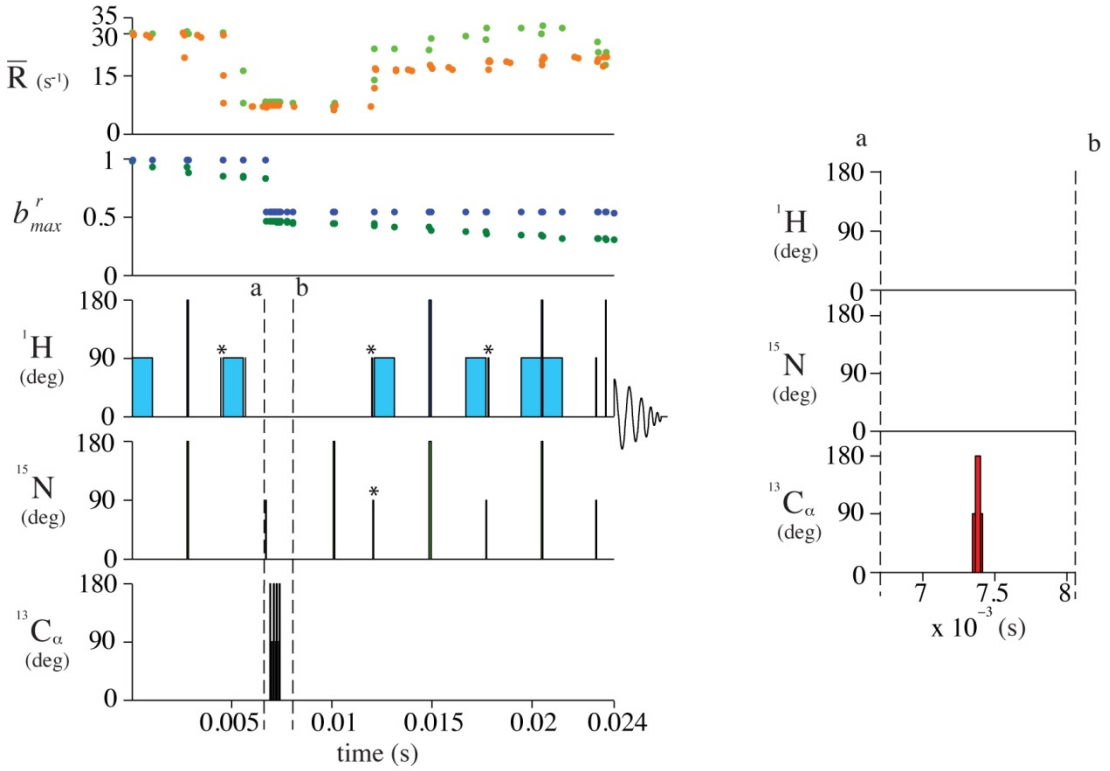


Figure 2.5: Performance annotation of the TROSY pulse sequences: Chemical shift evolution, events between (a) and (b), is shown, simplified, at the bottom. The Chemical shift offsets are as follows: $\omega(^1H) = 4.7$ ppm, $\omega(^{15}N) = 119$ ppm and $\omega(^{13}C) = 110$ ppm. The rest of the parameters and the Spinach input script can be found in NMR-RDB online at www.trosy.com. Spectrometer frequency was set to 700MHz. Chemical shift evolution is simulated for 4 t1-time incremental points. Pulse sequence events, amplitudes and relaxation rates were plotted versus time (milliseconds). Pulses on each channel are represented by bars with the height of the bar specifying the total angle of rotation and the width of the bar marks the pulse duration (some pulses are short enough to appear as thin lines). The norm of the density operator along the productive magnetization pathways, with (green) and without relaxation (blue) was plotted versus time. At the top panel, the relaxation rates, at the end of the corresponding events, are plotted in green and orange, corresponding to ST2-PT and CTE pulse sequence events. Adopted from (Senthamarai *et al.,* 2010)

In order to monitor the norm of the density operator and the apparent relaxation rates as a function of time, the pulse program events were plotted as a function of time. To be consistent, the vertical scale was used to semantically annotate the rotation of magnetization caused by RF pulses, wherein the height and the width represent the total angle of rotation and the event duration on the common time grid, respectively. In the corresponding magnetization transfer graph for ST2-PT highlighting the productive and the non-productive magnetization transfer pathways are highlighted Figure 2.6 & Figure 2.7.

For both experiments the initial density operator represents the Boltzmann equilibrium magnetization on $^1H^N$ and ^{15}N spins as it is shown in the magnetization transfer graph of Figure 2.7. In the absence of relaxation, only ½ of the norm of the density operator is retained after the ^{15}N transverse coherence $2N_z(E/2+H_z)$ is excited

by the first ^{15}N 90^{o} pulse due to selection of the TROSY single transition operator by the phase cycle. The Anti-TROSY transition is not transferred to the detectable magnetization and, thus, falls to nonproductive operators, which generate no detectable NMR signal, highlighted in yellow Figure 2.6. From the simulated relaxation profile, it is clear that the main relaxation losses stem from the three coherence transfer INEPTS, so that the improved CTE (Yang *et al.,* 1999a; Nietlispach 2005) contributes to higher sensitivity by providing the reduced apparent relaxation rates. However, since the improved CTE ends with a ^{1}H 90^{o} pulse, it requires additional delays for flipping water magnetization back to +z direction, which is usually combined with the echo-anti-echo PFG-based quadrature encoding scheme resulting in overall longer pulse sequence. Since ST2-PT TROSY (Pervushin *et al.,* 1998b) does not require such extra delays, it effectively delivers similar signal-to-noise.

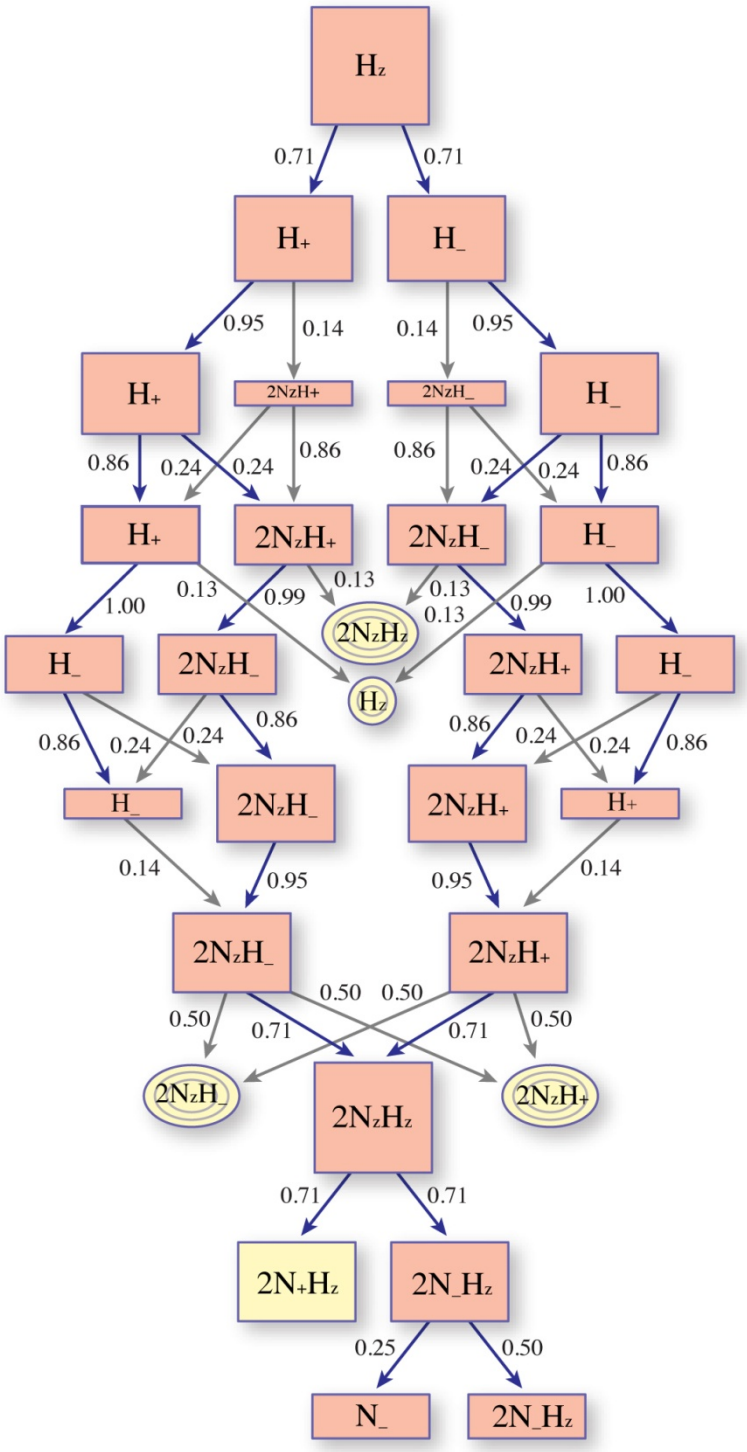


Figure 2.6: Magnetization flow graph of TROSY pulse sequence: Automatically generated flow of magnetization chart during the first INEPT block of TROSY-ST2-PT. Size of each box containing spin state is proportional to the magnetization transferred via that particular spin state during the corresponding event in the pulse sequence. Blue lines represent the strong magnetization flow channels. Productive spin-states are shown in magenta and non-productive are shown in yellow. Adopted from K. Pervushin and co-workers (Senthamarai *et al.,* 2010)

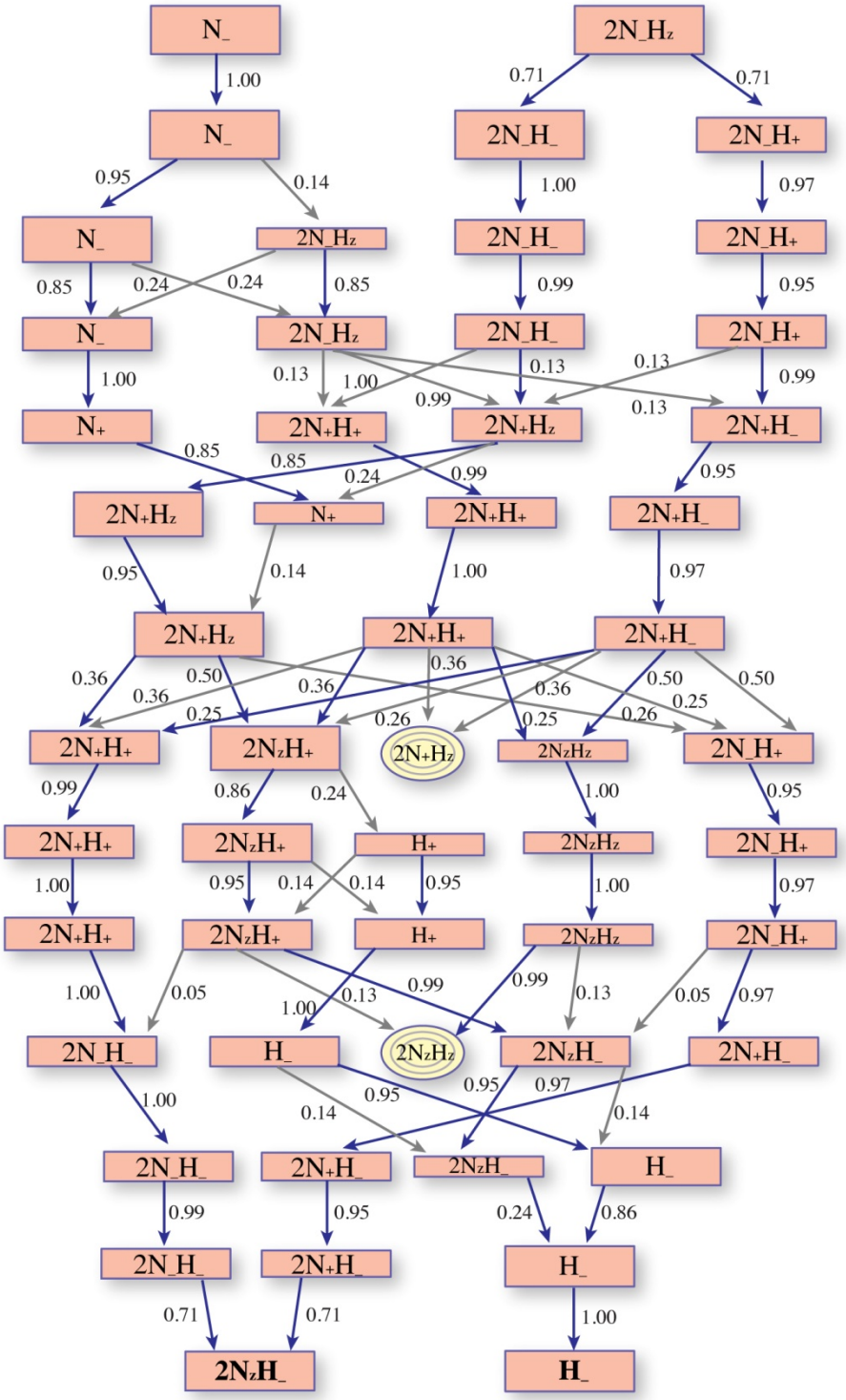


Figure 2.7: Magnetization flow graph of TROSY pulse sequence: Automatically generated flow of magnetization chart during the ST2-PT element of TROSY-ST2-PT. Size of each box containing spin state is proportional to the magnetization transferred via that particular spin state during the corresponding event in the pulse sequence. Blue lines represent the strong magnetization flow channels. Productive spin-states are shown in magenta and non-productive are shown in yellow. Adopted from K. Pervushin and co-workers (Senthamarai *et al*., 2010)

2.4.3 Sensitivity of multidimensional experiments

In order to more systematically compare performance of widely used 2D $^{1}H,^{15}N$-correlation experiments, the performance measures and experimentally measured S/N^{2D} in series of 2D spectra using two test proteins, 25 kDa $^{13}C,^{15}N$-labeled αXβ2 Integrin I domain, and 51 kDa $^{2}H,^{13}C,^{15}N$-labeled NS3 helicase from *Dengue* virus in a complex with 14 nucleotide single stranded RNA (Luo *et al.,* 2008b), were calculated and listed in Table 2.1a & 2.1b, repsectively. The rotational correlation times of 20 ns and 40 ns for αX-I and NS3 proteins, respectively, were estimated using protein solvent accessible surface area as well as known 3D shapes (Krishnan *et al.,* 1998).

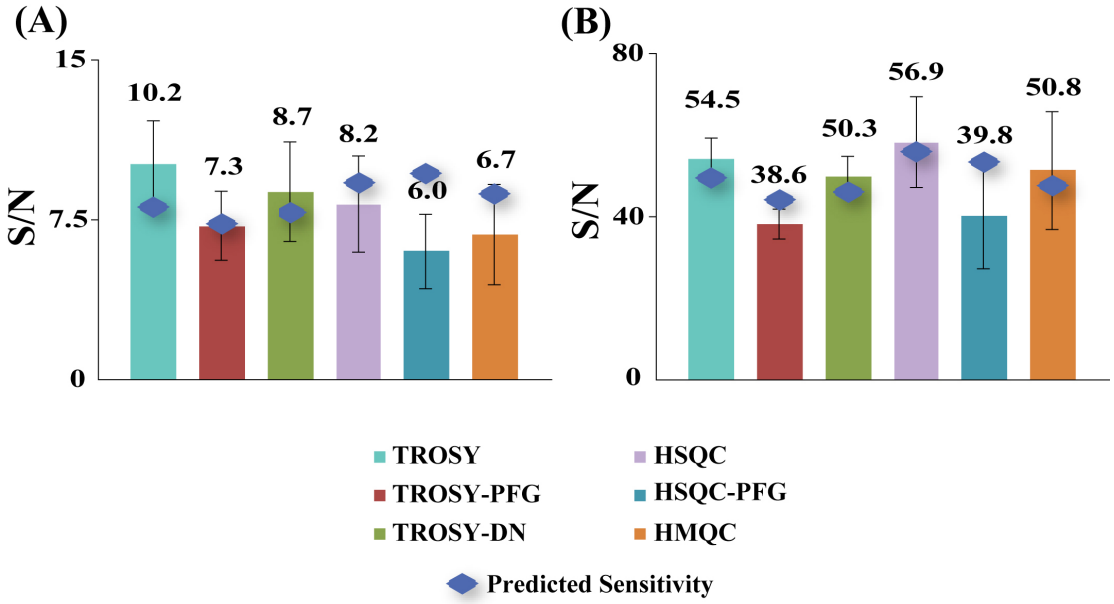


Figure 2.8: Predicted vs experimental S/N between a set of 2D $^{1}H,^{15}N$-correlation experiments: (A), $^{2}H,^{13}C,^{15}N$-labeled 51 kDa NS3 helicase in complex with RNA and (B), $^{13}C,^{15}N$-labeled 25 kDa αXβ2 Integrin I domain. A Value above each bar represents average S/N of defined number of peaks, common in all the spectra. Acquisition parameters: (A) number of scans = 8, number of complex time domain points sampled in the indirect dimension = 128 and spectral width (^{15}N) = 2838 Hz. (B) number of scans = 4, number of complex time domain points sampled in the indirect dimension = 100 and spectral width (^{15}N) = 2980 Hz. The predicted S/N^{2D} are calculated with Equation 2.11 using data of Table-2.1a&b and S/N^{1D} = 8.0 and S/N^{1D} = 2.0 for αXβ2 Integrin I domain and NS3 helicase, respectively. Adopted from (Senthamarai *et al.,* 2010)

Figure 2.8 shows S/N^{2D} in series of 2D $^{1}H,^{15}N$-correlation spectra and compares it with predicted *S/N*. Overall, a good agreement between predicted and experimental S/N is observed with notable exception of HSQC-PEP-PFG experiment for both proteins. This underperformance of the latter experiment has been traced back to saturation of water signal due to lack of water-flip-back (Piotto *et al.,* 1992) in our

implementation. Although the aim of the performance annotation of RDB is to predict signal-to-noise in nD spectra from seeding lower dimensional spectra, here the 2D correlation spectra was used as the reference and S/N^{1D} in the corresponding seeding 1D spectra was back-predicted. S/N^{1D} = 8.0 and S/N^{1D} = 2.0 for αXβ2 Integrin I domain and NS3 helicase, respectively. These values closely correspond to the experimentally observed *S/N* in the seeding 1D ^{1}H spectra, albeit the direct comparison is hampered by the absence of well-resolved individual protons in 1D spectrum.

Table-2.1a. Simulated performance of 2D ^{1}H-^{15}N correlation experiments for uniformly ^{13}C,^{15}N-labeled protein (αXβ2 Integrin I domain) with τ_c = 20 ns.

