UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

_J_-coupling predictor function building and large–scale NMR simulations of proteins and nucleic acids

by

Zenawi Teklay Welderufael

Thesis for the degree of Doctor of Philosophy

September 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

Thesis for the degree of Doctor of Philosophy

_J_-coupling predictor function building and large–scale NMR simulations of proteins and nucleic acids

Zenawi Teklay Welderufael

This thesis discusses large scale quantum mechanical NMR simulations and fitting works, and deals with biomolecular spin systems.

Chapter one introduces the theoretical background that underpins the simulation work. It describes spin and it properties, the interactions of spin with the magnetic field, the environment and it self and the Hamiltonians that describe those interactions. It also covers the Bloch equations that is normally used to predict the longitudinal and transverse relaxations. It further briefly describes the common relaxation mechanisms that are available for a spin system. It also included NMR experiments that are often used in the study of protein and nucleic acid systems.

Chapter 2 summarizes the building of *J*-coupling predictor function. It provides the detail on how the estimator function was built. The function is required for NMR simulations and incorporates the functionality of *J*-coupling estimation, which provides *J*-coupling values for proteins and nucleic acids. After building the estimator function NOESY spectra simulation of ubiquitin and RNA stem loops spin systems was performed. Chapter 3 then discusses our simulation method, treating the system quantum mechanically along with an advanced and detailed description of NOE using Redfield relaxation theory.

Chapter 4 reports on large–scale simulations and fitting of an exchanging ^{1}H-^{15}N HSQC spectra of calmodulin (PDB: 1CLL) upon stepwise addition of a ligand, that was analysed using a

formalism that operates in the direct product of spin state space and chemical state space. It also presents the least squares technique for the fitting procedure and the Nelder-Mead simplex minimization technique used in the process. The chapter also highlights the data collated from the fitting work.

Finally, chapter 5 concludes the thesis with the simulation of NOESY stem loops RNA spin system and fitting with respect to correlation time and frequency offset. It also explains how we fitted the data from theory to those from the experiment and why the diagonal peaks were eliminated from the fitting.

Acknowledgments

First and for most I would like to thank my supervisor Dr Ilya Kuprov for giving me the platform to grow as a scientist. I would also like to express my special thanks for the invaluable encouragement and support who provided me and for the excellent supervision and guidance throughout my PhD time. He was also an inspiration for me. I would also like to extend my thanks to Prof Jeremy Frey for agreeing to serve as my advisor and for lots of good advice.

I really enjoyed the collaborations I had with Joern Warner and Kelly Hooper from Southampton's Biological Sciences; and Gerhard Wagner, Scott Robson, and Imai Shunsuke of the Harvard Medical School. Their support is greatly appreciated.

I am thankful to the school of chemistry for the financial contribution towards a season ticket and latter towards rent when I relocated at Southampton.

Massive cheers goes to the IK group, both to previous members (Luke, Dima, Andreas, and Zihan) and current members (David, Liza, and Ahmed).

Big thanks also goes to Southampton's Magnetic Resonance research group for creating a comfortable, enjoyable working environment and being helpful.

The funding from the EPSRC via the doctoral training centre for integrated magnetic resonance based at the University of Warwick is highly acknowledged.

At last but not the least, I would like to express my special appreciation and thanks to Eden for the love and care she showed me throughout the last four hectic years and also for her prayers that helped me sustain thus far. My appreciation also goes to my lovely daughters Nael and Naomi for being good and cheerful girls. I would like to extend my warmest thanks to my parents for their love and support throughout my life. My siblings and cousins deserve my wholehearted thanks also. To Eritrean Bethel Church London, thank you for the spiritual support and encouragement. To Dr Biniam Tesfatsion, many thanks for the constant and invaluable advice he offered me during my PhD as well.

List of abbreviations

Abbreviations	Descriptions
B3LYP	Becke 3-term correlation, Lee-Yang-Parr exchange
BMRB	biological magnetic resonance bank
BRW	Bloch-Redfield-Wangsness
cc	correlation consistency
CC	coupled cluster
COSY	correlation spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CPU	central processing unit
CSA	chemical shift anisotropy
DD	dipole-dipole
DFT	density functional theory
DNA	deoxyribonucleic acid
DSO	diamagnetic spin orbit
E.COSY	exclusive correlation spectroscopy
FID	free induction decay
GB	gigabyte
GIAO	gauge independent atomic orbital
GTO	Gaussian type orbitals
HF	Hartree-Fock
HPC	high performance computing
HSQC	hetronuclear single quantum coherence
k_{ex}	chemical exchange rate constant
LCAO	linear combination of atomic orbitals
ms	millisecond (10−3 s)
NMR	nuclear magnetic resonance

NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
ns	nanosecond (10−9 s)
PCM	polarizable continuum model
PDB	protein data bank
ps	picosecond (10−12 s)
PSO	paramagnetic spin orbit
R1ρ	rotating frame spin relaxation
RAM	random access memory
RDC	residual dipolar coupling
Rex	transverse relaxation rate constant due to chemical exchange
rf	radiofrequency
s	second
SCF	self consistent field
STO	slater type orbitals
STO-nG	slater type orbitals with number of Gaussians
XC	exchange correlation
μs	microsecond (10−6 s)
τ_{ex}	chemical exchange timescale (= 1 / k_{ex})

Contents

1 Introduction

Spin dynamics simulation algorithms play a significant role in the design of experimental techniques, analysis of spectra and to extract structural and dynamic parameters.

Spin dynamics focuses in simulating the evolution of a spin system over time. NMR simulation of systems containing over 20 spins is difficult, because the CPU time for simulation in Liouville space scales exponentially [1] with system size.

A number of computational tools that simulate spin dynamics have emerged over the years [2-5]. Spin dynamics simulation packages are categorized as analytical and numerical tools. Almost all analytical tools are written in Mathematica, while numerical tools are scripted and coded in FORTRAN, Python, C and/or Matlab [3]. A class of polynomially scaling algorithms that allow computationally efficient simulation of systems with a large number of coupled spins are developed [3]. The development of Spinach [3], a spin dynamics simulation package, opens the way to the simulation of large bio-molecules such as proteins and nucleic acid as a whole in one go. But, this requires a *J*-coupling predictor function.

One bond and two bond coupling are largely known and available in literature [6-9]. Three bond couplings are known to depend at the intervening dihedral angle [10]. This relation is very well researched using Karplus equation. Thus, an electronic structure calculation package, DFT, is used to predict the change of *J*-couplings due to dihedral angle scan.

Even though DFT is used for predicting spin-spin couplings, it fails to scale to very large systems like proteins. The reason for the lack of scalability among others, are the introduction of orbitals in DFT methods as proposed by Kohn-Sham [11, 12]. This makes the computation complex by increasing the calculation at least cubically due to the increase in number of variables from 3 to 3*N*–with *N* the number electrons. Thus DFT methods are useful to predict physical or chemical properties of moderately sized spin systems [13]. However scalar coupling computation for protein spin system would put very high strain on modern computer if not crash it. Moreover, spin-spin coupling is more expensive compared to the calculation of chemical shielding and quadrupolar coupling constants. This is due to the computational dependence of *J*-coupling on the second derivative of electronic energy which is calculated as a second-order time-independent property [14].

As it is already noticed and will witness from the main body of the thesis, the terms quantum mechanical and large–scale are used very often. The “quantum mechanical” term indicates that the formalism employed for the simulation is the quantum mechanical treatment of the spin system and the simulated systems for purposes of this thesis are proteins and nucleic acids, hence the term “large–scale”.

This thesis reflects on how the *J*-coupling predictor function for proteins and nucleic acids were built and on large-scale simulations of NOESY NMR spectra for ubiquitin and stem loops RNA. It also considers large–scale NMR simulations and fitting of chemical exchange. It studied the simulation of ^{1}H-^{15}N HSQC spectra of calmodulin and of that with the subsequent addition of a ligand and least squarely fitted the data from theory to those from experiment as a function of exchange and transverse relaxation rates to investigate conformational change and calmodulin peptide binding.

2 Background

2.1 Spin and its properties

Spin is an intrinsic property of elementary particles. Algebraically spin has similar properties to angular momentum although spin is not related to rotations in physical space. Spin theory can be introduced using angular momentum. Classically, angular momentum of a particle is defined as the cross product of the coordinate vector $\vec{r} = (x, y, z)$ and momentum vector $\vec{p} = (p_x, p_y, p_z)$

$$\vec{L} = \vec{r} \times \vec{p} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x & y & z \\ p_x & p_y & p_z \end{vmatrix} = \begin{pmatrix} yp_z - zp_y \\ zp_x - xp_z \\ xp_y - yp_x \end{pmatrix} = \begin{bmatrix} L_x \\ L_y \\ L_z \end{bmatrix} \tag{1}$$

where $\vec{i}$, $\vec{j}$, and $\vec{k}$ are unit vectors.

Replacing the classical quantities with the corresponding quantum mechanical operators; angular momentum operators for the three components of the angular momentum vectors are obtained

$$\hat{L}_x = -i\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y} \right), \quad \hat{L}_y = -i\left(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z} \right), \quad \hat{L}_z = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} \right). \tag{2}$$

Another two important angular momentum operators exists. One of these operators is the total angular momentum operator given as

$$\hat{L}^2 = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 \tag{3}$$

This operator appears when a system undergoes rotational diffusion. The other is the raising and lowering operators defined as

$$\hat{L}_+ = \hat{L}_x + i\hat{L}_y \quad \hat{L}_- = \hat{L}_x - i\hat{L}_y \tag{4}$$

This operators are useful for the manipulation of angular momentum eigenfunctions.

Similar to angular momentum, the state of a spin, ½, can be described as a linear combination of two state vectors:

$$c_{½}\left|\alpha\right\rangle + c_{-½}\left|\beta\right\rangle = \begin{pmatrix} c_{½} \\ -c_{½} \end{pmatrix} \tag{5}$$

where $\left|\alpha\right\rangle$ and $\left|\beta\right\rangle$ are the basis state vectors that represent the state space of the spin. The matrix representation of $\hat{I}_z$ in the Hilbert space spanned by $\left|\alpha\right\rangle$ and $\left|\beta\right\rangle$ states is

$$\hat{I}_z = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix} \tag{6}$$

Similarly, there exist an x- and y-component spin angular momentum operators ($\hat{I}_x$ and $\hat{I}_y$) for a spin. The matrix representation of these operators in the $\left|\alpha\right\rangle$ and $\left|\beta\right\rangle$ basis is

$$\hat{I}_x = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \qquad \hat{I}_y = \begin{pmatrix} 0 & -i/2 \\ i/2 & 0 \end{pmatrix} \tag{7}$$

Another important spin operator is the total spin angular momentum operator $\hat{I}^2$. The spin quantum number I, is associated to the $\hat{I}^2$ operator via $\hat{I}^2\psi = I\left(I+1\right)\psi$.

The other operators of a spin are the raising and lowering operators

$$\hat{I}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad \hat{I}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{8}$$

$\hat{I}_+$ acts on the spin down state (β) and raises the state into α and would give zero if it acts on the up state (α). Likewise $\hat{I}_-$ would only lower the upper state and gives zero when acting on the lower state.

Spin is associated to magnetic moment, $\hat{\mu}$, by a proportionality constant, γ, known as the gyromagnetic ratio. The relationship holds for the projection operators and is defined as:

$$\hat{\mu}_x = \gamma\hat{I}_x \qquad \hat{\mu}_y = \gamma\hat{I}_y \qquad \hat{\mu}_z = \gamma\hat{I}_z \tag{9}$$

The gyromagnetic ratio is distinct for each atomic nuclei type in the periodic table. Atomic nuclei are composite particles that involve atomic sub-particles known as quarks, with each quark having spin ½. The orientation of the quarks determines the magnetic dipole moment and the spin quantum number of the nuclei. For example ^{13}C has spin ½. This arises from the seven antiparallel up and six parallel down alignment of the quarks. However, nuclei of the

same element can have different spin and dipole moment depending on the number of the quarks that makes up the nucleus. Finally, quark spin interaction is dependent on the magnetic field produced by the magnetic moment and the Pauli Exclusion Principle, which states that no two spins have the same quantum state at the same time.

The expectation value of any operator, such as $\hat{A}$, is given by

$$\left\langle \hat{A} \right\rangle = \mathrm{Tr}(\rho \hat{A}) \tag{10}$$

Tr is the trace of the matrix and is given by the sum of the elements in the main diagonal. It is a requirement that ρ - the density matrix, has the following property:

- It must be Hermitian
- The trace of ρ the density matrix is one
- The diagonal elements of ρ must be nonnegative

The derivation of the density matrix is available in many text books including [15-17]. The diagonal elements of a density matrix are populations while the off-diagonal are coherences. The elements of density matrix contain statistically the behaviour of spin system, and the result of an experimental measurement can be obtained using Equation (10).

The equation of motion, *i.e.* the development of ρ under a given Hamiltonian is determined by the Liouville-von Neumann equation:

$$\frac{\partial}{\partial t}\rho(t) = -i\left[\hat{H}, \rho(t)\right] = -i\hat{\hat{H}}\rho(t) \tag{11}$$

The first line in Equation (11) is the representation of the equation in the Hilbert space, where the evolution of the density matrix is given by the commutation of ρ with the Hamiltonian operator, $\hat{H}$. In the second line or the right part of Equation (11) we have relocated to the Liouville space, where ρ becomes a vector. The action of taking Hamiltonian superoperator, $\hat{\hat{H}}$, is a commutation that transforms $\hat{H}$ into a matrix in the Liouville space. Superoperators (recognized for their double hat) are defined as operations in the space of operators. In solving the Liouville-von Neumann equation, its representation in the Liouville space is convenient for the purposes of numerical computation. Detailed calculation of quantum dynamics in the Liouville space can be found in [18-20].

2.2 Spin interactions and their Hamiltonians

NMR mainly deals with nuclear spin interactions in diamagnetic systems; though there are areas of NMR that deals with paramagnetic samples, which are not part of the project work.

Spin interacts with its environment and with other spins via the associated magnetic moment. Although there are many types of spin interactions in NMR, their spin Hamiltonians can broadly be classified into linear, bilinear and quadratic. The interactions that involve the spin and the external field are *linear*, for example Zeeman interaction:

$$\hat{H} = \hat{\vec{S}} \cdot \mathbf{A} \cdot \vec{B} = \begin{pmatrix} \hat{S}_{\mathrm{x}} & \hat{S}_{\mathrm{y}} & \hat{S}_{\mathrm{z}} \end{pmatrix} \begin{pmatrix} a_{\mathrm{xx}} & a_{\mathrm{xy}} & a_{\mathrm{xz}} \\ a_{\mathrm{yx}} & a_{\mathrm{yy}} & a_{\mathrm{yz}} \\ a_{\mathrm{zx}} & a_{\mathrm{zy}} & a_{\mathrm{zz}} \end{pmatrix} \begin{pmatrix} B_{\mathrm{x}} \\ B_{\mathrm{y}} \\ B_{\mathrm{z}} \end{pmatrix} \tag{12}$$

where $\hat{\vec{S}}$ is a vector of spin operators, $\vec{B}$ is magnetic field vector, and $\mathbf{A}$ is the Zeeman interaction tensor.

When the interaction is a coupling between spins, it is known as *bilinear* in spin

$$\hat{H} = \hat{\vec{S}} \cdot \mathbf{A} \cdot \hat{\vec{I}} = \begin{pmatrix} \hat{S}_{\mathrm{x}} & \hat{S}_{\mathrm{y}} & \hat{S}_{\mathrm{z}} \end{pmatrix} \begin{pmatrix} a_{\mathrm{xx}} & a_{\mathrm{xy}} & a_{\mathrm{xz}} \\ a_{\mathrm{yx}} & a_{\mathrm{yy}} & a_{\mathrm{yz}} \\ a_{\mathrm{zx}} & a_{\mathrm{zy}} & a_{\mathrm{zz}} \end{pmatrix} \begin{pmatrix} \hat{I}_{\mathrm{x}} \\ \hat{I}_{\mathrm{y}} \\ \hat{I}_{\mathrm{z}} \end{pmatrix} \tag{13}$$

where $\hat{\vec{S}}$ and $\hat{\vec{I}}$ are spin operator vectors, and $\mathbf{A}$ is the coupling tensor.

Interactions that are caused indirectly by other interactions, but manifest as a coupling of spin to itself, are *quadratic* in spin. Nuclear quadrupolar interaction belongs to this category:

$$\hat{H} = \hat{\vec{S}} \cdot \mathbf{A} \cdot \hat{\vec{S}} = \begin{pmatrix} \hat{S}_{\mathrm{x}} & \hat{S}_{\mathrm{y}} & \hat{S}_{\mathrm{z}} \end{pmatrix} \begin{pmatrix} a_{\mathrm{xx}} & a_{\mathrm{xy}} & a_{\mathrm{xz}} \\ a_{\mathrm{yx}} & a_{\mathrm{yy}} & a_{\mathrm{yz}} \\ a_{\mathrm{zx}} & a_{\mathrm{zy}} & a_{\mathrm{zz}} \end{pmatrix} \begin{pmatrix} \hat{S}_{\mathrm{x}} \\ \hat{S}_{\mathrm{y}} \\ \hat{S}_{\mathrm{z}} \end{pmatrix} \tag{14}$$

where $\hat{\vec{S}}$ is a vector of spin operators and $\mathbf{A}$ is the quadrupolar coupling tensor.

The overall Hamiltonian is a sum of all the interaction Hamiltonians in the system.

2.2.1 Chemical shielding Hamiltonian

This interaction arises [21] from the effect of electron motion in the vicinity of the nuclear spin. The electron motion is induced by the externally applied static magnetic field $\vec{B}_0$ of the NMR spectrometer. The magnetic field at an arbitrary nucleus *k* is

$$\vec{B}(k) = \vec{B}_0 + \vec{B}_{\text{ind}}(k) \tag{15}$$

where the field $\vec{B}_{\text{ind}}$ is induced by the electronic structure surrounding the nucleus.

Magnetic interactions are weak compared to the electronic ones, hence the induced field is a linear function of the applied magnetic field $\vec{B}_0$

$$\vec{B}_{\text{ind}}(k) = \vec{B}_0 + \boldsymbol{\delta}(k) \cdot \vec{B}_0 = \vec{B}_0 (\mathbf{1} + \boldsymbol{\delta}) \tag{16}$$

where $\boldsymbol{\delta}$ is the shielding tensor.

The energy of magnetic dipole moment that results from this interaction is therefore

$$E = -\vec{\mu}_k \cdot \vec{B}_{\text{ind}}(k) = -\vec{\mu}_k \cdot \vec{B}_0 (\mathbf{1} + \boldsymbol{\delta}(k)) \tag{17}$$

The interaction is a coupling between the spin and the applied magnetic field. The Hamiltonian can be obtained by replacing the classical quantities by the quantum equivalents. It is given as a scalar product between the spin vector, the shielding tensor and applied magnetic field

$$\hat{H}_{\text{CS}} = \hbar \gamma B_0 \sum_K \begin{bmatrix} \hat{I}_{\text{x}} & \hat{I}_{\text{y}} & \hat{I}_{\text{z}} \end{bmatrix}_k \cdot \begin{bmatrix} \delta_{\text{xx}} & \delta_{\text{xy}} & \delta_{\text{xz}} \\ \delta_{\text{yx}} & \delta_{\text{yy}} & \delta_{\text{yz}} \\ \delta_{\text{zx}} & \delta_{\text{zy}} & \delta_{\text{zz}} \end{bmatrix}_k \cdot \begin{bmatrix} B_{\text{x}} \\ B_{\text{y}} \\ B_{\text{z}} \end{bmatrix} \tag{18}$$

2.2.2 Nuclear dipolar interaction Hamiltonian

To a very good approximation, nuclear spins can be considered point dipole moments [22]. The magnetic vector potential $\vec{A}$, of a dipole is

$$\vec{A} = \frac{\mu_0}{4\pi} \frac{\vec{\mu} \times \vec{r}}{r^3} \tag{19}$$

The magnetic field induced by the magnetic dipole $\vec{\mu}$ is the curl of the vector potential

$$\vec{B} = \vec{\nabla} \times \vec{A} \;\Rightarrow\; \vec{B} = \nabla \times \left(\frac{\mu_0}{4\pi} \frac{\vec{\mu} \times \vec{r}}{r^3} \right) = \ldots = \frac{\mu_0}{4\pi} \frac{3(\vec{\mu} \cdot \vec{r})\vec{r} - r^2 \vec{\mu}}{r^5} \tag{20}$$

Therefore for two nuclear spins separated by a distance r, the interaction energy between their respective magnetic dipoles is given by

$$E = -\vec{\mu}\cdot\vec{B} = -\frac{\mu_0}{4\pi}\frac{3\left(\vec{\mu}_\mathrm{a}\cdot\vec{r}\right)\left(\vec{r}\cdot\vec{\mu}_\mathrm{b}\right) - r^2\left(\vec{\mu}_\mathrm{a}\cdot\vec{\mu}_\mathrm{b}\right)}{r^5} \tag{21}$$

where the constant $\mu_0 = 4\pi\times10^{7}\,\mathrm{kg\cdot m\cdot s^{-2}\cdot A^{-2}}$ is the permeability constant and $\vec{r}$ is the vector that connects the two spins. The quantum mechanical energy operator can then be obtained by replacing the magnetic dipole vectors $\vec{\mu}_\mathrm{a}$ and $\vec{\mu}_\mathrm{b}$ by the corresponding spin operator vectors $\hat{\vec{I}}^{(\mathrm{a})}$ and $\hat{\vec{I}}^{(\mathrm{b})}$

$$\hat{H}_\mathrm{DD} = -\frac{\mu_0}{4\pi}\frac{\gamma_\mathrm{a}\gamma_\mathrm{b}\hbar}{r^5}\left(3\left(\hat{\vec{I}}_\mathrm{a}\cdot\vec{r}\right)\left(\vec{r}\cdot\hat{\vec{I}}_\mathrm{b}\right) - r^2\left(\hat{\vec{I}}_\mathrm{a}\cdot\hat{\vec{I}}_\mathrm{b}\right)\right) \tag{22}$$

where γ_a and γ_b are the gyromagnetic ratio of the respective spins. The full dipolar Hamiltonian matrix is

$$\begin{aligned}\hat{H}_\mathrm{DD} &= -\frac{\mu_0}{4\pi}\frac{\gamma_\mathrm{a}\gamma_\mathrm{b}\hbar}{r^5}\times\\ &\begin{pmatrix}\hat{I}_\mathrm{x}^{(\mathrm{a})}\\ \hat{I}_\mathrm{y}^{(\mathrm{a})}\\ \hat{I}_\mathrm{z}^{(\mathrm{a})}\end{pmatrix}\begin{pmatrix}3\left(x_\mathrm{a}-x_\mathrm{b}\right)^2 - r^2 & 3\left(x_\mathrm{a}-x_\mathrm{b}\right)\left(y_\mathrm{a}-y_\mathrm{b}\right) & 3\left(x_\mathrm{a}-x_\mathrm{b}\right)\left(z_\mathrm{a}-z_\mathrm{b}\right)\\ 3\left(y_\mathrm{a}-y_\mathrm{b}\right)\left(x_\mathrm{a}-x_\mathrm{b}\right) & 3\left(y_\mathrm{a}-y_\mathrm{b}\right)^2 - r^2 & 3\left(y_\mathrm{a}-y_\mathrm{b}\right)\left(z_\mathrm{a}-z_\mathrm{b}\right)\\ 3\left(z_\mathrm{a}-z_\mathrm{b}\right)\left(x_\mathrm{a}-x_\mathrm{b}\right) & 3\left(z_\mathrm{a}-z_\mathrm{b}\right)\left(y_\mathrm{a}-y_\mathrm{b}\right) & 3\left(z_\mathrm{a}-z_\mathrm{b}\right)^2 - r^2\end{pmatrix}\begin{pmatrix}\hat{I}_\mathrm{x}^{(\mathrm{b})}\\ \hat{I}_\mathrm{y}^{(\mathrm{b})}\\ \hat{I}_\mathrm{z}^{(\mathrm{b})}\end{pmatrix}\end{aligned} \tag{23}$$

where $\left(x_\mathrm{a}, y_\mathrm{a}, z_\mathrm{a}\right)$ and $\left(x_\mathrm{b}, y_\mathrm{b}, z_\mathrm{b}\right)$ are spin coordinates for $\hat{\vec{I}}^{(\mathrm{a})}$ and $\hat{\vec{I}}^{(\mathrm{b})}$ respectively. The trace *i.e.* the sum of diagonal elements of the interaction tensor and rhombicity, the difference between the smallest and second smallest eigenvalues, are zero making dipolar interaction traceless and axial. In solution state NMR nuclear dipolar interaction is averaged out by the tumbling of the molecules. In solids, however, it provides information on orientations and distances.

2.2.3 *J*-coupling

J-coupling is a bilinear through bond interaction. Being isotropic [23], spin-spin coupling does not average out upon molecular tumbling. Thus the interaction Hamiltonian takes the form

$$\hat{H}_J = 2\pi \hat{\vec{I}} \cdot J \cdot \hat{\vec{S}} = \begin{pmatrix} \hat{I}_x & \hat{I}_y & \hat{I}_z \end{pmatrix} \begin{pmatrix} 2\pi J & 0 & 0 \\ 0 & 2\pi J & 0 \\ 0 & 0 & 2\pi J \end{pmatrix} \begin{pmatrix} \hat{S}_x \\ \hat{S}_y \\ \hat{S}_z \end{pmatrix} \tag{24}$$

$$= 2\pi J \left(\hat{S}_x \hat{I}_x + \hat{S}_y \hat{I}_y + \hat{S}_z \hat{I}_z \right)$$

This interaction is responsible for the multiplicity of resonance signals in an NMR spectrum and indicates the number of neighbouring nuclei. Its magnitude depends on dihedral angles. In contrast to chemical shift, the coupling constant does not depend on the strength of the static magnet of the spectrometer. Scalar coupling is mediated by bonding electrons. Hence the dominant contribution for the Hamiltonian comes from Fermi contact interaction:

$$\hat{H}_{FC} = \frac{8\pi\hbar}{3} \sum_{kn} \gamma_k \gamma_n \delta(\vec{r}_{kn}) \hat{\vec{I}}_k \cdot \hat{\vec{S}}_n \tag{25}$$

J-coupling values are related to molecular conformation and this is useful in structure determination by NMR spectroscopy [24, 25]. The magnitude of coupling constant normally but not always depends on the number of bond that separate the nuclei *i.e.* the higher the number of bond the smaller the coupling becomes. Spin-spin coupling (specifically ^3J) dependence in dihedral angle is called Karplus equation [10].

2.2.4 Quadrupolar coupling

Nuclei with spin quantum number $I > \frac{1}{2}$ possess non–symmetric charge distributions, which generates an electric quadrupolar moment that interacts with an electric field gradient (EFG). Quadrupolar moment is aligned with the nuclear spin. The interaction can be described as

$$\hat{H} = \frac{eQ}{2I(2I-1)} \hat{\vec{I}} \cdot \mathbf{V} \cdot \hat{\vec{I}} \qquad V_{ij} = \frac{\partial^2 \varphi}{\partial r_i \partial r_j} \tag{26}$$

where eQ is the quadrupolar moment, I is the nuclear spin, and $\mathbf{V}$ is a symmetric and traceless second rank tensor of the EFG. The electric field gradients at the nucleus are generated by electronic charges present in the spin system outside the nuclei and are described as the second derivative of the electrostatic potential. In principal axis frame, an asymmetric parameter η and its magnitude eq are

$$\eta = \frac{V_{xx} - V_{yy}}{V_{zz}} \qquad eq = V_{zz} \tag{27}$$

The quadrupolar Hamiltonian in the above stated axis system is

$$\hat{H}_{\mathrm{Q}} = -\frac{e^2 qQ}{4I\left(2I-1\right)\hbar}\left[3\hat{I}_{\mathrm{z}}^2 - \hat{I}^2 + \frac{\eta}{2}\left(\hat{I}_{\mathrm{x}}^2 - \hat{I}_{\mathrm{y}}^2\right)\right] \tag{28}$$

The coupling of the quadrupolar moment and EFG affects the line width of NMR signals and makes them broad.

Detailed description of other spin interactions can be found in NMR books, in particular [22, 26, 27].

2.3 Bloch equations

The net magnetization vector, $\mathbf{M}$, that comes into existence due to magnetic moments of nuclear spins in NMR experience moment in the presence of magnetic field. This is given by the cross product between the magnetization $\mathbf{M}$ and the magnetic field $\mathbf{B}$.

$$\frac{d\mathbf{M}\left(t\right)}{dt} = \mathbf{M}\left(t\right)\times\gamma\mathbf{B}\left(t\right) \tag{29}$$

Equation (29) is fundamental for the formulation of Bloch equations. After, application of an rf pulses, nuclear spins relax to equilibrium with the transverse and longitudinal relaxation rates along the x-y plane and z-axis respectively. It is assumed these relaxation rates follow first-order kinetics. Thus, the full Bloch equations for free precession after the integration of the relaxation term is

$$\frac{d\mathbf{M}\left(t\right)}{dt} = \mathbf{M}\left(t\right)\times\gamma\mathbf{B}\left(t\right) - R\left(\mathbf{M}\left(t\right) - M_0\right). \tag{30}$$

where $M_0 = \dfrac{\gamma^2\hbar^2\left(I+1\right)\vec{B}_0}{3k_B\mathrm{T}}$ is the thermal equilibrium magnetization, and R is the relaxation rate. For each of the components Equation (30) becomes

$$\begin{aligned}
\frac{dM_{\mathrm{Z}}}{dt} &= \gamma\left(M_{\mathrm{x}}B_{\mathrm{y}} - M_{\mathrm{y}}B_{\mathrm{x}}\right) - \left(-\frac{\left(M_{\mathrm{Z}} - M_0\right)}{T_1}\right),\\
\frac{dM_{\mathrm{x}}}{dt} &= \gamma\left(M_{\mathrm{y}}B_{\mathrm{z}} - M_{\mathrm{z}}B_{\mathrm{y}}\right) - \frac{M_{\mathrm{x}}}{T_2},\\
\frac{dM_{\mathrm{y}}}{dt} &= \gamma\left(M_{\mathrm{z}}B_{\mathrm{x}} - M_{\mathrm{x}}B_{\mathrm{z}}\right) - \frac{M_{\mathrm{y}}}{T_2}.
\end{aligned} \tag{31}$$

A convention in NMR is that the static field $\vec{B}_0$ is treated as applied along the z-axis that gives $B_x = B_y = 0$, hence Equation (31) as of [28] simplifies to

$$\begin{aligned} \frac{dM_Z}{dt} &= -\frac{\left(M_Z - M_0\right)}{T_1}, \\ \frac{dM_x}{dt} &= \gamma M_y \vec{B}_0 - \frac{M_x}{T_2}, \\ \frac{dM_y}{dt} &= -\gamma M_x \vec{B}_0 - \frac{M_y}{T_2}. \end{aligned} \tag{32}$$

The first line of Equation (32) is used to predict the exponential relaxation of the longitudinal magnetization component M_z to its equilibrium value. The time that takes for this to happen is known as spin lattice relaxation time and is denoted as T_1. The solution for M_z after initial inversion of the longitudinal magnetization using 180° pulse at time $t = 0$ is given as

$$M_z = M_0\left(1 - 2\exp\left(-\frac{t}{T_1}\right)\right) \tag{33}$$

The second and third lines describes the coherent motion of the magnetization vector **M** and the exponential decay of the transverse magnetization respectively. The time process for the exponential decay of the transverse magnetization to happen is called spin-spin relaxation time and is abbreviated as T_2. After application of on-resonance 90° pulse, the evolution of the transvers magnetization can be expressed as

$$M_{x,y} = M_0 \exp\left(-\frac{t}{T_2}\right) \tag{34}$$

$\frac{1}{T_2}$ is associated to the line width and can be described using Equation (35)

$$\Delta\nu = \frac{1}{\pi T_2} \tag{35}$$

where $\Delta\nu$ is the line width at half height of the peak.

2.4 Brief description of NMR relaxation mechanisms

Nuclear spin systems experience five major relaxation mechanisms [29, 30]. These are dipolar, chemical shift anisotropy, scalar coupling relaxation of type one and two, quadrupolar

interactions, and spin rotation [25, 29-31]. Of all the available relaxation mechanisms for a spin system the dipolar relaxation mechanism is the most dominant for spin ½ systems.

Dipolar interaction depends in internuclear separation, which falls as the inverse cube of r_{IS}, and the angle of the vector that connects the interacting nuclei with the static field $\vec{B}_0$. Dipolar interaction can be described using Equation (36)

$$D_{IS} = -\frac{\mu_0}{4\pi}\frac{\gamma_I\gamma_S}{r_{IS}^3}\left[\frac{3\left(\hat{I}\cdot\mathbf{r}\right)\left(\hat{S}\cdot\mathbf{r}\right)}{r_{IS}^2} - \left(\hat{I}\cdot\hat{S}\right)\right] \tag{36}$$

While Equation (37) governs the relaxation of spin I due to dipolar interaction with a nearby spin S.

$$\begin{aligned}
\frac{1}{T_1^{DD}} &= \frac{1}{10}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\gamma_I^2\gamma_S^2}{r_{IS}^6}\left(\frac{\tau_C}{1+\left(\omega_I-\omega_S\right)^2\tau_C^2}+\frac{3\tau_C}{1+\omega_I^2\tau_C^2}+\frac{6\tau_C}{1+\left(\omega_I+\omega_S\right)^2\tau_C^2}\right),\\
\frac{1}{T_2^{DD}} &= \frac{1}{20}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\gamma_I^2\gamma_S^2}{r_{IS}^6}\left(\begin{array}{l}4\tau_C+\dfrac{\tau_C}{1+\left(\omega_I-\omega_S\right)^2\tau_C^2}\\ +\dfrac{3\tau_C}{1+\omega_I^2\tau_C^2}+\dfrac{6\tau_C}{1+\omega_S^2\tau_C^2}+\dfrac{6\tau_C}{1+\left(\omega_I+\omega_S\right)^2\tau_C^2}\end{array}\right),\\
\sigma_{IS} &= \frac{1}{10}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\gamma_I^2\gamma_S^2}{r_{IS}^6}\tau_C\left(\frac{6}{1+\left(\omega_I+\omega_S\right)^2\tau_C^2}-\frac{1}{1+\left(\omega_I-\omega_S\right)^2\tau_C^2}\right).
\end{aligned} \tag{37}$$

Equation (37) includes spectral density component that couples the system to the stochastic rotational dynamics of the molecules. The detailed description and derivation of the spectral density function is available in chapter four. Molecular motions – such as translation, rotation and vibration causes r_{IS} and θ to change in a complex manner over a period of time producing rapidly fluctuating magnetic fields. Dipolar interaction modulated by rotational tumbling in liquids induces fluctuating magnetic fields that causes a nearby nuclei to relax.