Experiments[1]	Performance Metrics[2]					
	[3] b_{max}/b^{r}_{max}	[4] T (ms)	[5] R_{PP} (s^{-1})	[6] R_{CS} (s^{-1})	[7] $\Delta\nu_{1/2}$ ^{15}N(Hz)	[8] S/N^{2D}
HMQC	0.69/0.49	12.7	30.5	36.0	20.8	19.6
HSQC	0.67/0.46	13.2	28.5	26.0	18.4	22.5
HSQC-PEP-PFG	0.69/0.32	26.5	26.0	26.0	18.4	22.1
TROSY	0.38/0.23	17.5	28.0	9.5	16.2	20.3
TROSY-PFG	0.38/0.21	22.8	25.0	9.5	16.2	18.4
TROSY-D.N	0.38/0.22	23.9	20.2	9.5	16.2	19.3

Table-2.1b. Simulated performance of 2D ^{1}H-^{15}N correlation experiments for uniformly ^{2}H,^{13}C,^{15}N-labeled protein (NS3 helicase) with τ_c = 40 ns.

Experiments[1]	Performance Metrics[2]					
	[3] b_{max}/b^{r}_{max}	[4] T (ms)	[5] R_{PP} (s^{-1})	[6] R_{CS} (s^{-1})	[7] $\Delta\nu_{1/2}$ ^{15}N(Hz)	[8] S/N^{2D}
HMQC	0.69/0.52	12.7	22.0	33.0	18.0	21.6
HSQC	0.67/0.52	13.2	19.2	32.4	18.0	23.3
HSQC-PEP-PFG	0.69/0.37	26.5	21.2	32.4	18.0	24.1
TROSY	0.38/0.24	17.5	25.6	10.6	15.2	20.7
TROSY-PFG	0.38/0.22	22.8	22.4	10.6	15.2	19.5
TROSY-D.N	0.38/0.24	23.9	19.0	10.6	15.2	20.7

[1]Experiments: HMQC(Mueller 1979) , HSQC(Bodenhausen *et al.,* 1980) , HSQC-PEP-PFG (Cavanagh *et al.,* 1991) (Palmer Iii *et al.,* 1991) (Kay *et al.,* 1992) , TROSY(Pervushin *et al.,* 1998b) , TROSY-PFG(Pervushin *et al.,* 1998b) , TROSY-D.N.(Nietlispach 2005) .

[2]In t_1 dimension 128 complex increments with spectral width sw(^{15}N) = 2980 Hz at B_0 = 700 MHz are simulated.

[3]Maximum transfer bounds computed in the absence and presence of relaxation.

[4]Time duration of the pulse sequence with chemical shift evolution periods set to the minimal values.

[5]Effective relaxation rate of the entire experiment with chemical shift evolution periods set to the minimal values.

[6]Effective relaxation rates for the coherences used to indirectly encode chemical shifts.

[7]Expected line-width in the indirectly detected dimension.

[8]Signal-to-noise ratio expected in 2D spectrum in the presence of relaxation with S/N =10 is the seeding 1D ^{1}H spectrum.

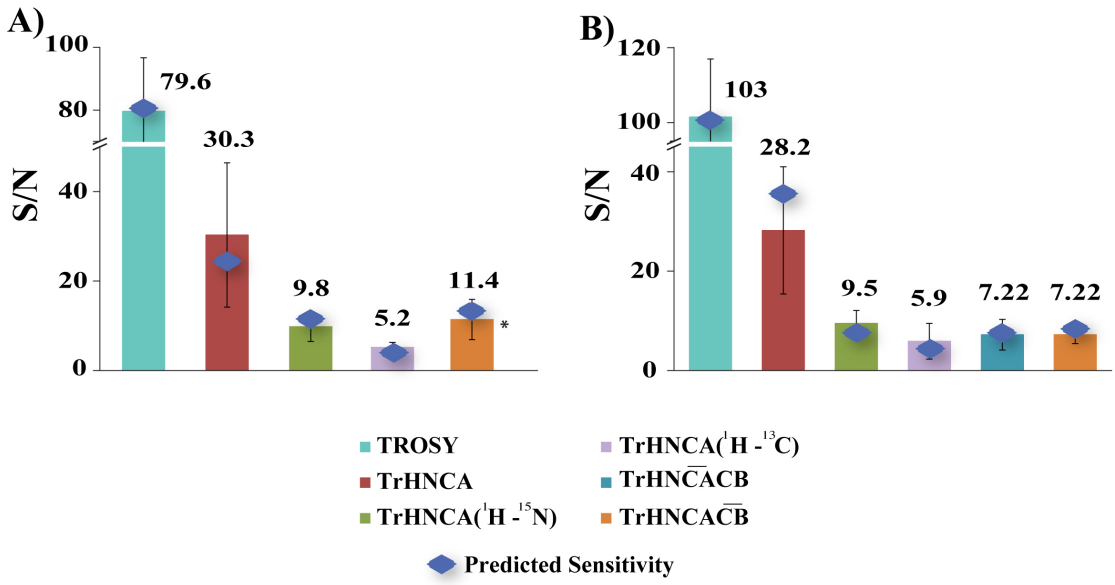


Figure 2.9: Sensitivity prediction across dimension: Comparison of predicted and experimental S/N in 2D [^{1}H,^{15}N]-TROSY, 2D [^{1}H,^{13}C]- and [^{1}H, ^{15}N]- planes of TROSY-HNCA, TROSY-HNCACAB, measured with (A) ^{2}H,^{13}C,^{15}N-labeled 51 kDa NS3 helicase and (B) ^{13}C,^{15}N-labeled 25 kDa αXβ2 Integrin I domain. A value above each bar represents the average S/N of those assigned peaks. The predicted S/N^{2D} are calculated with Equation 2.11 using data of Appendix-C and S/N^{1D} = 12.5 and S/N^{1D} = 13.7 for αXβ2 Integrin I domain and NS3 helicase, respectively. Note, that the used samples are different from those reported in Figure 2.8 (* T_{CACB} = 14.4 ms; full magnetization transfer from C_{α} to C_{β}). Adopted from (Senthamarai *et al.,* 2010)

To predict S/N in TROSY series of 2D and 3D experiments measured with both proteins, the corresponding rotational correlation times τ_c and S/N^{1D} in the seeding 1D ^{1}H spectra were used as inputs. The corresponding experimental parameters as well as calculated performance measures are listed in Appendix – C for NS3 helicase and αXβ2 Integrin I domain, respectively. Comparison between predicted and experimental S/N in 2D TROSY and 3D HNCA and 3D HNCACB as well as in individual 2D [^{1}H, ^{13}C] and [^{1}H, ^{15}N] shows that the sensitivities can be reasonably predicted thus demonstrating viability of the proposed benchmarking of NMR experiments and shown in Figure 2.9.

2.4.4 WIKI based online NMR-DB for the setup, execution, sharing and analytics of complex NMR experiments

An online database-NMR-RDB, have been created, depositing all the available solution NMR experiments. Each entry in the database encapsulates all the necessary auxiliary information such as, CPD schemes, waveforms, and calibration tables and thus forms the functional unit of the NMR-RDB as depicted in Figure 2.10.

(A)

(B)

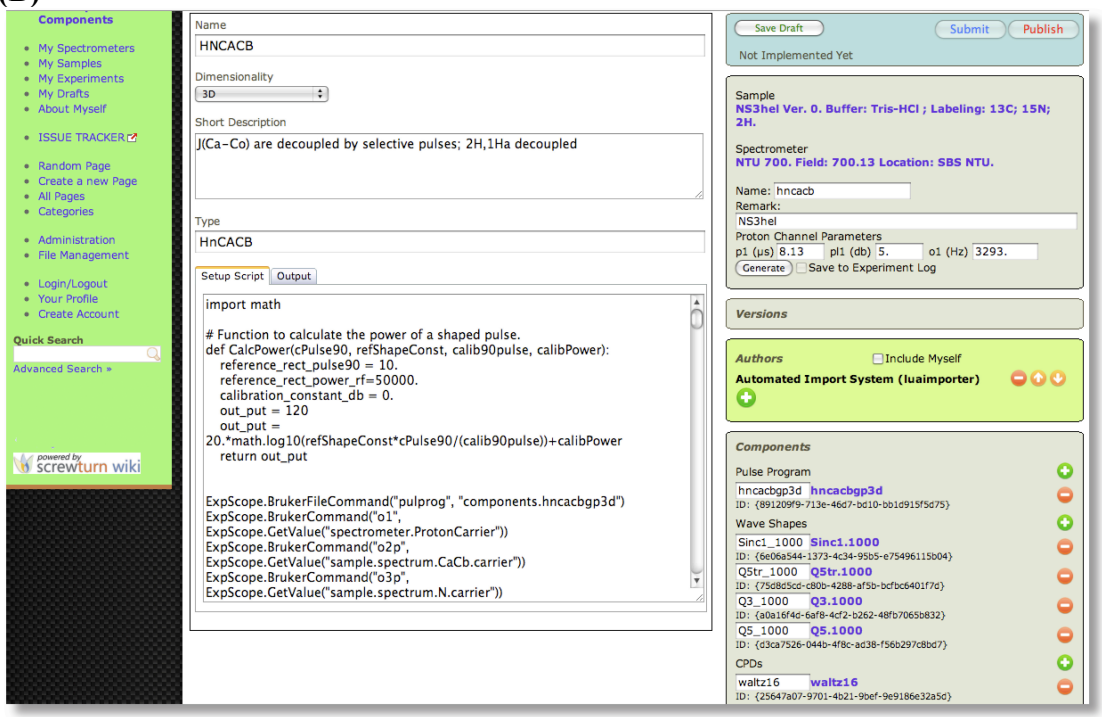


Figure 2.10: Screenshot of NMR-DB: (A) User interface screenshot of the NMR-DB showing the experimental entries. Each deposited experiment forms the functional unit of the NMR-DB encapsulating pulse program, wave shapes, decoupling schemes and algorithms for automated setup in spectromter running Topspin. (B) User interface screenshot showing a HNCACB experiment as example.

The interesting and unique feature of this online WIKI is that its ability to automatically setup any solution NMR experiments optimally (in an algorithmic manner) in any NMR spectrometer around the globe. Further, the developed Spinach library has also been interfaced with the NMR-DB offering direct simulation of NMR experiments and graphical visualization of the experimental analysis. A schematic functional architecture of the NMR-DB interfaced with spinach library and Spectrometer running Topspin is given in Figure 2.11.

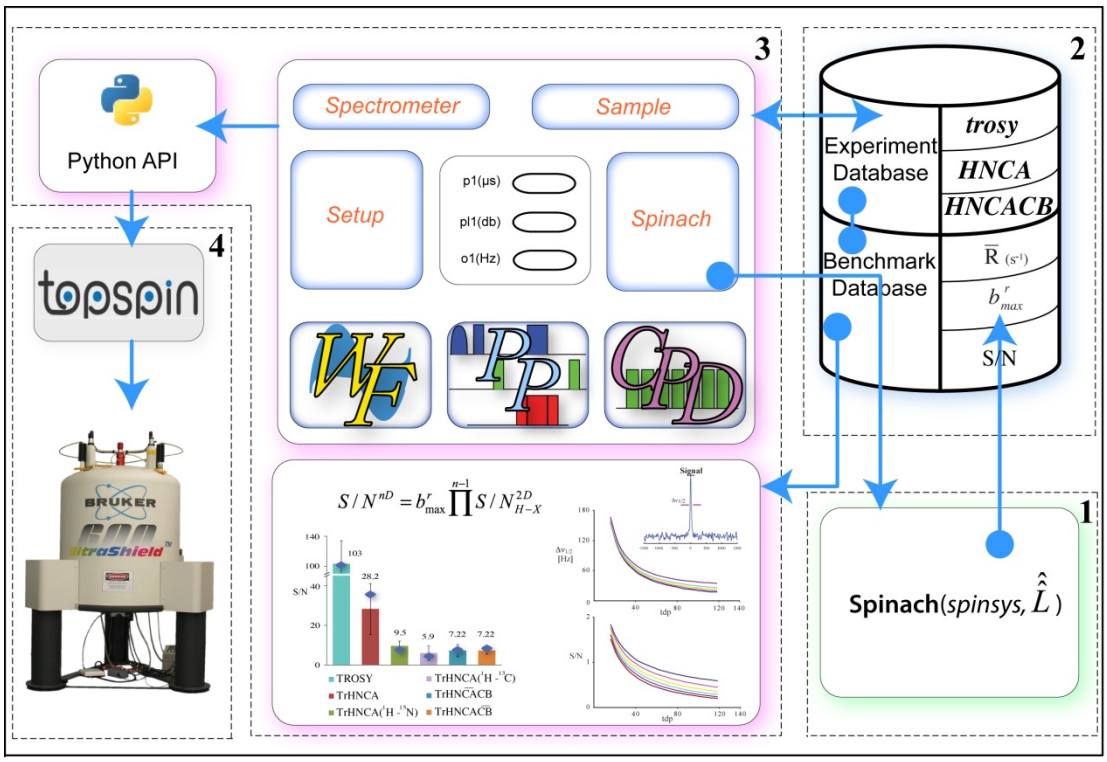


Figure 2.11: Functional architecture of NMR-DB: Schematic showing the functional architecture of the online NMR-DB. (1) The spinach library module for the simulation of NMR experiments. (2) The relational database, consisting the experiment entries and their associated performance metrics and values. (3) The user interface to the NMR-DB, which allows the user interaction for the supply of input parameters and the visualization of benchmarking results. (4) NMR spectrometer running Topspin™. The arrowheads indicate the directionality and the flow of information. The bold circles indicate the interconnection between functionally independent modules.

The WIKI based NMR-DB provides supportive information on formal description of NMR experiments, their analysis and the results. Thus serves as a central place not only for NMR spectroscopists, but also for the nonstructural biologists interested in NMR analysis of their proteins of interests.

DENV4 NS3 helicase: backbone resonance assignment

3. DENV4 NS3 helicase: backbone resonance assignment

3.1 Introduction

The large size of many protein complexes however still represents a challenge for NMR studies. With increasing molecular weight, a number of problems arise that limit the sensitivity and resolution of the NMR spectra: (1) faster transverse spin relaxation, (2) sensitivity when only limited concentrations can be achieved, and (3) signal overlap due to higher number of resonances (Wider *et al.,* 1999; Riek *et al.,* 2000).

Large molecular size of NS3 helicase makes it a challenging target for NMR spectroscopy. The rate of transverse relaxation increases with molecular size and results in line broadening and signal losses. The answer to this problem is the use of TROSY technique (Pervushin *et al.,* 1997a; Pervushin 2000). It allows to partially suppressing transverse relaxation during evolution and acquisition periods of the NMR experiment. Deuteration of protein sample could also be used to reduce relaxation by suppressing additional relaxation sources such as dipole-dipole coupling to remote protons. However, the conformational dynamics of DENV4 NS3 helicase pose a major challenge causing line broadening and disappearance of signals. This challenge is attributed to the highly dynamic functional nature of the protein itself.

Large molecular structure with molecular weight of 51.3 kDa, NS3 helicase is not soluble to typical concentrations used for NMR studies (~ 1mM). With the advent of cryo-probe head technology the sensitivity of NMR measurements with protein solution have improved paving way to work with lower concentrations.

Concerning the spectral overlap, a number of isotope labeling schemes have been proposed that result in the simplification of the spectra. Deuteration and recent trends in amino acid selective labeling provide strategies to reduce spectral complexity (Takeuchi *et al.,* 2007).

Paramagnetic relaxation enhancement (PRE) agents have been reported in delineating the surface- exposed residues in molecular structures (Pintacuda *et al.,* 2002). This is particularly helpful in reducing the spectral complexity.

Alternative to all the above, smaller fragments of NS3 helicase can be produced which would be amenable to solution NMR techniques while retaining their native NTP/nucleic acid binding sites. This strategy alleviates the signal loss due to rapid transverse relaxation and spectral complexity.

3.1.1 Dengue virus

Dengue virus (DENV) is a member of the *Flaviviridae* family, which has a positive sense RNA and known to cause a broad spectrum of illness in humans ranging from simple fever through Dengue hemorrhagic fever (DHF) to dengue shock syndrome (DSS) (Calisher *et al.,* 2003) (Lindenbach *et al.,* 2003). Dengue is caused by four antigenically distinct viruses designated as Dengue virus serotypes 1 to 4 (DENV 1–4) (Lindenbach *et al.,* 2003). The insect vector, female mosquito *Aedes agepti*, transmits the virus between humans, domestic animals and birds. Currently, no effective vaccine or antiviral drug therapy is available against Dengue viruses.

The Dengue viral genome consists of a single stranded, non-segmented, positive sense RNA of about 11 kb in length (Lindenbach *et al.,* 2003).

Following the attachment to the host cell surface (Chu *et al.,* 2005), virus transports via endocytotic path and the pH dependent structural changes occur in the surface envelope glycoprotein E (Roussel *et al.,* 2006; Lescar *et al.,* 2008) leading to the merging of viral and host lipid membranes and release of viral genome in to the host cell cytoplasm (Bressanelli *et al.,* 2004; Modis *et al.,* 2004). A schematic picture describing the life cycle of dengue virus is given in Figure 3.1. Upon capsid disassembly, host cell translation machinery translates the viral RNA in to single precursor polypeptide precursor of approximately 370 kDa, containing three structural and seven non-structural (NS) proteins in the following order C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5 (Umareddy *et al.,* 2007). The genome architecture of Dengue virus is shown in Figure 3.2.

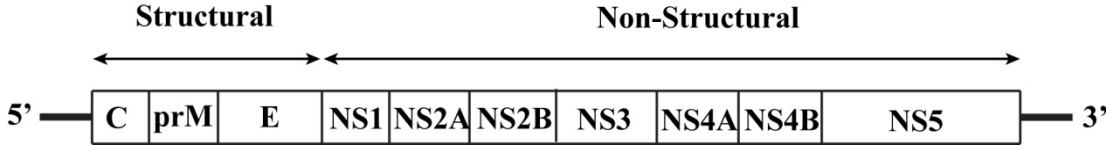


Figure 3.1: Dengue virus genome architecture: Viral genome annotated with structural and nonstructural protein coding genes. Structural proteins: C – Capsid, prM – Membrane precursor and E – Envelope protein. Non-structural proteins are numbered according to their precedence in the precursor-polypeptide.

Upon synthesis of genome length minus strand RNA by NS5 RNA-dependent RNA polymerase in tight association with NS3, the dsRNA need to be separated for further rounds of genome replication and transcription. It is generally thought that the NS3 helicase separates into individual strands the dsRNA intermediate, in an ATP-driven manner. However, recently J. Lescar and co-workers (Luo *et al.,* 2008b) have given a clear insight on the RNA unwinding coupled to the ATP hydrolysis. Thus,

NS3 helicase has an established role in the Flavivirus transcription and life cycle. A detailed account on the structural as well as functional aspects of Flavivirus NS3 helicase is given as follows.

3.1.1.1 NS3 protein

Dengue virus NS3 protein is a multifunctional protein, possessing the N-terminal protease activity and the C-terminal helicase activity and plays a crucial role in viral genome replication and transcription, as shown in Figure 3.2. In addition to its enzymatic activities, NS3 has also been reported to have a role in virus assembly and virion production (Liu *et al.,* 2002; Patkar *et al.,* 2008).