Another common cause of spin relaxation is the chemical shift anisotropy (CSA). Chemical shift is associated to the electronic environment in the vicinity of a nucleus and is normally anisotropic, *i.e.* its value depends on the orientation of the molecule relative to the applied magnetic field. As the molecule rotates, the magnetic field that the nucleus experiences keeps changing. This makes the chemical shift direction dependent. Therefore, the variation of the field as a function of the rotation time stimulates relaxation that is proportional to the square

of the applied field strength and the gyromagnetic ratio. The relaxation rates from such interaction are

$$\begin{aligned}\frac{1}{T_1^{\mathrm{CSA}}} &= \frac{2}{15}\gamma^2\vec{B}_0^2\left(\sigma_{\parallel}-\sigma_{\perp}\right)^2\left[\frac{\tau_{\mathrm{C}}}{1+\omega_{\mathrm{I}}^2\tau_{\mathrm{C}}^2}\right],\\ \frac{1}{T_2^{\mathrm{CSA}}} &= \frac{1}{45}\gamma^2\vec{B}_0^2\left(\sigma_{\parallel}-\sigma_{\perp}\right)^2\left[4\tau_{\mathrm{C}}+\frac{3\tau_{\mathrm{C}}}{1+\omega_{\mathrm{I}}^2\tau_{\mathrm{C}}^2}\right].\end{aligned} \tag{38}$$

where $\left(\sigma_{\parallel}-\sigma_{\perp}\right)$ is the anisotropy parameter, $\vec{B}_0$ is the applied magnetic field, and γ is the gyromagnetic ratio of the nucleus.

Spins that possesses spin quantum number $I > 1/2$ have quadrupolar moment. Their relaxation is dominated by quadrupolar relaxation mechanism. Electrons that surround such nuclei creates electric field gradient. The coupling between the electric field gradient and the quadrupolar moment is responsible for quadrupolar coupling. This interaction also couples the spin to the rotational motion of the molecule. The quadrupolar relaxation of a spin [32, 33] is given by

$$\frac{1}{T_1^{\mathrm{Q}}} = \frac{3}{50}\pi^2\frac{\left(2I+3\right)}{I^2\left(2I-1\right)}\left(1+\frac{\eta^2}{3}\right)\left(\frac{e^2qQ}{h}\right)^2\tau_{\mathrm{C}}\left[\frac{1}{1+\omega_{\mathrm{Q}}^2\tau_{\mathrm{C}}^2}+\frac{4}{1+4\omega_{\mathrm{Q}}^2\tau_{\mathrm{C}}^2}\right] \tag{39}$$

where η is the field asymmetry parameter, q is electric field gradient, Q is the quadrupolar moment, and τ_{C} is the rotational correlation time.

Quadrupolar relaxation is responsible for broad resonance signals. It effectively removes *J*-couplings. NOE enhancements for quadrupolar nuclei are diminishingly small or impossible to measure. For perfectly symmetric molecules, *i.e.* systems of tetrahedral and octahedral structure, the electric field gradient is small.

Spins can also relax by scalar coupling relaxation mechanism. Scalar coupling is an isotropic interaction, therefore its Hamiltonian is not affected by orientation of the molecule with respect to the external field. This interaction acts as a source of relaxation when its strength varies over time. The variation of the coupling strength over time on spin I, might be due to the relaxation of the coupling spin S. Therefore, the longitudinal and transverse relaxation of spin I, due to scalar coupling relaxation, according [30] is

$$\frac{1}{T_1^{SC}} = \frac{4}{3}\pi^2 J^2 S(S+1)\left[\frac{T_{2S}}{1+(\omega_I-\omega_S)^2 T_{2S}^2}\right],$$
$$\frac{1}{T_2^{SC}} = \left[\frac{4}{3}\pi^2 J^2 S(S+1)\right]\left[T_{2S}+\frac{T_{2S}}{1+(\omega_I-\omega_S)^2 T_{2S}^2}\right]. \quad (40)$$

When the scalar coupling relaxation arises from the time dependence of the *J*-coupling, which is the case for strongly coupled spins, the reader is recommended to read [32, 34] for a detailed discussion.

2.5 Common protein and nucleic acid NMR experiments

This project involves protein and nucleic acid NMR spectroscopy and modelling of exchange processes to extract kinetic information. Thus, it is sensible to include into the introduction some common protein and nucleic acid experimental methods along with experimental techniques that are necessary to study exchange processes. NOESY method for proteins and nucleic acids is covered in detail in chapters four and six respectively. HSQC for calmodulin is also discussed in chapter five. Here described are techniques used to study exchange processes in proteins and nucleic acids. Experimental methods employed to study exchange processes split into two [35]. These are magnetization transfer and lineshape analysis. Magnetization transfer experiments commonly applied to investigate systems with exchange process are:

- Saturation transfer [36, 37]
- EXSY spectroscopy [35, 38]
- CPMG relaxation dispersion [39-41]
- Rotating frame relaxation dispersion [42-44]

2.5.1 Saturation transfer experiment

This is a technique used to study exchange processes that are an order of magnitude slower than the ones that can be studied using conventional lineshape analysis [37]. In such a slow exchanging system, signals from one of the exchanging species is saturated using rf pulses that oscillate at the resonance frequency of the spin; this destroys the signal from the spin [36, 37]. For two site exchange processes available in detail in chapter five; the observed change in

intensity in the neighbouring spin indicates the spins are connected by chemical exchange. The pulse sequence of the experiment as of [45] is depicted in **Figure 1** (below)

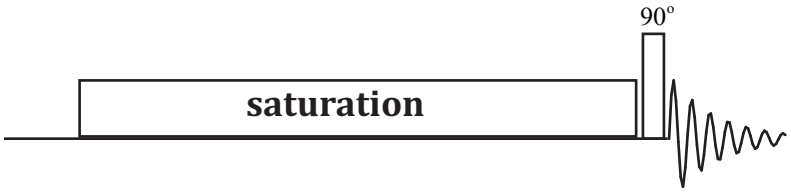

Figure 1 *Saturation transfer experiment pulse scheme. Illustrating selective saturation of one spin i.e. the populations of the* α *and* β *energy levels are equalized. Then this is followed by a* 90^{o} *pulse and detection of the signal.*

The time dependent longitudinal magnetization, M_{ZB}, in site B, can be described by

$$\frac{d}{dt}M_{ZB} = \left(M_{0B} - M_{ZB}\right)\frac{1}{T_{1B}} - \frac{M_{ZB}}{\tau_B} + \frac{M_{ZA}}{\tau_A} \tag{41}$$

where M_{ZA} is the longitudinal magnetization at site A, τ_A and τ_B are the lifetimes of states A and B, T_{1B} is the longitudinal relaxation of state B, and M_{0B} is the equilibrium z magnetization in site B.

After irradiation using an rf field that is sufficiently strong to saturate the magnetization at site A, the signal from state A, is destroyed, giving $M_A = 0$. The last term in Equation (41) disappears. After rearrangement Equation (41) becomes:

$$\frac{d}{dt}M_{ZB} = M_{0B}\left(\frac{1}{T_{1B}}\right) - M_{ZB}\left(\frac{1}{T_{1B}} + \frac{1}{\tau_B}\right) \tag{42}$$

Integration and simplification of Equation (42) gives the signal from site B as [5]

$$M_{ZB} = M_{oB}\left[\left(\frac{k_{-1}}{R_{1B} + k_{-1}}\right)\exp\left(-\left(R_{1B} + k_{-1}\right)t\right) + \frac{R_{1B}}{R_{1B} + k_{-1}}\right] \tag{43}$$

2.5.2 2D EXSY spectroscopy

EXSY [35, 38, 46] is a technique employed to map exchange pathways in dynamic processes. The technique has identical pulse sequence as the nuclear Overhauser effect spectroscopy. The cross peak signals in EXSY indicate the exchanging species in a multisite exchange. Even though the technique has advantage in determining exchange processes qualitatively; rate extraction needs acquisition of a series of EXSY spectra at different mixing time (τ_m). In

addition intensity integrals and longitudinal relaxation times of the exchanging species is required.

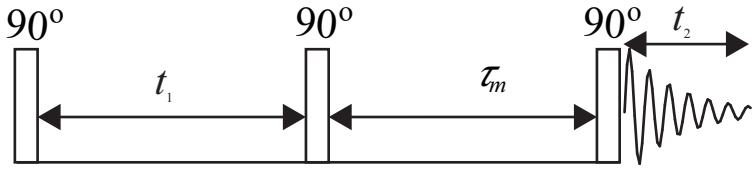

Figure 2 *2D EXSY pulse sequence. Rectangular boxes are π/2 pulses. The pulse sequence is identical to the NOESY experiment. During mixing time both exchange processes and dipolar couplings can transfer polarization. An alternative approach to get rid of NOE effects is CLEAN-EXSY experiment [47, 48].*

The solution to the Bloch-McConnell equation (covered in detail in chapter four) can be given by Equation (44) as of [35]

$$\mathrm{I}_{i,j}\left(\tau_{\mathrm{m}}\right) = M_j^0 \exp\left(-R\tau_{\mathrm{m}}\right)_{i,j} \tag{44}$$

where $\mathrm{I}_{i,j}$ is the cross peak volume at position ω_i and ω_j, τ_{m} –is the mixing time, and R –is the term that contains exchange rate and relaxation rate constants. Equation (44) makes the following assumptions:

$$\begin{aligned} M_{\mathrm{A}} &= M_{\mathrm{B}} = M \\ k_1 &= k_{-1} = k \\ R_{\mathrm{A}} &= R_{\mathrm{B}} = R \end{aligned}$$

This gives the solution to the diagonal and cross peak as

$$\begin{aligned} \mathrm{I}_{\mathrm{A,A}} = \mathrm{I}_{\mathrm{B,B}} &= \frac{1}{2}\left\{\left[1+\exp\left(-2k\tau_{\mathrm{m}}\right)\right]\exp\left(-\tau_{\mathrm{m}}R\right)\right\} \\ \mathrm{I}_{\mathrm{A,B}} = \mathrm{I}_{\mathrm{B,A}} &= \frac{1}{2}\left\{\left[1-\exp\left(-2k\tau_{\mathrm{m}}\right)\right]\exp\left(-\tau_{\mathrm{m}}R\right)\right\} \end{aligned} \tag{45}$$

The exchange rate can be obtained from the ratio of the diagonal to cross peak intensities at short mixing time.

2.5.3 Carr-Purcell-Meiboom-Gill relaxation dispersion

CPMG is an experimental technique that involves a spin-echo pulse sequence to investigate a change in transverse relaxation rate due to exchange process [39-41, 46, 49]. A series of spin-echo pulses is applied to the magnetization. The magnetization evolves under resonance frequency that differs due to presence of exchange. In these type of experiments the pulsing

frequency υCPMG is given by N_{cyc}/T. Where N_{cyc} is the number of cycle of the spin echoes, and T is the total time of CPMG element given by $4\tau N_{cyc}$, where τ - the time between consecutive refocusing pulses. The relaxation dispersion data are usually depicted by plotting R_2 and the pulse reputation rate (υ_{CPMG}) in the vertical and horizontal dimension of a 2D plot respectively.

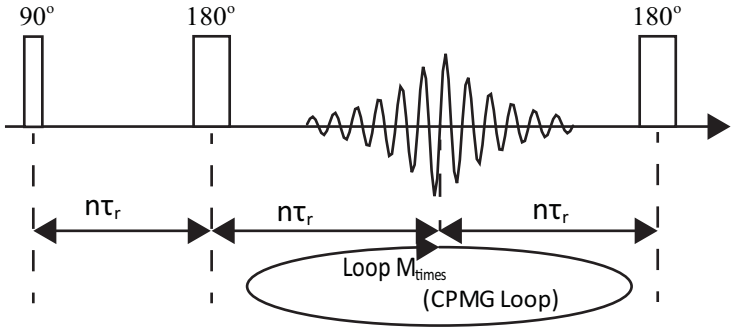

Figure 3 *CPMG pulse sequence.*

Experimentally the exchange process parameters such as exchange rate constant, populations, and resonance frequency are determined qualitatively by measuring the spin-spin relaxation rate at several different τ values. Dynamic parameters can also be quantified by fitting the relaxation data to exchange equations. Loria *et al* [49] showed the dependence of R_2 on the pulse range of CPMG experiment in the fast regime as:

$$R_2\left(1/\tau_{cp}\right) = R_2^0 + \frac{p_A p_B \Delta\omega^2}{k_{ex}\left[1 - \frac{2\tanh\left(\frac{k_{ex}\tau_{cp}}{2}\right)}{k_{ex}\tau_{cp}}\right]} \tag{46}$$

p_A and p_B are equilibrium populations, $\Delta\omega = \Omega_A - \Omega_B$, $k_{ex} = k_1 + k_{-1}$, $\tau_{cp} = 2n\tau_r$ is the delay between two consecutive 180° pulses and R_2^0 is the transverse relaxation in the absence of exchange. The populations and frequency differences are only determined after additional independent measurements are performed.

2.5.4 Rotating frame relaxation dispersion experiment R1ρ [42-44]

The evolution of magnetization under the presence of rf fields spin locked parallel to the effective field in the rotating frame of reference is known as spin-lattice relaxation time $T_{1\rho}$.

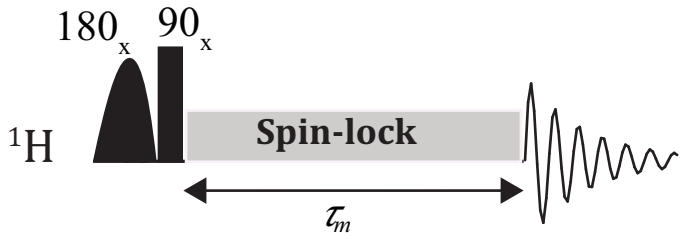

Figure 4 *Rotating frame relaxation rate $R_{1\rho}$ pulse sequence.*

The spin lattice relaxation rate $R_{1\rho}$ depends on the amplitude of the rf field hence the name relaxation dispersion [42]. Chemical exchange in the intermediate to fast regime is the effect that contributes to the $R_{1\rho}$ relaxation [43]. In $R_{1\rho}$ experiment, sample is subjected with rf field (**B1**) in the x-axis long enough to rotate magnetization vector **M** to the y-axis. At this field strength, **M** starts to precess around the y-axis with a precession frequency $\frac{\gamma B_1}{2\pi}$. If we now change the phase of the **B1** field by 90^{o} ($\pi/2$) and align it to the y-axis, *i.e.* rf is applied along the y-axis, the two fields are parallel and no force to rotate vector away from that axis, thus magnetization is spin locked along that axis. If we now look to the evolution of magnetization from the rotating frame point of view, the **B1** field would appear stationary and the relaxation of **M** appears similar to the spin lattice relaxation with rate $R_{1\rho}$. In the fast limit the relaxation rate becomes [49]

$$R_{1\rho} = R_1 \cos^2\theta + \left(R_2 - \left(\frac{\left(p_A p_B \Delta\omega^2\right) k_{ex}}{k_{ex}^2 + \omega_e^2} \right) \right) \sin^2\theta + \left(\frac{\left(p_A p_B \Delta\omega^2\right) k_{ex}}{k_{ex}^2 + \omega_e^2} \right) \sin^2\theta \qquad (47)$$

where $\omega_e = \sqrt{\omega_1^2 + \Omega^2}$ is the effective field, with $\theta = \arctan\left(\omega_1/\Omega\right)$ is the tilt angle. In the presence of R_1 relaxation data measured independently, R_2 can be extracted. The exchange parameters can be evaluated after $R_{1\rho}$ is measured at varying effective field strengths and tilt angles. Rotating frame relaxation dispersion experiment is useful in the study of physical processes such as loop motions, side chain motions, secondary structure element, and ligand binding.

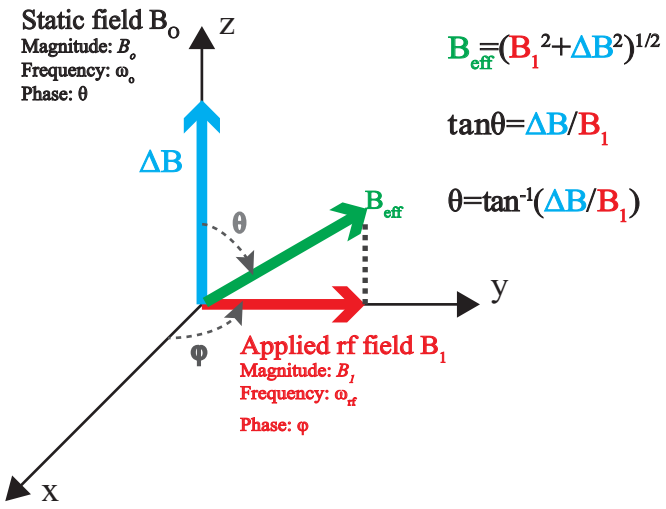

Figure 5 *Rotating frame representation, the effective field B_{eff} arises from the vector sum of ΔB and B_1, while the tilt angle, θ, is the angle between ΔB and B_{eff}.*

Table 1 *A comparisons between CPMG and $R_{1\rho}$ experimental techniques. These techniques are useful in study of exchange processes in the intermediate to fast exchange.*

CPMG	$\boldsymbol{R_{1\rho}}$
ms (milli-second)–exchange events	μs (micro-second) exchange events
R_2 – relaxation of transverse magnetization	$R_{1\rho}$ – relaxation of magnetization locked along the effective field
Laboratory frame of reference	Rotating frame of reference

2.6 A review of the recent research on polynomially scaling algorithms

A significant number of spin dynamics simulation methods cannot handle big spin systems [50-52]. This is due to the fact that the algorithms rely on numerical techniques such as matrix factorization and diagonalization [50-52]. These are well suited to study small systems and are not applicable to big systems that generate large matrices that are difficult to handle computationally. Numerical algorithms that depend on direct propagation of the density matrix in Liouville space scale exponentially with spin system size [3, 20].

Full Liouville space includes all correlations between all spins [3, 53]. But, in magnetic resonance simulation of large systems, it is found that a large number of spin states are not populated and can be left out of the state space. This led to the development of a numerical

algorithm that scales polynomially with the size of the spin system and makes use of the reduced state-space approximation method to identify and discard spin states that are not populated during spin dynamics [20]. Emsley *et al* also designed a similar algorithm that scales polynomially with the spin system for the simulation of strongly dipolar coupled spin ½ systems by using low-order correlation in Liouville space [53, 54]. The technique was implemented as a Suzuki-Trotter [55] numerical algorithm that avoids the storage of the Liouvillian and the propagator to make efficient use of memory. The idea of reduced space approximation is based on splitting spin system into clusters of connected spins, in which the Liouvillian matrix for each cluster is generated and recoupled during the recoupling stage to establish the overall reduced states. Hence the global Liouvillian works for the whole system.

It is believed [18, 20, 56] that high order spin correlations and high order coherences of spins that are remote in the interaction graph relax very fast or do not accumulate enough to give an observable signal. This is one of the reasons for the elimination of high order correlations and coherences of large spin systems such as proteins and nucleic acids during quantum mechanical spin dynamics simulations. Coherence order is determined from the difference between the number of the raising and lowering operators, where the spin states are given in the direct product operator form. In addition, any two spins that are remote in the interaction graph would not contribute to an observable signal and can be dropped from the simulation process. The state space reduction technique is adaptive, and can simulate local tight and loose couplings by altering cluster size.

The magnetic resonance simulations of large systems are further accelerated using on Krylov subspace technique [57] and zero track elimination [58]. This allows further compression of the state space and resulted into another algorithm been added to the package to further reduce simulation speed.

2.7 Objectives of the work

Time domain spin dynamics simulation in Liouville space is complicated due to the exponential scaling [18, 28] of the operator matrices dimensions (2^n for spin ½ nuclei, with n being the number of spins) involved that require large computational power and making it impractical to simulate spin dynamics of large biological systems. However emergence of algorithms [20, 58] that scale polynomially with system size makes it feasible for large-scale simulations of proteins and nucleic acids [3, 59]. The polynomial scaling algorithm is achieved by reducing

the dimension of the Liouvillian matrices accurately to significantly improve simulation time; making use of the reduced state space approximation technique to eliminate states that are not populated and insignificant. The exclusion of high order spin correlations and coherences is vindicated by the arguments that such states dephase or relax rapidly than the low order correlations does and do not accumulate to a significant extent. In addition two-spin order correlations that are topologically remote in the interaction graph accumulate slower; hence, excluded from the basis set.

This was the first hypothesis that led to subsequent development of sophisticated algorithms that form the bases for SPINACH, a software library, for spin dynamics simulation of large systems such as proteins and nucleic acids. This has helped to make possible quantum mechanical NMR simulation of proteins and nucleic acids in one go and allowed modelling of important biomolecular NMR pulse sequences including NOESY, HSQC and HNCO. As an illustration the quantum mechanical NMR modelling of the NOESY spectra of ^{13}C and ^{15}N isotopically labelled protein ubiquitin with over thousand coupled spins is available in Chapter four. The detailed description of the simulation algorithm that carefully prunes the density operator space along with the relaxation algorithm employed that implements the Bloch, Redfield, and Wangsness relaxation theory [60] is delineated in section three of chapter four and our published paper [59].

3 *J*–coupling predictor function for proteins

3.1 What has been done?

Three bond *J*-couplings of all the 20 standard amino acids are calculated using Gaussian. This was performed quantum mechanically using the DFT theory level B3LYP for structure optimization and M06 exchange correlation functional with basis set cc-pVTZ for the scalar couplings. The computation required the construction of the amino acid structure using Gauss View software - then optimize the structure before running the spin-spin coupling job. The Karplus coefficients [10] A, B, and C are then generated from the DFT calculations to establish the function that predicts *J*-couplings during protein NMR simulations. Such function did not exist before.

All the backbone and sidechain dihedral angles of the 20 amino acids are scanned making a constraint at a particular dihedral angle to extract the Karplus coefficients from the *J*-couplings and the associated dihedral angles using Karplus equation $\left({}^3J = A\cos^2\phi + B\cos\phi + C\right)$. Each and every dihedral angle is scanned for 36 steps at a size of 10 degrees.

When the scan of the energy is completed, the coordinates of each of the optimized conformations are extracted into a Gaussian input files to run the spin-spin coupling jobs and these were submitted to Southampton's HPC cluster, Iridis4, for the computation of magnetic interactions.

This produced a log files as an output of the calculation, which contains molecular properties along with the ${}^3J$ coupling values. The coupling between two spins that are three bond apart are then extracted and fitted to Karplus equation as described in section 3.8. Karplus curves and coefficients are then generated from the fitting and all the results are given in the results section of this chapter.

The dihedral angles computed using the DFT approach employed in the research work are the ϕ, ψ, and ω of the main chain, and χ_1, χ_2, χ_3, χ_4 and χ_5 of the sidechain torsion angles for all the 20 amino acids. From all the calculations and literature review, a *J*-coupling field was built and integrated into SPINACH as a *J*-coupling predictor function.

The scan of all the possible dihedral angles and computing of the ^3J coupling values as mentioned above for all the essential 20 amino acids - is a huge task, and required a significant chunk of CPU time.

To give an idea of the nature of the task, the ground state optimization of leucine, a 31 atom amino acid, elapsed 1128 CPU core hours. This gave 36 conformations of molecular structure from that single amino acid molecular structure; and are required to be optimized before running the computation of magnetic interactions for all the orientations of the molecular structure.

3.2 NMR methods for obtaining *J*-couplings in proteins

Experiments to measure *J*-couplings in protein NMR has advanced rapidly and saw an increase in applications due to the advancement of isotope labelling in hetronuclear 2D NMR experiments in the study of proteins and nucleic acids [61]. The series of experiments used to measure *J*-couplings are splint into several groups depending on the underlying principles that are employed to record the data. Some of the experiments to mention are the in-phase anti-phase (IPAP) [62] *J*-coupled technique, exclusive correlation spectroscopy [63] (E.COSY), and quantitative *J* correlation (QJ) [64]. In the in-phase and anti-phase (IPAP) [62] experiment, an in-phase and anti-phase NMR data are recorded and combined to provide an up-field and downfield component of a doublet in a spectrum. The time domain signals for the in-phase data are obtained from the evolution of an in-phase terms (operators) such as I_x and I_y, hence the name in-phase. The anti-phase experiment, however, evolves under an anti-phase operators (*e.g.* $2I_xS_z$, $2I_zS_x$), which are not observable, but over time evolve into an in-phase terms, which give an observable transverse magnetization in the time-domain signal. This shows the importance of the anti-phase components in multi-pulse and multi-dimensional experiments. The INEPT [65] technique is an example that makes use of anti-phase states to transfer coherences. The pulse sequence of the technique is heavily employed as a building block in a complex multi-pulse experiments. The IPAP technique is suitable to measure coupling constant for proteins and produces spin state separated spectra, which makes *J*-coupling splitting measurable. Unlike E.COSY discussed below, the IPAP is not sensitive to the size of the one bond coupling constant for separating the spin state signals.

For larger systems spins of active coupling (responsible for cross peak) are passively coupled to other spins gives a broad linewidth that exceed the size of the coupling value. Hence to

resolve and measure coupling values in such cases a method known as E.COSY was introduced. The technique was first proposed by Griesinger *et al* in 1986 and is used to measure coupling values in proteins and nucleic acids that are isotopically enriched. It is also helpful to determine the sign of the couplings [63]. The method simplifies cross peak multiplets by transferring coherence between exclusively connected transitions for spins that are weakly coupled. The method requires three spins represented as *ISX* in which one of the spins is actively coupled to the other two. Then excitation of the connected spins leaves the passive spins energy unperturbed and as a result the experiment generates reduced cross peak multiplets due to the actively coupled spins [66]. The utility of E.COSY experiment is hampered by the low sensitivity and limited digital resolution that the technique offers. To overcome the limitation of the E.COSY an experiment known as quantitative *J*-correlation have been developed. Quantitative *J* correlation experiments rely in the modulation of signal intensities due to the evolution of magnetization under the influence of *J*-coupling and are used to measure a large variety of couplings over two and three bond correlations [64]. There are a variety of this experiment. One type of the experiments quantitates the loss in signal that result when the coupling is active compared to the data obtained when the coupling is inactive. The second type of experiments runs in a series of measurements where the signal is modulated by the coupling evolution. In this type of experiments the coupling constant is extracted from the fitting of peak intensities from the $^{1}J_{NH}$ coupling modulated 2D spectra [67]. The other type of experiment determines the coupling constant from the intensity ratio of diagonal peak to that of the cross peak [68] in a 3D experiment. This was applied for the measurement of $^{3}J_{HnH\alpha}$ coupling values for ^{15}N labelled protein where the ratio of the intensity not affected by transverse relaxations of the amide and α proton spins, and are useful to determine intra-residue correlations.

3.3 Why are we doing this and what was done before?

Large-scale quantum mechanical protein and nucleic acid spectra simulation requires *J*-coupling predictor function. Protein and nucleic acid *J*-coupling predictor functions was built and implemented into Spinach. The functions incorporate the functionality of *J*-coupling estimation for protein and nucleic acid systems into Spinach, which will provide *J*-coupling values during NMR simulations of proteins, DNAs and RNAs.

Several commercial [69-72] and non-commercial [73, 74] *J*-coupling prediction methods are out there. The methods use different techniques to achieve the desired outcome. Lehtivarjo *et*

al. [73] for example have developed a universal *J*-coupling predictor method based on data driven approach. The method involves prediction parameters to take into account the effect of hydrogen bond and substituents, have on *J*-coupling. Juniper algorithm, developed as a result of the work, is designed to predict coupling for small molecules. The algorithm would take long time for proteins and would ruin structure, as the technique do geometry optimization based on small molecule force field. Due to conformational variability of proteins, the ^{3}J-couplings data Juniper algorithm have, might not estimate to a sufficient accuracy vicinal coupling for proteins, as the coupling values are obtained from small molecules. Another similar package is SPINUS. It predicts proton *J*-couplings of small organic molecule along with chemical shifts using Associated Neural Networks (ASNN) [75] from a given structure. The program involves a method of data collection from literature and also experimental data from collaborators. For description purposes hydrogen atoms split into aromatic, nonaromatic π, rigid aliphatic, and non-rigid aliphatic depending to which the proton is bonded. The descriptors are physicochemical, geometrical, and topological descriptors. A data set of couplings categorized into several classes depending on the type and number of bonds between protons is entered into the ASNN memory. These are then used to predict couplings for a new pair of coupled protons.

Wang *et al*. [76] also utilized DFT and computed ^{3}J-couplings for GB3 protein using automated fragmentation quantum mechanics/molecular mechanics approach that integrates protein environment effect. The approach avoids the computational intensive quantum chemical calculations by splitting the entire protein into non-overlapping fragments in which residues within certain range are assigned as a buffer region. The fragment and the buffer region are then treated quantum chemically while the rest of the protein approached empirically by point-charge model.

Another accounts of *J*-coupling predictor functions are the works of Xia and Roy *et al*. who discussed the implementation of the technique in their papers [77, 78] and [79] respectively.

3.4 Origin and theoretical study of spin-spin coupling

In electronic structure calculations, magnetic parameters, such as spin-spin coupling constants, can be computed using wave function methods and density functional methods. In this work we have investigated spin-spin coupling using density functional theory. The basic argument for density functional theory is that ground state energy is determined by electron density [80],

and the ground state electron density ρ of a chemical system with interacting electrons in the presence of external potential can be used to predict the properties of the system. DFT theory significantly reduces the complexity of computing system properties by making the problem three-dimensional [12, 81].

According to Hohenberg and Kohn [80], the energy functional of a system is given by

$$E[\rho] = \left\langle \psi_0 \mid \hat{T} + \hat{V} + \hat{U} \mid \psi_0 \right\rangle \tag{48}$$

where ρ is the ground state electron density, $\hat{T}$ is kinetic energy operator, $\hat{V}$ is the electron-nuclear attraction operator, and $\hat{U}$ is the electron-electron repulsion operator. The variational principle in quantum chemistry states that $E[\rho] = E_0$, *i.e.* there exists a solution for the electron density function with a global minimum at the ground state density. Kohn-Sham theory mimics wave function theory in its usage of orbitals [12]. This allows choosing a basis set and set up one-electron orbitals, and to generate trial density to optimize against energy E.

$$\rho(\vec{r}) = \sum_{k=1}^{N} \left| \chi_k(\vec{r}) \right|^2 \qquad \chi_k(\vec{r}) = \sum_n a_{kn} \varphi_n(\vec{r}) \tag{49}$$

where $\chi_k(\vec{r})$ are one electron orbitals, $\varphi_n(\vec{r})$ are the basis set functions and a_{kn} are coefficients.

The model, introduced by Kohn and Sham [11], has similar mathematics to the Hartree-Fock (HF) method, possessing the same formula for kinetic, electron-nuclear, and Coulomb energies apart from the exchange-correlation term. The equation for a single electron orbital is

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + \hat{V}(\vec{r}) \right] \chi_k(\vec{r}) = \varepsilon_k \chi_k(\vec{r}) \tag{50}$$

where $-\frac{\hbar^2}{2m}\nabla^2$ is the kinetic energy term, $\hat{V}(\vec{r})$ is the potential energy term and ε_k is the energy eigenvalue. The term which is not known in DFT is the exchange-correlation energy, in which the remaining energy is absorbed. This term may be written as

$$E_{\mathrm{XC}}[\rho] = \int \rho(\vec{r})\, \varepsilon_{\mathrm{XC}}[\rho(\vec{r})]\, dv \tag{51}$$

where $\rho(\vec{r})$ is the density that indicates the number of electrons per unit volume and $\varepsilon_{XC}\left[\rho(\vec{r})\right]$ is the exchange-correlation energy density which represents energy per electron.

An important research area within DFT is to derive approximations to the exchange-correlation functional. A functional takes a function and returns a scalar. In quantum chemistry functionals are denoted as $F[f]$. There are a variety of exchange-correlation functionals in DFT that employ different approximations and all aim to increase the accuracy of the method. The spin systems structure was optimizated using hybrid XC functional B3LYP [12, 81-83], which is a three-parameter functional defined by

$$E_{XC}^{B3LYP} = (1-a)E_X^{LDA} + aE_X^{HF} + b\Delta E_X^B + (1-c)E_C^{LDA} + cE_C^{LYP} \tag{52}$$

with typical values for a, b and c are 0.2, 0.7, and 0.8 respectively.