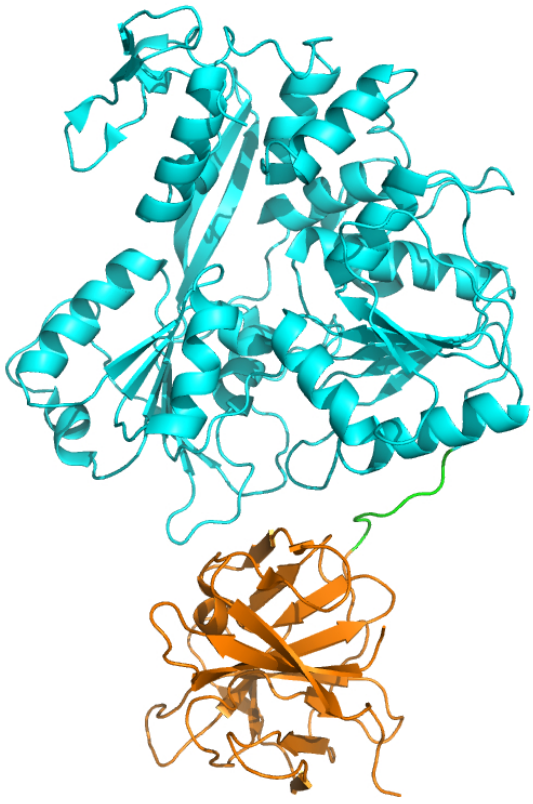


Figure 3.2: DENV4 NS3 helicase-protease: Cartoon representation of DENV4 NS3 helicase-protease. The bifunctional enzyme, helicase domain and protease domain are coloured in cyan and orange respectively. The linker region is coloured green. (PDB:2VBC.PDB)

3.1.1.2 NS3 protease domain

NS3 protease (N-terminal residues 1-180) is a serine protease with NS2B acting as a co-factor (Chambers *et al.,* 1991; Falgout *et al.,* 1991) for its proteolytic activity. In complex with NS2B NS3 plays a pivotal role in viral polyprotein processing. The open and closed conformation NS2B/NS3pro complex is shown in Figure 1.5. This complex is responsible for all cleavage between or within viral

proteins, including the non-structural proteins as well as structural proteins and itself (Lindenbach *et al.,* 2003).

3.1.1.3 NS3 helicase domain

NS3 helicase (C-terminal residues 170-618) is a DEAH RNA helicase belonging to SF2 superfamily (Gorbalenya *et al.,* 1993; Xu *et al.,* 2008). The helicase domain of the NS3 protein possesses (1) RNA helicase activity, (2) NTPase activity, (3) RTPase activity and (4) a role in viral assembly (Chambers *et al.,* 1993; Li *et al.,* 1999; Borowski P 2000; Benarroch *et al.,* 2004; Lescar *et al.,* 2008; Patkar *et al.,* 2008). Impaired replication processes have demonstrated the importance of the helicase by Flaviviruses with impaired helicase activity (Grassmann *et al.,* 1999; Matusan *et al.,* 2001).

It is known that the helicase activity is an energy driven process, utilizing the energy released by ATP hydrolysis for the separation of duplex oligonucleotides into single strands. The ATP hydrolysis become absolute requirement as the duplex region extends from 22 or 23 to 36 base-pairs (Li *et al.,* 1999; Benarroch *et al.,* 2004; Wang *et al.,* 2009) otherwise the NTPase and helicase activity can be uncoupled (Sampath *et al.,* 2006).

Flavivirus NS3 helicase' NTPase activity involves biding of NTP, hydrolysis of γ phosphoric anhydride bond and eventual release of NDP and inorganic phosphate. The conserved motifs among helicases and motor proteins, motif-I (binding to nucleotide phosphates), motif-II (coordination with divalent metal ion and activation of attacking water molecule) and motif-IV (binding of nucleotide phosphates), are directly involved in the hydrolysis of NTP (Li *et al.,* 1999; Benarroch *et al.,* 2004; Xu *et al.,* 2006).

RNA triphosphatase (RTPase) activity requires the γ-phosphoric anhydride bond hydrolysis of triphosphorylated RNA (Rice 2007; Dong *et al.,* 2008). Surprisingly, the RTPase activity shares a common active site in Flavivirus NS3 with NTPase activity (Bartelma *et al.,* 2002; Benarroch *et al.,* 2004; Wang *et al.,* 2009). Recently, it has also been reported that NS5 can also stimulate RTPase activity (Yon *et al.,* 2005).

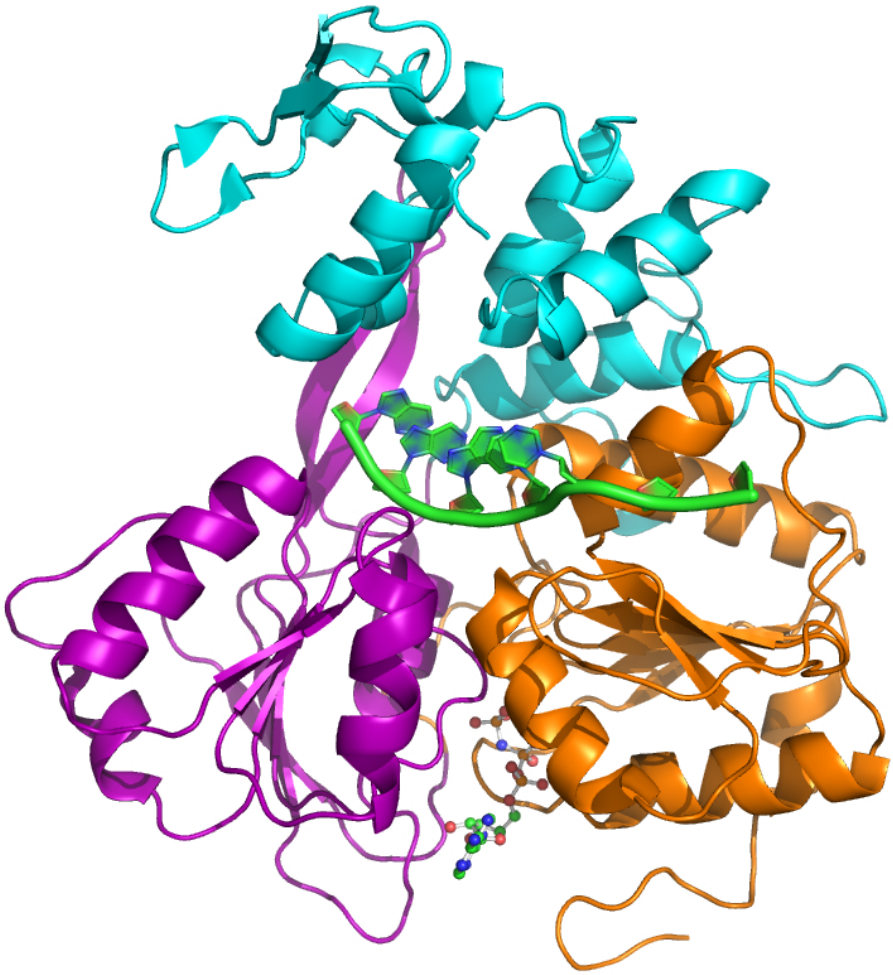


Figure 3.3: Flavivirus NS3 helicase – RNA – AMPPNP ternary complex: Cartoon representation of NS3 helicase bound with ssRNA (green colored tube representation) and non-hydrolizable ATP analogue (colored by elements in ball-stick representation) (PDB:2JLV.PDB). Domains one, two and three are colored orange, purple and cyan respectively.

Since the first reported crystal structures of NS3 helicase domains (Wu *et al.,* 2005; Xu *et al.,* 2005), a number of Flavivirus and other NS3 helicase structures have been published (Mastrangelo *et al.,* 2006; Mancini *et al.,* 2007; Lescar *et al.,* 2008; Luo *et al.,* 2008a; Luo *et al.,* 2008b; Speroni *et al.,* 2008; Yamashita *et al.,* 2008). A number of structural features were found. Firstly, each of these proteins comprises three subdomains. All the identified sequence motifs for the members of the SF2 superfamily helicases are located within Rec-A like subdomains 1 and 2 (Lescar *et al.,* 2008). The NTP binding site is present between the subdomains 1 and 2, whereas the RNA is accommodated in the tunnel separating the NTP binding subdomains 1 and 2 from subdomain 3 (Wu *et al.,* 2005; Xu *et al.,* 2005) as shown in Figure 3.3.

In the present study, we chose to work with DENV4 NS3 helicase since it is an important protein in the Flaviviral life cycle and thus a potential drug target (Lescar *et al.,* 2008). The drug discovery process for any target protein involves a number of stages; from target validation through high throughput screening of chemical libraries until design of a potential lead compound (Klages *et al.,* 2007). Over the last couple of decades, NMR spectroscopy has emerged as an inevitable tool that seeped into several stages of drug discovery process. Unique advantage is that, it offers dynamic details (protein-ligand interaction, dissociation constant) at atomic level. The primary requirement for any detailed study of protein structure and dynamics is the extensive backbone resonance assignment. This chapter discusses the strategies and results obtained in the process of backbone resonance assignment of DENV4 NS3 helicase protein.

3.2 Materials and Methods

3.2.1 Materials

3.2.1.1 Chemicals

All the chemicals used for the study were at least analytical grade. Chemicals were purchased from the following companies:

Buffers and Salts	Sigma (St. Louis, MO, USA)
	Fluka (Sigma, Buchs Germany)
	Merck (Darmstadt, Germany)
$^{15}NH_4Cl$	Cambridge Isotopes Lab (USA)
^{13}C Glucose	Cambridge Isotopes Lab (USA)
Media Components and Salts	MERK (Darmstadt, Germany)
	Applichem (GE Healthcare Lifesciences)
DTT	Applichem (GE Healthcare Lifesciences)
1-^{13}C Alanine , Leucine and Valine	Cambridge Isotopes Lab (USA)
Antibiotics	Applichem (GE Healthcare Lifesciences)
IPTG	Fermentas

3.2.1.2 Electrophoresis Reagents

AA/BAA, SDS, APS	BioRad (CA, USA)

3.2.1.3 Molecular Biology Materials

Plasmid Miniprep Kit	QIAGEN (Hilden, Germany)
Escherichia coli expression strains	DH5α, BL21 Codon Plus
pET-32b (+) modified	(Luo *et al.,* 2008b)
pNIC28-Bsa4	SGC, Toronto

3.2.1.4 Chromatography

IMAC	
His-Trap columns	GE (USA)
Gel-filtration	
Superdex-75 HR (16/60)	Amersham (GE Healthcare Lifesciences)

3.2.1.5 Instrumentation and accessories

Akta FPLC, Akta Prime	GE Healthcare Lifesciences
Sartorius sterdim	
Synriges, Needles and accessories	BD Biosciences
NanoDrop – 1000 Spectrophotometer	ThermoScientific (MA, USA)
Gel documentation	BioRad (CA, USA)

3.2.1.6 Computer Software

MatLab7.6	MathWorks® (MA, USA)
CARA-1.8.4	(Keller)
Topspin2.1	Bruker Biospin
PyMol v1.2	DeLano Scientific LLC, USA

3.3 Methods:

To perform studies of large macromolecular systems by solution NMR, a number of requirements must first be met at the level of sample preparation. The first important requirement is a highly pure and homogeneous sample. Typically, protein concentration of 1-5 mM is ideal. With the advent of cryo-probe techniques, high quality spectra can be obtained within reasonable measurement time for sample concentration as low as 100 μM. The second critical step is the production of stable isotope labeled protein sample. An organism bearing the abilities to produce isotope labeled proteins (mostly ^{15}N and ^{13}C) with minimal growth requirements (M9 – minimal media) (Sambrook *et al.,* 2001), thus becomes mandatory. For proteins with molecular sizes above 35 kDa, the requirement of uniform ^{2}H-labeling, adds another level of complexity in the protein production and sample preparation. Besides these requirements, the sample must be stable against aggregation and degradation during several days of NMR data acquisition.

3.3.1 Stable isotope labeling of DENV4 NS3 helicase

The aim of the isotope labeling of NS3 helicase is to increase the sensitivity of the experiments and resolution of the spectra. Routine use of heteronuclear multi-dimensional NMR experiments is possible with the incorporation of ^{15}N and ^{13}C isotopic labels. Due to its high molecular weight nature of NS3 helicase, deuteration becomes necessary. Amino acid selective labeling is applied to reduce the spectral complexity and help the assignment procedure. This section describes the various isotope-labeling schemes employed in addressing the resonance assignment of NS3 helicase.

3.3.1.1 Uniform ^{2}H, ^{15}N and ^{2}H, ^{15}N, ^{13}C labeling of NS3 helicase

Expression of NS3 helicase was achieved in *E. coli* BL21 *Codonplus* cells (Stratagene). However, stepwise adaption of the cells to increasing levels of D_2O was necessary to achieve efficient growth and expression of NS3 helicase. While pre-cooling was necessary before inoculation to avoid sequestration of the recombinant protein into inclusion bodies. Care was taken for the maintenance of aseptic environment throughout the procedure.

Supply of ^{15}N labeled ammonium chloride (NH_4Cl) during the stepwise adaptation of cells was found to be unnecessary. Hence, to achieve uniform ^{15}N labeling of backbone amides (unavoidably side chain amides of Asparagine and Glutamine) and to reduce the production cost, ^{15}N labeled ammonium chloride was supplemented as a sole source of nitrogen only in the D_2O based M9-minimal growth medium.

For ^{2}H, ^{15}N labeled samples, protonated glucose was used as the sole source of carbon, irrespective of protein expression stage. Moreover, for ^{2}H, ^{15}N, and ^{13}C labeled samples uniformly ^{13}C labeled glucose was used as the sole carbon source without deuteration. The major advantage of the medium composition, used here are efficient growth and less cost. The medium composition are tabled in Table-3.1a and Table 3.1b

Table-3.1a: M9 medium composition

M9 – medium Salts	Quantity (gm L^{-1})
Disodium hydrogen phosphate	12.8
Potassium dihyrogen phosphate	3.0
Sodium chloride	0.5
Ammonium chloride‡	1.0
Glucose†	2.0
ddH_2O§	q.s

‡ Substituted with ^{15}N labeled ammonium chloride in the case of ^{15}N-labeling.
† Substituted with uniformly ^{13}C labeled Glucose in the case of ^{13}C labeling.
§ Substituted with D_2O, when deuterated environment is needed.

Table-3.1b: Medium supplements

Trace Elements and Vitamins	Quantity (gm L^{-1})
$MgSO_4$. 7H_2O	0.4940
$CaCl_2$.2H_2O	0.0152
$FeSO_4$.7H_2O	0.0100
Thiamine hydrochloride	0.0100

3.3.1.2 Selective 1-^{13}C-Leu, Ala, Val and Phe labeled NS3 helicase

For any selective amino acid labeling the knowledge of the desired amino acid biosynthetic pathway is necessary. The major concerns when ^{15}N labeling specific types of amino acids are isotopic dilution and transamination. However, supplementing a minimal growth medium with excess of labeled amino acids can reduce cross labeling and that this procedure works best for amino acids that are the

terminal products of the biosynthetic pathway. But still, there is possibility of scrambling across amide nitrogen due to stressed growth condition in minimal deuterated medium. Using rich synthetic media supplemented with excess of unlabeled amino acids other than the intended amino acid or by supplementing excess of amino acid of intention, could result in huge production cost.

And thus, using 1-^{13}C labeled amino acids instead of ^{15}N labeled (Takeuchi *et al.,* 2007) since the ^{13}C of carbonyl are less prone to scrambling than ^{15}N at the α-amino position was adopted. 1-^{13}C labeling is also a cost effective method compared to labeling using synthetic rich media. For our purpose, NS3 helicase with 1-^{13}C labeled amino acids namely, leucine, alanine, valine and phenylalanine were produced by supplementing 5 fold excess of the respective amino acid with reference to the amount present in a typical Celtone media to the deuterated M9-minimal medium in ^{2}H-^{15}N background.

3.3.2 DENV4 NS3 helicase expression

3.3.2.1 NS3 helicase protein expression protocol

Table-3.2: Growth condition[†]

Growth parameters	Optimal conditions
Shaker speed	180 rpm
Incubation temperature[Œ]	37° C and 16° C

[†] Throughout the expression procedure, a combination of antibiotics, ampicillin and chloremphenicol, were used in concentrations 100 µg ml^{-1} and 34 µg ml^{-1} (final concentration) respectively.
[Œ] the temperature was reduced to 16° C after induction with IPTG

Previously cloned, modified plasmid pET32b(+) containing thioredoxin (Trx) protein, hexa-histidine-tag, a thrombin cleavage site and NS3h (in order) was transformed in to the BL21 *Codonplus* competent cells by heat shocking. Following overnight growth at 37° C, single colony was picked and inoculated in to LB medium. After ~12 hours of growth, the cells were pelleted by centrifugation at 4000 rpm at 4° C for 10 minutes. Subsequently, the pellett was resuspended in 100 ml of M9-H_2O medium. Following ~12 hours of growth, the cells were pelleted by centrifugation at 4000 rpm at 4° C for 10 minutes. The pellet was washed in Phosphate buffer saline followed by resuspension into a volume of 5 ml of the same buffer. The OD_{600} was measured and the inoculation volume was calculated to achieve an initial OD_{600} of ~0.1 in 200 ml of M9-D_2O medium. Upon reaching OD_{600} of ~0.4-0.5, the culture was

diluted five times with sufficient volume of M9-D_2O medium. When the OD_{600} reaches 0.6 - 0.8, induction was done with 0.4 mM (to final concentration) of IPTG and the growth temperature was reduced. For 1-^{13}C selectively amino acid labeled NS3 helicase preparations, the respective amino acid of interest was supplemented twice, first after induction followed by ~10-11 hrs after induction. Following ~21 hours of incubation, cells were harvested by centrifugation at 7500 rpm at 4° C for 20 minutes. The cell pellet was subsequently washed with phosphate buffer saline for two to three times before storing at -20° C.

3.3.3 Purification of DENV4 NS3 helicase

3.3.3.1 NS3 helicase purification protocol

(A) The protocol described previously (Luo *et al.,* 2008b) was adopted for DENV4 full length NS3 helicase purification. Before lysis by sonication, cells were resuspended in buffer A (20 mM Na_3PO_4, pH 7.4, 0.5 M NaCl and 40 mM imidazole). The lysate was clarified by centrifugation at 30000 g for 60 minutes at 4° C. The resulting supernatant was purified by metal-affinity using HisTrap HP column (GE Healthcare Lifesciences), equilibrated a priori with buffer A. Linear gradient of imidazole from 40 to 500 mM was used, for eluting the protein. Subsequently, the fraction containing Trx-$(His)_6$-NS3 helicase was dialysed against buffer B (200 mM NaCl, 10 mM Phosphate, 2.7 mM KCl and pH 7.4) with concomitant cleavage of thioredoxin tag by thrombin (10 enzyme units per mg of protein) at 4° C for approximately 24 hrs. The thioredoxin-$(His)_6$ protein from the mixture was removed by loading onto a HisTrap HP column equilibrated with buffer C (20mM Na_3PO_4, pH 7.4, 0.2M NaCl). The collected protein factions were pooled together and concentrated in Vivaspin (Sartorius stedim) 20 filter concentraters (10K MWCO). Subsequently, the concentrated protein was purified by gel-filteration purification step through a HiPrep 16/60 Superdex 75 column (Amersham Biosciences) in buffer D (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM DTT). Finally, the collected pure protein was aliquoted into smaller fractions, flash freezed in liquid nitrogen followed by storage in -80° C.