The basis set employed in the calculation of the indirect spin-spin coupling magnetic parameters is cc-pVTZ, which stands for correlation consistent polarized Valence Triple Zeta [84]. By convention correlation consistent is denoted as (cc) lower case letters to avoid confusion with coupled cluster (CC) method. Correlation consistent [84, 85] basis sets are developed by Dunning *et al*. and employs a smaller set of primitives to achieve similar or better accuracy than the "STO-nG" type basis sets. STO-nG – stands for Slater Type Orbitals with n is the number of Gaussian Type Orbitals that are linearly combined to mimic an STO. The Slater Type and Gaussian Type functions for STO and GTO are

$$s(\vec{\mathbf{r}}) = e^{-\zeta|\mathbf{r}|} \qquad g(\vec{\mathbf{r}}) = e^{-\zeta\mathbf{r}^2} \tag{53}$$

where $\vec{\mathbf{r}}$ is coordinate vector and ζ is zeta and controls the width of orbitals. Correlation consistent basis sets are designed in a way that recovers correlation energy of valence electrons. However, with the addition of core functions, nuclear properties such as magnetic shielding and spin-spin couplings, can also be calculated. The cc-pVTZ basis set incorporates the basis functions listed in **Table 2** (below) for hydrogen, carbon, and nitrogen in the amino acids investigated for magnetic parameters in this work.

Accurate spin-spin coupling requires large basis set [86]. This is partly due to the Fermi Contact contribution requirements of accurate nuclear regions that the normal basis set would not offer [86, 87]. The magnetic parameters were computed using the GIAO method, which stands for Gauge-Independent Atomic Orbitals. It is a method that employs basis functions that explicitly

possesses field dependence [88, 89]. As the NMR shielding calculation requires, the structure is optimized using B3LYP hybrid functional before the spin-spin coupling was computed using the option "spinspin=mixed" in Gaussian 09 that request a two-step coupling calculation to improve the accuracy of spin-spin coupling calculation [90, 91].

In order to reflect the potential environment of proteins in biological systems the studied system was treated in polarizable continuum solvent water and using SMD model [92]. In quantum chemistry solvent models are broadly classified into two categories. These are implicit and explicit models. There are a variety of implicit methods. Generally implicit method is defined as the model in which explicit solvent model are removed and replaced by a polarizable continuum that considers the medium as a continuous isotropic and thermally averaged medium [93]. SMD model belongs to the implicit treatment.

Table 2 *The contracted correlation consistent basis set basis functions for hydrogen, carbon, and nitrogen employed in the computation of indirect spin-spin coupling. For comparison cc-pVDZ and cc-pVQZ basis sets are also included.*

Basis	**Hydrogen**	**Carbon**	**Nitrogen**
cc-pVDZ	2s1p	3s2p1d	3s2p1d
cc-pVTZ	3s2p1d	4s3p2d1f	4s3p2d1f
cc-pVQZ	4s3p2d1f	5s4p3d2f1g	5s4p3d2f1g

3.5 Ramsey theory: spin-spin coupling mechanism

In non-relativistic DFT theory, spin-spin couplings are obtained from individual contributions of the following four mechanisms [14, 94]:

1. Diamagnetic spin orbit (DSO) term
2. Paramagnetic spin-orbit (PSO) term
3. The spin-dipole term and
4. Fermi contact term

Of all the mechanisms, the Fermi-contact term is the most prominent contribution. The spin-spin coupling tensor is the second derivative of the total electronic energy with respect to magnetic moments of the nuclei

$$J_{nm} = \hbar \frac{\gamma_n}{2\pi} \frac{\gamma_m}{2\pi} K_{nm}; \qquad K_{nm} = \left. \frac{d^2 E}{d\vec{\mu}_n d\vec{\mu}_m} \right|_{\mu_n = \mu_m = 0} \tag{54}$$

where J_{nm} is the coupling tensor between the spins n and m that have magnetic moments of $\vec{\mu}_n$ and $\vec{\mu}_m$. The experimentally observed coupling tensor in liquids and gases *i.e.* systems that tumble freely may be obtained using the trace as

$$J = \frac{1}{3} \mathrm{Tr}(\mathbf{J}) \tag{55}$$

According to Ramsey's non-relativistic theory, the coupling tensor may be calculated using second order perturbation theory [95, 96]

$$\begin{aligned} K_{nm} = \left\langle 0 \middle| \hat{H}_{nm}^{\mathrm{DSO}} \middle| 0 \right\rangle &- 2 \sum_{n_s \neq 0} \frac{\left\langle 0 \middle| \hat{H}_n^{\mathrm{PSO}} \middle| n_s \right\rangle \left\langle n_s \middle| \left(\hat{H}_m^{\mathrm{PSO}} \right)^{\mathrm{T}} \middle| 0 \right\rangle}{E_{n_s} - E_0} \\ &- 2 \sum_{n_T} \frac{\left\langle 0 \middle| \hat{H}_n^{\mathrm{FC}} + \hat{H}_n^{\mathrm{SD}} \middle| n_T \right\rangle \left\langle n_T \middle| \left(\hat{H}_m^{\mathrm{FC}} \right)^{\mathrm{T}} + \left(\hat{H}_m^{\mathrm{SD}} \right)^{\mathrm{T}} \middle| 0 \right\rangle}{E_{n_T} - E_0} \end{aligned} \tag{56}$$

The first summation is over all excited singlet states with energy E_{n_s}, while the second summation is over all triplet states with energy E_{n_T}. The operators for the four mechanisms in atomic units are:

$$\begin{aligned} \hat{H}_{nm}^{\mathrm{DSO}} &= \alpha^4 \sum_i \frac{\mathbf{r}_{in}^{\mathrm{T}} \mathbf{r}_{im} \mathbf{I}_3 - \mathbf{r}_{in} \mathbf{r}_{im}^{\mathrm{T}}}{\mathbf{r}_{in}^3 \mathbf{r}_{im}^3}, \\ \hat{H}_n^{\mathrm{PSO}} &= -i\alpha^2 \sum_i \frac{\mathbf{r}_{in} \times \nabla_i}{\mathbf{r}_{in}^3}, \\ \hat{H}_n^{\mathrm{FC}} &= \frac{8\pi\alpha^2}{3} \sum_i \delta(\mathbf{r}_{in}) s_i, \\ \hat{H}_n^{\mathrm{SD}} &= \alpha^2 \sum_i \frac{3\mathbf{r}_{in}^{\mathrm{T}} s_i \mathbf{r}_{in} - \mathbf{r}_{in}^2 s_i}{\mathbf{r}_{in}^5}, \end{aligned} \tag{57}$$

where α is the fine structure constant [14], $\mathbf{r}_{in}$ is the position of electron with respect to nucleus, $\delta(\mathbf{r}_{in})$ is the Dirac delta function, $\mathbf{I}_3$ is a three-by-three unit matrix, and s_i is spin of electron. All the summations in Equation (57) are over all electrons in the system. The superscript T shows that the vectors are transposed.

3.6 Karplus law derivation – dependence on dihedral angle

Vicinal or three bond ($^3J(X, Z) = {}^3J(X\text{-}Y\text{-}W\text{-}Z)$) spin-spin couplings in magnetic resonance are noted for their dependence on the local dihedral or torsion angle. The dihedral angle is defined as $\phi = X - W - Y - Z$, as depicted in **Figure 6** (below). To obtain the dihedral angle, first, it is necessary to compute three intermediate vectors; $a_{\mathrm{n}} = W - X$, $b_{\mathrm{n}} = Y - W$ and $c_{\mathrm{n}} = Z - Y$. Their unit vectors are obtained as

$$a = \frac{a_{\mathrm{n}}}{|a_{\mathrm{n}}|}, \qquad b = \frac{b_{\mathrm{n}}}{|b_{\mathrm{n}}|}, \qquad c = \frac{c_{\mathrm{n}}}{|c_{\mathrm{n}}|}; \tag{58}$$

The dihedral angle ϕ is then derived from the unit vectors using an inverse tangent function with two arguments as an input:

$$\phi = 180^{\mathrm{o}} \cdot \frac{\arctan(y, x)}{\pi} \tag{59}$$

where $x = (|b| \cdot a) \cdot (b \times c)$ and $y = (a \times b) \cdot (b \times c)$. The use of the atan2 function rather than the inverse cosine function method [97] is recommended, because this avoids a singularity.

The dependence of NMR vicinal coupling values on dihedral angle is given to a good approximation by Karplus equation [10]:

$$^3J = A\cos^2\phi + B\cos\phi + C \tag{60}$$

This equation is useful in determining the geometry of spin systems from coupling constants and for deducing the coupling constants from atomic coordinates.

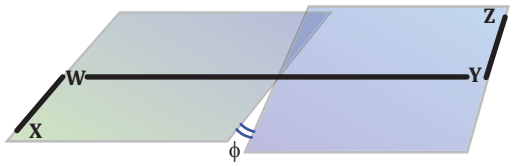

Figure 6 *Visualization of dihedral angle that connects arbitrary nuclei X and Z.*

The coefficients *A*, *B*, and *C* of the Karplus equation are parameters that depend on the types of atoms that are coupled through the three bonds *i.e. X* and *Z* in this case and the substituents that are attached to atoms *Y* and *W*, atoms that define the torsion angle. Vicinal spin-spin coupling constants depend on the cosine of the angle ϕ and its square. Complicated relations between vicinal coupling constant and dihedral angle exist that include substituent effects [98, 99]. In particular, Abraham and Pachler derived an expression that included the electronegativity effect [100]. In protein NMR spectroscopy the ^3J coupling constant is linked to backbone and side chain angles. The backbone angle is described using ϕ, ψ, and ω torsion angles, while the side chain conformation can be defined by up to five dihedral angles, namely χ_1, χ_2, χ_3, χ_4 and χ_5.

The fragment of a protein chain in **Figure 7** (below) illustrates the ^3J coupling network in the backbone and side-chain torsion angles that characterize the conformation.

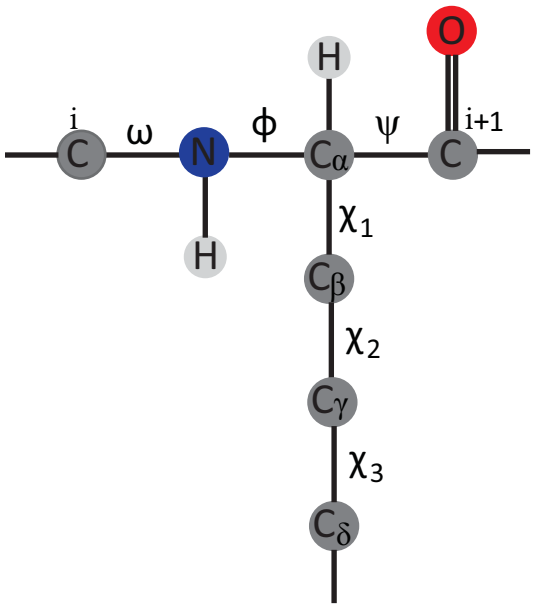

Figure 7 *Schematic illustration protein backbone ϕ, ψ, ω and side-chain dihedral angles χ_1 through to χ_3.*

The ϕ dihedral angle is associated with up to six ^3J coupling constants. These are $^3J_{HnHa}$, $^3J_{HnCB}$, $^3J_{HnCO}$, $^3J_{COCB}$, $^3J_{COHA}$, and $^3J_{COCO}$.

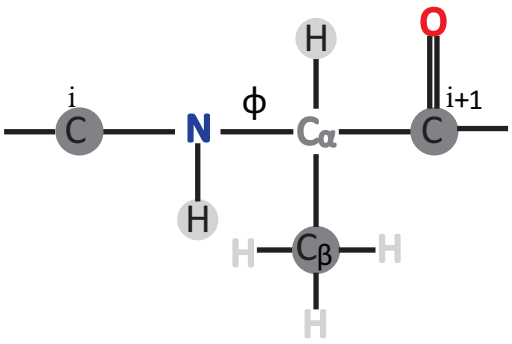

Figure 8 *ϕ dihedral angle illustration – a rotation around the bond NH and Cα in protein backbone. This dihedral angle is associated with six ^3J coupling constants. The spins connected by these coupling constants are spins of the circled atoms.*

The second most important backbone dihedral angle in the conformation of protein is the ψ torsion angle. Although in principle there are six ^3J values that are related to ψ, the coupling constants $^3J_{HaiNi+1}$ and $^3J_{CBiNi+1}$ are the only interactions with sufficient contribution that would provide information to describe dependence of the local geometry on the coupling values. **Figure 7** (above) gives a schematic of the χ_1, angle and the associated spins are indicated in **Figure 9** (below). Glycine is the only amino acid, which has no χ torsion angle.

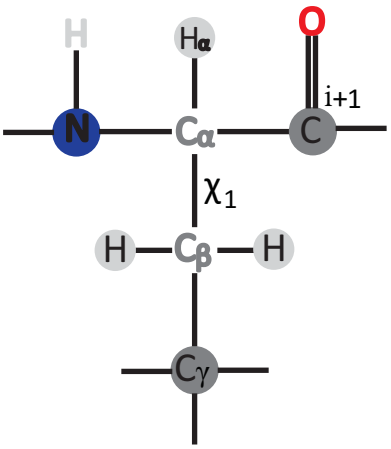

Figure 9 *Side-chain angle χ_1 - a rotation around the bond between C_α and C_β. Circled atoms illustrate ^3J coupled spins that are affected by the rotation about the χ_1 dihedral angle.*

Side chain sequence determines the unique properties of a particular protein. This includes the biological activity and the three dimensional structures of proteins [97]. Theoretically there exist up to five χ dihedral angles to define the side chain conformations; in practice the χ_1 and χ_2 are well studied. The dihedral angle χ_1 determines up to nine ^3J coupling constants. The four atoms that define the dihedral angle are N, C_α, C_β, and C_γ.

Accurate parameterization of Karplus equation is necessary in order to associate ^3J couplings to dihedral angle. However calibration of Karplus equation and determination of dihedral angle are complicated by relaxation induced self-decoupling, which reduces couplings in larger proteins and intra-molecular motions. It also averages ^3J values over distribution of dihedral angle [25].

3.7 Complexity scaling of DFT calculations

The DFT approach for computing indirect spin-spin coupling is not realistic for proteins. As it is demonstrated in **Figure 10** (below) the complexity increases steeply as the system size increases, hence it is not feasible to run coupling calculations even for small proteins such as ubiquitin and GB3. The main reason for this is the increase in the number of orbitals in DFT, which increases the complexity by increasing the calculation time at least cubically [11, 12, 101]. Therefore, we have created an empirical parameterization that looks similar to a molecular dynamics force field.

Table 3 *List of the amino acids studied for J-coupling using GIAO B3LYP/ M06/cc-pVTZ and the total CPU time in hours elapsed for the computation of spin-spin coupling. B3LYP was employed for geometry optimization while M06 was used for the computation of spin-spin coupling.*

Amino acid	**number of Atoms**	**Calculation time in hours**	**CPU core count**	**Total core-hours**
Glycine	19	10	12	120
Alanine	22	15	12	180
Arginine	36	35.5	24	852
Asparagine	26	54	12	648
Aspartic acid	25	69	12	828

Glutamine	29	73.5	12	882
Glutamic acid	28	17.5	24	420
Histidine	30	143.5	16	2296
Isoleucine	31	90	12	1080
Leucine	31	94	12	1128
Lysine	34	28	24	672
Methionine	29	20.5	24	492
Phenylalanine	32	192	24	4608
Proline	26	19	16	304
Serine	23	20	12	240
Threonine	26	79.5	16	1272
Tryptophan	36	46	24	1104
Tyrosine	33	31.5	24	756
Valine	28	104.5	16	1672

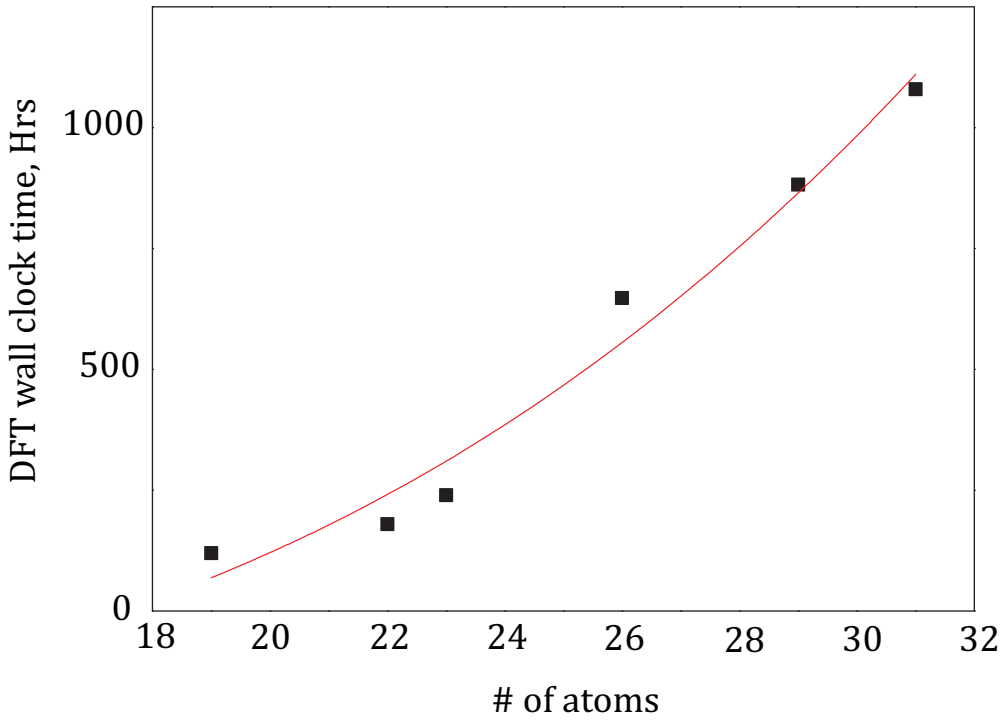

Figure 10 *Wall clock DFT time elapsed in the computation of spin-spin coupling constant for single amino acids using GIAO M06/cc-pVTZ in SMD water.*

3.8 Empirical parameterization of *J*–couplings

In Spinach protein module the function "`protein.m`" takes as input a PDB (Protein Data Bank) and BMRB (Biological Magnetic Resonance Bank) files. The PDB file should contain coordinates of all the atoms (along with those of protons). In a case where the file possesses multiple structures, Spinach by default reads the first geometry. The BMRB file should also contain chemical shift data of the spins assigned. If there are unassigned nuclear spins, these are used in the *J*-coupling prediction, but do not appear in the simulated spectrum. The function uses the command below to take the two files as input and spits out Spinach input data structures.

```
[sys, inter]=protein('pdb_file','bmrb_file')
```

The following fields are returned from the command above:

```
sys.isotopes
```

```
sys.labels
```

```
inter.zeeman.scalar

inter.zeeman.matrix

inter.coupling.scalar

inter.coordinates
```

The field names are user friendly and straightforward with each isotope names, PDB labels are assigned to `sys.isotopes` and `sys.labels`. The Zeeman interaction, chemical shift anisotropy (CSA), and estimated *J*-couplings are assigned to the `inter.zeeman.scalar`, `inter.zeeman.matrix`, and `inter.coupling.scalar` respectively. The Cartesian coordinates of the atoms from the PDB are placed into the `inter.coordinates`.

The actions that undertaken by Spinach during protein data import are:

I. PDB file is parsed, amino acid numbers and types along with PDB atom ids and coordinates are read in.

II. BMRB file is parsed, amino acid numbers and types are extracted in addition to the BMRB atom ids and chemical shifts.

    The function removes atoms that are not required in the simulation processes oxygens, sulphurs, and terminal atoms such as hydroxyl groups. It then matches BMRB chemical shifts to the PDB coordinates. Then it prints out missing atoms that will not appear in the simulation.

III. The *J*-couplings are estimated using the following procedure.

    The molecular bonding graph is split into connected sub-graphs. The angles and dihedrals are then computed from atomic coordinates.

    One bond coupling

    The molecular bonding graph is split into connected subgraphs of size two. The *J*-couplings are assigned from a complete database of nuclear pairs. One bond *J*-couplings are either found from literature or are estimated using DFT for individual amino acids.

    Two bond couplings

The molecular bonding graph is partitioned into connected subgraphs of size three, and two bond *J*-couplings are assigned from a database of connected triples of nuclei. The couplings are assigned from experimental data and DFT calculations.

Three bond couplings

The molecular bonding graph is split into subgraphs of sequentially connected four atoms and the dihedral angle is computed from atomic coordinates [59]. DFT calculations are used to compute ^{3}J couplings, by scanning dihedrals and the data fitted to Equation (60) to extract the coefficients *A*, *B*, and *C*, which are then used in the simulation of proteins.

Couplings across more than three bonds are ignored as the couplings are of vanishingly small size. However, care has been taken for amino acids that contain aromatic side chain where long range couplings would provide useful structural information. An estimated ^{4}J coupling values between two proton spins in aromatic systems falls in the range of 2 to 3 Hz. Examples of these amino acids are tyrosine, tryptophan, histidine, and phenylalanine. Electrostatic environment was also left out as its effect is small enough to not affect the outcome to the accuracy required for the quantum mechanical simulations of protein [7, 102].

IV. CSA estimation procedure is run for amide nitrogen atoms only with the local directions determined from the coordinates of N, H, C and CA spins. CSA tensor is directed along ZZ, YY, and XX, which are assumed as collinear to the N-CO bond, perpendicular to the peptide plane, and perpendicular to the other two planes respectively.

Once the import function completes running the *sys* and *inter* structures are outputted and used in the subsequent simulation of the protein system by Spinach.

3.9 Results and dihedral angle scan plots for all amino acids

The Karplus coefficients A, B, and C are obtained by least squares fit of Equation (60) to the DFT computed *J*-coupling value scans. Least squares is a mathematical approach that minimizes the sum of the squares of the deviation of the data points from the line of best fit

[103]. Spinach function takes a series of *J*-coupling values from a Gaussian *J*-coupling calculation logs that differ in the value of dihedral angle and numbers of the four consecutive spins that make the dihedral angle. The form of curve fitting employed in our work is the linear least squares fitting.

The least square fitting of Karplus equation given by ${}^3J = A\cos^2\phi + B\cos\phi + C$, where the solution for the constants A, B and C is given by the formulas:

$$\begin{pmatrix} \cos^2\phi_1 & \cos\phi_1 & 1 \\ \cos^2\phi_2 & \cos\phi_2 & 1 \\ \cos^2\phi_3 & \cos\phi_3 & 1 \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ \cos^2\phi_{35} & \cos\phi_{35} & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} J_1 \\ J_2 \\ J_3 \\ . \\ . \\ . \\ . \\ J_{35} \end{pmatrix} \tag{61}$$

If the matrix and the vector on the left are represented as Ω and x, and the vector on the right as y ; Equation (61) would be of the form:

$$\Omega x = y \tag{62}$$

Thus the solution to Equation (62) is given by

$$x = \Omega^{-1} y \tag{63}$$

where the pseudoinverse Ω^{-1} is computed using Matlab function.

A large database of amino acid specific of the Karplus coefficients *A*, *B*, *C* has been generated and incorporated into Spinach to predict the coupling between spins in magnetic resonance simulations of proteins. The result is presented below as Karplus curves and Boltzmann probability distribution histograms of all the dihedral angle scans for the sites designated as flexible. For the Boltzmann probability distribution histograms, the four atoms for which the dihedral angle is calculated, are exactly the same four atoms that the *J*-coupling is propagated across. Proline is not scanned as it has no part that deems to be scanned at a dihedral angle however, it has a five membered ring that do isomerize between cis and trans forms. Side chain scan for Alanine and Glycine was not performed as these amino acids possess no side chain with relevant dihedral angles.

3.9.1 Karplus coefficients, ALA

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
N	C	CA	N	0.05	0.29	-0.04	0.17	0.02	0.15
HA	CA	C	N	1.09	0.38	-0.02	0.17	0.03	0.15
HA	CA	C	N	-0.88		-0.61		-0.27 [104]	
CB	CA	C	N	0.18	0.33	0.03	0.14	0.18	0.18
H	N	CA	HA	10.80	0.36	-0.11	0.20	0.03	0.20
H	N	CA	HA	6.51		-1.76		1.60 [105]	
C	N	CA	HA	5.67	0.27	-1.68	0.13	0.70	0.15
H	N	CA	C	6.81	0.36	-1.23	0.16	-0.30	0.22
H	N	CA	CB	5.54	0.34	-0.65	0.30	-0.33	0.15
H	N	CA	CB	3.06		-0.07		0.13 [106]	
C	N	CA	C	2.64	0.26	-1.00	0.13	0.35	0.16
C	N	CA	C	1.33		-0.88		0.06 [107]	
C	N	CA	CB	2.04	0.26	-0.70	0.14	0.31	0.16

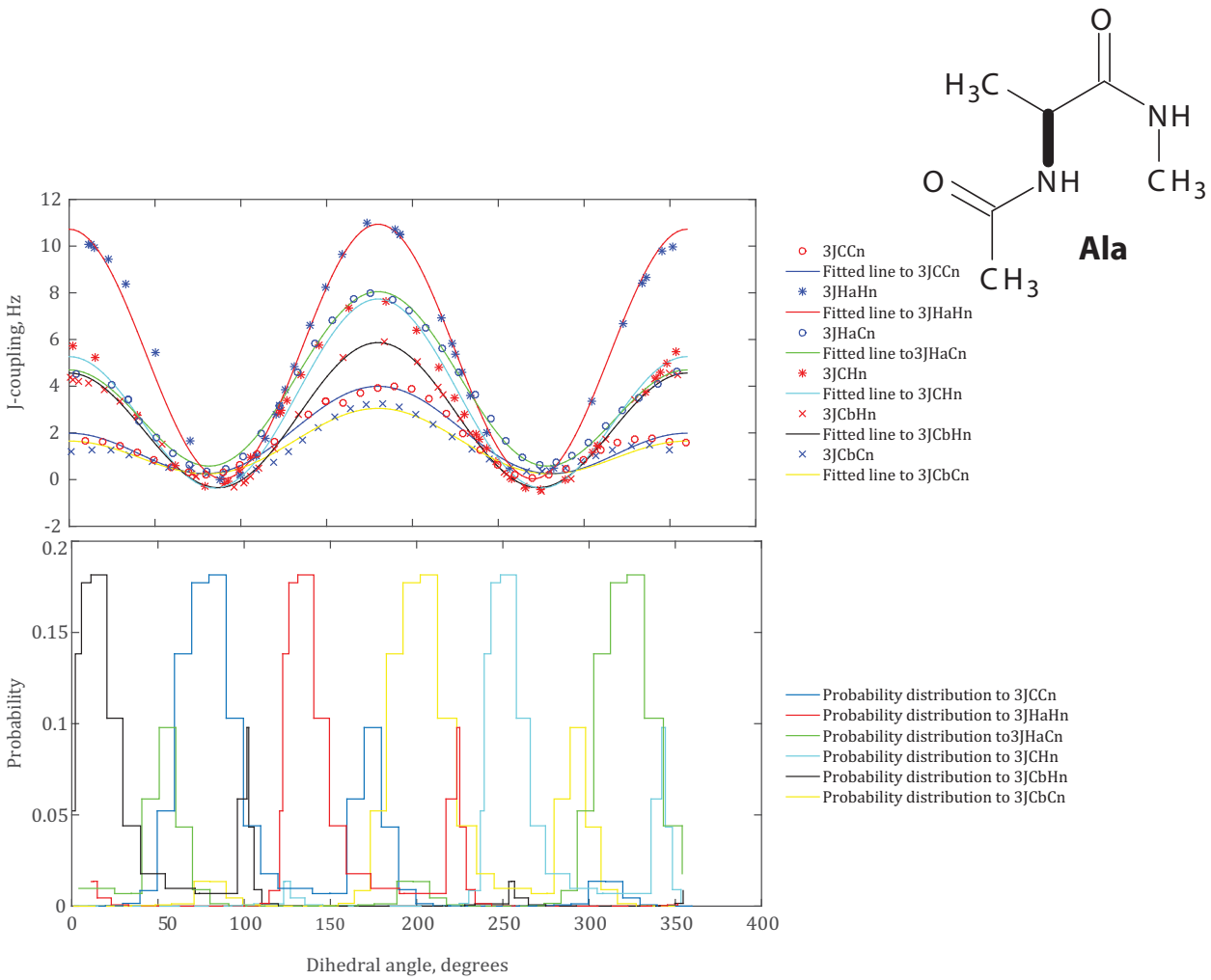

Figure 11 *ALA ϕ dihedral angle scan and the Karplus curves of all the torsion angles related to the dihedral angle (top panel) along with Boltzmann probability distribution graphs of the respective three bond coupled spins (bottom panel). References in the table indicate that the Karplus coefficient values that are obtained from that literature and integrated into Spinach as that values came from experiment, however all the Karplus curves in the graph are acquired from theory.*

The result of the fitting of ^3J coupling data to Karplus curve for six torsion angles is given in **Figure 11** (above). Top panel shows the Gaussian dihedral angle scan data nicely follow the function for all Karplus curves. The bottom panel Boltzmann probability distribution histograms are drawn as stairs. The probability distribution is determined using Equation (64) *i.e.* dividing the Boltzmann factor by the partition function:

$$P_i = \frac{\exp\left(-\frac{E_i}{RT}\right)}{\sum_k \exp\left(-\frac{E_k}{RT}\right)} \tag{64}$$

where P_i and E_i are the probability and energy at state i, while R and T are the ideal gas constant and temperature of the system respectively.

In the generation of the dihedral graphs in Matlab, all that was required were Spinach function "`dihedral`" to compute dihedral angles connecting an arbitrary binary spins and another function "`karplus_fit`" that fits Karplus equation to coupling data, from electronic structure calculation at the above mentioned method, at a specific dihedral angle and plots the Karplus curves, and the probability distribution as histograms.

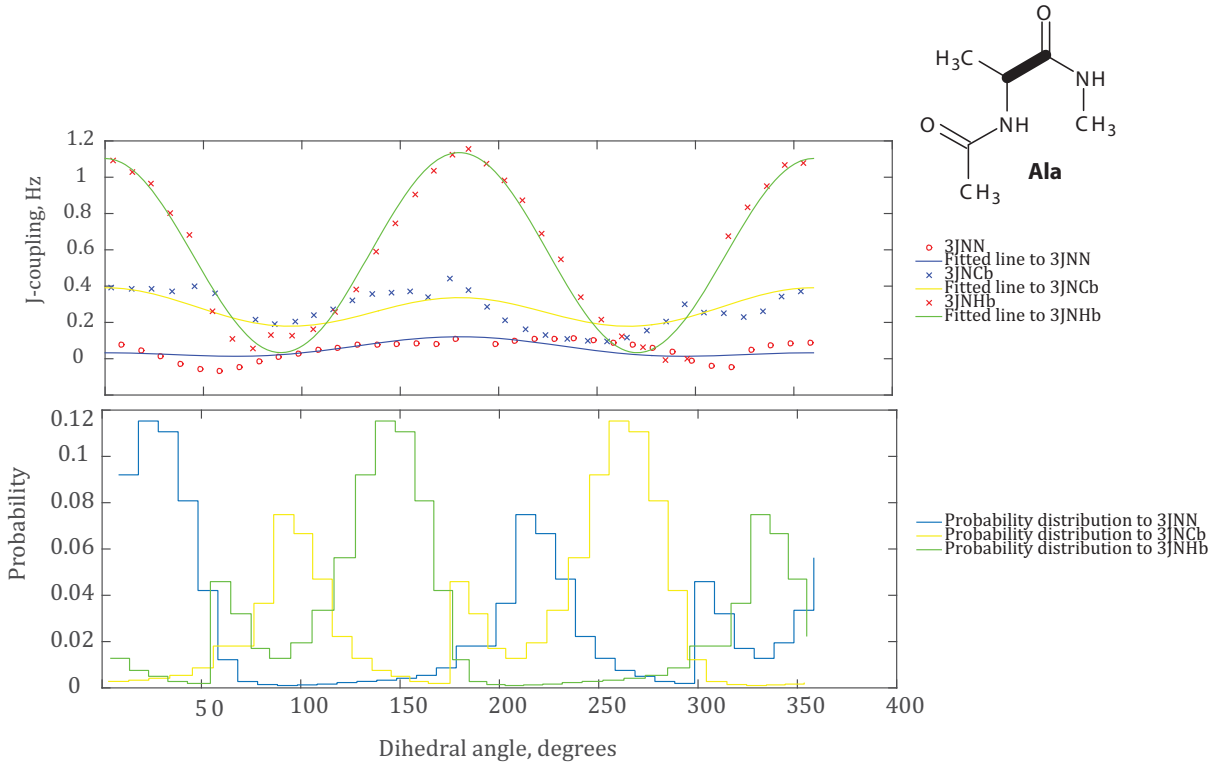

Figure 12 *ALA ψ dihedral angle scan and the Karplus curves of all the torsion angles associated to the ψ dihedral angle along with the Boltzmann probability distribution graphs of the respective coupled spins top and bottom panels respectively.*

Three bond scalar couplings (^3J) between two nitrogen spins is not very well studied. The amplitudes are found to be very small. As it is depicted in **Figure 12** (above) the ^3J between

nitrogen and proton at alpha position shows the dependence of the coupling values on dihedral angle or the Karplus relation while the coupling between the above mentioned spins do not. Therefore, a pattern of under and over prediction specifically for the $^3J_{NCb}$ coupling data is clearly visible in **Figure 12** (above). The probability density distribution graph shows, the probability distribution in the range of 10° to 60° and 170° to 250° for $^3J_{NN}$ of the ψ backbone dihedral angle. There is also probability distribution of just over 0.04 at 310° which goes down to around 0.01 that went up to nearly 0.05 at 360°. For the $^3J_{NCb}$ the distribution is mostly in the range 200° to 300° with another pronounced peak in between 50° and 150° along with a small blip from 180° to 200°. Similar pattern for the $^3J_{NHb}$ dihedral angle however the probability distribution is at different conformations, please refer to probability distribution graph in **Figure 12** (above) to find out the range of coverage.

3.9.2 Karplus coefficients, ALA

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water M06/cc-pVDZ optimized structure

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
H	N	CA	HA	10.88	0.41	-0.23	0.20	-0.03	0.22
C	N	CA	HA	5.82	0.27	-1.67	0.15	0.68	0.14
H	N	CA	C	7.01	0.37	-1.21	0.16	-0.35	0.23
H	N	CA	CB	5.51	0.27	-0.71	0.19	-0.33	0.16
C	N	CA	C	2.65	0.27	-1.01	0.13	0.40	0.16
C	N	CA	CB	2.10	0.29	-0.69	0.14	0.30	0.17

3.9.3 Karplus coefficients, ARG

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
N	C	CA	N	0.01	0.37	-0.03	0.21	-0.02	0.23
HA	CA	C	N	1.03	0.32	0.06	0.15	-0.13	0.16
HA	CA	C	N	-0.88		-0.61		-0.27 [104]	
CB	CA	C	N	0.13	0.36	-0.04	0.16	0.03	0.17
HA	CA	CB	CG	7.89	0.32	0.35	0.12	-0.08	0.22
N	CA	CB	CG	1.18	0.42	-0.08	0.17	-0.00	0.23

C	CA	CB	CG	3.26	0.29	-0.67	0.12	0.12	0.19
HA	CA	CB	HB1	11.60	0.38	0.73	0.15	0.49	0.20
HA	CA	CB	HB1	9.5		-1.60		1.80 [108]	
N	CA	CB	HB1	2.77	0.35	-0.09	0.13	0.06	0.22
C	CA	CB	HB1	7.20	0.39	-0.53	0.17	0.09	0.22
C	CA	CB	HB1	7.2		-2.04		0.60 [109]	
HA	CA	CB	HB2	12.14	0.31	1.68	0.15	0.75	0.19
N	CA	CB	HB2	3.44	0.32	-0.36	0.11	0.06	0.22
C	CA	CB	HB2	7.51	0.31	-0.78	0.12	0.17	0.19
H	N	CA	HA	11.20	0.30	0.10	0.12	-0.40	0.20
C	N	CA	HA	5.81	0.38	-1.90	0.19	0.34	0.21
H	N	CA	CB	5.61	0.34	-0.61	0.17	-0.34	0.14
C	N	CA	CB	2.02	0.30	-0.87	0.15	0.13	0.15
H	N	CA	C	6.39	0.37	-1.24	0.18	-0.28	0.15
C	N	CA	C	2.73	0.25	-0.86	0.13	0.06	0.14
CA	CB	CG	CD	5.92	0.29	-0.11	0.16	0.14	0.16
HB1	CB	CG	CD	12.54	0.39	0.02	0.22	-0.03	0.20
HB2	CB	CG	CD	13.48	0.33	-0.10	0.22	-0.18	0.17
CA	CB	CG	HG1	7.93	0.40	-0.03	0.23	0.07	0.23
HB1	CB	CG	HG1	13.12	0.31	1.08	0.13	0.32	0.17
HB2	CB	CG	HG1	12.90	0.34	0.91	0.15	0.63	0.15
CA	CB	CG	HG2	7.65	0.38	0.31	0.15	0.08	0.26
HB1	CB	CG	HG2	11.76	0.34	0.75	0.14	0.76	0.18
HB2	CB	CG	HG2	13.09	0.31	1.57	0.12	0.50	0.22
CB	CG	CD	NE	1.49	0.32	-0.20	0.18	-0.01	0.15
HG1	CG	CD	NE	5.18	0.28	-0.74	0.13	-0.02	0.18
HG2	CG	CD	NE	4.25	0.34	-0.09	0.12	0.09	0.22
CB	CG	CD	HD1	7.45	0.35	-0.50	0.14	0.16	0.20
HG1	CG	CD	HD1	12.79	0.36	0.41	0.15	0.83	0.18
HG2	CG	CD	HD1	13.48	0.39	0.82	0.18	0.15	0.22
CB	CG	CD	HD2	6.74	0.33	-0.10	0.12	0.17	0.20

HG1	CG	CD	HD2	13.28	0.36	0.86	0.19	0.31	0.20
HG2	CG	CD	HD2	11.31	0.35	0.38	0.13	0.83	0.18
CG	CD	NE	CZ	2.63	0.35	-0.76	0.20	0.22	0.23
HD1	CD	NE	CZ	7.37	0.28	-2.24	0.15	0.41	0.15
HD2	CD	NE	CZ	7.57	0.34	-2.65	0.19	0.58	0.12
CG	CD	NE	HE	5.46	0.33	-0.62	0.15	-0.38	0.22
HD1	CD	NE	HE	11.93	0.42	-0.51	0.18	0.17	0.14
HD2	CD	NE	HE	12.33	0.33	-0.24	0.14	0.17	0.19
CD	NE	CZ	NH1	0.55	0.42	0.22	0.22	0.20	0.16
HE	NE	CZ	NH1	1.47	0.47	-0.52	0.10	0.73	0.43
CD	NE	CZ	NH2	0.60	0.30	0.22	0.16	0.11	0.14
HE	NE	CZ	NH2	1.60	0.36	-0.52	0.11	0.57	0.32

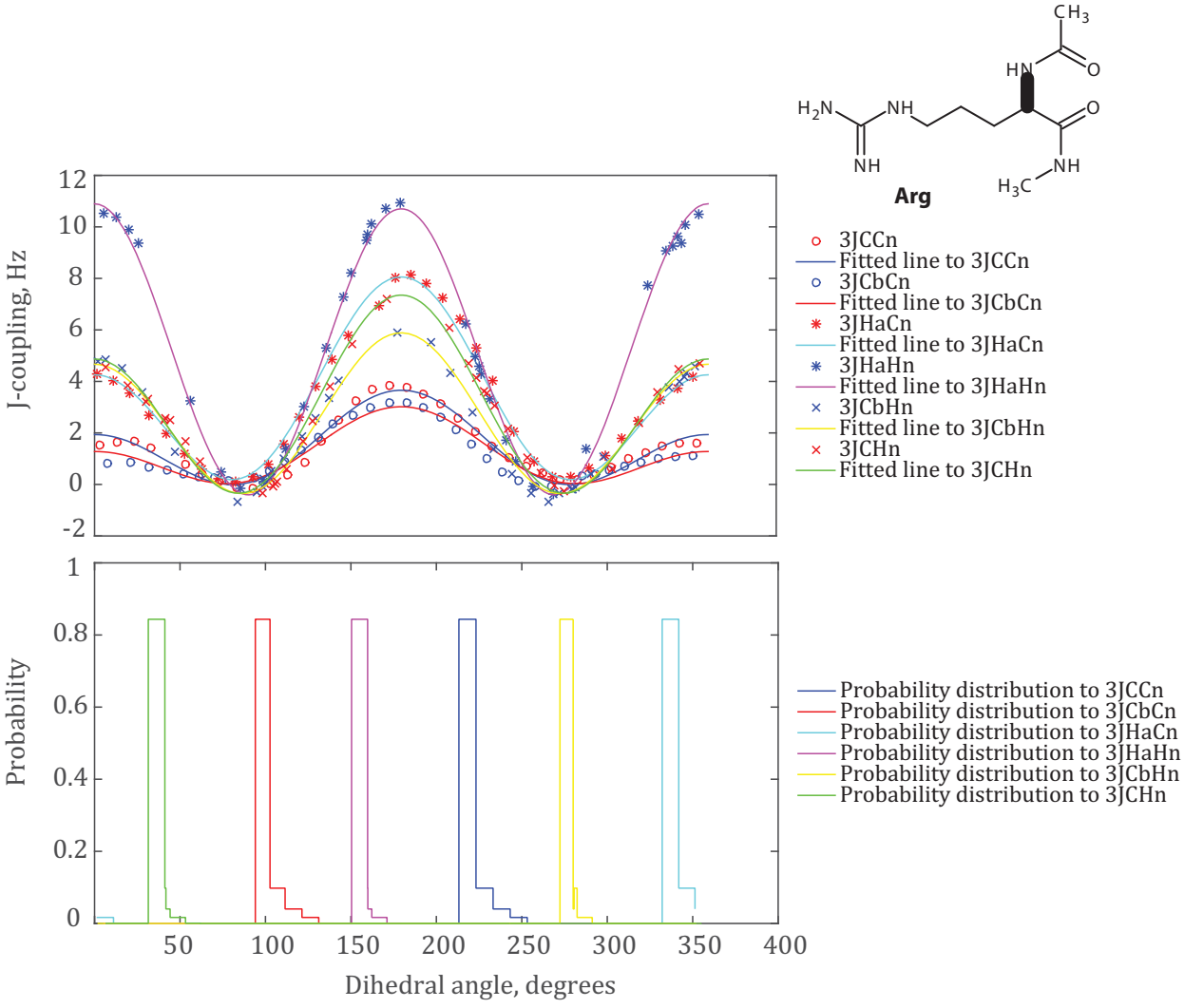

Figure 13 *ARG ϕ backbone dihedral angle scan. Top panel Karplus curves for the torsion angles linked to the ϕ dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The Karplus curve for the $^{3}J_{HaHn}$ torsion angle is broken due to the presence of hydrogen bond. This also affected the appearance of the probability distribution histograms for all the torsion angles linked to the ϕ backbone dihedral angle, making them look sharp.

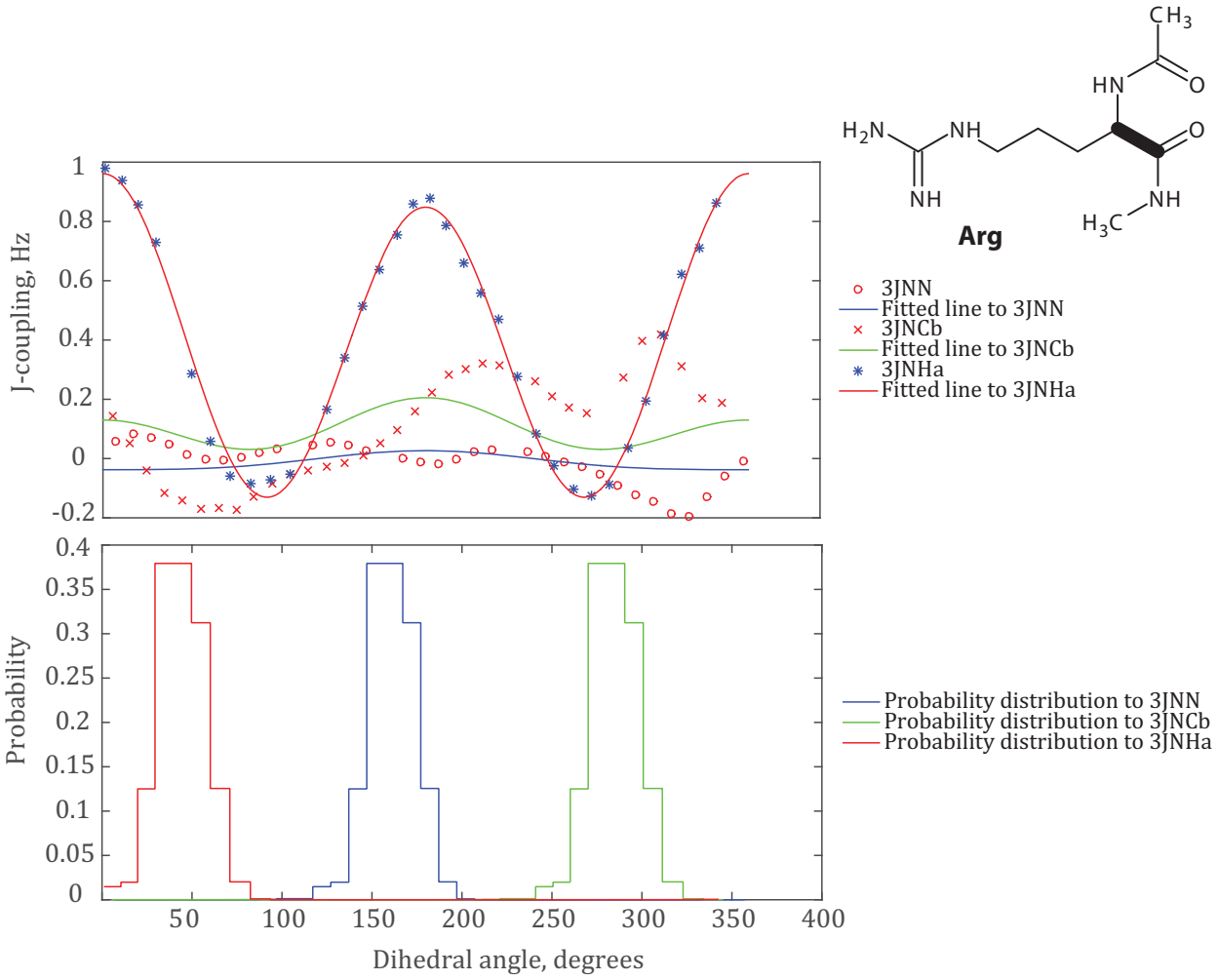

Figure 14 *ARG ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and well behaved Boltzmann probability distribution histograms for the torsion angles associated to the ψ dihedral angle scan respectively.*

Three bond couplings between two nitrogen and between nitrogen and carbon spins is of vanishingly small values. This corresponds to small amplitude numerical noise, hence the

emergence of anomalous Karplus curve graphs. As it is depicted in **Figure 14**, the vicinal coupling between nitrogen and the proton at α position shows the Karplus relation between the coupling values and the dihedral angle *i.e.* couplings are largest at 0, 180, and 360 degrees with zero at 90 and 270 with the values in between the extremes reflecting the deduction of the values from Karplus equation.

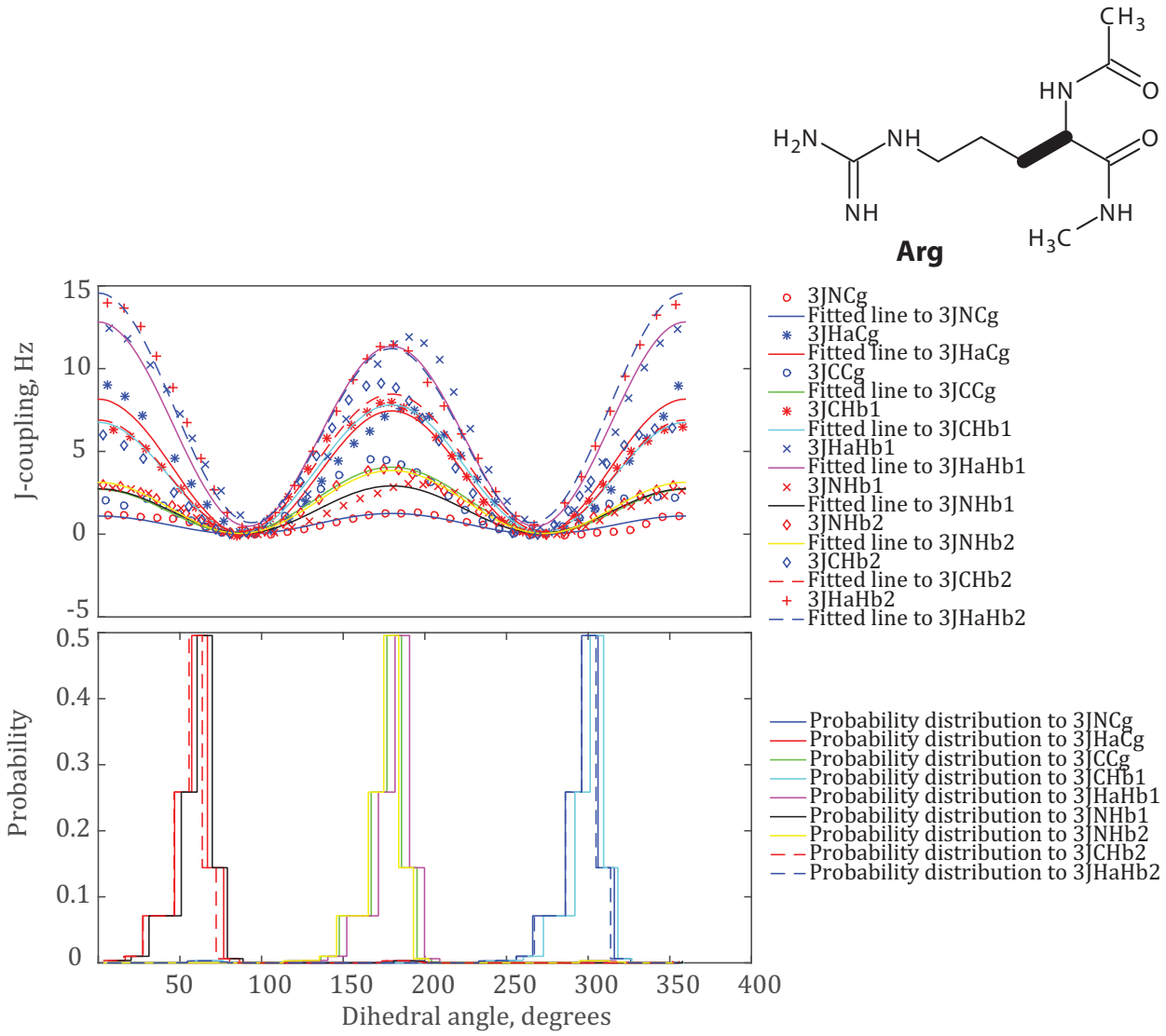

Figure 15 *ARG* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for torsion angles linked to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

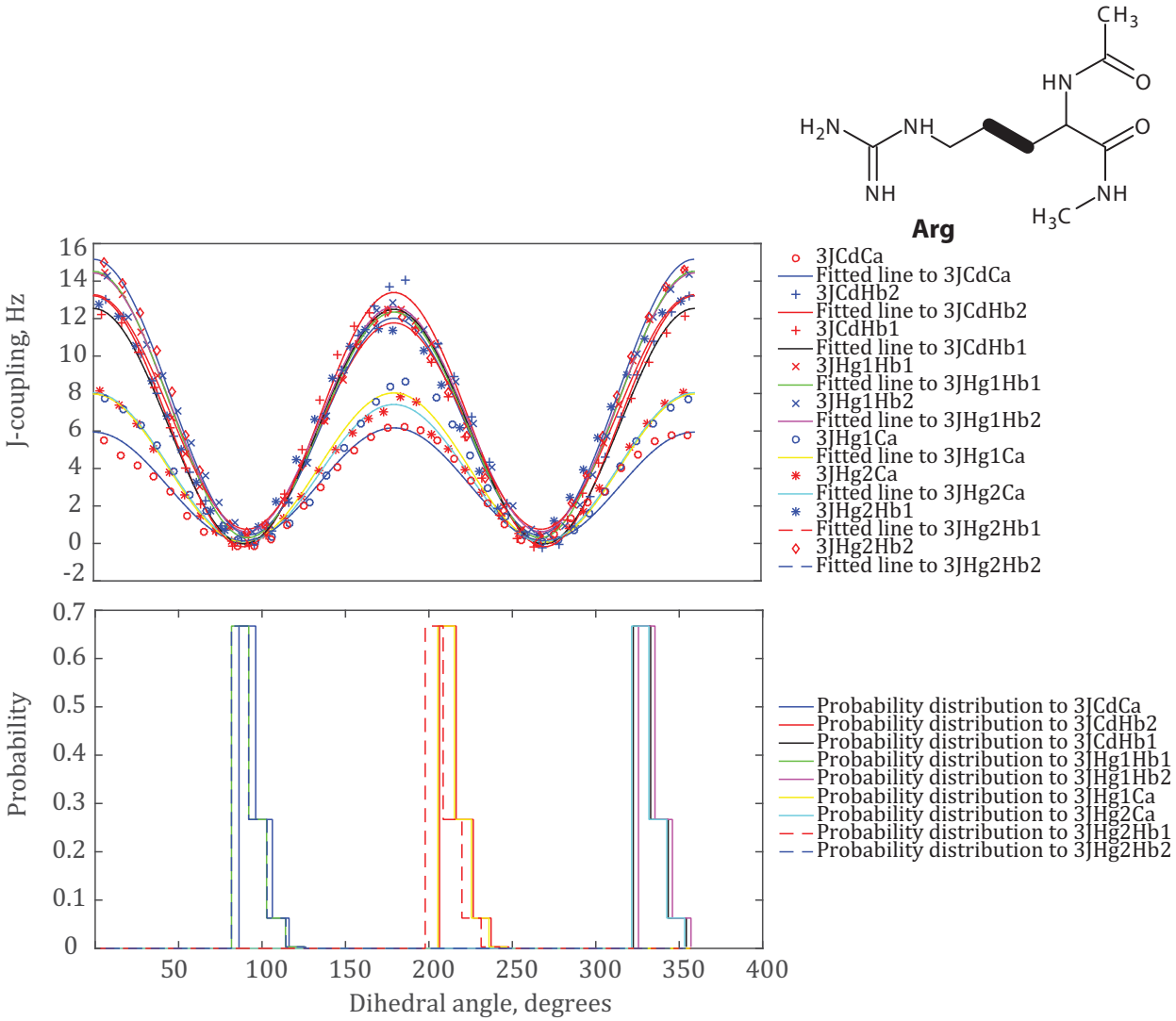

Figure 16 *ARG* χ_2 *side-chain dihedral angle scan. Top panel fitting of Karplus curves associated to the* χ_2 *dihedral angle to theoretical NMR vicinal coupling data and bottom panel probability distribution histograms for the respective torsion angles.*

The existence of hydrogen bond is responsible for the jumps in the Karplus curves and sharp appearance of the Boltzmann's probability distribution graphs.

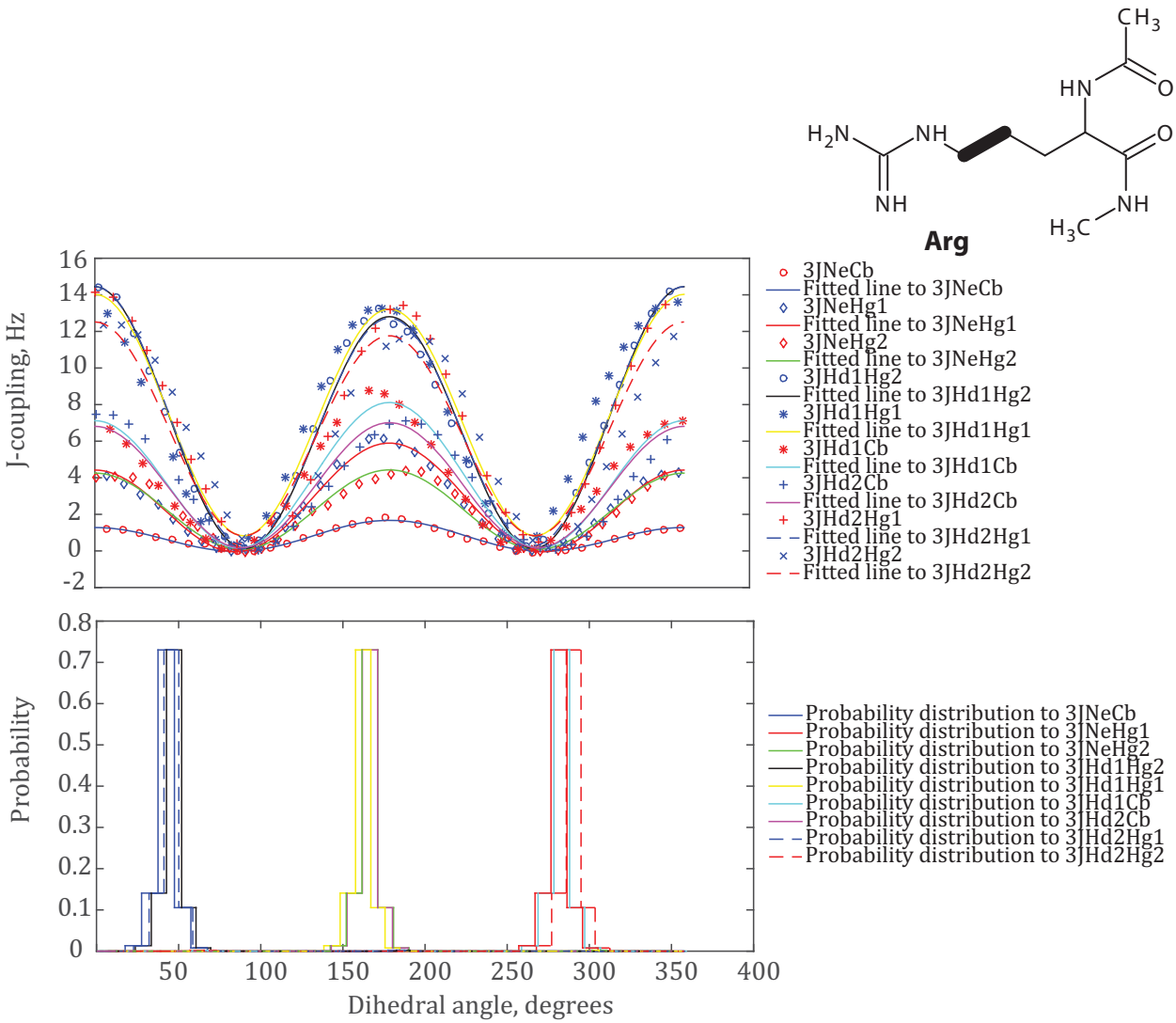

Figure 17 *ARG* χ_3 *side-chain dihedral angle scan. Top panel Karplus curves for torsion angles related to the* χ_3 *dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

Hydrogen bond effect is also responsible for the sharp appearance of the Boltzmann probability distribution histograms and for some of the Karplus curves that appear broken in **Figure 17** upper panel.

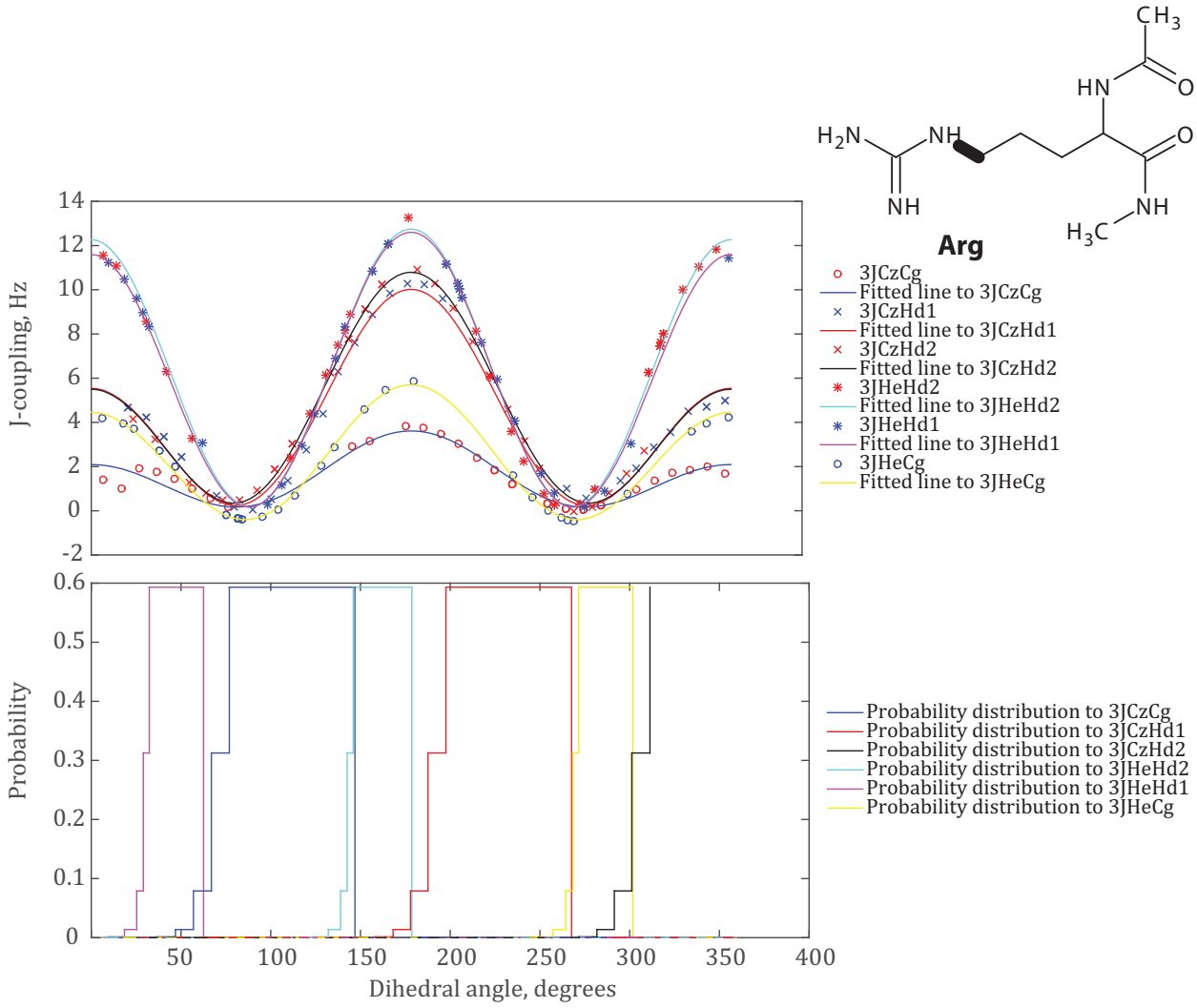

Figure 18 *ARG* χ_4 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles related to the* χ_4 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The jump and discontinuity observed on the Karplus curves is due to hydrogen bond between a proton from the sidechain and oxygen from the backbone amino acid residue and also possibly from Van der Walls interactions from the presence of the large hydrocarbon sidechain of the amino acid arginine. It is thought that these effects are also likely to be responsible for the broad probability distribution histograms depicted in **Figure 18** (above). The Karplus curve for the $^3J_{CzCg}$ couplings (blue colour legend) is broken and its Boltzmann probability distribution histogram along with that of the $^3J_{CzHd1}$ torsion angle are broad. Arginine as basic amino acids favour charge-charge interactions.

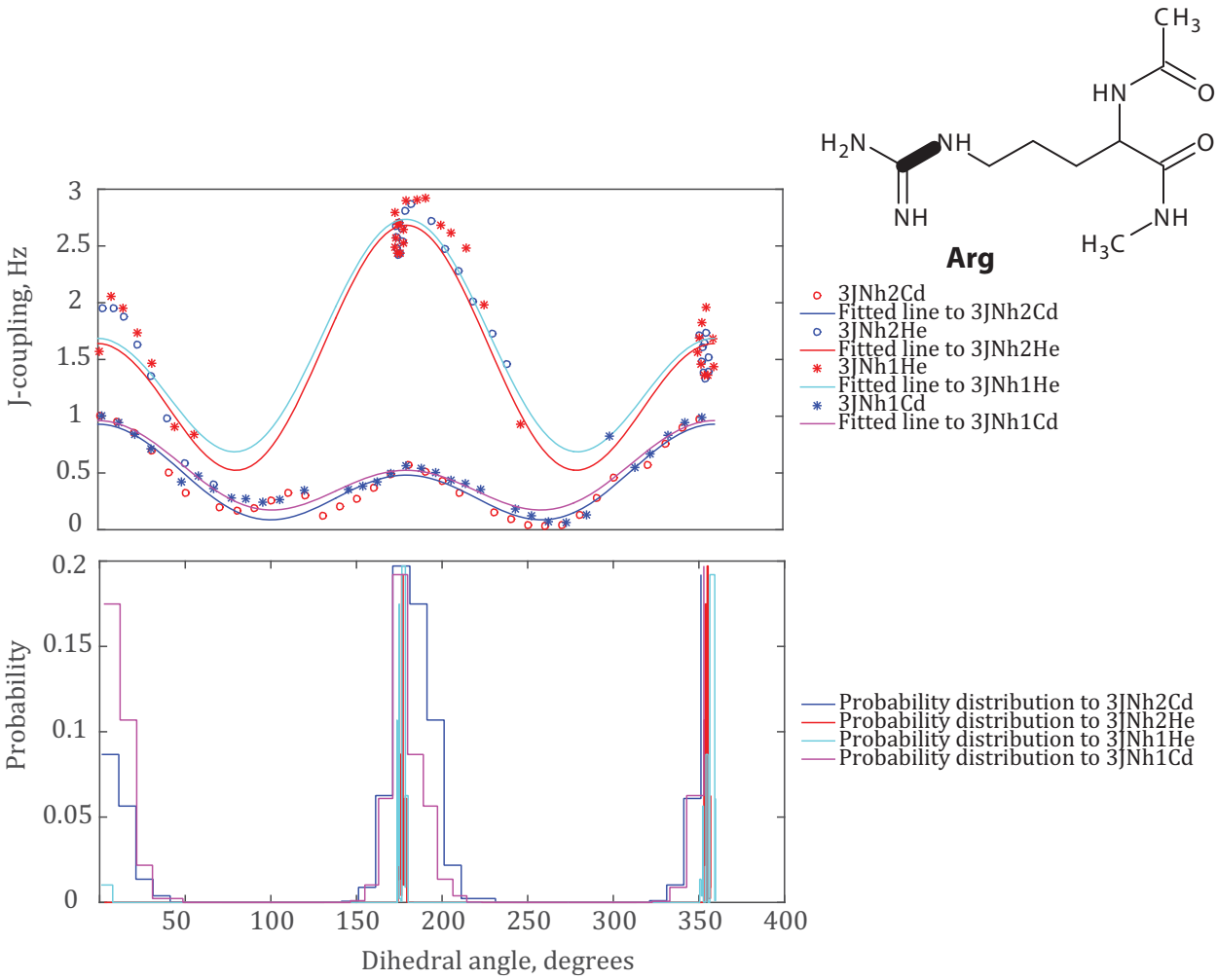

Figure 19 *ARG* χ_5 *side-chain dihedral angle scan. Top panel* χ_5 *Karplus curves for the torsion angles and bottom panel Boltzmann probability distribution histograms related to the* χ_5 *dihedral angle scan.*

The broken curve for $^3J_{Nh2He}$ and $^3J_{Nh1He}$ is likely due to the hydrogen bond being pulled away as a result of dihedral angle scan. This pulling of hydrogen bonded atoms disturbs the spatial arrangement and destroys the Karplus relation hence causes jumps in coupling values. The other two three bond couplings are $^3J_{Nh2Cd}$ and $^3J_{Nh1Cd}$. The values for these couplings are small, and small amplitudes are numerical noise hence the breakage of the coupling curve.