(B) For the purification of fragments of DENV4 NS3 helicase, the following protocol was used. Cells were lysed by sonication following resuspension in the buffer A (20 mM Na_3PO_4, pH 7.4, 0.5 M NaCl and 40 mM imidazole). The lysate was clarified by centrifugation at 30000 g, 4°C for 45 minutes. The resulting supernatant

was filtered through 20 μm filter and subjected to purification by metal-affinity chromatography using HisTrap HP column (GE Healthcare lifesciences), equilibrated a priori with buffer A (20 mM Na_3PO_4, pH 7.4, 0.5 M NaCl and 40 mM imidazole). Linear gradient of imidazole from 40 to 500 mM was used, for eluting the protein. Subsequently, the eluted fractions were pooled together and concentrated in Vivaspin (Sartorius stedim) filter concentrators (10K MWCO). The concentrated protein was loaded on to HiPrep 16/60 Superdex 75 column (Amersham Biosciences), pre-equilibrated with buffer E (20 mM HEPES, 300 mM NaCl, 5 mM DTT, pH 7.5), for gel-filtration purification. Finally, the collected pure protein batch was aliquoted into smaller fractions, flash freezed in liquid nitrogen followed by storage in -80°C.

The purified fraction of NS3 helicase was purity tested by SDS-PAGE analysis using 15% v/v polyacrylamide gels.

3.4 Results and Discussion

3.4.1 Expression and purification of NS3 helicase

A qualitative analysis of purified NS3 helicase by SDS-PAGE revealed its high purity, observation of prominent band corresponding to its molecular size of 51.3 kDa is shown in Figure 3.4**.** The adopted procedure for purification yielded 5-8 mg ml^{-1} (varies from batch to batch) of pure NS3 helicase as estimated by Nanodrop using extinction coefficient of 65890 $M^{-1}cm^{-1}$ with a molecular weight of 51.3 kDa.

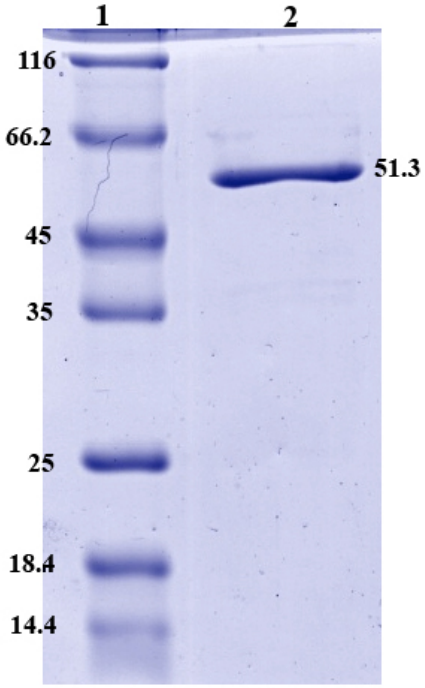


Figure 3.4: Purification analysis of DENV4 NS3 helicase:15% v/v polyacrylamide gel showing high purity DENV4 NS3 helicase protein after Gel filtration chromatography. Lane 1: Standard, Lane 2: NS3 helicase

Dengue virus 4 NS3 helicase is a monomer of 451 residues. The amino acid sequence of DENV4 NS3 helicase is given in Figure 3.5. The purity and proper assembly of subunits was confirmed by 1D-^{1}H spectrum as shown in Figure 3.6.

```
  1 GSAMGEPDYE VDEDIFRKKR LTIMDLHPGA GKTKRILPSI VREALKRRLR
 51 TLILAPTRVV AAEMEEALRG LPIRYQTPAV KSDHTGREIV DLMCHATFTT
101 RLLSSTRVPN YNLIVMDEAH FTDPCSVAAR GYISTRVEMG EAAAIFMTAT
151 PPGSIDPFPQ SNSPIEDIER EIPERSWNTG FDWITDYQGK TVWFVPSIKA
201 GNDIANCLRK SGKRVIQLSR KTFDTEYPKT KLTDWDFVVT TDISEMGANF
251 RAGRVIDPRR CLKPVILTDG PERVILAGPI PVTPASAAQR RGRIGRNPAQ
301 EDDQYVFSGD PLKNDEDHAH WTEAKMLLDN IYTPEGIIPT LFGPEREKTQ
351 AIDGEFRLRG EQRKTFVELM RRGDLPVWLS YKVASAGISY KDREWCFTGE
401 RNNQILEENM EVEIWTREGE KKKLKPRLID ARVYADPMAL KDFKEFASGR
451 K
```

Figure 3.5: Amino acid sequence of DENV4 NS3 helicase: Amino acid sequence of DENV4 NS3 helicase indicated in the one-letter code (Luo *et al.,* 2008a).

3.4.2 NMR spectroscopy of DENV4 NS3 helicase

DENV4 NS3 helicase is highly dynamic, as observed from the line-broadening induced by conformational exchange, as shown in Figure 3.7. However, in complex with a 12-mer ssRNA the quality of the [^{1}H, ^{15}N]-TROSY-HSQC spectrum was much improved with the appearance of new peaks and better dispersion of resonances. This is due to the formation of NS3 helicase ssRNA complex. Addition of non-hydrolysable ATP analogue-AMP-PNP (Adenosine 5'-(β,γ-imido) triphosphate tetralithium salt hydrate) to NS3 helicase results in better chemical shift dispersion with the appearance of new peaks as shown in Figure 3.8. However, this complex was observed to be precipitate rapidly in NMR buffer. Conversely, addition of AMP-PNP in combination with ssRNA results in stable protein sample during long duration NMR experiments.

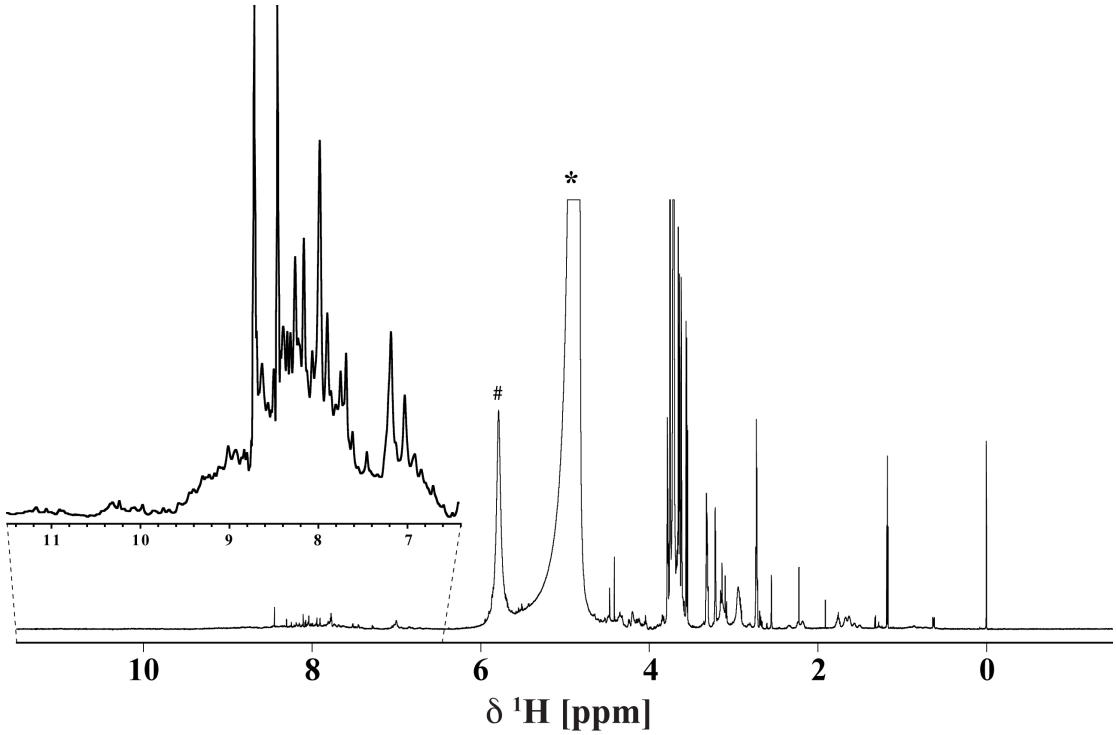


Figure 3.6: 1D ^{1}H NMR spectrum of DENV4 NS3 helicase: 1D ^{1}H spectrum of DENV4 NS3 helicase with ^{2}H, ^{15}N and ^{13}C labelled sample in buffer (20 mM Tris-Hcl, 300 mM NaCl, 5 mM DTT and 2% v/v glycerol pH-7.4), at 298K in complex with ssRNA (equimolar concentration with protein). Expanded region shows sharp peaks in the backbone amide proton region indicates the proper folding and the broadened peak base is characteristic for a high molecular size protein. 950 MHz filed at 298K. (# - resonance arise from ssRNA; * - water resonance).

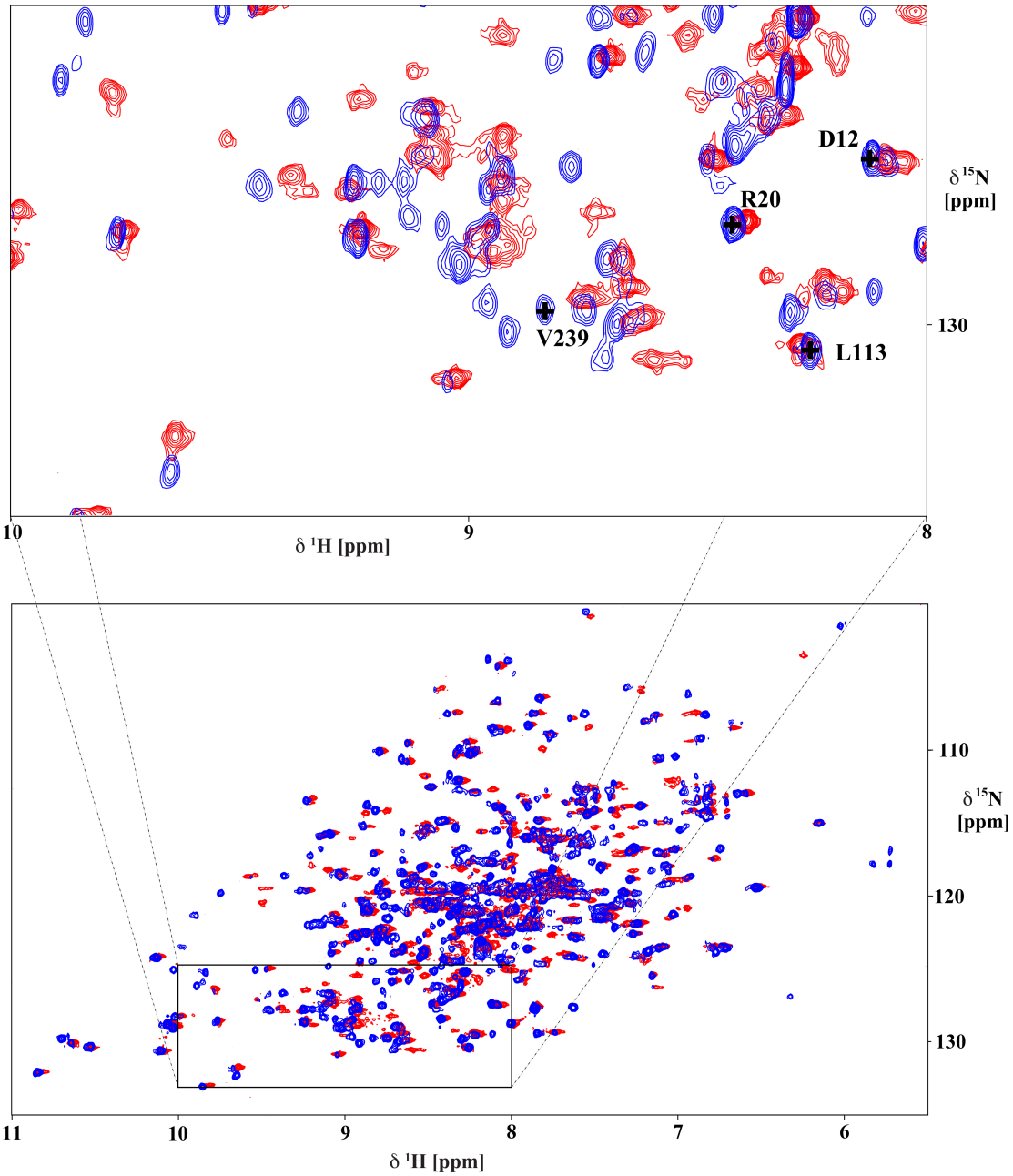


Figure 3.7: Conformational exchange induced line-broadening: [^{1}H, ^{15}N]-TROSY-HSQC spectrum of DENV4 NS3 helicase (red) and in complex with (blue) ssRNA (in equimolar concentration with protein) in Tris-HCl buffer (20 mM Tris-Hcl, 300 mM NaCl, 5 mM DTT, 2% v/v glycerol, pH-7.4) at 298K. Upon RNA binding NS3 helicase undergoes a conformational change, resulting in the appearance of new peaks and better dispersion of resonances in comparison with the spectrum without RNA. Assigned residues are marked in the zoomed region.

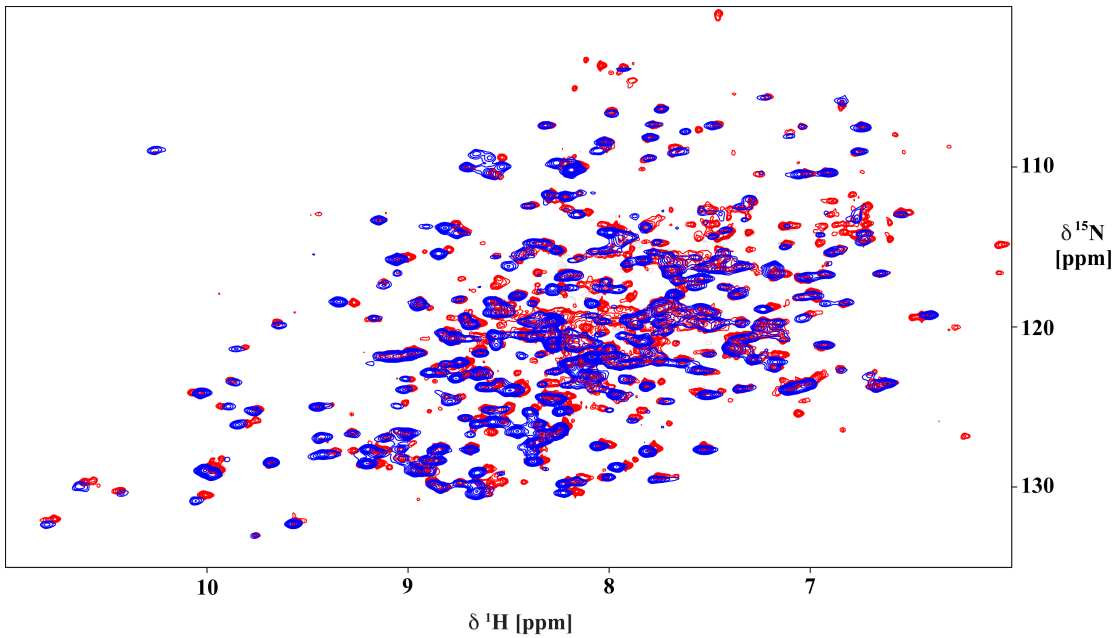


Figure 3.8: Effect of non-hydrolisable ATP analogue: [^{1}H, ^{15}N]-TROSY-HSQC spectrum of DENV4 NS3 helicase with (red) and without (blue) the non-hydrolisable ATP analogue (AMP-PNP). The sample conditions were same as for Figure 3.7. The spectra were measured at 700 MHz spectrometer frequency.

3.4.2.1 Resonance assignment process

The resonance assignment process was carried out using Computer Aided Resonance Assignment (CARA) program (Keller). Assignment process was started with spin-system identification using 2D [^{1}H, ^{15}N]-TROSY-HSQC spectrum, as shown in Figure 3.12 of [U-^{2}H; U-^{15}N]-NS3 helicase, resulted in the identification of 356 out of the 420 expected resonances (i.e., total 451 residues minus 30 prolines and one N-terminal residues). Localized conformational exchange-induced line broadening as well as incomplete back- $^{2}H^{N}/^{1}H^{N}$ exchange in NS3 helicase result in disappearance of cross-peaks in the 2D [^{1}H,^{15}N]-TROSY-HSQC spectrum. Further, the sequence specific assignment was carried with triple resonance experiments [^{1}H, ^{15}N]-TROSY-HNCA, ^{13}C-CT-[^{1}H, ^{15}N]-TROSY-HNCA, [^{1}H, ^{15}N]-TROSY-HNCACB and [^{1}H, ^{15}N]-TROSY-HN(CA)CB spectrum. (Mr. Wong Leo E of our laboratory prepared Protein samples and NMR data were recorded at The Biomolecular Magnetic Resonance Center (BMRZ), Frankfurt, Germany by Prof. Konstantin Pervushin,).