3.9.4 Karplus coefficients, ASN

Method: GIAO DFT B3LYP/ M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
CG	CB	CA	N	1.64	0.37	–0.60	0.27	–0.12	0.19
CG	CB	CA	C	3.40	0.36	–0.94	0.14	0.27	0.19
CG	CB	CA	HA	9.19	0.28	-0.57	0.13	0.01	0.20
HA	CA	CB	HB1	12.85	0.32	1.45	0.12	0.58	0.20
HA	CA	CB	HB2	12.37	0.33	0.70	0.12	0.88	0.22
N	CA	CB	HB1	3.65	0.27	0.31	0.14	0.14	0.14
N	CA	CB	HB2	4.19	0.49	-0.01	0.17	-0.07	0.36
C	CA	CB	HB1	7.86	0.33	0.21	0.12	-0.17	0.20
C	CA	CB	HB2	6.56	0.34	-0.18	0.13	0.31	0.20
CA	CB	CG	ND	0.26	0.34	–0.33	0.13	–0.04	0.20
H1	CB	CG	ND	1.46	0.42	–0.41	0.22	–0.22	0.25
H2	CB	CG	ND	1.36	0.34	–0.41	0.19	–0.30	0.20
H	N	CA	HA	11.48	0.34	-0.41	0.13	-0.09	0.22
C	N	CA	HA	5.83	0.45	-1.67	0.19	0.51	0.23
H	N	CA	C	6.26	0.46	-1.44	0.25	-0.34	0.18
H	N	CA	CB	5.62	0.36	-0.70	0.21	-0.34	0.16
C	N	CA	C	2.89	0.25	-1.03	0.14	0.08	0.14
C	N	CA	CB	1.93	0.29	-0.73	0.13	0.05	0.17
N	C	CA	N	0.09	0.29	-0.05	0.14	0.00	0.18
N	C	CA	HA	0.94	0.35	-0.16	0.16	-0.16	0.17
HA	CA	C	N	-0.88		-0.61		-0.27 [104]	
CB	CA	C	N	0.10	0.38	-0.14	0.21	0.00	0.17

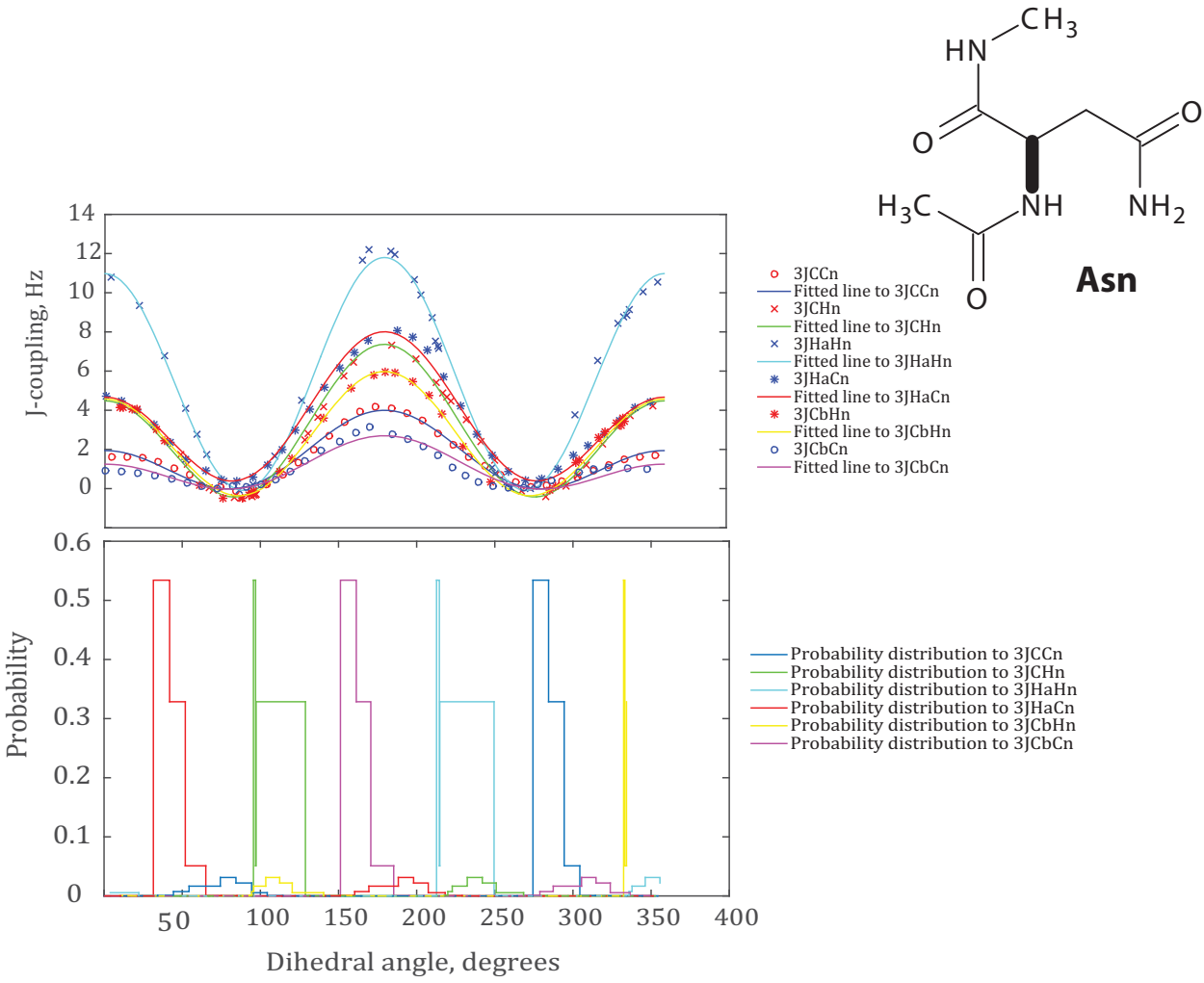

Figure 20 *ASN ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann's probability distribution graphs for torsion angles associated to the ϕ backbone dihedral angle scan.*

The presence of hydrogen bond between the backbone amide proton and sidechain oxygen atom bonded to C_δ is very likely to be responsible for the broken Karplus curves that involve the backbone amide participated in couplings. The hydrogen bond breakage might also be responsible for the sharp appearance of the Boltzmann probability distribution histograms of the torsion angles that involved the backbone amide proton. The affected torsion angles are ${}^3J_{CHn}$, ${}^3J_{HaHn}$ and ${}^3J_{CbHn}$ coloured as green, cyan and yellow respectively. There, also exist other atoms that are involved in hydrogen bond interactions (example: backbone amide proton and backbone oxygen atom and sidechain amide proton and oxygen atoms), however these interactions are not affected by the scan of the ϕ dihedral angle.

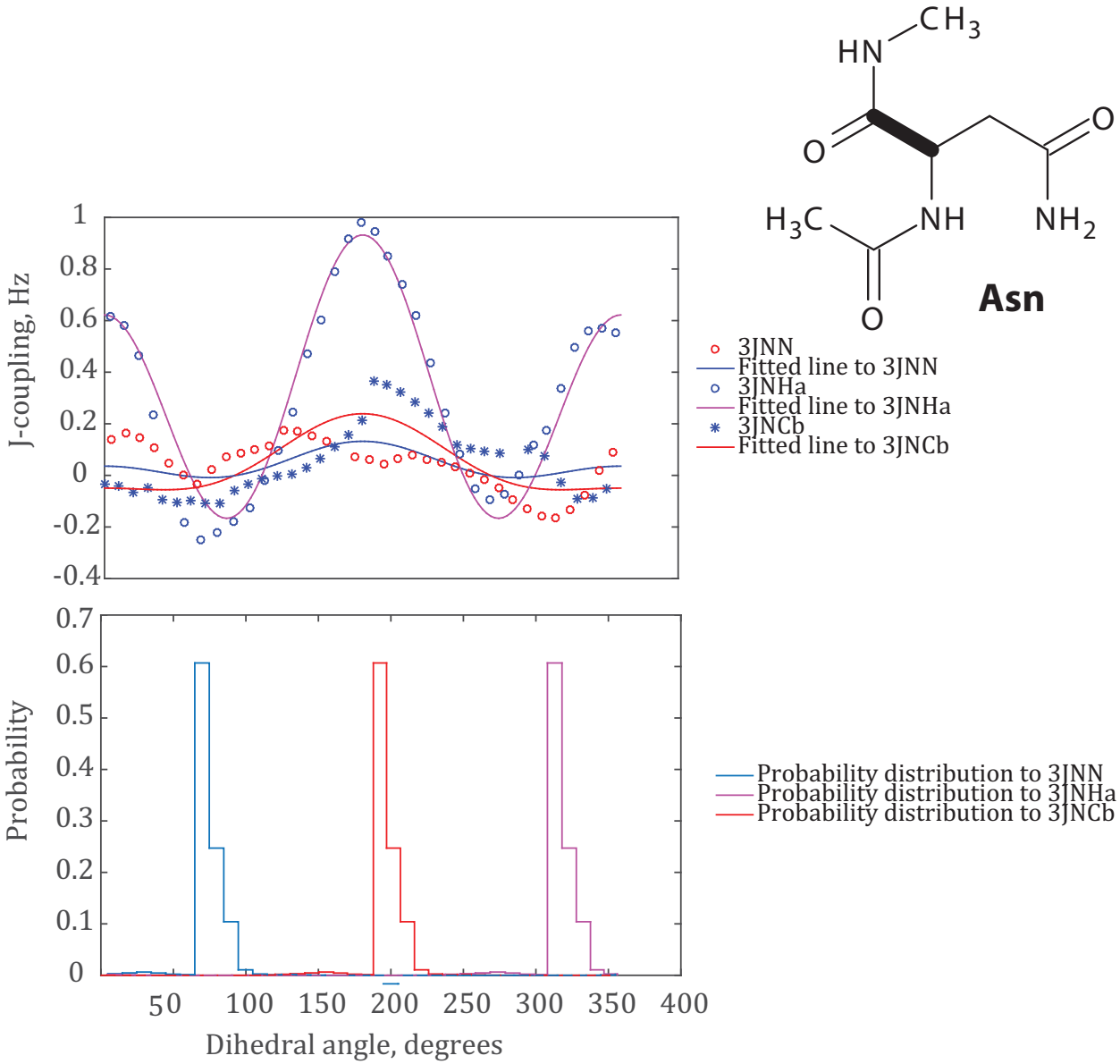

Figure 21 *ASN ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution histograms for the torsion angles linked to ψ dihedral angle scan.*

The broken curves are due to the very small coupling values between two nitrogen atoms, and nitrogen and carbon nuclei separated by three bonds. Please note that these are very small amplitudes hence the numerical noise.

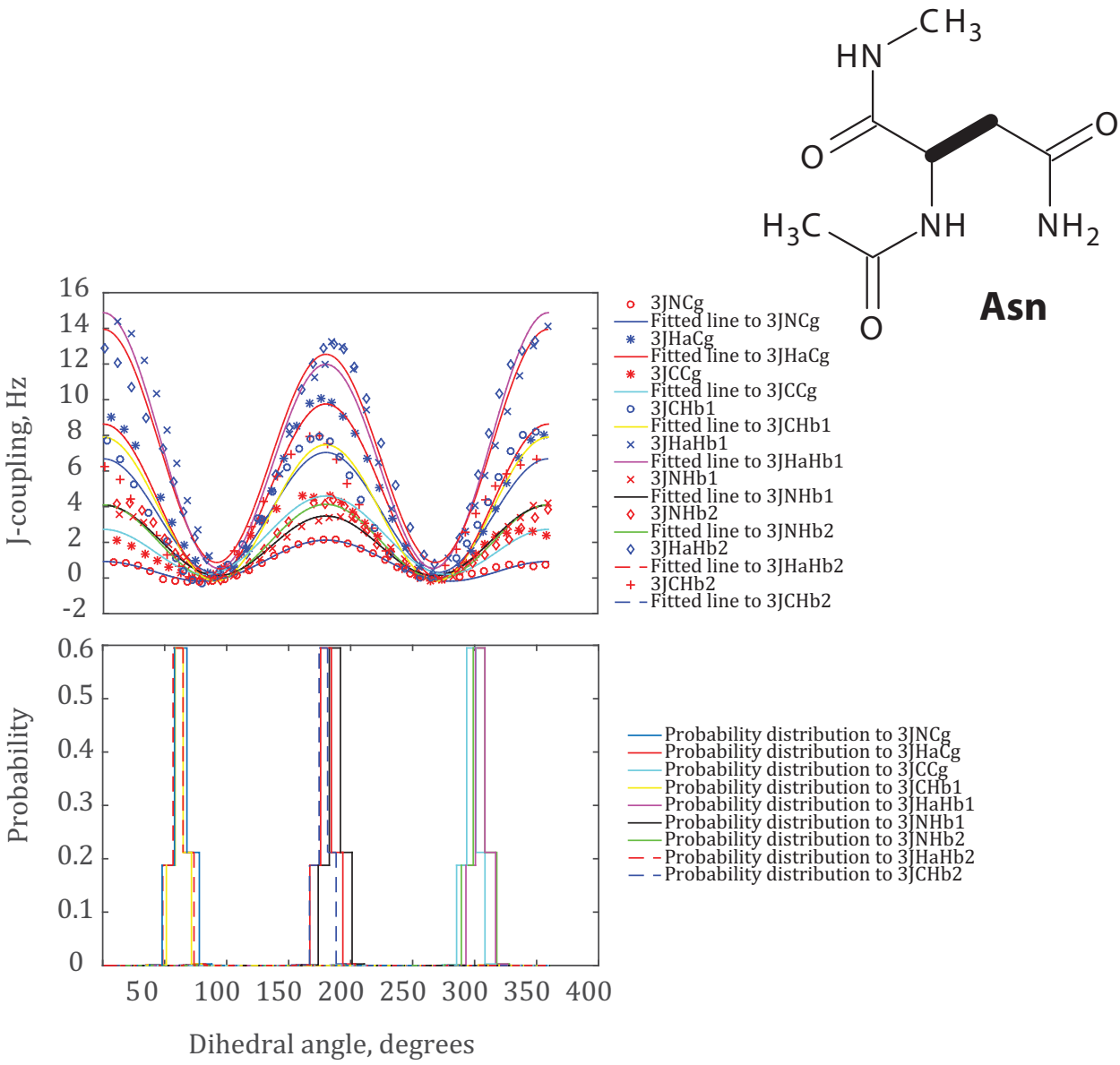

Figure 22 *ASN χ_1 side-chain dihedral angle scan. Top panel Karplus curves for torsion angles linked to the χ_1 dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

Hydrogen bond between oxygen atom at C_γ position and the amide backbone proton is responsible for the jump in some of the Karplus curves and the thin appearance of the probability distribution histograms.

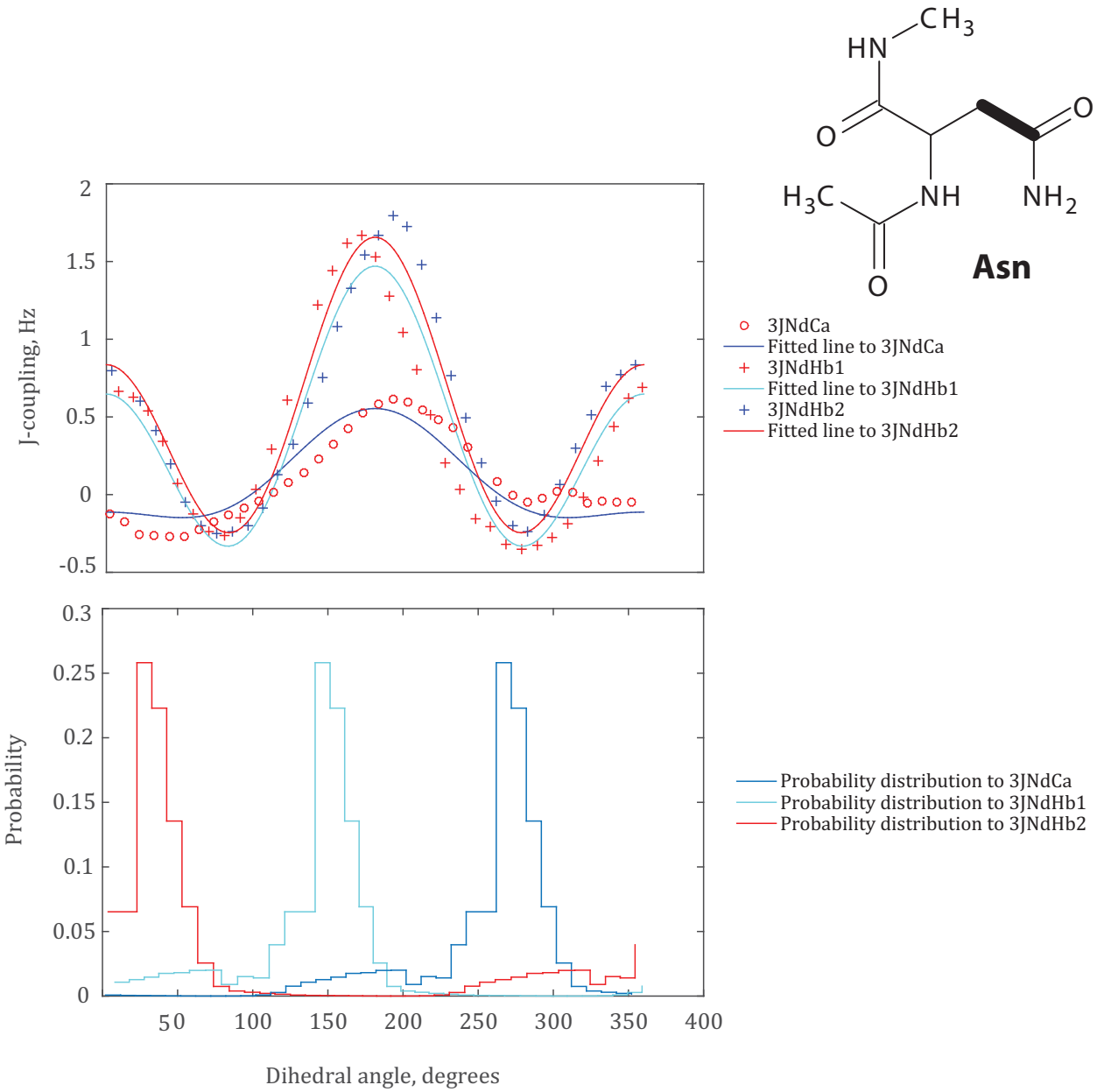

Figure 23 *ASN* χ_2 *side-chain dihedral angle scan. Top panel Karplus curves associated to the* χ_2 *dihedral angle scan and bottom panel probability distribution graphs for the respective torsion angles.*

The $^3J_{NdCa}$ coupling values are of vanishingly small size hence the curve do not fit the data properly. The coupling between these nuclei is very small, hence the numerical noise. The other Karplus curves involve couplings between sidechain nitrogen atom at δ position and hydrogen atoms of C_β.

3.9.5 Karplus coefficients, ASP

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
CG	CB	CA	N	1.39	0.32	-0.86	0.18	0.01	0.16
CG	CB	CA	C	4.13	0.30	-1.13	0.14	0.23	0.16
CG	CB	CA	HA	9.63	0.27	–0.94	0.16	-0.01	0.13
HA	CA	CB	HB1	12.53	0.36	1.13	0.15	1.09	0.20
HA	CA	CB	HB2	13.22	0.26	1.14	0.12	0.29	0.17
N	CA	CB	HB1	4.27	0.58	0.13	0.24	-0.09	0.40
N	CA	CB	HB2	3.76	0.37	0.21	0.12	0.09	0.23
C	CA	CB	HB1	7.96	0.33	0.13	0.14	-0.15	0.19
C	CA	CB	HB2	7.59	0.36	-0.11	0.13	0.10	0.22
H	N	CA	HA	10.81	0.44	-0.18	0.11	-0.08	0.31
C	N	CA	HA	5.52	0.44	-1.56	0.22	0.37	0.21
H	N	CA	C	6.82	0.35	-1.19	0.13	-0.45	0.23
H	N	CA	CB	5.72	0.34	-0.63	0.21	-0.52	0.18
C	N	CA	C	2.76	0.27	-0.94	0.13	0.07	0.18
C	N	CA	CB	2.18	0.28	-0.87	0.13	0.05	0.17
N	C	CA	N	0.09	0.35	-0.07	0.15	-0.06	0.24
N	C	CA	HA	0.86	0.53	-0.20	0.20	-0.17	0.28
HA	CA	C	N	-0.88		-0.61		-0.27 [104]	
CB	CA	C	N	0.05	0.34	-0.17	0.17	0.04	0.21

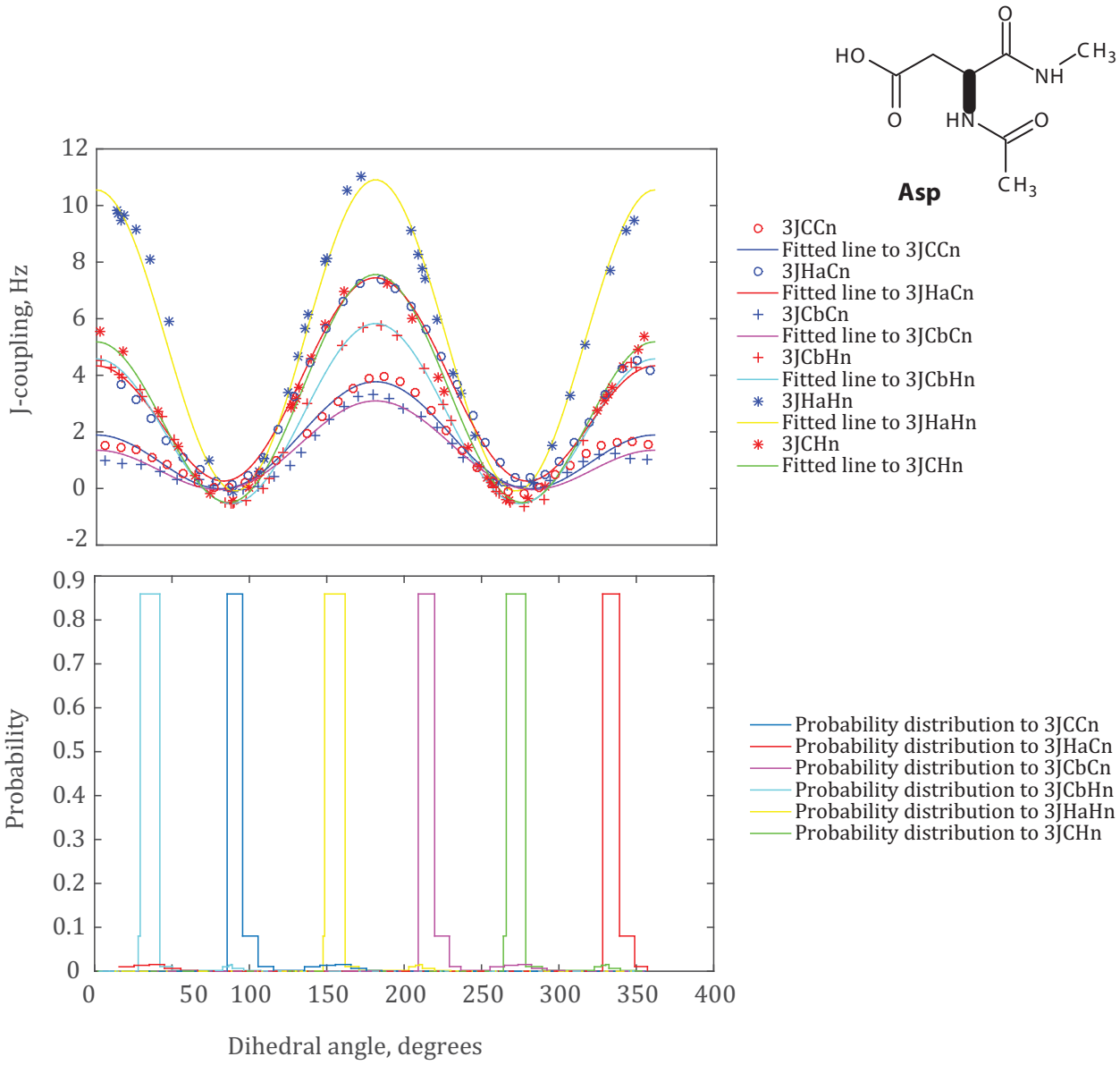

Figure 24 *ASP ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution histograms for torsion angles related to the ϕ dihedral angle respectively.*

The sharp appearance of the probability distribution graphs is attributed to hydrogen bond.

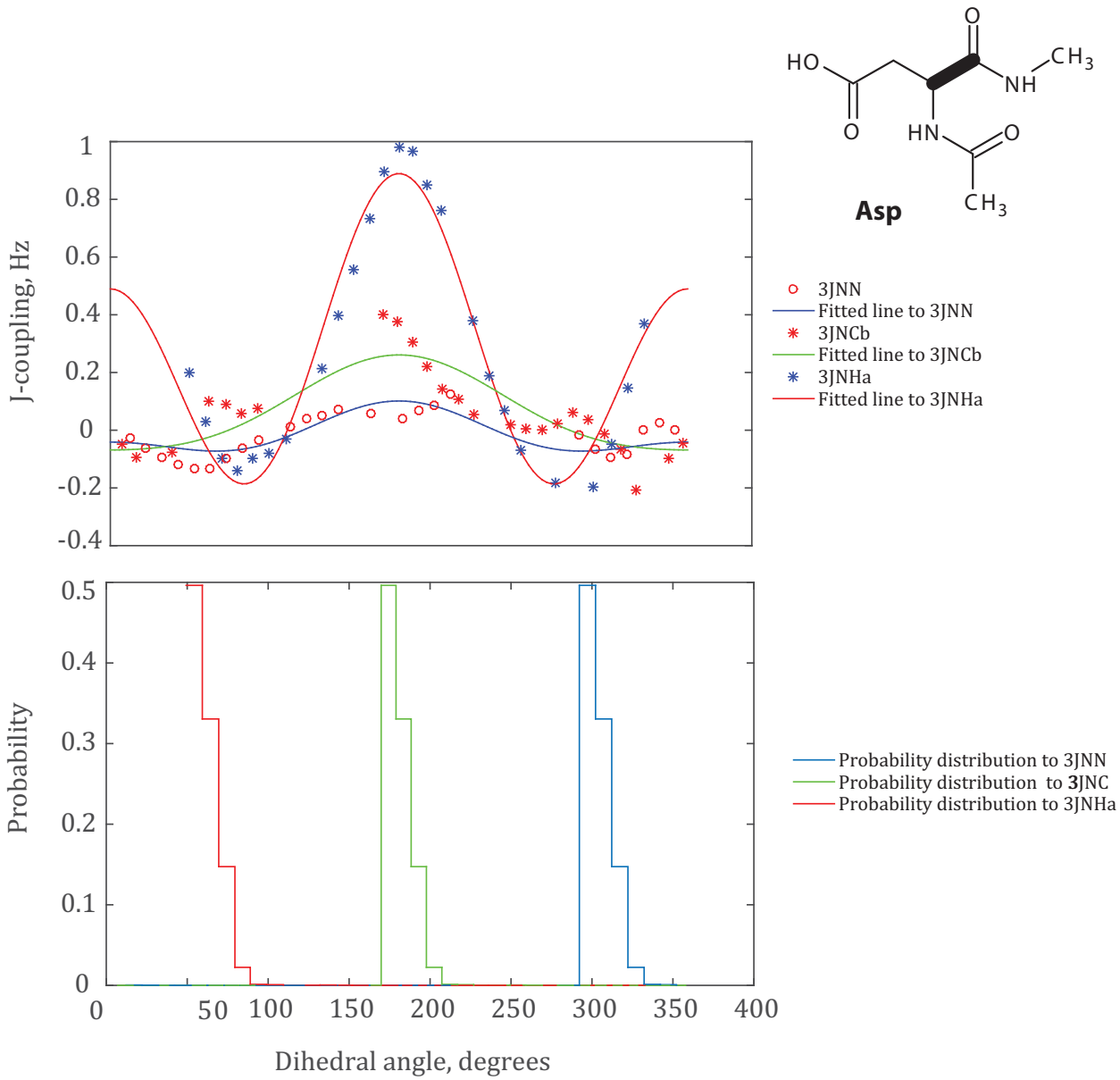

Figure 25 *ASN ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and probability distribution graphs for torsion angles related to the ψ dihedral angle respectively.*

Karplus curve is broken for $^{3}J_{NN}$ and $^{3}J_{NCb}$ as these are very small couplings and the $^{3}J_{NHa}$ coupling curve has jumps likely to be due to the hydrogen bond breakage.