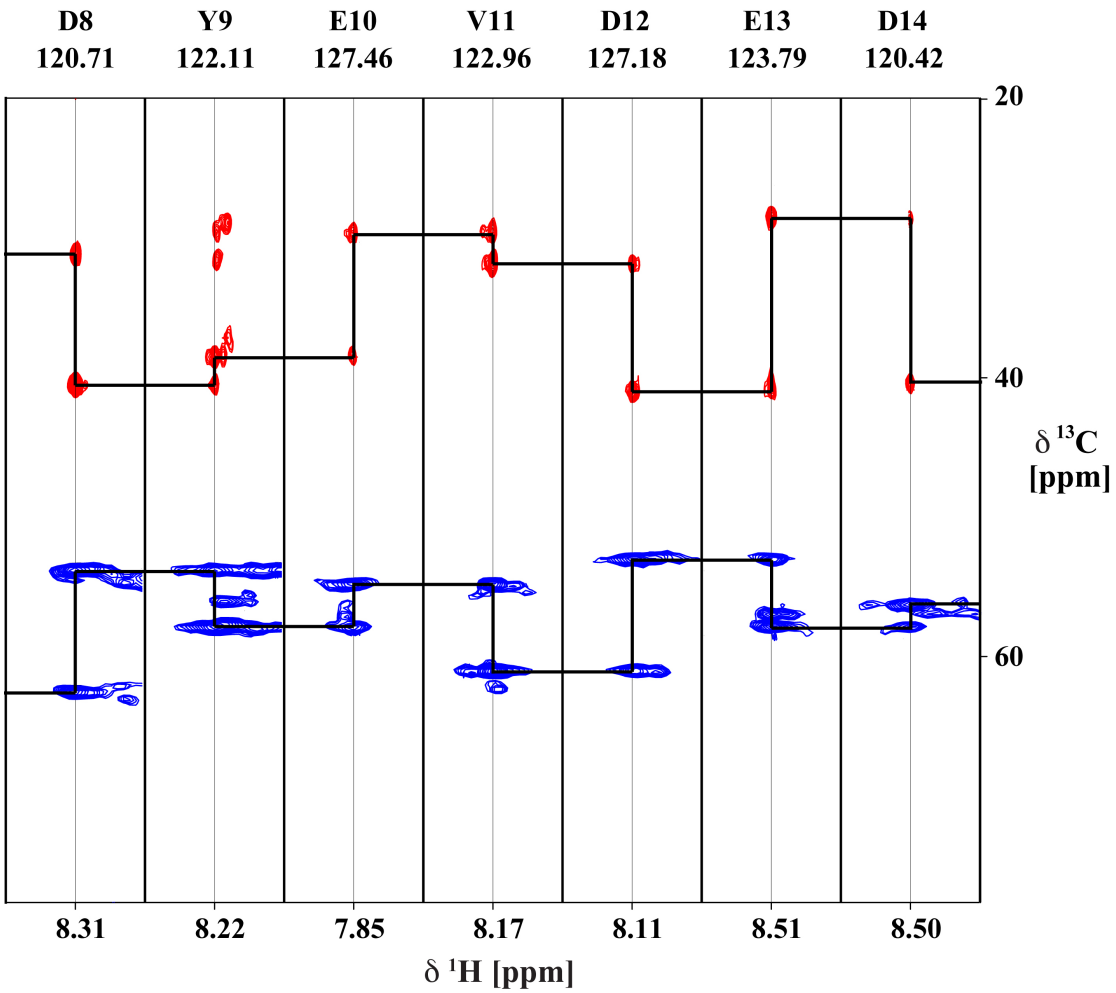


Figure 3.9: Strips from 3D TROSY-HNCA and TROSY-HN(CA)CB showing CA (blue) and CB (red) sequential connectivities, respectively: [^{1}H,^{13}C] strips from 3D [^{1}H,^{15}N]-TROSY-HNCA and TROSY-HN(CA)CB spectrum of ^{2}H, ^{15}N- and ^{13}C-labeled NS 3helicase. Sequential connectivities indicated by bold lines. Assignments obtained are indicated as the one-letter code with residue numbers.

Sequential connectivity information of the NS3 helicase was primarily obtained from the 3D [^{1}H, ^{15}N]-TROSY-HNCA and [^{1}H, ^{15}N]-TROSY-HNCACB spectra, using AutoLink (Masse *et al.,* 2005). Autolink works in tight association with CARA making use of the intra- and inter-residue data that are incorporated in to "spin-system based object model" of the CARA program itself (Keller). Figure 3.9 shows sequentially connected ^{13}Cα strips from 3D [^{1}H, ^{15}N]-TROSY-HNCA. Cases where connectivity information could not be obtained from 3D [^{1}H, ^{15}N]-TROSY-HNCA, due to ^{13}Cα chemical shift degeneracy and the absence of inter-residue cross peaks, ^{13}C-CT-[^{1}H, ^{15}N]-TROSY-HNCA spectra was used for the establishment of connectivities, when available. Furthermore, the ^{13}Cβ chemical shifts were used from the [^{1}H, ^{15}N]-TROSY-HNCACB and [^{1}H, ^{15}N]-TROSY-HN(CA)CB spectra to extend the sequential assignment and confirmation of the ^{13}Cα connectivities where available.

Figure 3.9 shows sequentially connected $^{13}C\beta$ strips from 3D [^{1}H, ^{15}N]-TROSY-HN(CA)CB, as well . The statistics of the resonance assignment process using the triple resonance experiments is given:

Total number of identified spin-systems	Spin-systems with Cα atoms identified	Spin-systems with Cβ atoms identified
356	309	170

Experiment	Strips with two peaks	Strips with one peak	Strips with no peaks	Strips with more than two peaks
[^{1}H, ^{15}N]-TROSY-HNCA	168 (47%)	118(33%)	47(13%)	23(7%)
[^{1}H, ^{15}N]-TROSY-HN(CA)CB	62(17%)	108(30%)	156¶ (44%)	-

¶ Glycines (29 nos) without Cβ atoms were accounted

Approximately half of the identified spin-systems could not be assigned due to either lack of connectivity information or the absence of the corresponding Cα peak using the 3D [^{1}H, ^{15}N]-TROSY-HNCA spectrum. With respect to the 3D [^{1}H, ^{15}N]-TROSY-HN(CA)CB spectrum, nearly half of the Cβ peaks of the identified spin-systems were absent. The presence of ambiguous peaks and/or absence of corresponding Cα and Cβ peaks for a single spin-system, adds additional complexity to the assignment process. On the whole, triple resonance experiments contributed less than 10% assignment of the total number of residues calling for complementary strategies for further assignment as discussed in Chapter – 1 and Chapter - 2.

3.4.2.2 1-^{13}C-Leu, 1-^{13}C-Ala selective labeling in the ^{2}H, ^{15}N background

In order to continue the resonance assignment, by filling in with sequential connectivity, an amino acid selective isotope labeling was adopted. Resulting [^{1}H, ^{15}N]-TROSY-HN(CO) spectra obtained with 1-^{13}C-Leucine and 1-^{13}C-Alanine selectively labeled NS3 helicase were used to continue the sequential assignment as this method of labeling provides the succeeding residue of the 1-^{13}C selectively labeled Alanine and Leucine as shown in Figure 3.10. Despite, providing the connectivity information, inability to connect a significant stretch of residues, further hampers the sequence specific assignment. However, with this information less than 2% of the resonances were assigned out of the total spin-systems identified. The exact

number of residues assigned, with the total number of Lecucine and Alanine in DENV4 NS3 helicase given in parenthesis, is given:

1-^{13}C-Leucine	1-^{13}C-Alanine
3(31)	1(34)

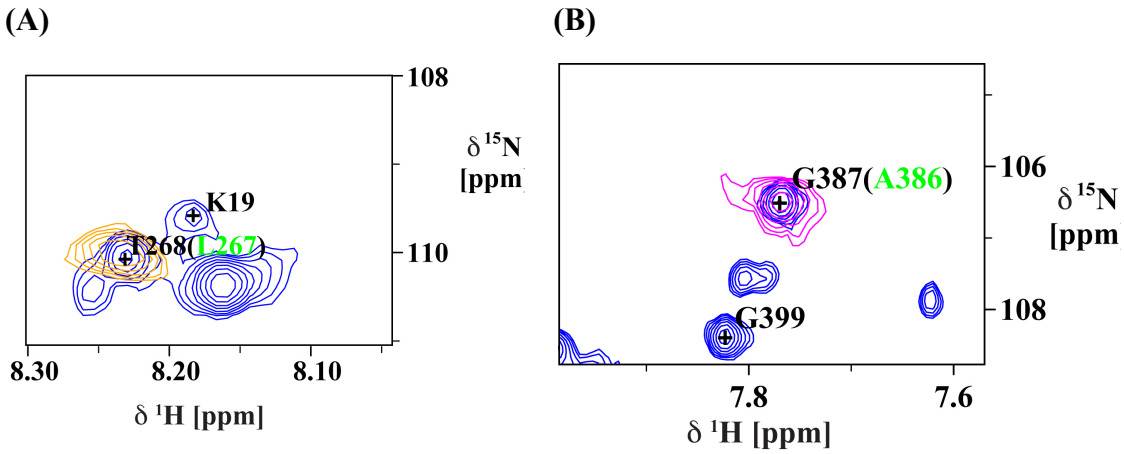


Figure 3.10: Sequential information from selectively 1-^{13}C amino acid labelled NS3 helicase: (A) Section of 2D [^{1}H,^{15}N]-TROSY-HN(CO) spectrum (orange) of 1-^{13}C-Leucine selectively labeled NS3 helicase, overlayed on [^{1}H,^{15}N]-TROSY-HSQC spectrum (blue). Resonance shaded orange correspond to amino acid succeeding 1-^{13}C labeled Lecuine. (B) Section of 2D [^{1}H,^{15}N]-TROSY-HN(CO) spectrum (purple) of 1-^{13}C-Alanine selectively labeled NS3 helicase, overlaid on [^{1}H,^{15}N]-TROSY-HSQC spectrum (blue). Resonance shaded purple correspond to amino acid succeeding 1-^{13}C labeled Alanine.

3.4.3 Preliminary PRE studies

With the aim of alleviating spectral overlap and delineate the surface residues of the NS3 helicase, PRE studies were performed. It is expected that, in the absence of specific binding, Gd(DTPA-BMA), enhance the relaxation rates of protein protons as a function of their surface exposure and distance from the surface. The main advantage of Gd(DTPA-BMA) is, lesser hydrophobicity than many agents in its category. In addition, its stronger paramagnetic nature allows the use at low concentrations. DTPA-BMA is octadentate ligand and the ninth coordinate shell is filled with water, which is difficult to replace with carboxyl or amino groups. Therefore, Gd(DTPA-BMA) does not present a site for protein binding. And thus, it becomes a suitable PRE agent for delineation of surface residues.

In the presence of 0.5 mM Gd(DTPA-BMA), a strong intensity attenuation was observed as shown in Figure 3.11A. With subsequent increase in Gd(DTPA-BMA) concentration protein precipitation was observed, yielding an uninferable spectrum. However, tentative assignment has been made for Gly387 and Gly399, as shown in Figure 3.11B. Further examination of the X-ray crystal structure, Gly387 and Gly399

are found to be surface exposed. Thus, confirming the assignment of Gly387 and Gly399 with the sequence specific assignment information obtained from triple resonance experiments.

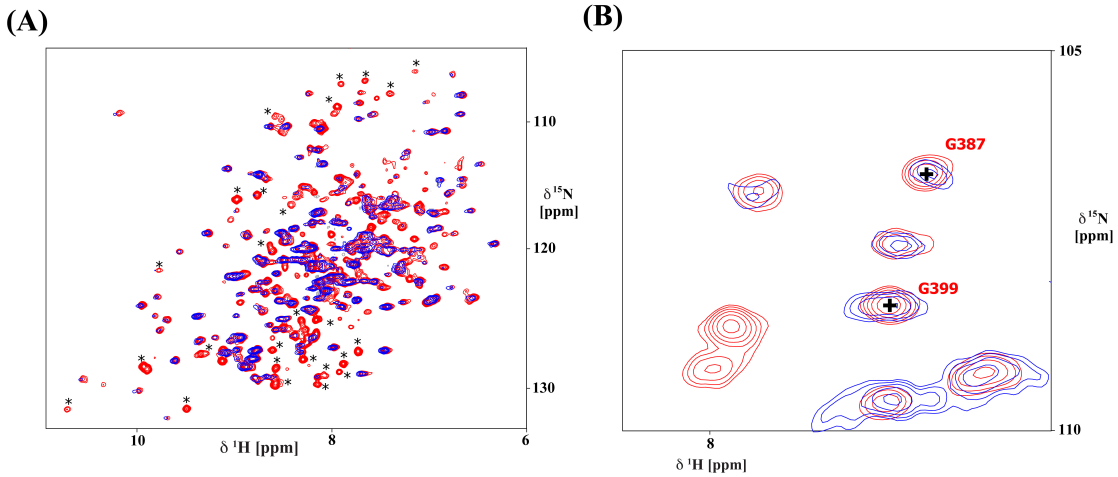


Figure 3.11: Paramagnetic relaxation enhancement study of NS3 helicase: (A) [^{1}H,^{15}N]-TROSY-HSQC spectrum of DENV4 NS3 helicase in the presence (blue) and in the absence of 0.05 mM Gd(DTPA-BMA). Strongly attenuated signals are marked with asterisk. (B) Residues marked with the assignment, using the information from PRE study.

Figure 3.12 shows the [^{1}H, ^{15}N]-TROSY-HSQC spectrum marked with the assigned residues. Appendix D lists the ^{1}H, ^{15}N, ^{13}Cα and ^{13}Cβ (when available) chemical shifts of the assigned residues.

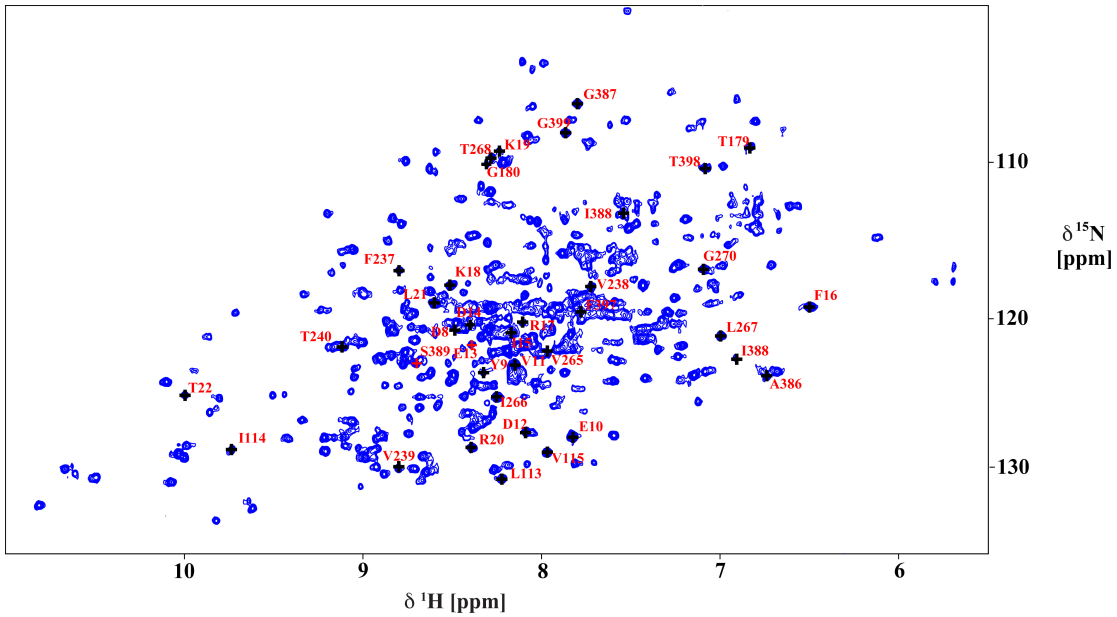


Figure 3.12: [^{1}H, ^{15}N]-TROSY-HSQC spectrum of DENV4 NS3 helicase in 20 mM Tris HCl buffer with 300 mM NaCl, 5 mM DTT and equimolar concetration of ssRNA measured at 298K.

3.4.4 Design, high-level expression and purification of fragments of DENV 4 NS3 helicase

Nevertheless, to address the problem of high molecular size limitation and eventual line broadening in addition to the resonance overlap due to large number of resonances, an alternative way of designing smaller fragments of NS3 helicase was adopted. With the knowledge of X-ray crystal structure, three fragments: derived only from sub domains 1 and 2, were designed. The sequences of the designed fragments were given in Appendix-E. While the removal of C-terminal subdomain 3 reduces the effective molecular size from 51.3 kDa to ~38 kDa, inspection of the three dimensional structure of NS3 helicase revealed that the conserved residues responsible for the NTP binding and energy coupling associated with the NTP hydrolysis are mainly located in the interface between domain 1 and 2. In addition, 20 out of 22 residues involved in the ssRNA binding are lining the surface of the ssRNA-binding groove present in sub-domain 1 and 2 (Luo *et al.,* 2008b). Thus, the removal of C-terminal subdomain 3 is rational in the aspect of NMR resonance assignment.

3.4.4.1 Design of a truncated β hairpin in subdomain 2

However, the designed fragment only with the N – terminal subdomains 1 and 2, exposes a β-hairpin associated with the subdomain 2 to the solvent surface, which is likely to adopt a non-native structure. The solvent exposed hydrophobic residues, namely Val432, Ile433, Leu434, Val441, Ile442, Leu443 and Ala444, are expected to cause protein instability and aggregation. To overcome this problem, a β-turn, Ser-Asp-Gly-Lys, was introduced in place of the solvent exposed β-hairpin, as shown in Figure 3.13, following the discussion by D. Wyss and co-workers (Gesell *et al.,* 2001). This design effectively reduces the molecular weight by ~3 kDa.

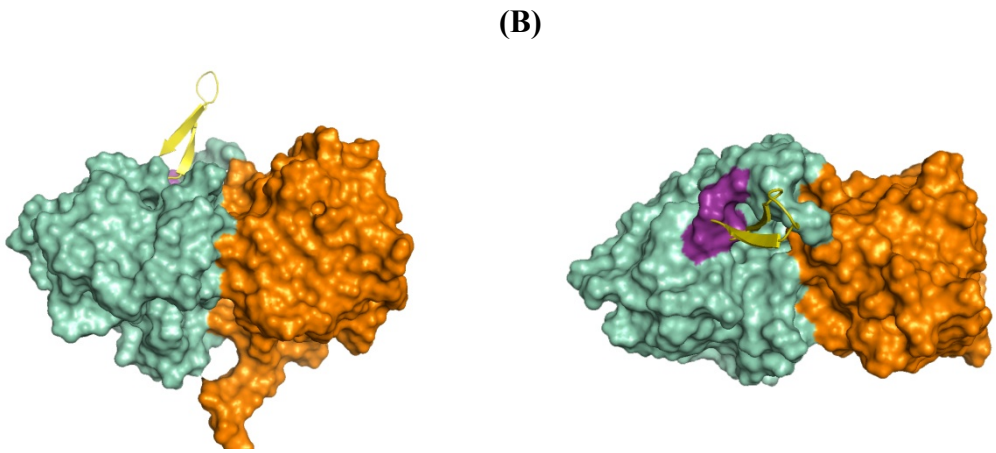


Figure 3.13: Subdomain 1 and 2 of DENV4 NS3 helicase: **(A)** Frontal view: Surface representation of subdomain 1 and 2 of DENV4 NS3 helicase. The solvent exposed β-hairpin was depicted in cartoon representation (yellow). **(B)** Apical view: The designed β-turn (purple). The subdomains 1 and 2 were colored in orange and cyan, respectively, for clarity.

3.4.4.2 High-level expression and purification of the designed fragments of DENV4 NS3 helicase

The designed gene sequence for d1,2Δ-NS3h was commercially ordered and obtained from *Genscript Inc*. The wild type fragments, d1,2a-NS3h$_{wt}$ and d1,2b-NS3h$_{wt}$, as cloned in to pNIC28-Bsa4 (SGC, GenBank accession EF198106) were obtained from Prof Julien Lescar's laboratory. The plasmids were, then, transformed in to BL21 *Codonplus* competent cells and the expression was carried out as explained in Section 3.3.2.1. The purification procedure for wild type (d1,2a-NS3h$_{wt}$ and d1,2b-NS3h$_{wt}$) and mutant (d1,2Δ-NS3h) construct was purified as described under Section 3.3.3.1. Distinct bands corresponding to their respective molecular weight appeared, upon examination of polyacrylamide gels, indicate the over-expression of the desired recombinant DENV4 NS3 helicase fragments, as shown in Figure 3.14.

The purity and the proper folding of the truncated fragments were unable to be checked, as visible precipitation was observed immediately at concentrations required for Circular Dichroism spectroscopy and 1D ^{1}H NMR spectroscopy. Thus, suggesting the removal of any of the three subdomains could affect the proper folding of the DENV4 NS3 helicase, unlike HCV NS3 helicase (Gesell *et al.,* 2001).

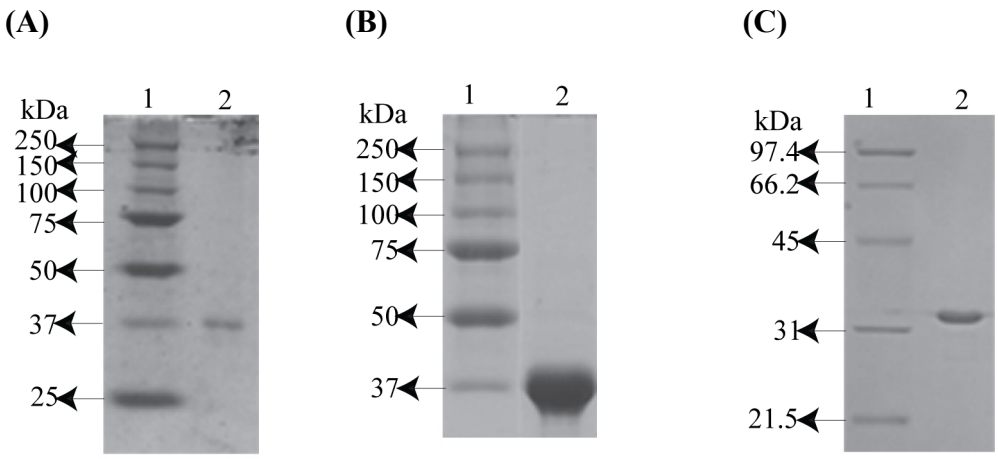


Figure 3.14: Purification analysis of designed fragments: 12% v/v polyacrylamide gels showing the purified NS3 helicase fragments. **(A)** d1,2a-NS3hwt in 20mM HEPES, 300mM NaCl, pH 7.4 and 5mM DTT, **(B)** d1,2b-NS3h and in 20mM HEPES, 300 mM NaCl, pH 7.4 and 5mM DTT **(C)** d1,2Δ-NS3h in 20mM Tris-HCl, 200mM NaCl, pH 7.4 and 5mM DTT. Prominent band appears corresponding to their respective molecular weight. Lane 1 – standard, Lane 2 – purified protein.

Conclusion

4. Conclusion and future perspectives

The results presented in this work establish that the developed NMR-DB and the Spinach theoretical tool enable the systematic and uniform numerical analysis of all the deposited NMR experiments using a single generic spin system representing a three-peptide part of a protein. It has been demonstrated here, as well, that the computed performance metrics: average relaxation rate during the evolution period, average relaxation rate for the complete pulse sequence, line width and sensitivity, for every entry in the database can be directly compared and the deposited experiments can be benchmarked based on their performance. Comparison of experimentally measured sensitivity from different ^{1}H, ^{15}N correlation schemes with the predicted sensitivity by Spinach signify its validity on real situations. Finally, it is demonstrated that seeding the sensitivity from one dimension it is possible to predict the sensitivity in higher dimensions (Senthamarai *et al.,* 2010).

It is important to provide quantitative estimation of the threshold of sensitivity of multi-dimensional experiments predicting the success in resonance assignment and structure determination. Since it is difficult to estimate the impact of missing peaks or peaks broadened by conformational exchange processes on the completeness of assignment, the numerical values of this threshold for minimal expected S/N shall be defined in the statistical sense. Thus we refrain from providing threshold of sensitivity and rather qualitatively recommend to execute experiment with the potential to provide S/N > 10 providing detectability of 90% of peaks.

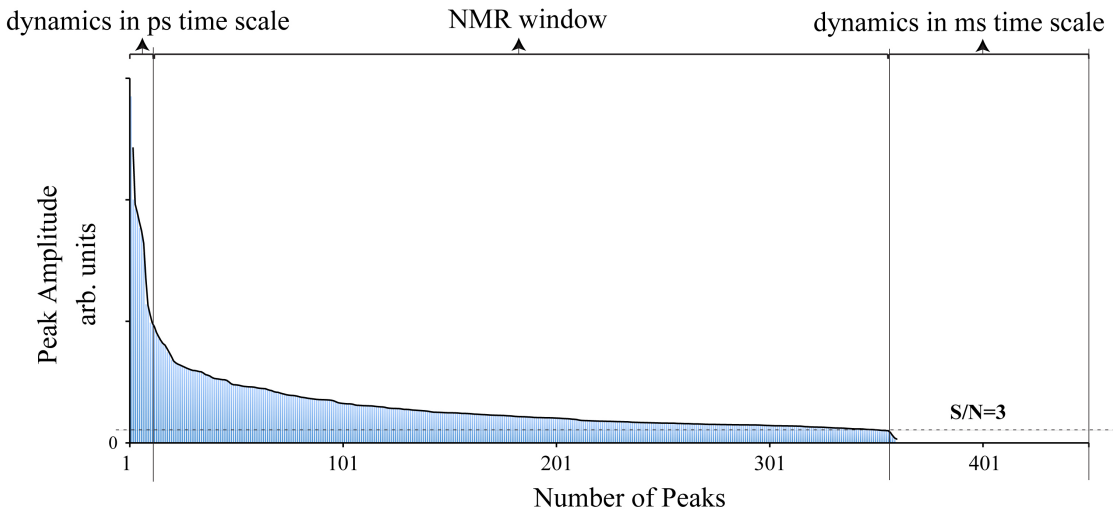


Figure 4.1: Statistical distribution of peak amplitudes for DENV4 NS3 helicase. Choosing S/N=2 as a threshold of sensitivity, we pick 14 cross peaks with dynamics outside of NMR window. Still 95 cross peaks will be outside of detectability.

We now posses a central repository of solution NMR experiments with a theoretical tool (Spinach) opening avenue for the complete and systematic analysis of all the deposited NMR experiments and benchmark the same on their performance. An online WIKI based service have been developed making the NMR-DB available to the users. The present status is the, progressing simulation and analysis of all the deposited experiments in the database. Upon successful completion, the user will be able to predict the sensitivity of any deposited NMR experiment providing the effective rotational correlation time, when not available the effective molecular size.

In the exploration of the limits of solution NMR spectroscopy, a challenging target, the backbone resonance assignment of DENV4 NS3 helicase was started. NS3 helicase protein is a single polypeptide of 451 amino acids protein endowed with three functionally important subdomains. NS3 helicase plays an important role in the Flavivirus replication. Thus, it is a potential drug target (Lescar *et al.,* 2008). The fundamental requirement of any drug discovery process involving NMR spectroscopy is the backbone resonance assignment.

Conformational exchange induced line broadening is the supreme challenge in the process of resonance assignment of DENV4 NS3 helicase. Conformational exchange was observed to be significant to the extent that in the absence of its natural ligand ssRNA, the NMR spectrum of DENV4 NS3 helicase was unsuitable for assignment. However, complex formation with ssRNA produced a spectrum with better uniform peak intensities. Unlike DENV NS2B-NS3pro complex, where the molecular weight is less than 30kDa, DENV4 NS3 helicase has a molecular weight of 51.3kD. The latter poses additional complexity in recording and interpreting the high quality spectra. However, attempts have been made here addressing such a large molecular weight by NMR spectroscopy with positive outlook for the extensive resonance assignment.

Firstly, the routine protocol for the backbone resonance assignment was altered considering the challenging molecular size of the target. High quality spectra recorded were used for the sequence specific assignments. However, the resonance assignment process was severely hampered due to combinatorial explosion of possibilities for a single spin-system. Addressing these problems, selective amino acid labeling was introduced. Although the issue of DENV4 NS3 helicase is the disappearance of resonances due to the conformational exchange problem, selective labeling still might help to contrast/enhance visibility of smaller peaks obscured by the visibility of many

larger peaks. For example, in the work on MjCM protein in its molten globule state Phenylalanine residues cannot be assigned in the uniformly labelled samples due to presence of background of other much stronger peaks. However, selective Phenylalanine labeling eliminated strong background making way to assign the Phenylalanine peaks (Pervushin *et al.,* 2007). Yet another example where weak peaks stemming from peripheral and NH_2 amide groups were clearly observed through selective deuterium labeling in the form DNH (Pervushin *et al.,* 1997b). The price involved in the selective amino acid labeling to the information content obtained was not comparable, forcing the use of an alternative strategy.

As an alternative to selective amino acid labeling, for the alleviation of spectral crowding, Paramagnetic Relaxation Enhancement study has been performed with the aim of delineating residues lining the surface of DENV4 NS3 helicase. Gd(DTPA-BMA) at 0.5mM molar ratio with protein inferred to be exerting much stronger effect than anticipated, suggesting the use of lower ratio.

Alternatively to reduce the spectral complexity and line broadening due to conformational exchange and rapid transverse relaxation, DENV4 NS3 helicase was truncated into fragments of lower molecular size. Care was taken to design functionally active fragments although functional studies were not in the goal list. However, the fragments produced were unstable under standard conditions for NMR spectroscopy as well as other biophysical techniques such as Circular Dichroism spectroscopy. This suggests the extremely dynamical behaviour of DENV4 NS3 helicase and the necessity of all the three domains for its stability and proper folding.

With respect to the question of limits of solution NMR experiments in application to challenging proteins, limitations have been encountered in this work. First, the conformational exchange in association with large molecular size has effects to an extent that they limit detection and resolution of cross peaks. Although similar problem was solved before (De La Cruz *et al.,* 2011), DENV4 NS3 helicase is further complicated by its large molecular size, severely hampering the resonance assignment process. Second is the lack of predictability on the outcome of NMR experiments. The later limitation has been succeeded in the present work by providing an analytical tool and a unified benchmark platform for the analysis and performance prediction of any chosen NMR experiment while the former needs to be resolved.

Future directions on the WIKI based NMR-DB and to complete the backbone resonance assignment of DENV4 NS3 helicase are:

- Development of an online based NMR database, where the presented analytical tools enables the performance prediction of any chosen NMR experiment.
- A GUI mediated interaction with the spin-dynamics tool will allow most effective use of the same.
- Readily available predicted performance metrics for any NMR experiments would be very helpful in choosing the optimal NMR experiment for the protein in hand.
- It is important to have a stable protein inside NMR tube especially when it comes to large molecular sizes. Engineering DENV4 NS3 helicase by introducing mutations at non-conserved regions among its close orthologs could be a viable option in the direction of achieving a stable protein.
- It would be interest to perform PRE/PCS analysis by site-directed spin labeling of DENV4 NS3 helicase.

Abbott, D. W., Higgins, M. A., Hyrnuik, S., Pluvinage, B., Lammerts Van Bueren, A. and Boraston, A. B. (2010) The molecular basis of glycogen breakdown and transport in Streptococcus pneumoniae. *Mol. Microbiol.* **77**, 183-199

Abragam, A. (1961) The principles of nuclear magnetism. *International series of monographs on physics*, Oxford University Press, Oxford, England

Allard, P., Helgstrand, M. and Hard, T. (1998) The complete homogeneous master equation for a heteronuclear two-spin system in the basis of cartesian product operators. *J. Magn. Reson.* **134**, 7-16

Allman, T., Bain, A. D. and Garbow, J. R. (1996) SIMPLTN, a Program for the Simulation of Pulse NMR Spectra. *J. Magn. Reson, Series A.* **123**, 26-31

Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M. J., Gans, P. and Boisbouvier, J. (2012) An optimized isotopic labelling strategy of isoleucine-gamma(2) methyl groups for solution NMR studies of high molecular weight proteins. *Chem. Commun.* **48**, 1434-1436

Bak, M., Rasmussen, J. T. and Nielsen, N. C. (2000) SIMPSON: a general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.* **147**, 296-330

Bartelma, G. and Padmanabhan, R. (2002) Expression, purification, and characterization of the RNA 5'-triphosphatase activity of dengue virus type 2 nonstructural protein 3. *Virology*. **299**, 122-132

Benarroch, D., Selisko, B., Locatelli, G. A., Maga, G., Romette, J. L. and Canard, B. (2004). The RNA helicase, nucleotide 5'-triphosphatase, and RNA 5'-triphosphatase activities of Dengue virus protein NS3 are Mg2+-dependent and require a functional Walker B motif in the helicase catalytic core. *Virology*. **328**, 208-218

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic. acids. Res.* **28**, 235-242

Blanton, W. B. (2003) BlochLib: a fast NMR C++ tool kit. *J. Magn. Reson.* **162**, 269-283

Bock, I. and Rosch, P. (1987) Simulation of the cross-peak fine structure in 2D NMR spectroscopy by analytical calculation of the density operator in a product operator basis. *J. Magn. Reson., (1969)*. **74**, 177-183

Bodenhausen, G. and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185-189

Borowski, P., Mueller, O., Niebuhr, A., Kalitzky, M., Hwang, L. H., Schmitz, H., Siwecka, M. A., Kulikowsk, T. (2000) ATP-binding domain of NTPase/helicase as a target for hepatitis C antiviral therapy. *Acta. Biochim. Pol.* **47**,173-180

Bressanelli, S., Stiasny, K., Allison, S. L., Stura, E. A., Duquerroy, S., Lescar, J., Heinz, F. X. and Rey, F. A. (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO. J.* **23**, 728-738

Brett, D. L., Heinz-J. T., Charles, M. R. (2007) Flaviviridae: The Viruses and Their Replication. Fields Virology (Eds: Knipe, D. M. and Howley, P. M.) 5th Ed, Lippincott-Raven Publishers, Philadelphia

Brystov, V. F. (1976) Spin-spin coupling and the conformational states of peptide systems. *Prog. Nucl. Mag. Res. Sp.* **10**, 41-82

Calisher, C. H. and Gould, E. A. (2003) Taxonomy of the virus family Flaviviridae. *Adv. Virus. Res*. **59**,1-19

Cavanagh, J., Palmer III, A. G., Wright, P. E. and Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear relay spectroscopy. *J. Magn. Reson., (1969)*. **91**, 429-436

Chambers, T. J., Grakoui, A. and Rice, C. M. (1991) Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J. Virol.* **65**, 6042-6050

Chambers, T. J., Nestorowicz, A., Amberg, S. M. and Rice, C. M. (1993) Mutagenesis of the yellow fever virus NS2B protein: effects on proteolytic processing, NS2B-NS3 complex formation, and viral replication. *J. Virol.* **67**, 6797-6807

Christophe, S., Robert, V., David, B., Gottfried, O, and Hubera, T. (2012) Protein Structure Determination from Pseudocontact Shifts Using ROSETTA. *J. Mol. Biol.* **416**, 668-677

Chu, J. J. H., Rajamanonmani, R., Li, J., Bhuvanakantham, R., Lescar, J. and Ng, M. L. (2005) Inhibition of West Nile virus entry by using a recombinant domain III from the envelope glycoprotein. *J. Gen. Virol.* **86**, 405-412

Clark, M. and Thrasher, J. S. (1990) LAOCOON PC: NMR simulation on a personal computer. *J. Chem. Edu.* **67**, 235

Clore, G. M. and Gronenborn, A. M. (1991) Two-, three-, and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *Annu. Rev. Biophys. Biophys. Chem.* **20**,29-63

Clore, G. M. and Gronenborn, A. M. (1998) NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* **2**, 564-570

Constantine, K. L., Mueller, L., Goldfarb, V., Wittekind, M., Metzler, W. J., Yanchunas, J., Jr., Robertson, J. G., Malley, M. F., Friedrichs, M. S. and Farmer II, B. T., (1997) Characterization of NADP+ binding to perdeuterated MurB: backbone atom NMR assignments and chemical-shift changes. *J. Mol. Biol.* **267**, 1223-1246

De Bouregas, F. S. and Waugh, J. S. (1992) ANTIOPE, a program for computer experiments on spin dynamics. *J. Magn. Reson., (1969)*. **96**, 280-289

De La Cruz, L., Nguyen, T. H., Ozawa, K., Shin, J., Graham, B., Huber, T. and Otting, G. (2011) Binding of Low Molecular Weight Inhibitors Promotes Large Conformational Changes in the Dengue Virus NS2B-NS3 Protease: Fold Analysis by Pseudocontact Shifts. *J. Am. Chem. Soc.* **133**, 19205-19215

Dong, H., Zhang, B. and Shi, P. Y. (2008) Flavivirus methyltransferase: A novel antiviral target. *Antiviral. Res.* **80**, 1-10

Ellison, A. (1983) Computer simulation of NMR spectra. *J. Chem. Edu.* **60**, 425

Emsley, J. W. and Feeney, J. (2007) Forty years of Progress in Nuclear Magnetic Resonance Spectroscopy. *Prog. Nucl. Mag. Res. Sp.* **50**, 179-198

Ernst, R. R., Bodenhause, G., Wokaun, A. (Sep 1990) Principles of Nuclear Magnetic Resonance in One and Two Dimensions. *International Series of Monographs on Chemistry*. **14**, 640

Falgout, B., Pethel, M., Zhang, Y. M. and Lai, C. J. (1991) Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J. Virol.* **65**, 2467-2475

Farmer, B. T., 2nd and Mueller, L. (1994) Simultaneous acquisition of [13C,15N]- and [15N,15N]-separated 4D gradient-enhanced NOESY spectra in proteins. *J. Biomol. NMR*. **4**, 673-687

Gans, P., Hamelin, O., Sounier, R., Ayala, I., Dura, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J. and Boisbouvier, J. (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew. Chem.* **49**, 1958-1962

Gardner, K. H. and Kay, L. E. (1997) Production and Incorporation of 15N, 13C, 2H (1H-δ1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J. Am. Chem. Soc.* **119**, 7599-7600

Gardner, K. H., Zhang, X., Gehring, K. and Kay, L. E. (1998) Solution NMR Studies of a 42 KDa Escherichia Coli Maltose Binding Protein/β-Cyclodextrin Complex: Chemical Shift Assignments and Analysis. *J. Am. Chem. Soc.* **120**, 11738-11748

Gesell, J. J., Liu, D., Madison, V. S., Hesson, T., Wang, Y. S., Weber, P. C. and Wyss, D. F. (2001) Design, high-level expression, purification and characterization of soluble fragments of the hepatitis C virus NS3 RNA helicase suitable for NMR-based drug discovery methods and mechanistic studies. *Protein. Eng.* **14**, 573-582