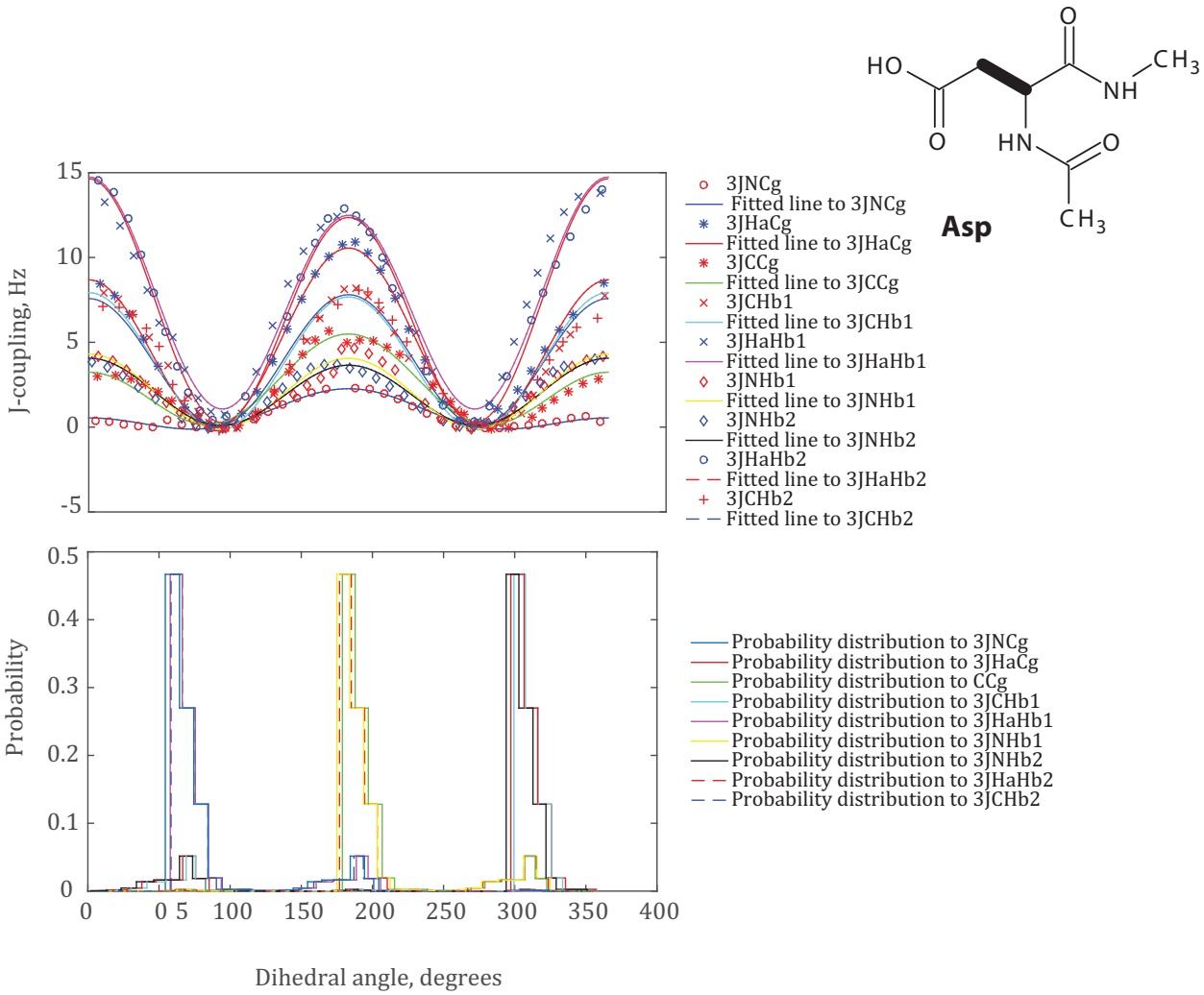

Figure 26 *ASP* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for torsion angles associated to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

3.9.6 Karplus coefficients, CYS

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	0.03	0.26	-0.08	0.14	0.00	0.14
HA	CA	C	N	0.91	0.32	-0.14	0.12	-0.19	0.20
HA	CA	C	N	-0.88		-0.61		-0.27 [104]	
CB	CA	C	N	-0.07	0.32	-0.35	0.13	0.07	0.17

C	CA	CB	HB1	7.30	0.34	-0.39	0.12	0.02	0.20
C	CA	CB	HB2	7.14	0.34	-0.28	0.12	0.16	0.21
HA	CA	CB	HB1	12.80	0.33	1.00	0.15	0.65	0.16
HA	CA	CB	HB2	11.89	0.39	0.96	0.17	0.80	0.24
N	CA	CB	HB2	4.06	0.45	0.16	0.18	-0.12	0.35
N	CA	CB	HB1	3.86	0.35	0.24	0.11	0.04	0.23
H	N	CA	HA	11.14	0.39	0.07	0.12	-0.19	0.27
C	N	CA	HA	5.69	0.29	-1.88	0.14	0.30	0.20
H	N	CA	C	6.57	0.44	-1.11	0.19	-0.36	0.23
H	N	CA	CB	6.90	0.30	-0.55	0.20	-0.66	0.15
C	N	CA	C	2.66	0.25	-1.00	0.13	0.10	0.16
C	N	CA	CB	2.41	0.26	-1.04	0.13	0.01	0.15

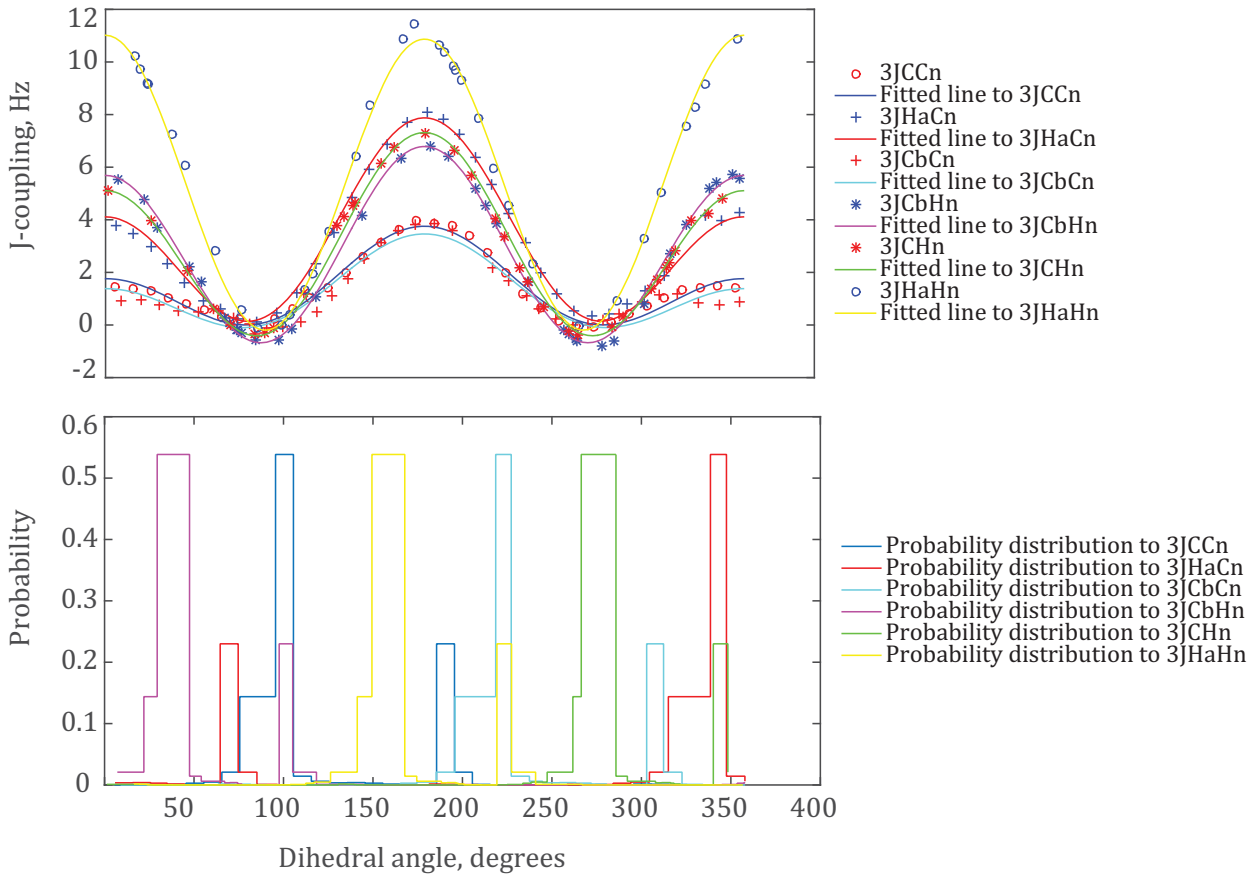

Figure 27 *CYS ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for torsion angles associated to the ϕ dihedral angle scan.*

The existence of hydrogen bond between the backbone amide proton and oxygen atom bonded to carbon next to C_α is likely to be responsible for the broken Karplus curves that involve the backbone amide participating in *J*-couplings. The hydrogen bond making and breaking as the ϕ dihedral angle rotated might also be responsible for the anomalous appearance of the Boltzmann probability distribution histograms of the torsion angles that involved the backbone amide proton.

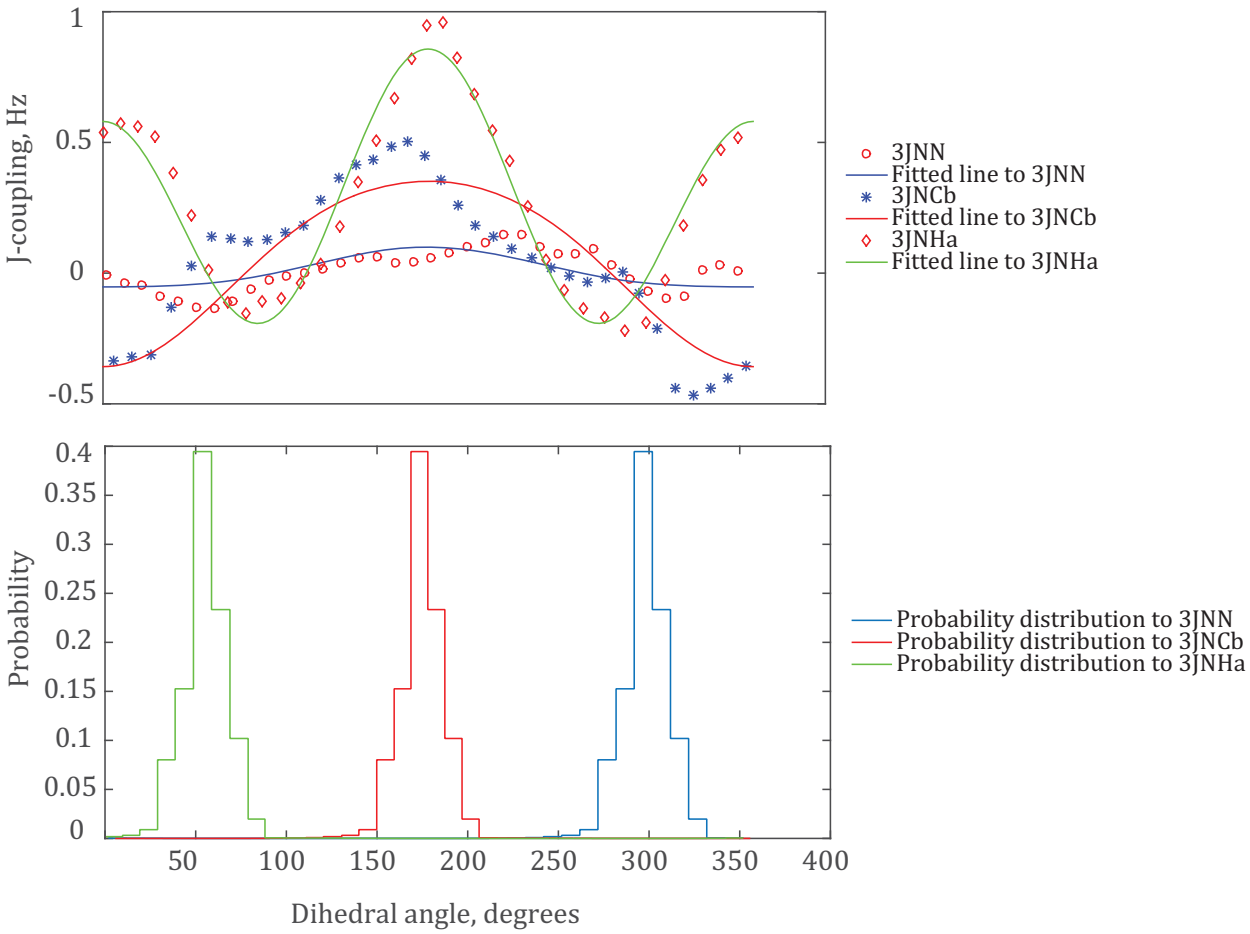

Figure 28 *CYS ψ backbone dihedral angle scan. Top panel Karplus curves of the torsion angles associated to the ψ dihedral angle and bottom panel are the respective probability distribution histograms.*

Similar to the above mentioned, this dihedral angle affects three ^{3}J couplings, namely $^{3}J_{NN}$, $^{3}J_{NCb,}$ and $^{3}J_{NHa}$. The first two are very small, however $^{3}J_{NHa}$ has higher coupling values that shows a Karplus like pattern. The broken curves are attributed to the small coupling values. Hydrogen bond breakage could also be responsible for the anomalous coupling curves.

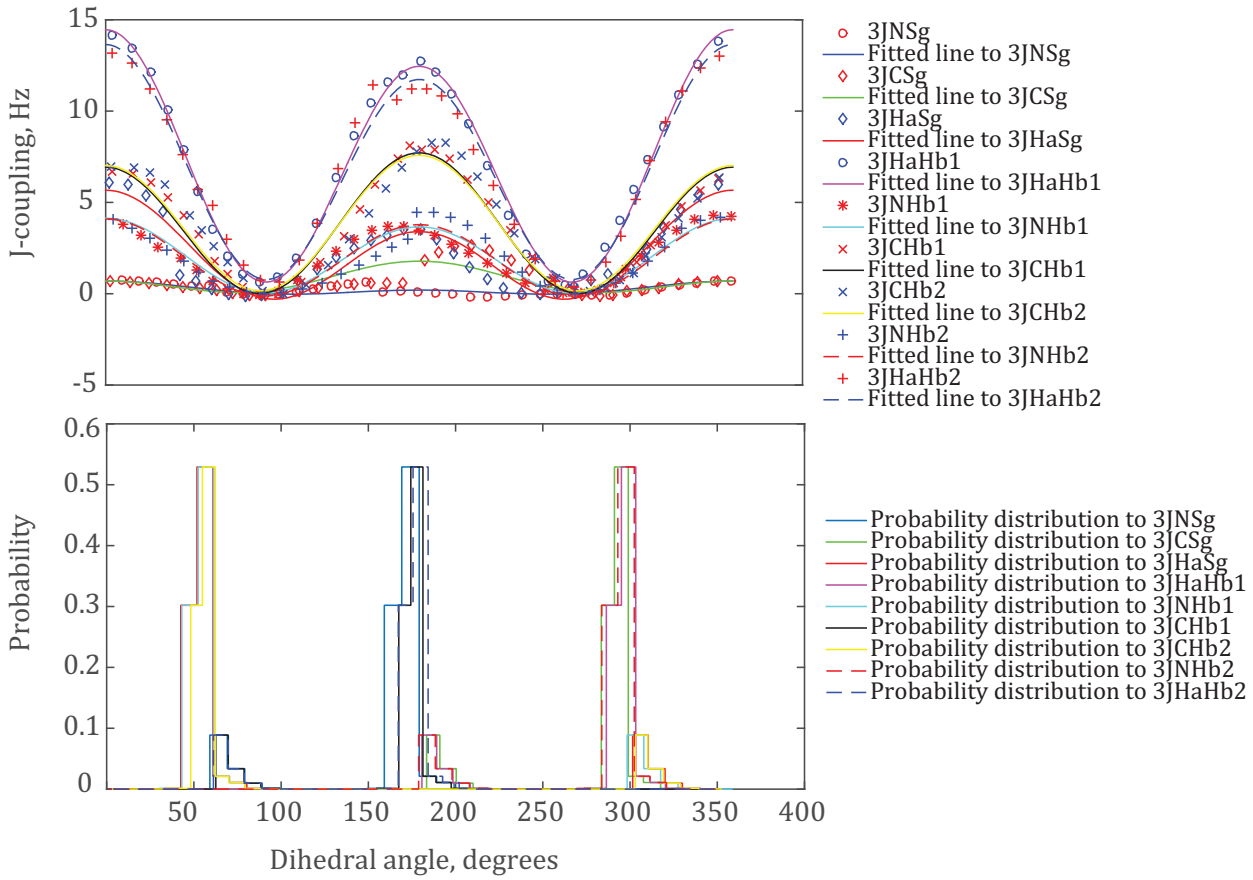

Figure 29 *CYS* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for torsion angles associated to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

χ_1 dihedral angle scan brings the sidechain sulphur atom closer the backbone amide hydrogen to form a hydrogen bond. Thus, the making and breaking of the hydrogen bond affects the coupling data and fitting of Karplus curves that involved sulphur atom appears to be broken.

3.9.7 Karplus coefficients, GLU

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
N	C	CA	N	0.04	0.30	-0.04	0.18	-0.03	0.19
HA	CA	C	N	1.18	0.32	-0.00	0.16	-0.18	0.17
HA	CA	C	N	-0.88		-0.61		-0.27 [104]	

CB	CA	C	N	0.23	0.35	-0.08	0.14	0.00	0.18
CG	CB	CA	N	2.03	0.32	0.09	0.13	-0.01	0.20
CG	CB	CA	C	3.63	0.35	-0.44	0.16	0.12	0.18
CG	CB	CA	HA	8.34	0.39	0.64	0.21	0.03	0.20
C	CA	CB	HB1	7.73	0.38	-0.20	0.22	-0.03	0.21
C	CA	CB	HB2	7.42	0.35	-0.38	0.12	0.24	0.22
HA	CA	CB	HB1	12.63	0.42	1.18	0.16	0.73	0.20
HA	CA	CB	HB2	12.67	0.34	1.76	0.13	0.63	0.24
N	CA	CB	HB2	4.59	0.46	0.16	0.12	-0.07	0.33
N	CA	CB	HB1	4.20	0.37	0.25	0.12	0.06	0.24
H	N	CA	HA	10.80	0.46	-0.16	0.12	-0.06	0.35
C	N	CA	HA	5.61	0.39	-1.62	0.23	0.33	0.20
H	N	CA	C	6.94	0.38	-1.09	0.14	-0.42	0.24
H	N	CA	CB	5.42	0.41	-0.60	0.38	-0.32	0.17
C	N	CA	C	2.73	0.27	-0.94	0.14	0.12	0.16
C	N	CA	CB	2.24	0.45	-0.64	0.20	0.04	0.24
CA	CB	CG	CD	5.17	0.95	-0.79	0.86	0.06	0.29
H1	CB	CG	CD	9.52	0.61	-1.32	0.22	0.09	0.41
H2	CB	CG	CD	9.99	1.87	-0.94	2.11	0.23	0.56
CA	CB	CG	HG1	8.12	2.50	-0.14	2.82	0.13	0.76
H1	CB	CG	HG1	14.19	0.71	1.70	0.46	0.18	0.24
H2	CB	CG	HG1	13.28	0.43	0.43	0.23	0.81	0.25
CA	CB	CG	HG2	8.29	0.50	-0.35	0.23	0.01	0.35
H1	CB	CG	HG2	9.67	0.99	2.18	1.02	1.67	0.34
H2	CB	CG	HG2	14.53	0.86	1.13	0.69	0.19	0.22

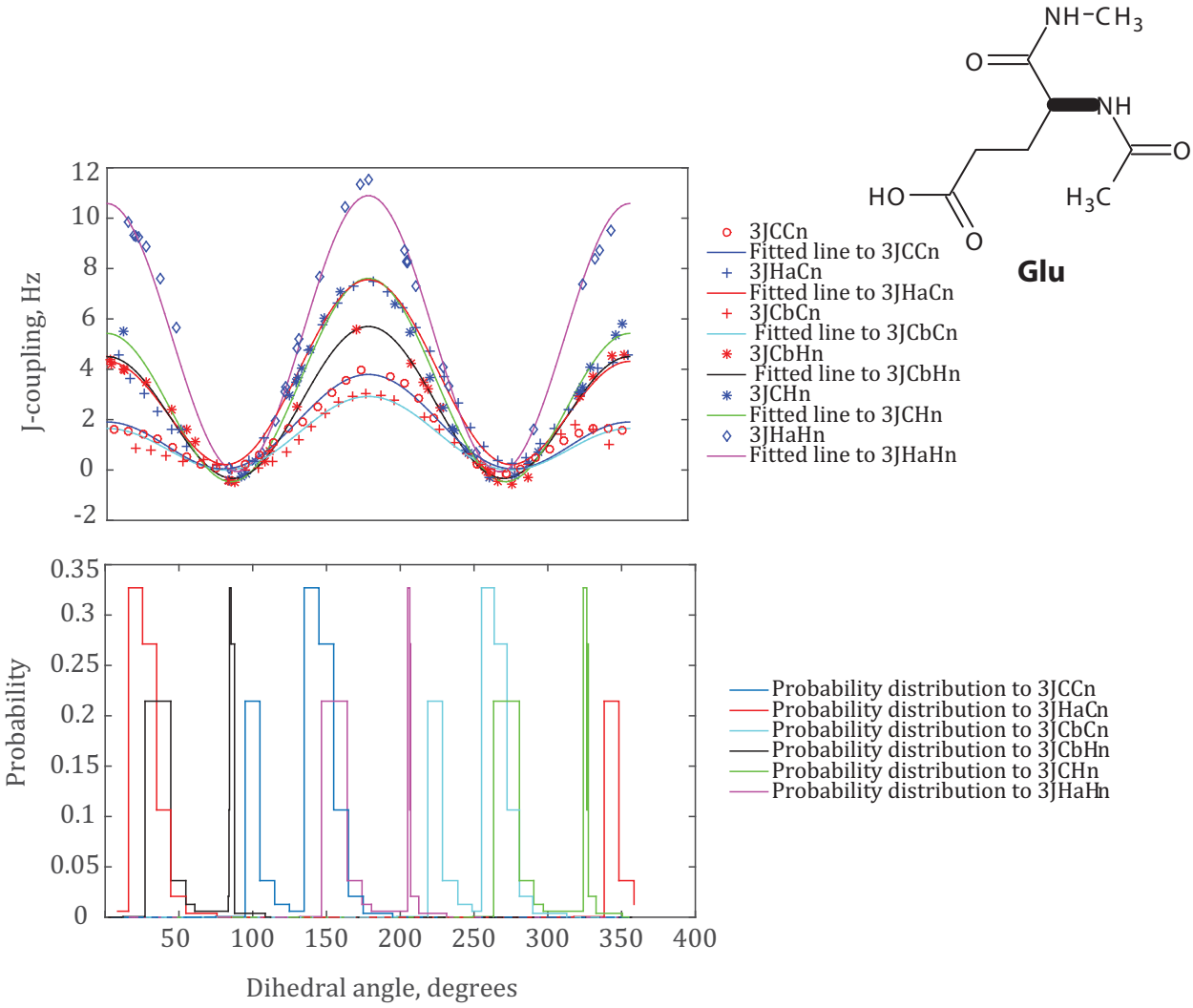

Figure 30 *GLU ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles related to the ϕ dihedral angle scan respectively.*

Torsion angles that involve the proton (H_α) bonded to C_α have broken Karplus curves and anomalous probability distribution graphs that are sharp. This may be, due to steric effects that come from longer methyl sidechain groups and the formation of zwitterion as a result of proton transfer from the carboxylic acid group.

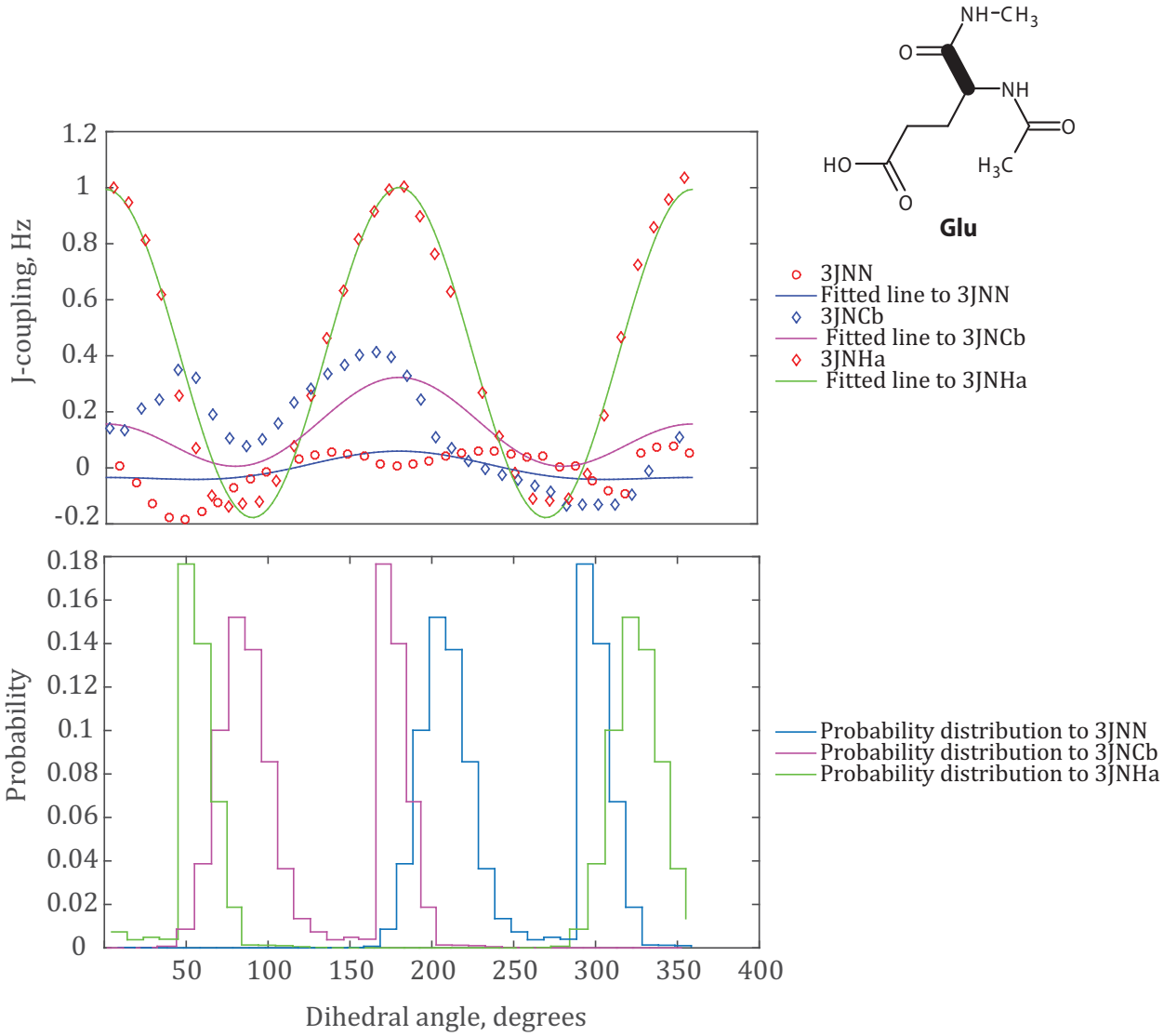

Figure 31 *GLU ψ backbone dihedral angle scan. Top panel Karplus curves for torsion angles associated to the ψ dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

The Karplus curves for the torsion angles that involve carbon and nitrogen as well as nitrogen and nitrogen are numerically very small, thus the fitting of the Karplus curves to the vicinal coupling data looks random and noisy. The fitting of the Karplus curve for the torsion angle involving nitrogen and the C_α proton also shows jumps, which may be attributed to steric effect.

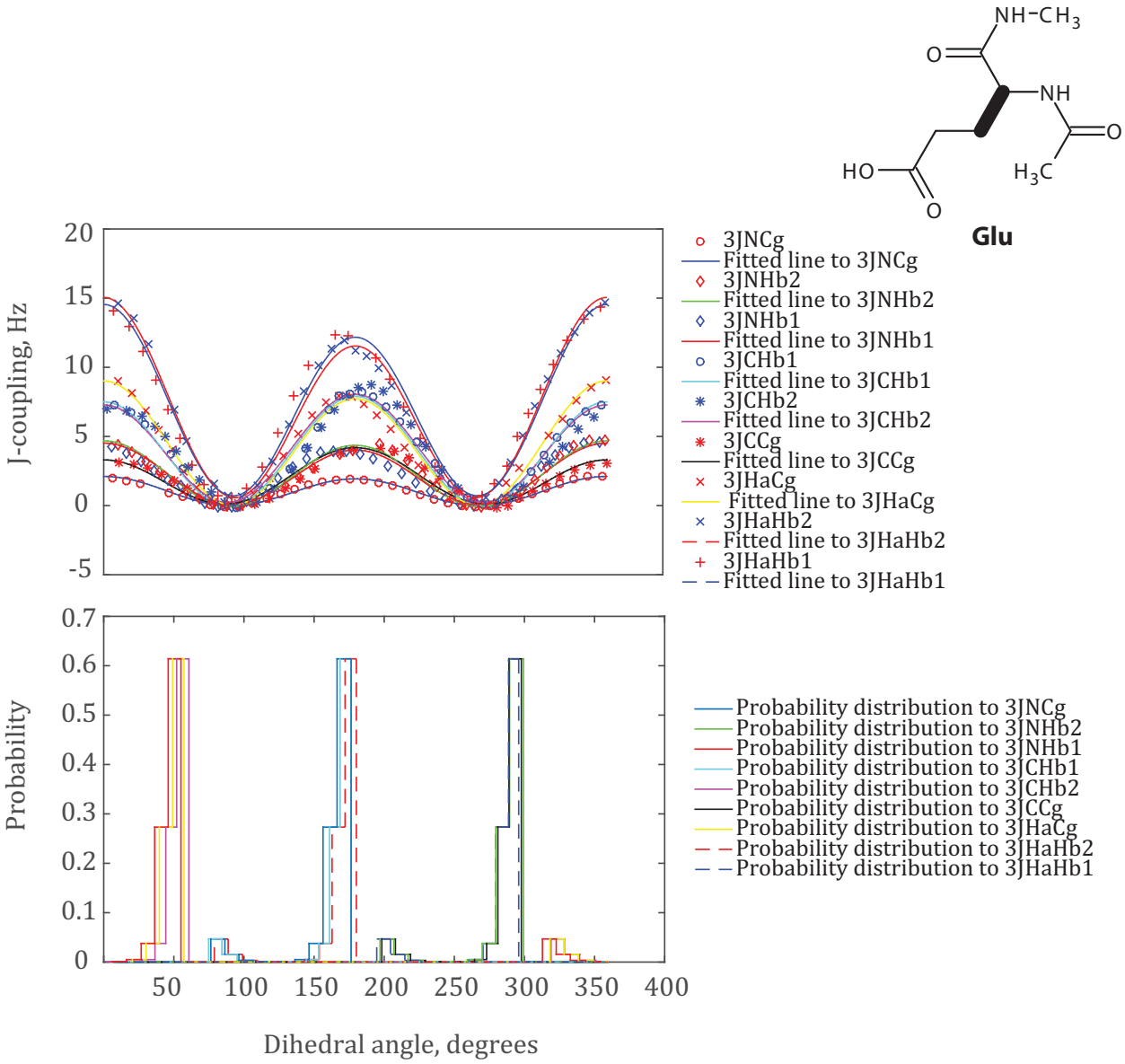

Figure 32 *GLU* χ_1 *side-chain dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles related to the* χ_1 *dihedral angle scan respectively.*

The sharp appearance and tiny bumps present in the Boltzmann probability distribution histograms might be down to the presence of longer hydrocarbon sidechains of the glutamic acid. There exist hydrogen bonds in the backbone chain, but this appears not to affect the Karplus curves and distribution graphs.

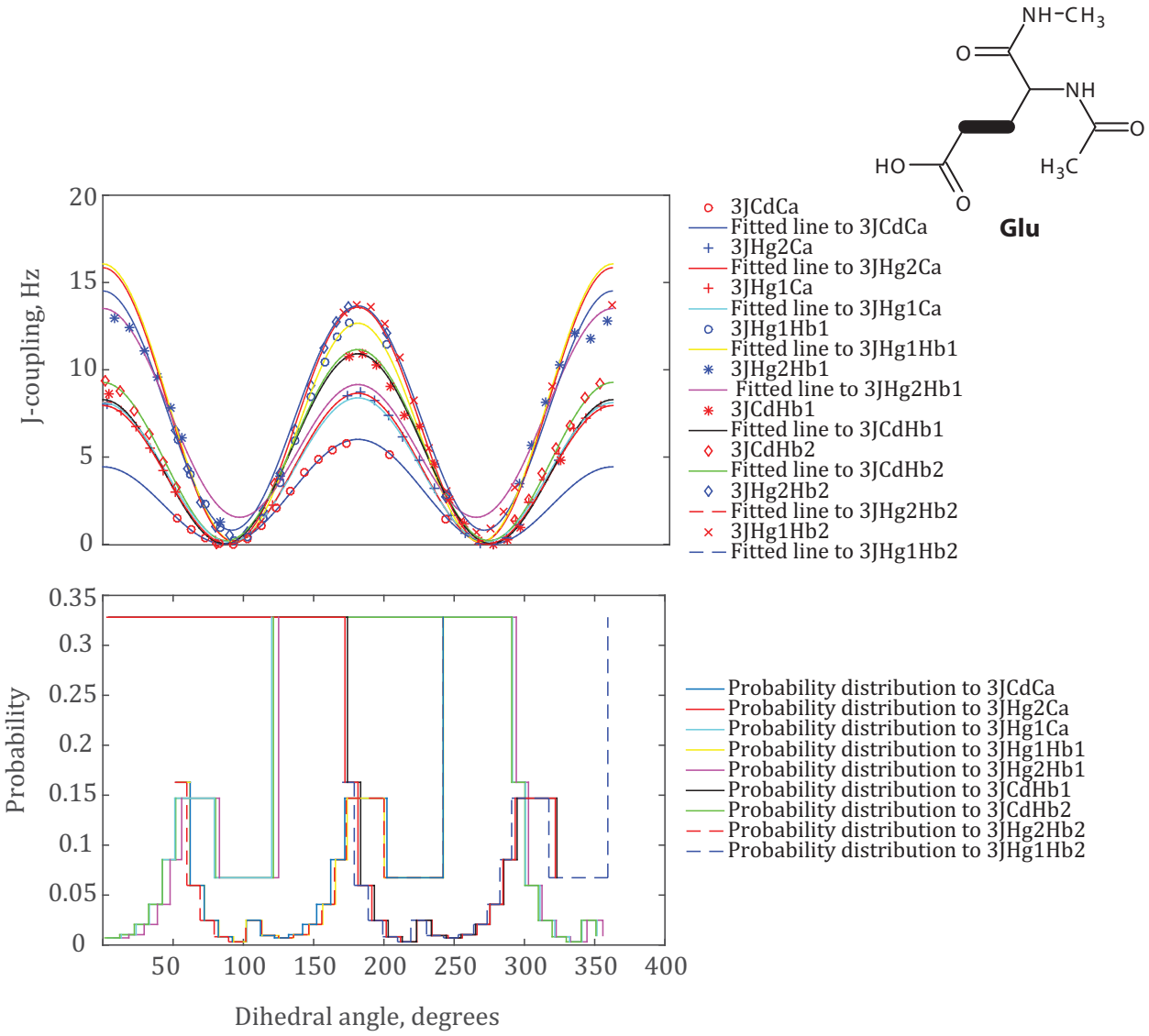

Figure 33 *GLU χ_2 side-chain dihedral angle scan. Top panel Karplus curves for torsion angles related to χ_2 dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The scan of the χ_2 dihedral angle brings the sidechain oxygen atom attached to Cδ closer to the backbone amide proton to form a hydrogen bond. Thus, the making and breaking of the hydrogen bond is likely to be responsible for the unusual feature of the Boltzmann's probability distribution graphs and jumps in the Karplus curves.