Glaser, S. J., Schulte-Herbruggen, T., Sieveking, M., Schedletzky, O., Nielsen, N. C., Sorensen, O. W. and Griesinger, C. (1998) Unitary control in quantum ensembles: maximizing signal intensity in coherent spectroscopy. *Science*. **280**, 421-424

Goldman, M. (2001) Formal theory of spin--lattice relaxation. *J. Magn. Reson.* **149**, 160-187

Gong, J. K. and Hinze, S. (1995) LAOCOON. An NMR Simulation Program. *J. Chem. Inf. Comput. Sci*. **35**, 651-651

Gorbalenya, A. E. and Koonin, E. V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419-429

Graham, B., Loh, C. T., Swarbrick, J. D., Ung, P., Shin, J., Yagi, H., Jia, X., Chhabra, S., Barlow, N., Pintacuda, G., Huber, T. and Otting, G. (2011) DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. *Bioconj. Chem.* **22**, 2118-2125

Grassmann, C. W., Isken, O. and Behrens, S. E. (1999) Assignment of the multifunctional NS3 protein of bovine viral diarrhea virus during RNA replication: an in vivo and in vitro study. *J. Virol.* **73**, 9196-9205

Grishaev, A., Tugarinov, V., Kay, L. E., Trewhella, J. and Bax, A. (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J. Biomol. NMR*. **40**, 95-106

Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993a) Carbon-13 line narrowing by deuterium decoupling in deuterium/carbon-13/nitrogen-15 enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. *J. Am. Chem. Soc.* **115**, 4369-4370

Grzesiek, S. and Bax, A. (1993b) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J. Biomol. NMR*. **3**, 185-204

Guntert, P. (2011) Calculation of Structures from NMR Restraints. *Protein NMR Spectroscopy: Practical Techniques and Applications* (eds L.-Y. Lian and G. Roberts), John Wiley & Sons, Ltd, Chichester, UK

Guntert, P., Schaefer, N., Otting, G. and Wuthrich, K. (1993) POMA: A Complete Mathematica Implementation of the NMR Product-Operator Formalism. *J. Magn. Reson., Series A*. **101**, 103-105

Helgstrand, M. and Allard, P. (2004) QSim, a program for NMR simulations. *J. Biomol. NMR*. **30**,71-80

Hiller, S., Garces, R. G., Malia, T. J., Orekhov, V. Y., Colombini, M. and Wagner, G. (2008) Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science*. **321**, 1206-1210

Hogben, H. J., Krzystyniak, M., Charnock, G. T., Hore, P. J. and Kuprov, I. (2011) Spinach--a software library for simulation of spin dynamics in large spin systems. *J. Magn. Reson.* **208**, 179-194

Howard, B. R., Endrizzi, J. A. and Remington, S. J. (2000) Crystal structure of Escherichia coli malate synthase G complexed with magnesium and glyoxylate at 2.0 A resolution: mechanistic implications. *Biochemistry*. **39**, 3156-3168

Jaravine, V. A. and Orekhov, V. Y. (2006) Targeted acquisition for real-time NMR spectroscopy. *J. Am. Chem. Soc.* **128**, 13421-13426

Jaravine, V. A., Zhuravleva, A. V., Permi, P., Ibraghimov, I. and Orekhov, V. Y. (2008) Hyperdimensional NMR spectroscopy with nonlinear sampling. *J. Am. Chem. Soc.* **130**, 3927-3936

Jerschow, A. (2005) MathNMR: spin and spatial tensor manipulations in Mathematica. *J. Magn. Reson.* **176**, 7-14

Jia, X., Yagi, H., Su, X. C., Stanton-Cook, M., Huber, T. and Otting, G. (2011) Engineering [Ln(DPA)3] 3- binding sites in proteins: a widely applicable method for tagging proteins with lanthanide ions. *J. Biomol. NMR*. **50**, 411-420

John W, S. (1991) NMR product-operator calculations in mathematica. *J. Magn. Reson., (1969)*. **94**, 612-616

Kanters, R. P. F., Char, B. W. and Addison, A. W. (1993) A Computer-Algebra Application for the Description of NMR Experiments Using the Product-Operator Formalism. *J. Magn. Reson., Series A*. **101**, 23-29

Kay, L., Keifer, P. and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665

Kay, L., Xu, G. Y. and Yamazaki, T. (1994) Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H2O Saturation. *J. Magn. Reson., Series A*. **109**,129-133

Kay, L. E., Clore, G. M., Bax, A. and Gronenborn, A. M. (1990) Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution. *Science*. **249**, 411-414

Kay, L. E. and Gardner, K. H. (1997) Solution NMR spectroscopy beyond 25 kDa. *Curr. Opin. Struct. Biol.* **7**,722-731

Kay, L. E., Scarsdale, J. N., Hare, D. R. and Prestegard, J. H. (1986) Simulation of two-dimensional cross-relaxation spectra in strongly coupled spin systems. *J. Magn. Reson., (1969)*. **68**, 515-525

Keller, R. (2001) *Optimizing The Process Of Nuclear Magnetic Resonance Spectrum Analysis And Computer Aided Resonance Assignment*. Ph.D. Thesis, Eidgenossische Technische Hochschule

Kim, M. K., Cho, M. K., Song, H. E., Kim, D., Park, B. H., Lee, J. H., Kang, G. B., Kim, S. H., Im, Y. J., Lee, D. S. and Eom, S. H. (2007) Crystal structure of UDP-N-acetylenolpyruvylglucosamine reductase (MurB) from Thermus caldophilus. *Proteins*. **66**, 751-754

Klages, J., Coles, M. and Kessler, H. (2007) NMR-based screening: a powerful tool in fragment-based drug discovery. *The Analyst*, **132**, 693-705

Konrat, R., Yang, D. and Kay, L. E. (1999) A 4D TROSY-based pulse scheme for correlating 1HNi,15Ni,13Cαi,13C'i-1 chemical shifts in high molecular weight, 15N,13C, 2H labeled proteins. *J. Biomol. NMR*. **15**, 309-313

Krishnan, V. V. and Cosman, M. (1998) An empirical relationship between rotational correlation time and solvent accessible surface area. *J. Biomol. NMR*. **12**, 177-182

Kumar, A. and Madhu, P. K. (1996a) Cross-correlations in multispin relaxation. *Concept. Magn. Reson.* **8**, 139-160

Kumar, P. and Kumar, A. (1996b) Effect of Remote Cross Correlations on Transverse Relaxation. *J. Magn. Reson.*, *Series A*. **119**, 29-37

Kuprov, I. (2008) Polynomially scaling spin dynamics II: further state-space compression using Krylov subspace techniques and zero track elimination. *J. Magn. Reson.* **195**, 45-51

Kuprov, I., Wagner-Rundell, N. and Hore, P. J. (2007) Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction. *J. Magn. Reson.* **189**, 241-250

Lee, D., Hilty, C., Wider, G. and Wuthrich, K. (2006) Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.* **178**, 72-76

Lescar, J., Luo, D., Xu, T., Sampath, A., Lim, S. P., Canard, B. and Vasudevan, S. G. (2008) Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. *Antiviral. Res.* **80**, 94-101

Li, H., Clum, S., You, S., Ebner, K. E. and Padmanabhan, R. (1999) The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. *J. Virol.* **73**, 3108-3116

Lindenbach, B. D. and Rice, C. M. (2003) Molecular biology of flaviviruses. *Adv. Virus. Res*. **59**, 23-61

Liu, D., Xu, R. and Cowburn, D. (2009) Segmental isotopic labeling of proteins for nuclear magnetic resonance. *Meth. Enzymol.* **462**, 151-175

Liu, W. J., Sedlak, P. L., Kondratieva, N. and Khromykh, A. A. (2002) Complementation analysis of the flavivirus Kunjin NS3 and NS5 proteins defines the minimal regions essential for formation of a replication complex and shows a requirement of NS3 in cis for virus assembly. *J. Virol.* **76**, 10766-10775

Loth, K., Pelupessy, P. and Bodenhausen, G. (2005) Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy. *J. Am. Chem. Soc.* **127**, 6062-6068

Luo, D., Xu, T., Hunke, C., Gruber, G., Vasudevan, S. G. and Lescar, J. (2008a) Crystal structure of the NS3 protease-helicase from dengue virus. *J. Virol.* **82**, 173-183

Luo, D., Xu, T., Watson, R. P., Scherer-Becker, D., Sampath, A., Jahnke, W., Yeong, S. S., Wang, C. H., Lim, S. P., Strongin, A., Vasudevan, S. G. and Lescar, J. (2008b) Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO. J.* **27**, 3209-3219

Man, B., Su, X. C., Liang, H., Simonsen, S., Huber, T., Messerle, B. A. and Otting, G. (2010) 3-Mercapto-2,6-pyridinedicarboxylic acid: a small lanthanide-binding tag for protein studies by NMR spectroscopy. *Chemistry*. **16**, 3827-3832

Mancini, E. J., Assenberg, R., Verma, A., Walter, T. S., Tuma, R., Grimes, J. M., Owens, R. J. and Stuart, D. I. (2007) Structure of the Murray Valley encephalitis virus RNA helicase at 1.9 Angstrom resolution. *Protein. Sci*. **16**, 2294-2300

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. and Clore, G. M. (1989) 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 1β. *Biochemistry*. **28**, 6150-6156

Masse, J. E. and Keller, R. (2005) AutoLink: automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. *J. Magn. Reson.* **174**, 133-151

Mastrangelo, E., Bollati, M., Milani, M., Brisbarre, N., De Lamballerie, X., Coutard, B., Canard, B., Khromykh, A. and Bolognesi, M. (2006) Preliminary crystallographic characterization of an RNA helicase from Kunjin virus. *Acta. Cryst. F.* **62**, 876-879

Matsuo, H., Kupce, Ƒ., Li, H. and Wagner, G. (1996) Use of Selective CαPulses for Improvement of HN(CA)CO-D and HN(COCA)NH-D Experiments. *J. Magn. Reson., Series B*. **111**, 194-198

Matusan, A. E., Pryor, M. J., Davidson, A. D. and Wright, P. J. (2001) Mutagenesis of the Dengue virus type 2 NS3 protein within and outside helicase motifs: effects on enzyme activity and virus replication. *J. Virol.* **75**, 9633-9643

Modis, Y., Ogata, S., Clements, D. and Harrison, S. C. (2004) Structure of the dengue virus envelope protein after membrane fusion. *Nature*. **427**, 313-319

Mueller, L. (1979) Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence. *J. Am. Chem. Soc.* **101**, 4481-4484

Muona, M., Aranko, A. S., Raulinaitis, V. and Iwai, H. (2010) Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro. *Nature protocols*. **5**, 574-587

Nakashima, T. T. and Mcclung, R. E. D. (1986) Simulation of two-dimensional NMR spectra using product operators in the spherical basis. *J. Magn. Reson., (1969)*. **70**, 187-203

Nicholas, P., Fushman, D., Ruchinsky, V. and Cowburn, D. (2000) The virtual NMR spectrometer: a computer program for efficient simulation of NMR experiments involving pulsed field gradients. *J. Magn. Reson.* **145**, 262-275

Nietlispach, D. (2005) Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. *J. Biomol. NMR*. **31**, 161-166

Otting, G. (2010) Protein NMR Using Paramagnetic Ions. *Annu. Rev. Biophys.* **39**, 387-405

Palmer III, A. G., Cavanagh, J., Wright, P. E. and Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson., (1969)*. **93**, 151-170

Patkar, C. G. and Kuhn, R. J. (2008) Yellow Fever virus NS3 plays an essential role in virus assembly independent of its known enzymatic functions. *J. Virol.* **82**, 3342-3352

Pervushin, K. (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q. Rev. Biophys.* **33**,161-197

Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1998a) Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems in 13C-Labeled Proteins. *J. Am. Chem. Soc.* **120**, 6394-6400

Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997a) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Nat. Acad. Sci. USA*, **94**, 12366-12371

Pervushin, K., Vamvaca, K., Vogeli, B. and Hilvert, D. (2007) Structure and dynamics of a molten globular enzyme. *Nat. Struct. Mol. Biol.* **14**, 1202-1206

Pervushin, K., Vogeli, B., Heinz, T. N. and Hunenberger, P. H. (2005) Measuring 1H-1H and 1H-13C RDCs in methyl groups: example of pulse sequences with numerically optimized coherence transfer schemes. *J. Magn. Reson.* **172**, 36-47

Pervushin, K., Wider, G. and Wuthrich, K. (1997b) Deuterium Relaxation in a Uniformly 15N-Labeled Homeodomain and Its DNA Complex1. *J. Am. Chem. Soc.* **119**, 3842-3843

Pervushin, K. V., Wider, G. and Wuthrich, K. (1998b) Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. *J. Biomol. NMR*. **12**, 345-348

Pintacuda, G. and Otting, G. (2002) Identification of protein surfaces by NMR measurements with a pramagnetic Gd(III) chelate. *J. Am. Chem. Soc.* **124**, 372-373

Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*. **2**, 661-665

Pucci, M. J., Discotto, L. F. and Dougherty, T. J. (1992) Cloning and identification of the Escherichia coli murB DNA sequence, which encodes UDP-N-acetylenolpyruvoylglucosamine reductase. *J. Bacteriol.* **174**, 1690-1693

Riek, R., Pervushin, K. and Wuthrich, K. (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends. Biochem. Sci.* **25**, 462-468

Riek, R., Wider, G., Pervushin, K. and Wuthrich, K. (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc. Nat. Acad. Sci.* **96**, 4918-4923

Roussel, A., Lescar, J., Vaney, M.-C., Wengler, G., Wengler, G. and Rey, F. A. (2006) Structure and Interactions at the Viral Surface of the Envelope Protein E1 of Semliki Forest Virus. *Structure*. **14**, 75-86

Saibil, H. R. (2000) Macromolecular structure determination by cryo-electron microscopy. *Acta. Cryst. F*. **56**, 1215-1222

Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Walchli, M., Smith, B. O., Shirakawa, M., Guntert, P. and Ito, Y. (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature*. **458**, 102-105

Salzmann, M., Pervushin, K., Wider, G., Senn, H. and Wuthrich, K. (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Nat. Acad. Sci. USA*. **95**, 13585-13590

Salzmann, M., Wider, G., Pervushin, K., Senn, H. and Wuthrich, K. (1999) TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **121**, 844-848

Sambrook, J. and Russell, D. (2001) Molecular Cloning: A Laboratory Manual.

Sampath, A., Xu, T., Chao, A., Luo, D., Lescar, J. and Vasudevan, S. G. (2006) Structure-based mutational analysis of the NS3 helicase from dengue virus. *J. Virol.* **80**, 6686-6690

Schenker, K. V. and Von Philipsborn, W. (1985) Optimization of INEPT and DEPT experiments for spin systems with hetero- and homonuclear couplings. *J. Magn. Reson., (1969)*. **61**, 294-305

Senthamarai, R. R., Kuprov, I. and Pervushin, K. (2010) Benchmarking NMR experiments: a relational database of protein pulse sequences. *J. Magn. Reson.* **203**, 129-137

Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H. and Kay, L. E. (1996) Assignment of 15N, 13Cα, 13Cβ, and HN Resonances in an 15N,13C,2H Labeled 64 kDa Trp Repressor-Operator Complex Using Triple-Resonance NMR Spectroscopy and 2H-Decoupling. *J. Am. Chem. Soc.* **118**, 6570-6579

Smith, S. A., Levante, T. O., Meier, B. H. and Ernst, R. R. (1994) Computer Simulations in Magnetic Resonance. An Object-Oriented Programming Approach. *J. Magn. Reson., Series A*. **106**, 75-105

Speroni, S., De Colibus, L., Mastrangelo, E., Gould, E., Coutard, B., Forrester, N. L., Blanc, S., Canard, B. and Mattevi, A. (2008) Structure and biochemical analysis of Kokobera virus helicase. *Proteins*. **70**, 1120-1123

Sun, H., Sanders, L. K. and Oldfield, E. (2002) Carbon-13 NMR shielding in the twenty common amino acids: comparisons with experimental results in proteins. *J. Am. Chem. Soc.* **124**, 5486-5495

Takeuchi, K., Ng, E., Malia, T. J. and Wagner, G. (2007) 1-13C amino acid selective labeling in a 2H15N background for NMR studies of large proteins. *J. Biomol. NMR*. **38**, 89-98

Tugarinov, V., Choy, W. Y., Orekhov, V. Y. and Kay, L. E. (2005a) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Nat. Acad. Sci. USA*. **102**, 622-627

Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. and Kay, L. E. (2003a) Cross-Correlated Relaxation Enhanced 1H,àí13C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J. Am. Chem. Soc.* **125**, 10420-10428

Tugarinov, V., Kanelis, V. and Kay, L. E. (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature protocols*. **1**, 749-754

Tugarinov, V. and Kay, L. E. (2003b) Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. *J. Am. Chem. Soc.* **125**, 13868-13878

Tugarinov, V., Kay, L. E., Ibraghimov, I. and Orekhov, V. Y. (2005b) High-Resolution Four-Dimensional 1H-13C NOE Spectroscopy using Methyl-TROSY, Sparse Data Acquisition, and Multidimensional Decomposition. *J. Am. Chem. Soc.* **127**, 2767-2775

Tugarinov, V., Muhandiram, R., Ayed, A. and Kay, L. E. (2002) Four-Dimensional NMR Spectroscopy of a 723-Residue Protein: Chemical Shift Assignments and Secondary Structure of Malate Synthase G. *J. Am. Chem. Soc.* **124**, 10025-10035

Tugarinov, V., Sprangers, R. and Kay, L. E. (2004) Line Narrowing in Methyl-TROSY Using Zero-Quantum 1H-13C NMR Spectroscopy. *J. Am. Chem. Soc.* **126**, 4921-4925

Umareddy, I., Pluquet, O., Wang, Q., Vasudevan, S., Chevet, E. and Gu, F. (2007) Dengue virus serotype infection specifies the activation of the unfolded protein response. *Virol. J.*, 10.1186/1743-422X-4-91

Veshtort, M. and Griffin, R. G. (2006) SPINEVOLUTION: a powerful tool for the simulation of solid and liquid state NMR experiments. *J. Magn. Reson.* **178**, 248-282

Villinger, S., Briones, R., Giller, K., Zachariae, U., Lange, A., De Groot, B. L., Griesinger, C., Becker, S. and Zweckstetter, M. (2010) Functional dynamics in the voltage-dependent anion channel. *Proc. Nat. Acad. Sci. USA*. **107**, 22546-22551

Wagner, G. (1993) Prospects for NMR of large proteins. *J. Biomol. NMR*. **3**, 375-385

Waldherr, G., Beck, J., Steiner, M., Neumann, P., Gali, A., Frauenheim, T., Jelezko, F. and Wrachtrup, J. (2011) Dark states of single nitrogen-vacancy centers in diamond unraveled by single shot NMR. *Phys. Rev. Lett.* **106**, 157601, 10.1103/PhysRevLett.106.157601

Wang, C. C., Huang, Z. S., Chiang, P. L., Chen, C. T. and Wu, H. N. (2009) Analysis of the nucleoside triphosphatase, RNA triphosphatase, and unwinding activities of the helicase domain of dengue virus NS3 protein. *FEBS. Lett.* **583**, 691-696

Wider, G. and Wuthrich, K. (1999) NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr. Opin. Struct. Biol.* **9**, 594-601

Wong, L. E., Masse, J. E., Jaravine, V., Orekhov, V. and Pervushin, K. (2008) Automatic assignment of protein backbone resonances by direct spectrum inspection in targeted acquisition of NMR data. *J. Biomol. NMR*. **42**, 77-86

Wu, J., Bera, A. K., Kuhn, R. J. and Smith, J. L. (2005) Structure of the Flavivirus helicase: implications for catalytic activity, protein interactions, and proteolytic processing. *J. Virol.* **79**, 10268-10277

Wuthrich, K. (1998) The second decade--into the third millenium. *Nat. Struct. Biol.* **5 Suppl**, 492-495

Xu, R., Ayers, B., Cowburn, D. and Muir, T. W. (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc. Nat. Acad. Sci. USA*. **96**, 388-393

Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Chene, P., Vasudevan, S. G. and Lescar, J. (2005) Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A. *J. Virol.* **79**, 10278-10288

Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Luo, D., Chene, P., Vasudevan, S. G. and Lescar, J. (2006) Towards the design of flavivirus helicase/NTPase inhibitors: crystallographic and mutagenesis studies of the dengue virus NS3 helicase catalytic domain. *Novartis Foundation symposium*. **277**, 87-97, discussion 97-101

Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Luo, D., Chene, P., Vasudevan, S. G. and Lescar, J. (2008) Towards the Design of Flavivirus Helicase/NTPase Inhibitors: Crystallographic and Mutagenesis Studies of the Dengue Virus NS3 Helicase Catalytic Domain. *New Treatment Strategies for Dengue and Other Flaviviral Diseases*. 87-101

Yagi, H., Banerjee, D., Graham, B., Huber, T., Goldfarb, D. and Otting, G. (2011) Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR. *J. Am. Chem. Soc.* **133**, 10418-10421

Yagi, H., Loscha, K. V., Su, X. C., Stanton-Cook, M., Huber, T. and Otting, G. (2010) Tunable paramagnetic relaxation enhancements by [Gd(DPA)(3)] (3-) for protein structure analysis. *J. Biomol. NMR*. **47**, 143-153

Yamashita, T., Unno, H., Mori, Y., Tani, H., Moriishi, K., Takamizawa, A., Agoh, M., Tsukihara, T. and Matsuura, Y. (2008) Crystal structure of the catalytic domain of Japanese encephalitis virus NS3 helicase/nucleoside triphosphatase at a resolution of 1.8 A. *Virology*. **373**, 426-436

Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. and Nakamura, H. (1998) Segmental Isotope Labeling for Protein NMR Using Peptide Splicing. *J. Am. Chem. Soc.* **120**, 5591-5592

Yang, D. and Kay, L. E. (1999a) Improved 1HN-detected triple resonance TROSY-based experiments. *J. Biomol. NMR*. **13**, 3-10

Yang, D. and Kay, L. E. (1999b) TROSY Triple-Resonance Four-Dimensional NMR Spectroscopy of a 46 ns Tumbling Protein. *J. Am. Chem. Soc.* **121**, 2571-2575

Yon, C., Teramoto, T., Mueller, N., Phelan, J., Ganesh, V. K., Murthy, K. H. and Padmanabhan, R. (2005) Modulation of the nucleoside triphosphatase/RNA helicase and 5'-RNA triphosphatase activities of Dengue virus type 2 nonstructural protein 3 (NS3) by interaction with NS5, the RNA-dependent RNA polymerase. *J. Biol. Chem.*, **280**, 27412-27419

Zhu, G., Xia, Y., Lin, D. and Gao, X. (2004) TROSY-based correlation and NOE spectroscopy for NMR structural studies of large proteins. *Methods. Mol. Biol.* **278**, 57-78

Appendix – A

Spatial coordinates and chemical shielding parameters of the generic spin-system constructed for the simulation of solution state NMR experiments

Isotopes	Spin Names	Offsets (ppm)	Coordinates			CSA (ppm)		
			X	Y	Z	σ_{xx}	σ_{yy}	σ_{zz}
^{15}N	N-1	119.80	-3.138	-16.334	5.186	19.2	91.6	-110.8
^{1}H	H-1	8.20	-2.456	-17.063	5.031	1.0	6.4	-7.5
^{13}C	Cα-1	56.60	-3.030	-15.529	6.404	0.4	14.3	-14.7
^{1}H	Hα-1	4.30	-3.925	-14.912	6.487	5.0	5.0	-10.0
^{13}C	C-1	176.30	-1.814	-14.631	6.357	97.6	-4.8	-92.8
^{15}N	N	122.20	-1.984	-13.367	6.738	43.3	82.9	-124.1
^{1}H	H	8.20	-2.896	-13.043	7.027	5.0	6.4	-6.9
^{13}C	Cα	55.60	-0.868	-12.424	6.749	2.9	12.7	-15.4
^{1}H	Hα	4.30	0.029	-13.000	6.522	5.0	5.0	-10.0
^{13}C	Cβ	42.30	-1.029	-11.366	5.656	-2.0	11.7	-9.8
^{1}H	Hβ1	1.60	-1.945	-10.820	5.881	5.0	5.0	-10.0
^{1}H	Hβ2	1.50	-0.175	-10.695	5.742	5.0	5.0	-10.0
^{13}C	Cγ	26.80	-1.115	-11.782	4.187	-2.0	11.7	-9.8
^{1}H	Hγ	1.50	-2.023	-12.370	4.050	5.0	5.0	-10.0
^{13}C	Cδ1	24.70	-1.201	-10.547	3.315	-2.0	11.7	-9.8
^{1}H	Hδ11	0.75	-1.291	-10.842	2.269	5.0	5.0	-10.0
^{1}H	Hδ12	0.75	-2.074	-9.961	3.602	5.0	5.0	-10.0
^{1}H	Hδ13	0.75	-0.301	-9.947	3.446	5.0	5.0	-10.0
^{13}C	Cδ2	24.10	0.061	-12.628	3.769	-2.0	11.7	-9.8
^{1}H	Hδ21	0.73	-0.039	-12.919	2.723	5.0	5.0	-10.0
^{1}H	Hδ22	0.73	0.980	-12.058	3.897	5.0	5.0	-10.0
^{1}H	Hδ23	0.73	0.100	-13.524	4.388	5.0	5.0	-10.0
^{13}C	C	176.90	-0.644	-11.752	8.105	-5.6	95.0	-90.1
^{15}N	N+1	120.70	0.626	-11.697	8.508	-136.0	76.7	59.1
^{1}H	H+1	8.30	1.333	-12.157	7.952	-4.6	4.1	0.5
^{13}C	Cα+1	57.40	1.064	-11.002	9.722	-12.5	11.7	0.7
^{1}H	Hα+1	4.30	0.318	-10.251	9.982	5.0	5.0	-10.0
^{1}H	H1	1.980	-6.635	-14.491	7.235	5.0	5.0	-10.0
^{1}H	H2	0.80	3.330	-12.701	8.884	5.0	5.0	-10.0
^{1}H	H3	7.05	2.550	-13.892	6.500	5.0	5.0	-10.0
^{1}H	H4	0.82	-5.600	-15.990	6.600	5.0	5.0	-10.0
^{1}H	H5	7.02	-4.900	-13.199	7.550	5.0	5.0	-10.0

Appendix – B

The total and ^{1}H-^{1}H dipolar densities in the generic spin system of Figure 2.4

Protons	Total dipolar densities ($x10^{-9}$, s^{-2})	^{1}H-^{1}H dipolar densities ($x10^{-9}$, s^{-2})
$^{1}H_{N-1}$	0.92	0.175
$^{1}H_{\alpha-1}$	5.35	2.609
$^{1}H_{N}$	2.84	2.000
$^{1}H_{\alpha}$	5.26	2.424
$^{1}H_{\beta1}$	6.29	3.471
$^{1}H_{\beta2}$	6.49	3.655
$^{1}H_{\gamma}$	4.98	2.140
$^{1}H1_{\delta1}$	8.05	5.314
$^{1}H1_{\delta2}$	8.10	5.345
$^{1}H1_{\delta3}$	8.20	5.434
$^{1}H2_{\delta1}$	8.11	5.363
$^{1}H2_{\delta2}$	8.22	5.447
$^{1}H2_{\delta3}$	8.53	5.760
$^{1}H_{N+1}$	2.67	1.840
$^{1}H_{\alpha+1}$	2.87	0.162

The dipolar density at the position of the spin j is computed as,

$$\sum_{i}^{N}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\left(\gamma_j\gamma_i\right)^2}{r_{ji}^6},i\neq j$$

where the sum goes over all or selected spins in the protein.

Averaged dipolar densities computed for NS3 helicase (PDB 2JLQ)

Proton types	Total number of each proton type in NS3Helicase (2JLQ.PDB)	^{1}H-^{1}H dipolar densities±SD ($x10^{-9}$, s^{-2})
$^{1}H_{N}$	421	3.707±6.44
$^{1}H_{\alpha}$	422	2.148±0.74
$^{1}H1_{\beta}$	334	4.461±1.06
$^{1}H2_{\beta}$	334	4.347±0.99
$^{1}H1_{\gamma}$	187	4.804±4.31
$^{1}H2_{\gamma}$	154	4.235±0.72
$^{1}H1_{\gamma2}$	88	6.477±0.62
$^{1}H1_{\gamma1}$	55	5.552±1.03
$^{1}H1_{\delta}$	129	4.482±3.16
$^{1}H2_{\delta}$	129	4.475±3.16
$^{1}H1_{\delta1}$	62	6.397±0.63
$^{1}H1_{\delta2}$	44	5.912±1.40
$^{1}H1_{\varepsilon}$	80	3.684±2.28
$^{1}H2_{\varepsilon}$	71	3.822±1.88

Appendix – C

Simulated performance of TROSY, Tr-HNCA and Tr-HNCACB experiments for uniformly $^{2}H,^{13}C,^{15}N$-labeled protein with tc = 40 ns at 950 MHz.

Experiments	Performance Metric						
	(3) b_{max}/b^{r}_{max}	(4) T (ms)	(5) R_{PP} (s^{-1})	(6) R_{CS} (s^{-1})	(7) S/N^{2D}_{H-N}	(8) S/N^{2D}_{H-C}	(9) S/N^{3D}_{HNC}
^{1}H-^{15}N TROSY	0.380/0.184	17.8	40.0	20.1 (^{15}N)	14.8	-	-
TROSY - HNCA	0.110/0.020	68.8	24.8	9.0 ($^{13}C\alpha$)	1.38	0.95	5.6
TROSY – HNCACB, T_{CaCb} = 14.4ms	0.108/0.016	97.6	19.6	7.0 ($^{13}C\beta$)	0.91	0.82	2.7

Experimental parameters:
TROSY: ns = 16, sw(^{15}N) = 3650 Hz, time domain = 128 complex points.
TROSY-HNCA: ns = 16, sw(^{15}N) = 3650 Hz, sw(^{13}C) = 6288 Hz, time domain = 48 (^{15}N) and 50 (^{13}C) complex points.
TROSY-HNCACB: ns = 16, sw(^{15}N) = 3650 Hz, sw(^{13}C) = 18864 Hz, time domain = 48 (^{15}N) and 80 (^{13}C) complex points. TCaCb = 14.4 ms.

Simulated performance of TROSY, Tr-HNCA and Tr-HNCACB experiments for uniformly 13C,15N-labeled protein with tc = 20 ns at 700 MHz.

Experiments (1)	Performance Metric (2)						
	(3) b_{max}/b^{r}_{max}	(4) T (ms)	(5) R_{PP} (s^{-1})	(6) R_{CS} (s^{-1})	(7) S/N^{2D}_{H-N}	(8) S/N^{2D}_{H-C}	(9) S/N^{3D}_{HNC}
^{1}H-^{15}N TROSY	0.38/0.23	17.8	28.0	9.5(^{15}N)	20.2	-	-
TROSY - HNCA	0.110/0.04	68.8	15.13	29.4($^{13}C\alpha$)	2.02	1.06	8.2
TROSY – HNCACB T_{CaCb} = 7.2ms	0.047/0.01	83.2	16.00	31.4($^{13}C\alpha$)	0.75	0.55	1.7
TROSY – HNCACB T_{CaCb} = 7.2ms	0.062/0.01	83.2	18.12	48.4($^{13}C\beta$)	0.80	0.60	2.0

Experimental parameters:
TROSY: ns = 8, sw(^{15}N) = 2838 Hz, time domain = 128 complex points.
TROSY-HNCA: ns = 16, sw(^{15}N) = 2838 Hz, sw(^{13}C) = 7394 Hz, time domain = 30 (^{15}N) and 50 (^{13}C) complex points.
TROSY-HNCACB: ns = 16, sw(^{15}N) = 2838 Hz, sw(^{13}C) = 12323 Hz, time domain = 40 (^{15}N) and 50 (^{13}C) complex points. TCaCb = 7.2 ms.
Sensitivity in the seeding 1D 1H is exchanged with the experimental 1D 1H sensitivity for sensitivity prediction, in order to compare the experimental sensitivities in higher Dimensionality experiments.

[1]Experiments: HMQC(Mueller 1979) , HSQC(Bodenhausen *et al.,* 1980) , HSQC-PEP-PFG (Cavanagh *et al.,* 1991) (Palmer III *et al.,* 1991) (Kay *et al.,* 1992) , TROSY(Pervushin *et al.,* 1998b) , TROSY-PFG(Pervushin *et al.,* 1998b) , TROSY-D.N.(Nietlispach 2005) .
[2]In t1 dimension 128 complex increments with spectral width sw(^{15}N) = 2980 Hz at B_0 = 700 MHz and 950 MHz are simulated.
[3]Maximum transfer bounds computed in the absence and presence of relaxation.
[4]Time duration of the pulse sequence with chemical shift evolution periods set to the minimal values.
[5]Effective relaxation rate of the entire experiment with chemical shift evolution periods set to the minimal values.
[6]Effective relaxation rates for the coherences used to indirectly encode chemical shifts.
[7]Expected S/N in ^{1}H-^{15}N dimension.
[8]Expected S/N in ^{1}H-^{13}C dimension.
[9]Expected S/N in 3D ^{1}H-^{15}N-^{13}C spectrum.

Appendix – D

Residue	$^{1}H_{N}$[ppm]	^{15}N[ppm]	$^{13}C\alpha$[ppm]	$^{13}C\beta$[ppm]
D8	8.51	120.71	53.85	40.48
Y9	8.32	122.11	57.85	38.51
E10	7.85	127.46	54.78	29.57
V11	8.17	122.96	61.13	31.74
D12	8.11	127.18	52.98	40.94
E13	8.51	123.09	57.85	28.39
D14	8.5	120.42	56.29	40.41
I15	8.2	120.84	54.97	
F16	6.52	119.3	57.87	38.84
R17	8.19	120.99	59.38	29.24
K18	8.64	117.33	57.4	27.27
K19	8.19	109.4	63.38	31.08
R20	8.41	128.3	53.88	32.2
L21	8.79	119.3	56.95	40.15
T22	10.01	124.92	60.28	
L113	8.24	130.27	54.6	42.12
I114	9.75	129.07	59.86	41.07
V115	7.99	128.58	53.19	28.65
T179	6.89	109.17	56.29	64.47
G180	8.31	110.7	43.94	
F237	8.81	117.04	55.89	44.22
V238	9.06	123.24	60.28	32.79
V239	8.20	129.71	60.49	31.93
T240	9.18	122.83	56.74	69.72
V265	7.97	120.99	61.82	
I266	8.23	126.05	60.28	37.32
L267	6.9	120.98	53.48	42.38
T268	8.23	109.73	60.04	68.34
S385	8.31	112.55	59.67	

A386	6.76	123.8	51.87	18.27
G387	7.87	106.36	45.26	
I388	6.91	123.52	60.7	37.98
S389	8.79	122.96	57.85	63.87
F397	7.57	118.2	52.32	
T398	7.20	113.67	58.51	62.82
G399	7.8	108.19	44.81	

Appendix - E

Fragment ID	Amino acid Sequence
d1,2a-NS3h$_{wt}$	GSAMGEPDYEVDEDIFRKKRLTIMDLHPGAGKTKRILPSIVREALKRRLRTLI LAPTRVVAAEMEEALRGLPIRYQTPAVKSDHTGREIVDLMCHATFTTRLLSS TRVPNYNLIVMDEAHFTDPCSVAARGYISTRVEMGEAAAIFMTATPPGSIDP FPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAGNDIANCLRK SGKRVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDPRRCL KPVILTDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSG
d1,2b-NS3h$_{wt}$	DPFPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAGNDIANCL RKSGKRVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDPRR CLKPVILTDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLK NDEDHAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKT FVELMRRGDLPVWLSYKVASAGISYKDREWCFTGERNNQILEENMEVEIWT REGEKKKLRPKWLDARVYADPMALKDFKEFASGRK
d1,2Δ-NS3h	GSDMGEPDYEVDEDIFRKKRLTIMDLHPGAGKTKRILPSIVREALKRRLRTLI LAPTRVVADEMEEALRGLPIRYQTPAVKSDHTGREIVDLMCHATFTTRLLSS TRVPNYNLIVMDEAHFTDPCSVAARGYISTRVEMGEAAAIFMTATPPGSIDP FPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAGNDIANCLRK SGKRVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDPRRCL KPVILTDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGFSGDPL KNDEDLE

Author's Publication:

Senthamarai, R. R., Kuprov, I. and Pervushin, K. (2010) Benchmarking NMR experiments: a relational database of protein pulse sequences. *J. Magn. Reson.* **203**, 129-137

Poster Presentations:

Alistair Irvine, **Russell Senthamarai**, Konstantin Pervushin, Social Network for setup, execution, sharing and analytics of complex NMR experiments. *4th Asia-Pacific NMR conference - 2011*. Beijing, China.

Russell R. P. Senthamarai, Ilya Kuprov, Konstantin Pervushin, Unified framework for comparison of performance of libraries of NMR experiments in applications to challenging proteins. *World Wide Magnetic Resonance - 2010.* Florence, Italy.

Russell R. P. Senthamarai, Konstantin Pervushin, Annotation of relaxation pathways in solution NMR experiments, *Indian Institute of Science Centenary Symposium on Future Directions in NMR*, October 19-22, 2008. Bangalore, India.

Oral Presentation:

NMRGen: Automated setup and benchmarking of solution NMR experiments. Homi Bhabha Centenary School on Relaxation in NMR, February16-20, 2009. Mumbai, India.

Awards:

Article entitled "Benchmarking NMR experiments: a relational database of protein pulse sequences" ranked No:1 among the top 25 articles of Journal of Magnetic Resonance from Jan 2010 to Mar 2010.

Article entitled "Benchmarking NMR experiments: a relational database of protein pulse sequences" ranked No: 9 among the top 25 articles of Journal of Magnetic Resonance from Oct 2009 to Sep 2010.