3.9.8 Karplus coefficients, GLN

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	0.02	0.33	-0.12	0.15	0.00	0.20
HA	CA	C	N	1.21	0.32	-0.08	0.12	-0.22	0.21
HA	CA	C	N	-0.88		-0.61		-0.27 [104]	
CB	CA	C	N	0.22	0.30	0.00	0.17	0.04	0.18
CG	CB	CA	N	1.37	0.26	-0.15	0.17	-0.04	0.13
CG	CB	CA	C	2.94	0.35	-0.42	0.19	0.18	0.19
CG	CB	CA	HA	8.40	0.28	0.74	0.13	0.25	0.16
C	CA	CB	HB1	7.53	0.33	-0.92	0.13	-0.19	0.18
C	CA	CB	HB2	6.80	0.34	-1.27	0.13	0.27	0.20
HA	CA	CB	HB1	13.20	0.34	1.52	0.13	0.27	0.18
HA	CA	CB	HB2	12.67	0.29	1.70	0.15	0.45	0.15
N	CA	CB	HB2	3.39	0.38	-0.35	0.19	0.09	0.24
N	CA	CB	HB1	3.36	0.34	-0.15	0.14	0.03	0.23
H	N	CA	HA	10.87	0.34	0.37	0.11	-0.17	0.26
C	N	CA	HA	5.45	0.29	-1.86	0.15	0.39	0.14
H	N	CA	C	6.44	0.48	-1.24	0.19	-0.33	0.22
H	N	CA	CB	5.54	0.31	-0.39	0.18	-0.36	0.13
C	N	CA	C	2.50	0.28	-0.89	0.13	0.16	0.18
C	N	CA	CB	2.07	0.28	-0.87	0.15	0.18	0.16
CA	CB	CG	CD	3.72	0.31	-0.97	0.17	0.13	0.13
H1	CB	CG	CD	8.91	0.33	-0.90	0.14	0.01	0.18
H2	CB	CG	CD	8.84	0.30	-0.62	0.15	-0.03	0.16
CA	CB	CG	HG1	8.65	0.40	-0.21	0.25	-0.14	0.18
H1	CB	CG	HG1	12.75	0.30	1.28	0.13	0.41	0.17
H2	CB	CG	HG1	11.92	0.36	1.4	0.34	0.94	0.16
CA	CB	CG	HG2	8.20	0.31	0.06	0.13	0.07	0.20
H1	CB	CG	HG2	12.04	0.27	0.48	0.13	0.94	0.18
H2	CB	CG	HG2	13.00	0.41	1.63	0.12	0.39	0.31
CB	CG	CD	NE	0.25	0.36	-0.17	0.18	-0.01	0.22
HG1	CG	CD	NE	1.34	0.36	-0.30	0.23	-0.26	0.21

HG2	CG	CD	NE	1.37	0.35	-0.27	0.22	-0.24	0.21

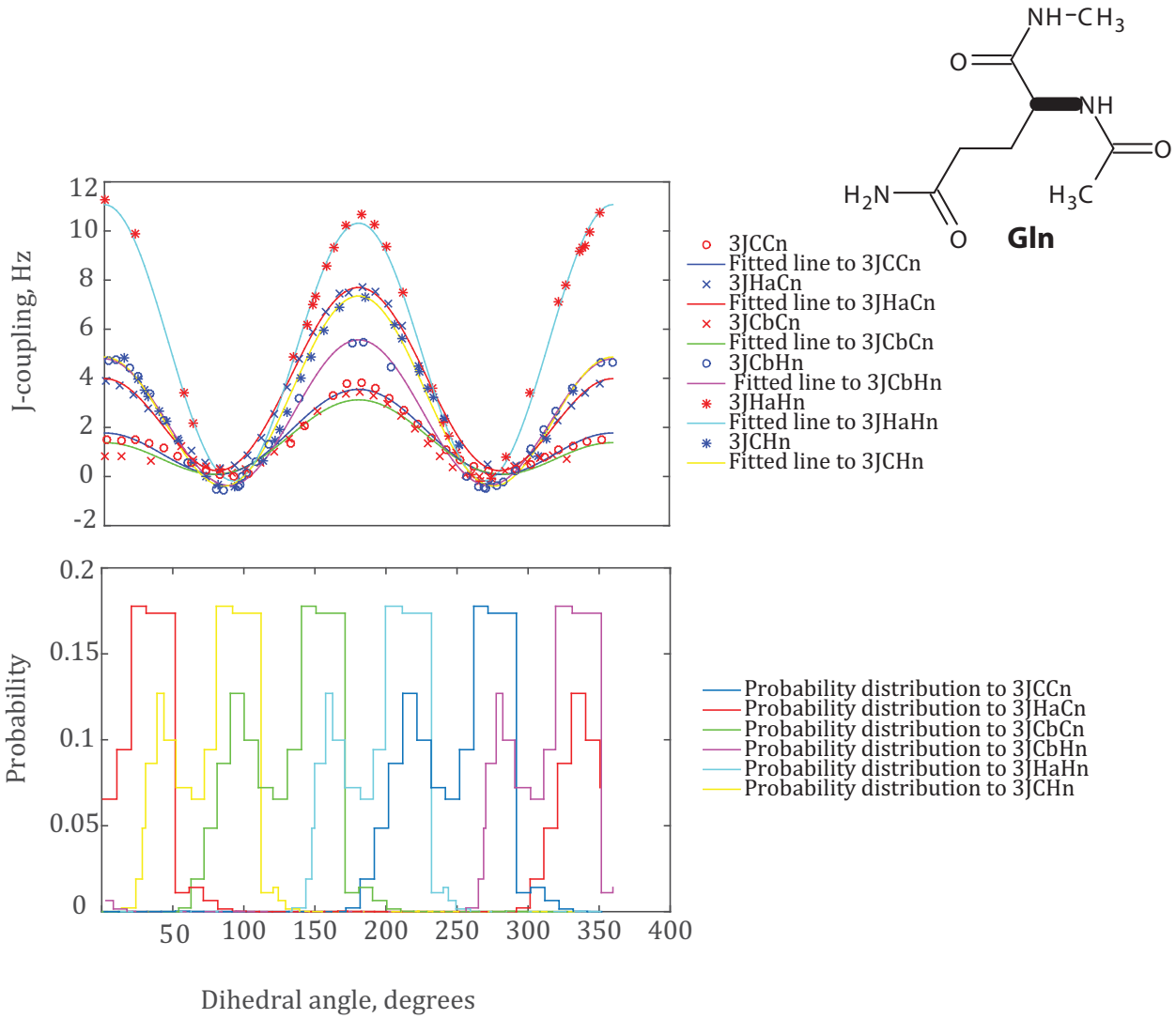

Figure 34 *GLN ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann's probability distribution graphs for torsion angles associated to the ϕ dihedral angle scan respectively.*

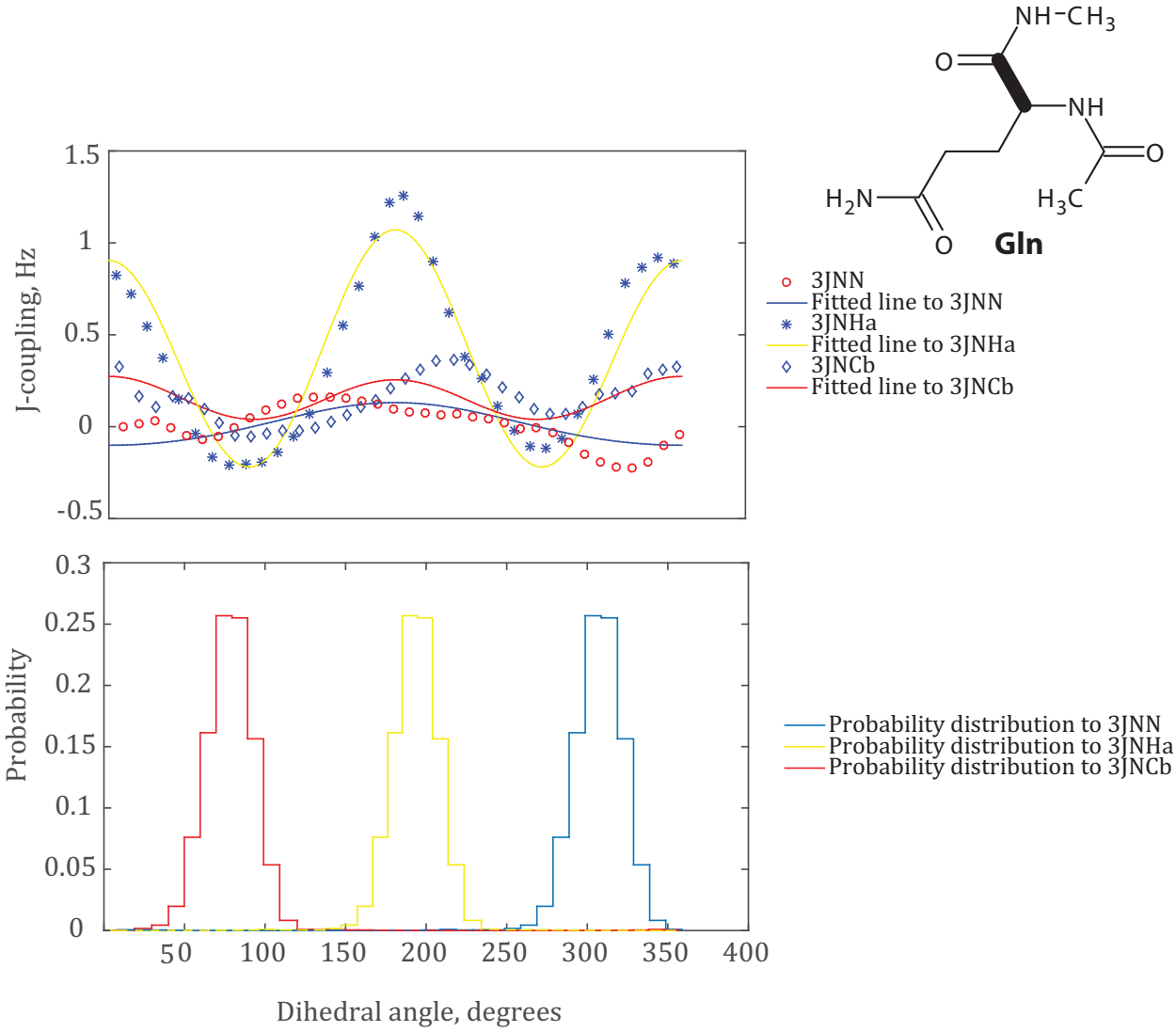

Figure 35 *GLN ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution histograms for the torsion angles related to the ψ dihedral angle scan respectively.*

The ψ dihedral angle is associated to three torsion angles with two of the *J*-couplings involving nitrogen with nitrogen and carbon nuclei. The vicinal coupling of these spins appears to have small amplitudes. The fitting of the Karplus curve to these data turns out to be random noise.

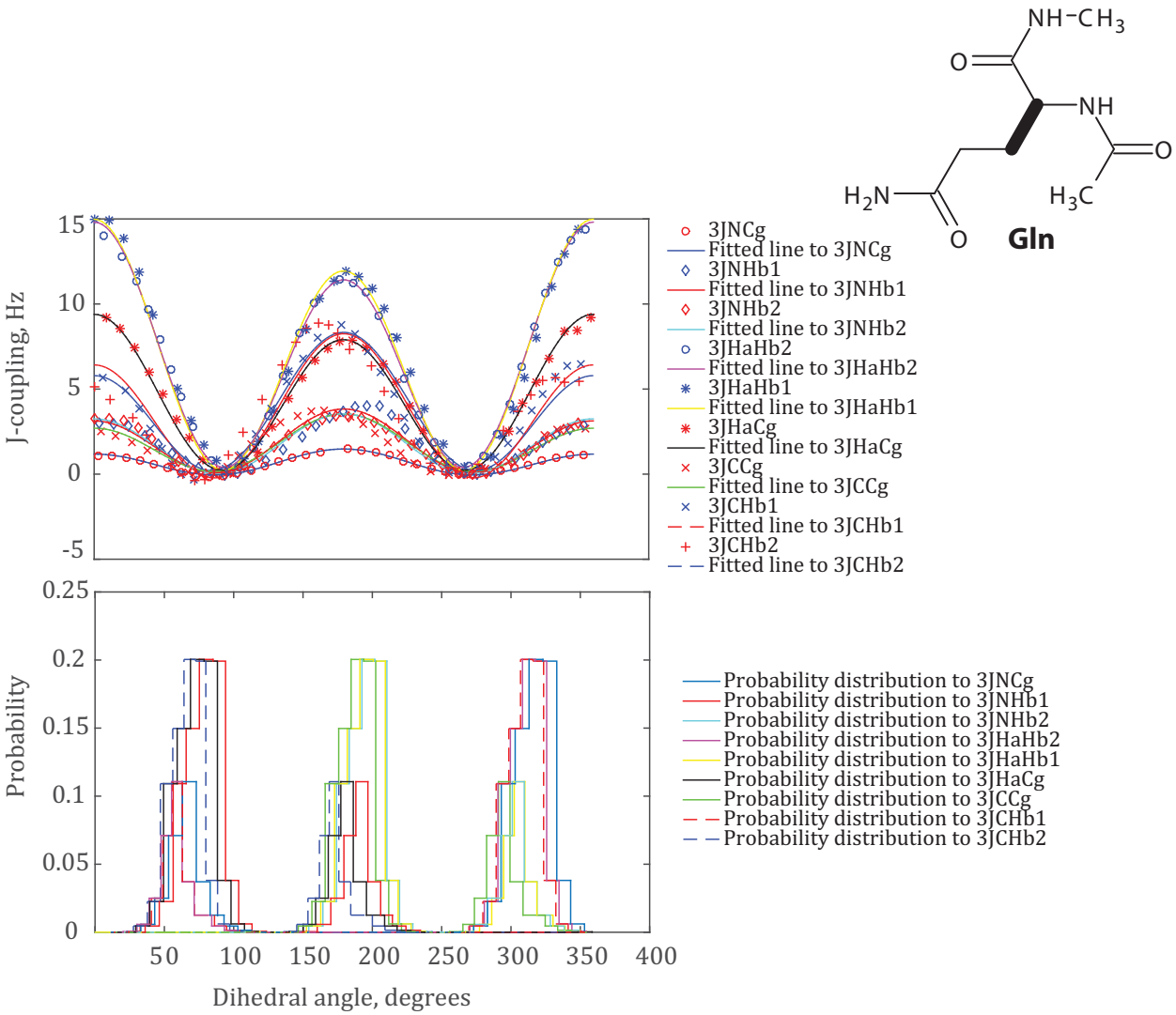

Figure 36 *GLN* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles linked to the* χ_1 *dihedral angle scan and bottom panel probability distribution histograms for the respective torsion angles.*

The staggered along the bumps appearance in the Boltzmann probability distribution histograms might be due to the presence of longer hydrocarbon sidechains of the glutamine amino acid. There also exist hydrogen bond interaction between the sidechain chain amide proton and oxygen attached to the C_δ, but this interaction appears not to affect the Karplus curves and distributions.

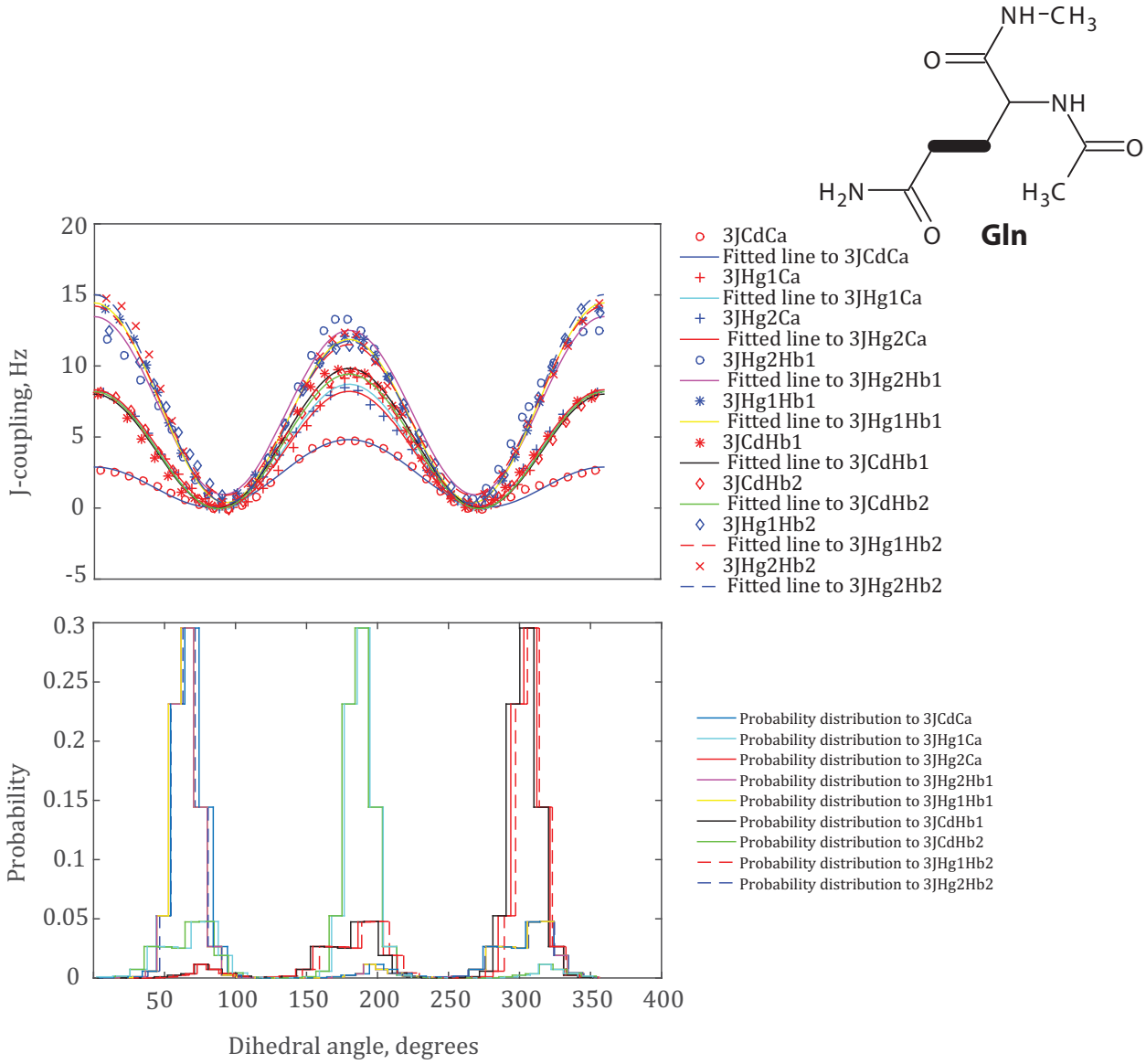

Figure 37 *GLN* χ_2 *side-chain dihedral angle scan. Top panel Karplus curves of torsion angles associated to the* χ_2 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs of the respective torsion angles.*

Hydrogen bond between sidechain oxygen atom attached to the Cδ and the backbone amide proton that arises as the χ_2 dihedral angle is rotated along with the sidechain hydrocarbon steric effect is likely to be responsible for the presence of bumps in the probability distribution graph and jumps of the Karplus curves.

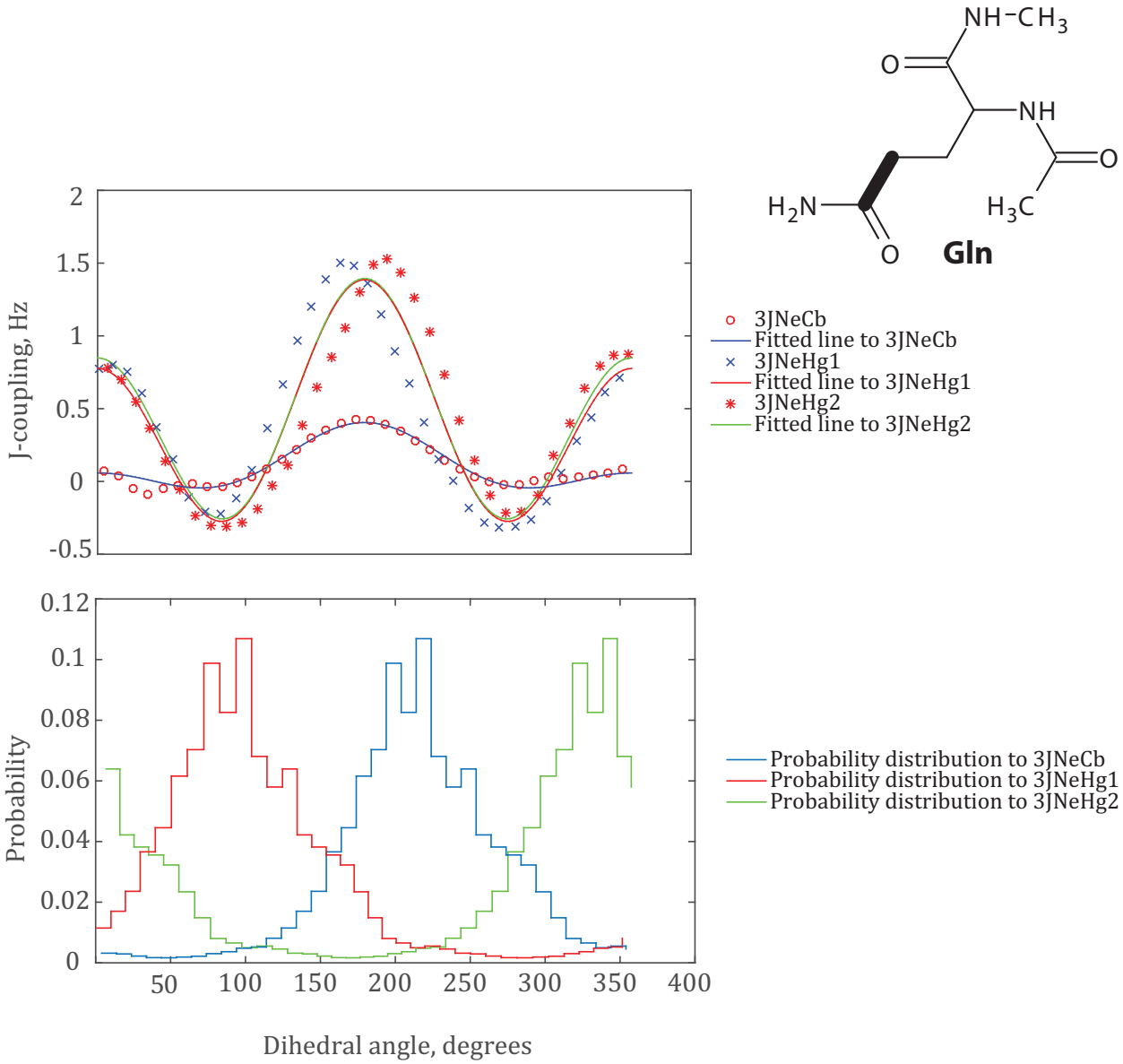

Figure 38 *GLN* χ_3 *side-chain dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles linked to the* χ_3 *dihedral angle scan respectively.*

The Karplus curves for the $^3J_{NeCb}$ coupling data involves very small amplitudes, hence the fitting of the curve resembles random noise. The curves for the other torsion angles involve *J*-coupling between nitrogen and amide protons. The fitting of these Karplus curves over and under predicts the data.

3.9.9 Karplus coefficients, GLY

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
N	C	CA	N	0.00		0.00		0.00	
HA	CA	C	N	0.91		-0.14		-0.19	
CB	CA	C	N	0.00		0.00		0.00	
H	N	CA	H3	13.07	0.51	-0.50	0.15	-0.21	0.35
C	N	CA	H3	6.97	0.29	-1.62	0.15	0.05	0.14
H	N	CA	C	7.73	0.33	-1.73	0.23	-0.72	0.14
H	N	CA	H2	13.25	0.40	-0.46	0.12	-0.47	0.31
C	N	CA	C	3.65	0.32	-0.99	0.17	-0.19	0.25
C	N	CA	H2	7.37	0.31	-1.78	0.16	-0.08	0.15

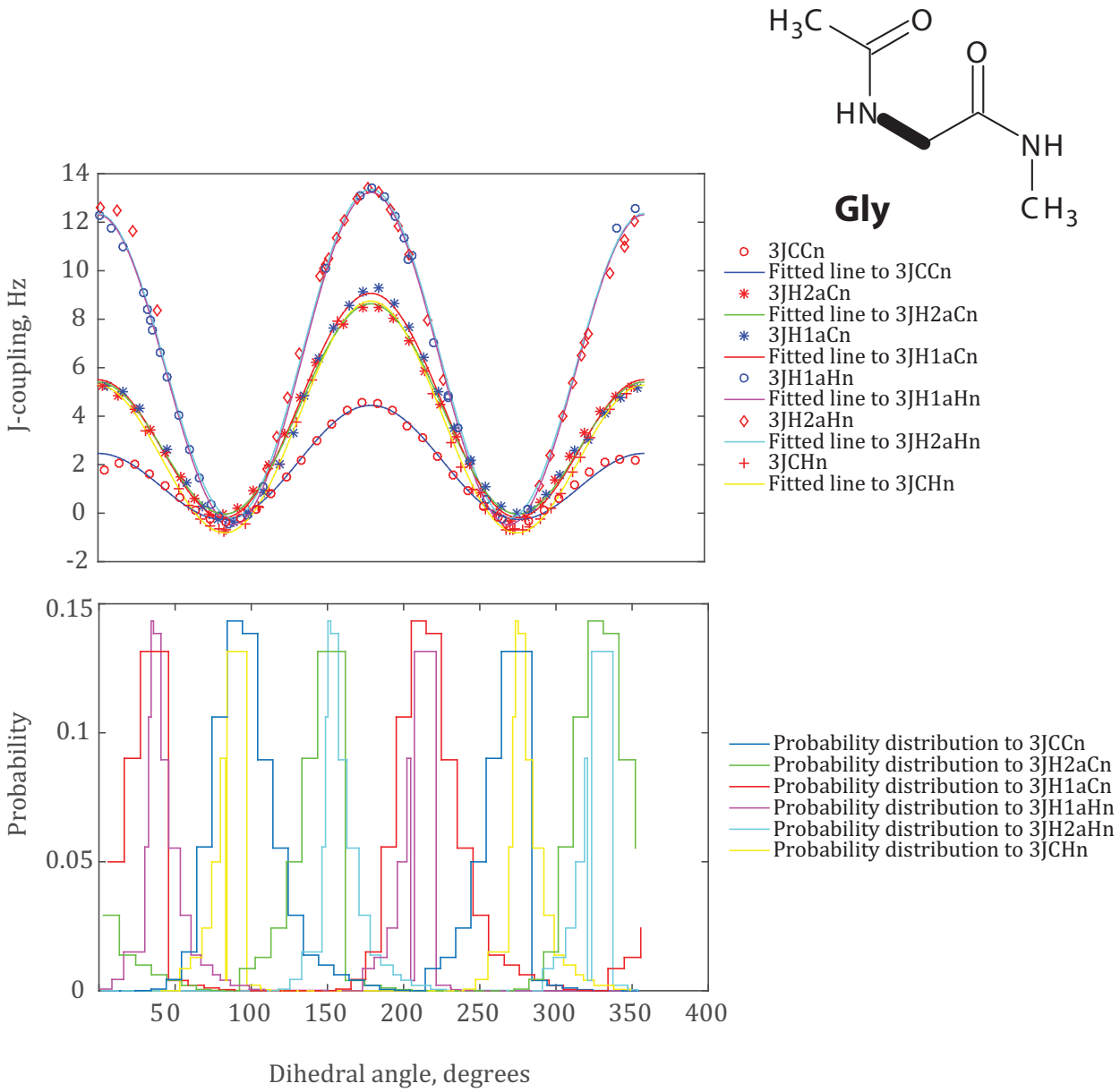

Figure 39 *GLY ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ϕ dihedral angle scan respectively.*

Hydrogen bond between the backbone nitrogen nuclei part of the ϕ dihedral angle and the oxygen atom bonded to carbon that is attached to the nitrogen is likely to be responsible for the anomalous appearance of the Boltzmann probability graphs and the jumps on the Karplus curves.

3.9.10 Karplus coefficients, HIS

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	-0.02	0.39	-0.04	0.25	-0.01	0.23
HA	CA	C	N	0.94	0.31	0.15	0.14	-0.18	0.17
CB	CA	C	N	0.09	0.35	-0.09	0.15	0.04	0.16
H	N	CA	HA	11.45	0.40	-0.27	0.11	-0.27	0.28
C	N	CA	HA	5.66	0.34	-1.65	0.13	0.37	0.22
H	N	CA	C	6.58	0.40	-0.84	0.13	-0.46	0.30
H	N	CA	CB	6.42	0.29	-0.70	0.12	-0.57	0.18
C	N	CA	C	2.46	0.33	-0.92	0.18	0.19	0.18
C	N	CA	CB	2.52	0.39	-0.60	0.26	-0.05	0.20
HA	CA	CB	CG	8.47	0.40	-0.79	0.12	0.34	0.25
HA	CA	CB	HB1	11.94	0.33	1.42	0.14	0.43	0.18
HA	CA	CB	HB2	11.55	0.33	1.31	0.14	0.70	0.16
N	CA	CB	CG	1.26	0.30	-0.24	0.13	0.01	0.17
N	CA	CB	HB1	2.82	0.30	-0.30	0.11	0.03	0.21
N	CA	CB	HB2	2.49	0.35	-0.07	0.12	0.08	0.22
C	CA	CB	CG	3.88	0.33	-0.59	0.13	0.09	0.18
HA	CA	CB	HB1	7.07	0.34	-0.39	0.13	0.03	0.20
C	CA	CB	HB2	6.62	0.32	-0.21	0.18	0.13	0.17

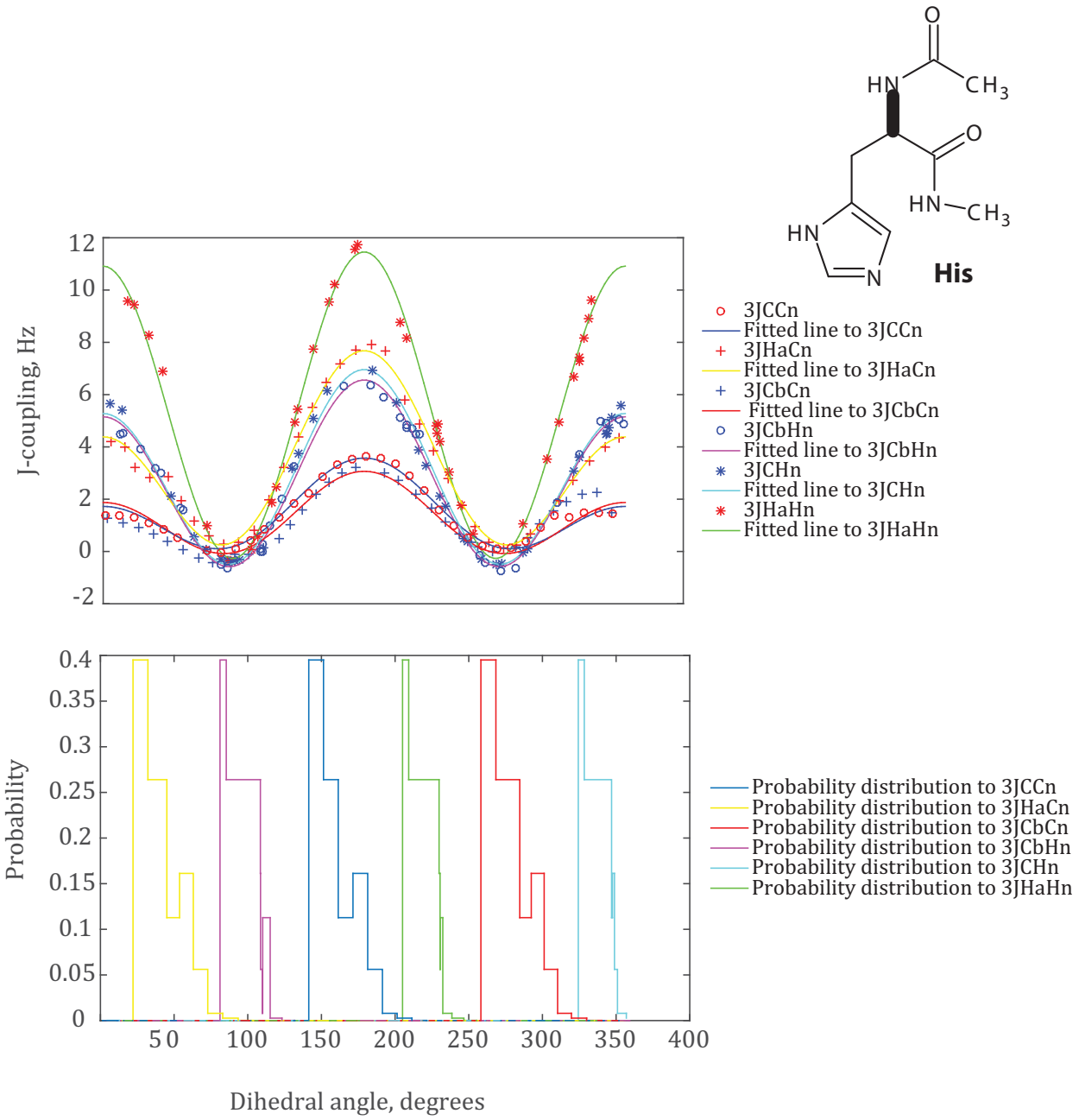

Figure 40 *HIS ϕ backbone dihedral angle scan. Top panel Karplus curves for the torsion angles associated to the ϕ dihedral angle scan. Bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The presence of hydrogen bond between amide backbone proton and oxygen attached to carbon that is bonded to C_α along with interaction between the amide proton of the cyclic sidechain group of histidine with oxygen attached to carbon atom next to amide nitrogen of the backbone are likely to be responsible for the anomaly of the Boltzmann energy probability distribution

graphs. The broken Karplus curves are those of the torsion angles that involved backbone amide proton as one of the three bond separating *J*-coupled nuclei. This is mainly due to the involvement of the amide proton in hydrogen bond interaction. Thus the breaking of hydrogen bond as the dihedral angle rotated breaks the coupling hence the fitting of the Karplus curves.

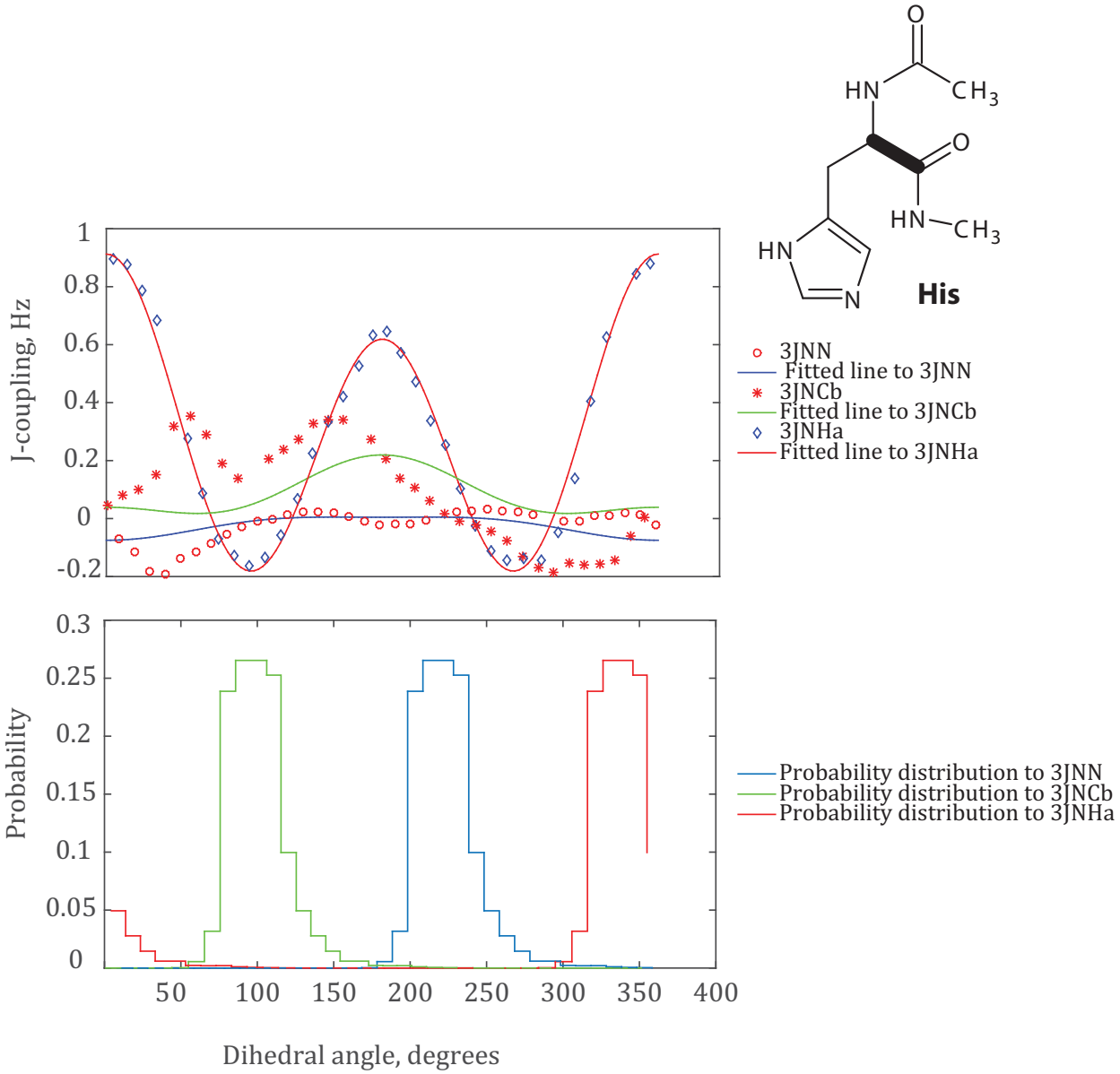

Figure 41 *HIS ψ backbone dihedral angle scan. Top panel Karplus curves for the torsion angles that are related to the ψ dihedral angle scan. Bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

As noted earlier, the amplitudes of the $^{3}J_{NN}$ and $^{3}J_{NCb}$ are small hence the fitting of the Karplus curves looks random noise. However, the fitting of the $^{3}J_{NHa}$ coupling data from electronic

structure calculation at the above mentioned theory approximated by Karplus curve that shows Karplus like pattern.

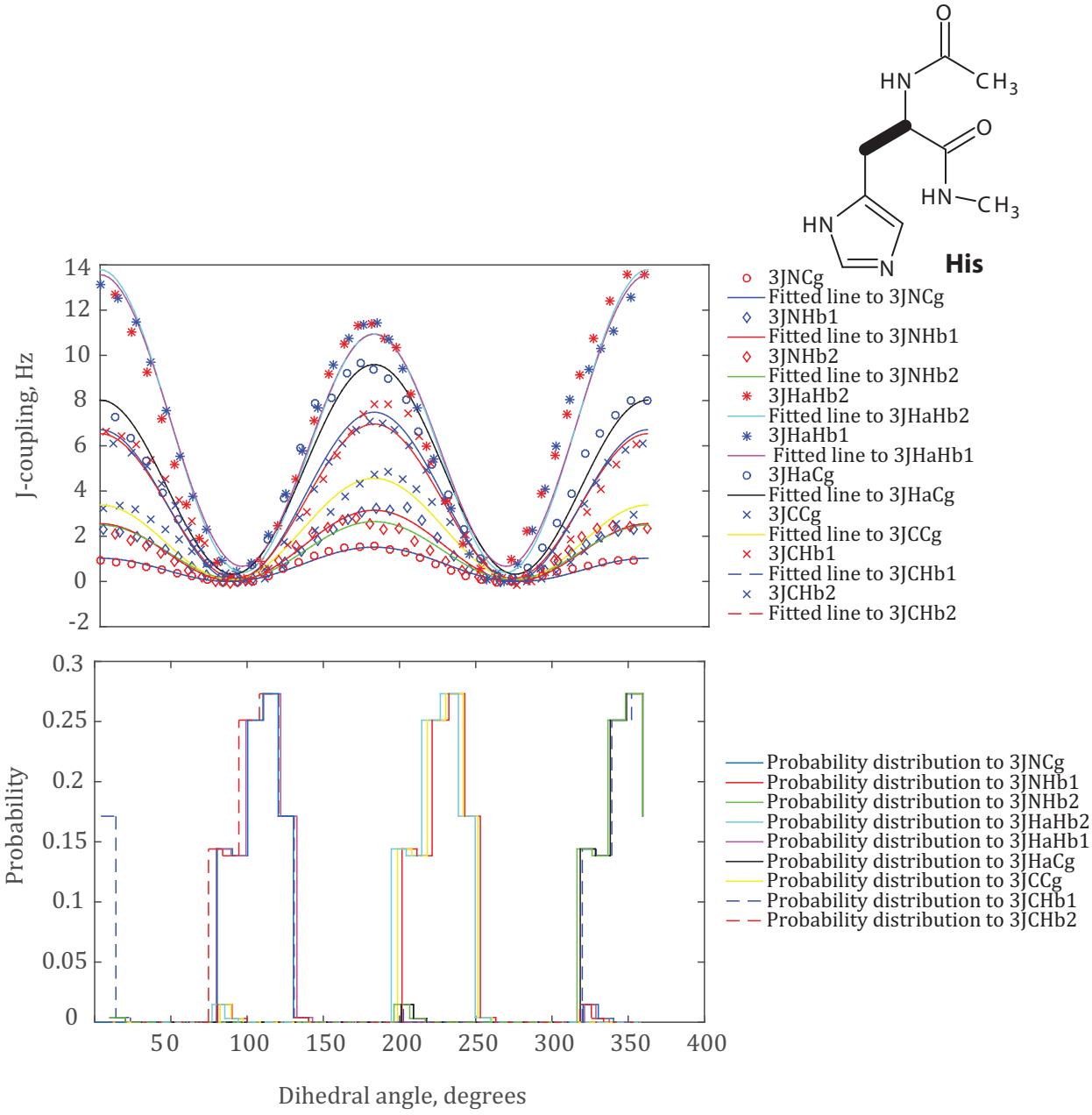

Figure 42 *HIS* χ_1 *side-chain dihedral angle scan. Top panel Karplus curve for the torsion angles associated to the* χ_1 *dihedral angle and bottom panel Boltzmann's probability distribution graphs for the respective torsion angles.*

The making and breaking of a hydrogen bond between the cyclic sidechain amide proton and backbone oxygen atom attached to the carbon that is bonded to C_α on the other residue as the

χ_1 dihedral scanned is likely to be responsible for the anomaly that is present in the Boltzmann probability distribution graphs. The fitting of some of the Karplus curves shows an under and over prediction due to the above mentioned hydrogen bond effect. There is another hydrogen bond present between atoms of the backbone chain and this appears not to influence the Karplus curves and distribution graphs.

3.9.11 Karplus coefficients, ILE

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
N	C	CA	N	0.00		0.00		0.00	
HA	CA	C	N	1.31		0.00		-0.11	
CB	CA	C	N	0.00		0.00		0.00	
CG1	CB	CA	N	1.12	0.34	0.07	0.14	0.12	0.20
CG1	CB	CA	N	2.64		0.26		-0.22 [110]	
CG1	CB	CA	C	3.31	0.35	-0.12	0.14	0.12	0.23
CG1	CB	CA	C	3.42		-0.59		0.17 [110]	
CG1	CB	CA	HA	6.95	0.35	0.72	0.12	0.13	0.22
C	CA	CB	HB	7.01	0.34	-0.65	0.14	0.03	0.18
C	CA	CB	CG2	3.22	0.36	-0.31	0.14	0.16	0.24
C	CA	CB	CG2	3.42		-0.59		0.17 [110]	
HA	CA	CB	HB	10.75	0.30	1.92	0.13	0.68	0.17
HA	CA	CB	CG2	6.87	0.31	1.27	0.12	0.18	0.23
N	CA	CB	CG2	1.34	0.32	0.04	0.15	0.02	0.17
N	CA	CB	CG2	2.64		0.26		-0.22 [110]	
N	CA	CB	HB	2.93	0.34	0.04	0.11	0.05	0.23
H	N	CA	HA	11.92	0.29	0.07	0.13	-0.10	0.21
C	N	CA	HA	5.96	0.26	-1.87	0.15	0.51	0.15
H	N	CA	C	7.16	0.46	-1.47	0.32	-0.45	0.17
H	N	CA	CB	5.02	0.31	-0.76	0.17	-0.17	0.13

C	N	CA	C	3.07	0.25	-1.05	0.13	-0.02	0.15
C	N	CA	CB	1.96	0.30	-0.74	0.15	0.03	0.15
CA	CB	CG1	CD	3.51	0.32	-0.30	0.16	0.03	0.19
HB	CB	CG1	CD	8.21	0.37	0.54	0.13	0.06	0.22
CG2	CB	CG1	CD	5.36	0.33	0.93	0.13	0.18	0.20
CA	CB	CG1	HG1	7.57	0.36	-0.00	0.13	0.11	0.22
HB	CB	CG1	HG1	10.85	0.35	1.39	0.13	0.73	0.19
CG2	CB	CG1	HG1	8.04	0.38	0.54	0.18	0.23	0.19
CA	CB	CG1	HG2	7.33	0.32	-0.63	0.15	0.30	0.19
HB	CB	CG1	HG2	11.17	0.40	1.57	0.25	0.57	0.22
CG2	CB	CG1	HG2	8.64	0.33	0.91	0.17	-0.08	0.18

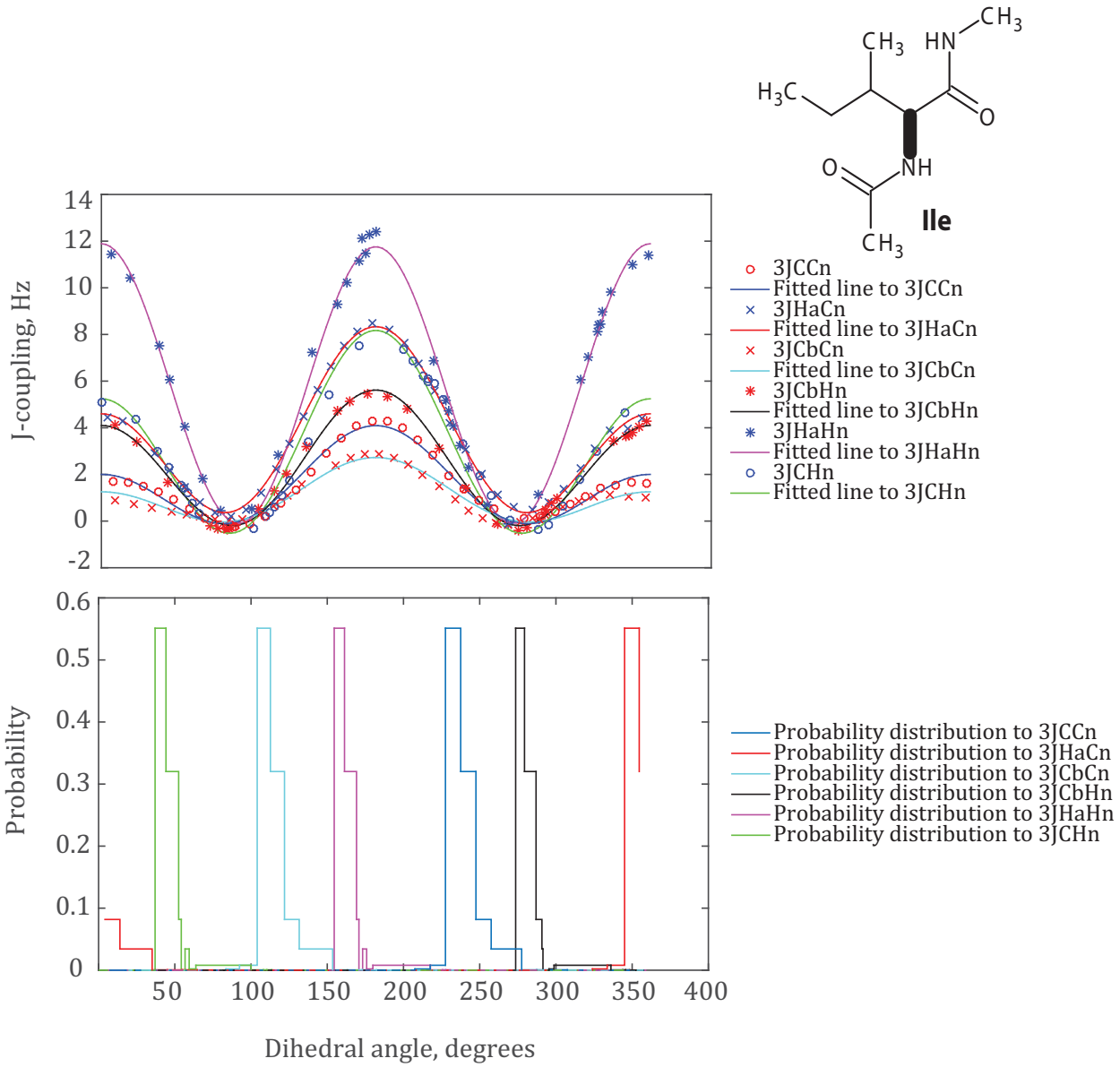

Figure 43 *ILE ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and probability distribution graphs for the torsion angles related to the ϕ dihedral angle.*

Karplus curves of the nuclei that involved the backbone amide proton are broken, this is due to the participation of the atom in a hydrogen bond. This is the likely reason for the thin sharp appearance of the Boltzmann probability distribution graphs and jumps of the Karplus curves.

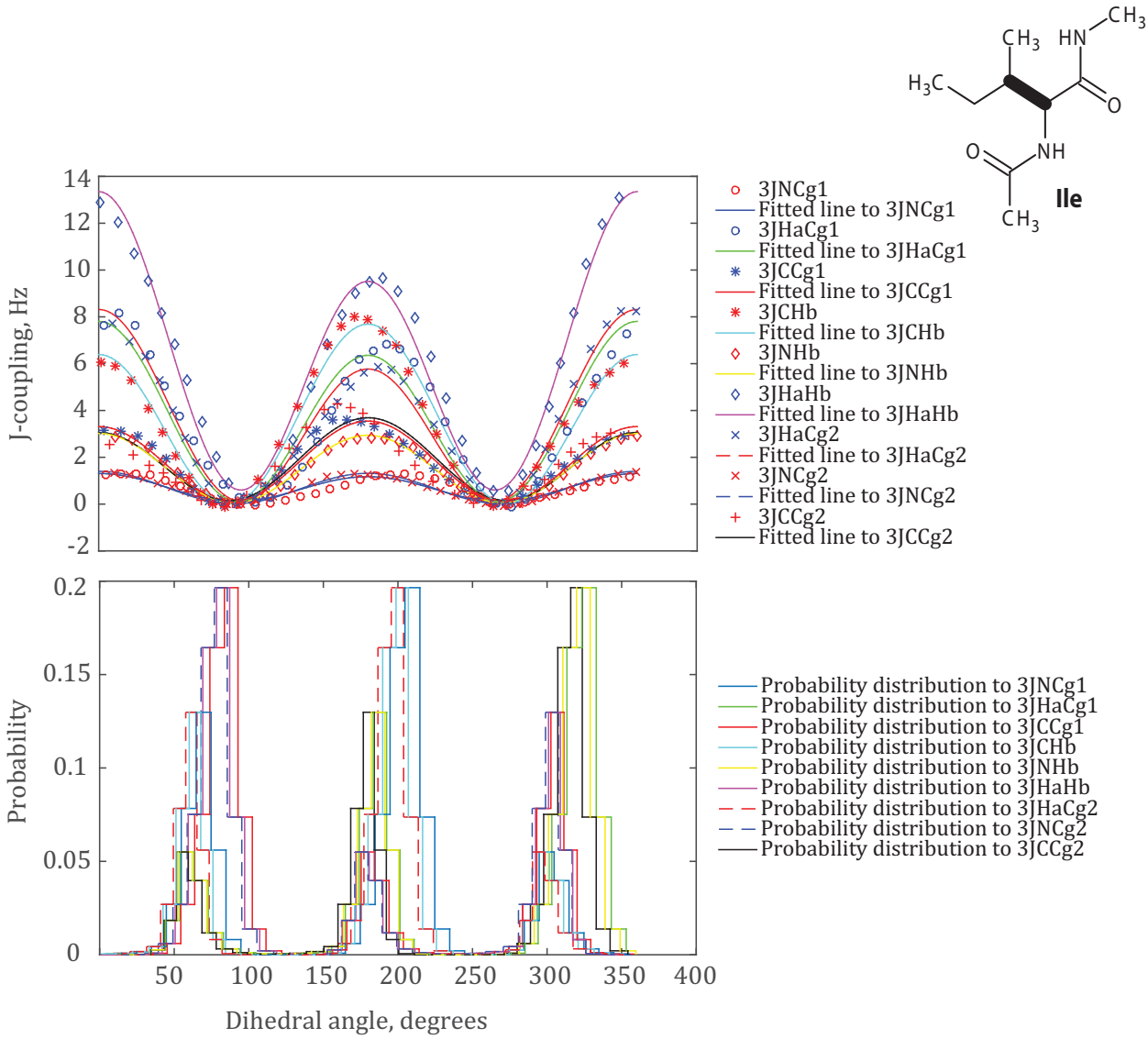

Figure 44 *ILE* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles associated to the* χ_1 *dihedral angle. Bottom panel Boltzmann's probability distribution graphs for the respective torsion angles.*

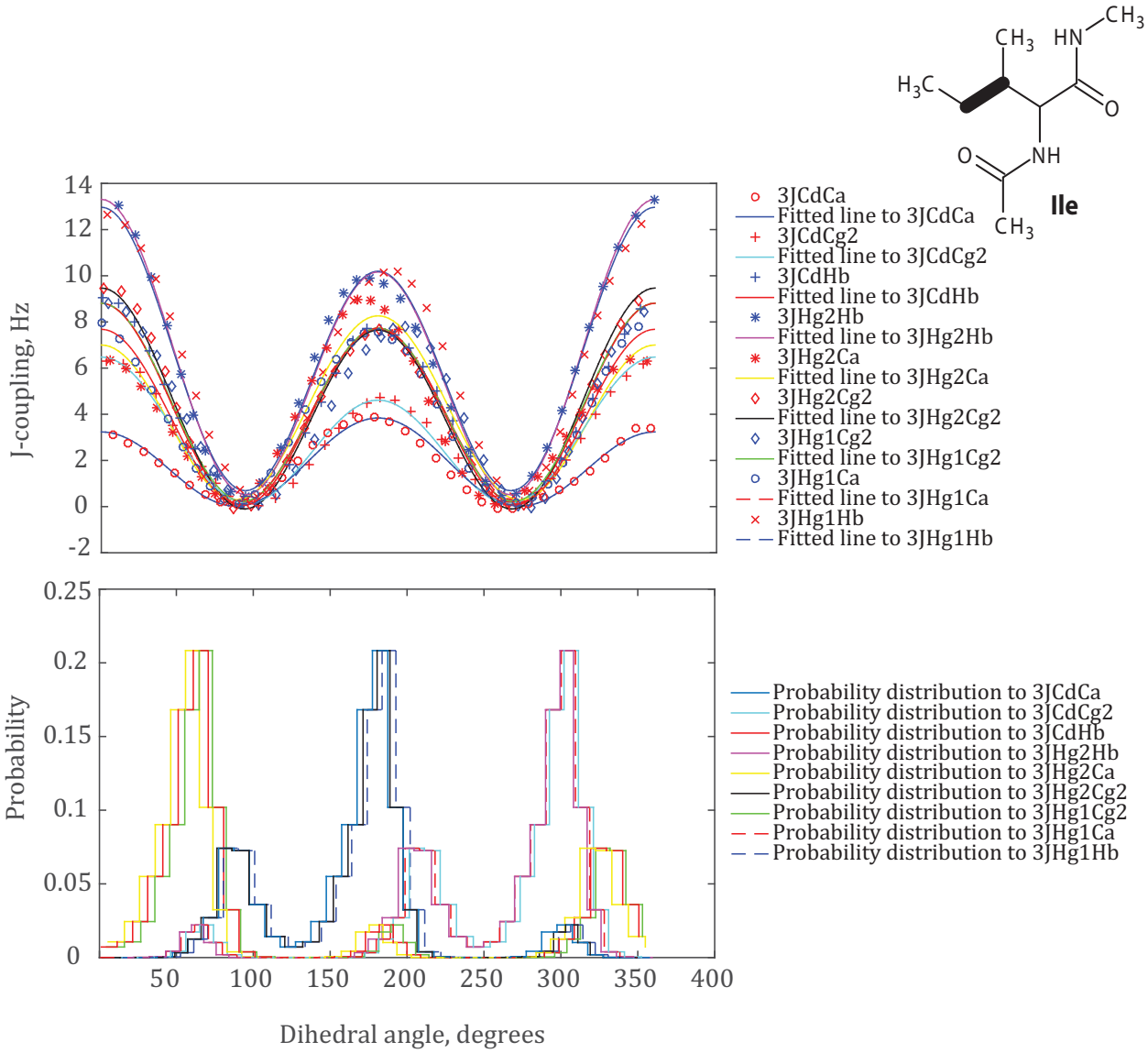

Figure 45 *ILE* χ_2 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles related to the* χ_2 *dihedral angle and bottom panel probability distribution for the respective torsion angles.*

3.9.12 Karplus coefficients, LEU

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	0.04	0.30	-0.11	0.15	0.00	0.14
HA	CA	C	N	1.17	0.35	-0.25	0.16	-0.14	0.20
CB	CA	C	N	0.02	0.33	-0.36	0.14	0.02	0.18

CG	CB	CA	N	1.86	0.40	0.14	0.21	-0.03	0.22
CG	CB	CA	C	3.00	0.52	-0.22	0.17	0.29	0.35
CG	CB	CA	HA	7.75	0.49	-0.23	0.39	0.19	0.23
C	CA	CB	HB1	6.94	0.54	-0.28	0.25	-0.09	0.30
C	CA	CB	HB2	7.76	0.66	-0.51	0.25	0.13	0.28
HA	CA	CB	HB1	12.06	0.62	1.96	0.38	0.54	0.22
HA	CA	CB	HB2	12.32	0.58	1.64	0.20	0.86	0.33
N	CA	CB	HB1	4.21	0.61	0.22	0.17	-0.05	0.45
N	CA	CB	HB2	4.27	0.47	0.16	0.29	0.29	0.27
H	N	CA	HA	11.98	0.33	0.05	0.14	-0.12	0.18
C	N	CA	HA	6.28	0.29	-1.62	0.17	0.44	0.14
H	N	CA	CB	5.16	0.31	-0.52	0.12	-0.27	0.20
C	N	CA	CB	1.95	0.32	-0.84	0.17	0.03	0.16
H	N	CA	C	6.48	0.44	-1.52	0.22	-0.38	0.20
C	N	CA	C	2.77	0.29	-1.09	0.15	0.12	0.15
CA	CB	CG1	CD1	3.97	0.30	-0.80	0.12	-0.07	0.20
HB1	CB	CG1	CD1	8.98	0.35	0.67	0.23	-0.12	0.22
HB2	CB	CG1	CD1	8.51	0.40	0.28	0.22	0.23	0.20
CA	CB	CG1	HG	7.59	0.32	-0.90	0.11	0.03	0.21
HB1	CB	CG1	HG	11.14	0.31	1.46	0.13	0.62	0.20
HB2	CB	CG1	HG	11.15	0.30	1.52	0.14	0.74	0.17
CA	CB	CG1	CD2	3.77	0.48	-0.43	0.36	0.01	0.24
HB1	CB	CG1	CD2	8.31	0.42	0.49	0.21	0.22	0.20
HB2	CB	CG1	CD2	9.13	0.36	0.60	0.16	-0.05	0.19

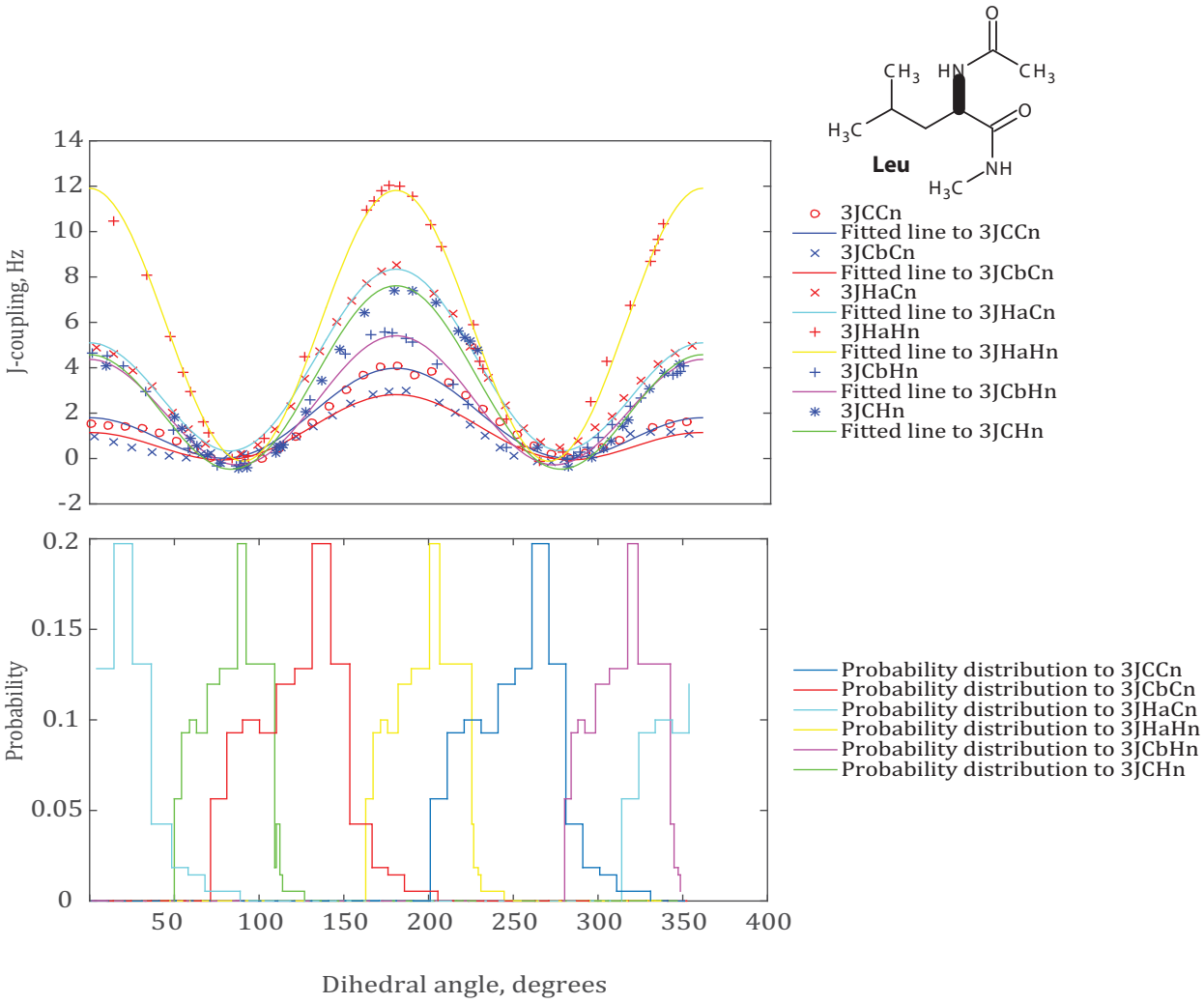

Figure 46 *LEU ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution for the torsion angles associated to the ϕ dihedral angle.*

Hydrogen bod making and breaking as the ϕ dihedral angle is rotated might be responsible for unusual feature of the probability distribution graphs and jumps of the Karplus curves.

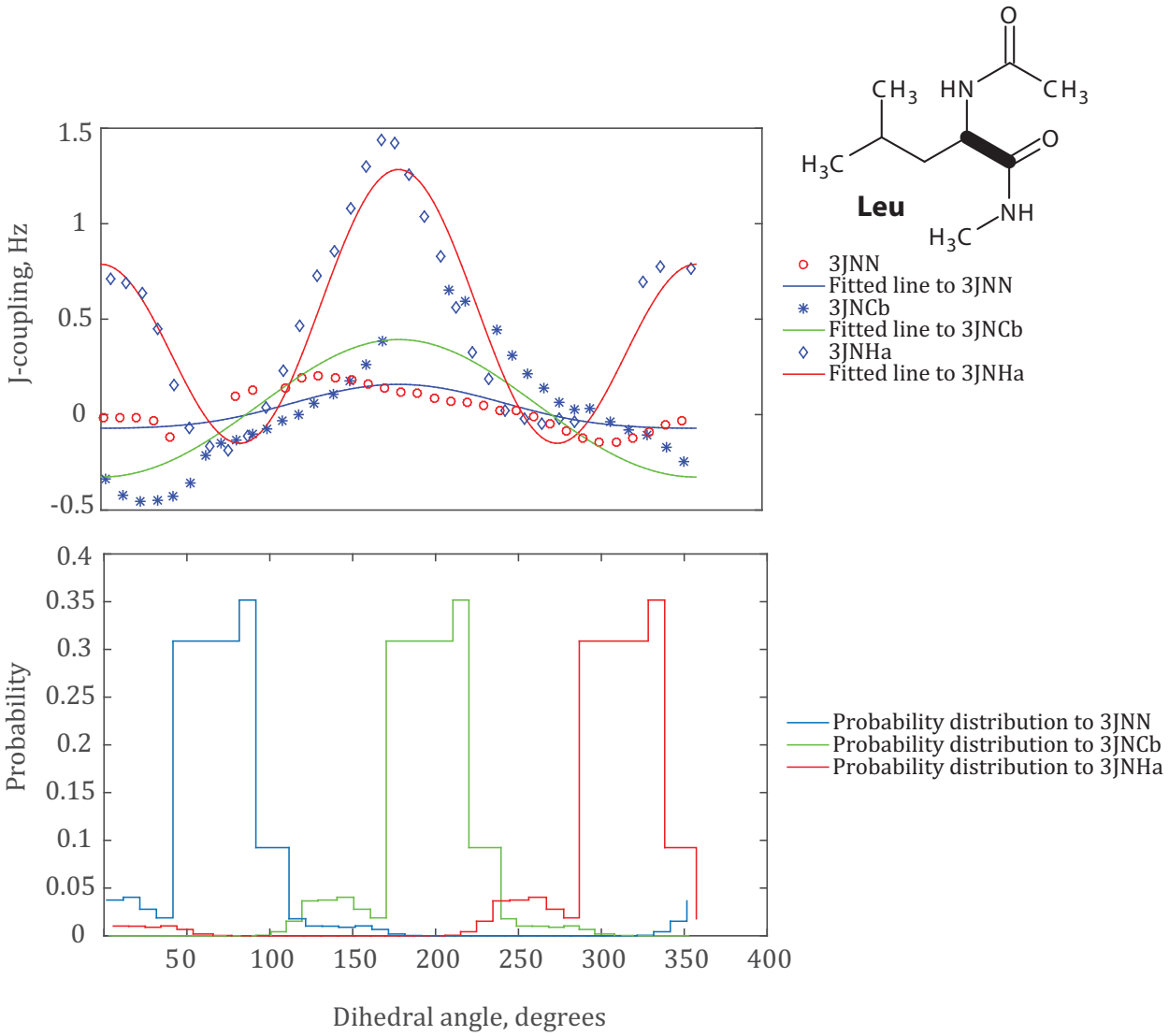

Figure 47 *LEU ψ backbone dihedral angle scan. Top and bottom are Karplus curves and probability distribution graphs for torsion angles linked to the ψ dihedral angle.*

The broken Karplus curves are due the small amplitudes of the *J*-couplings between the atoms involved in the respective torsion angles ($^3J_{NN}$ and $^3J_{NC}$).

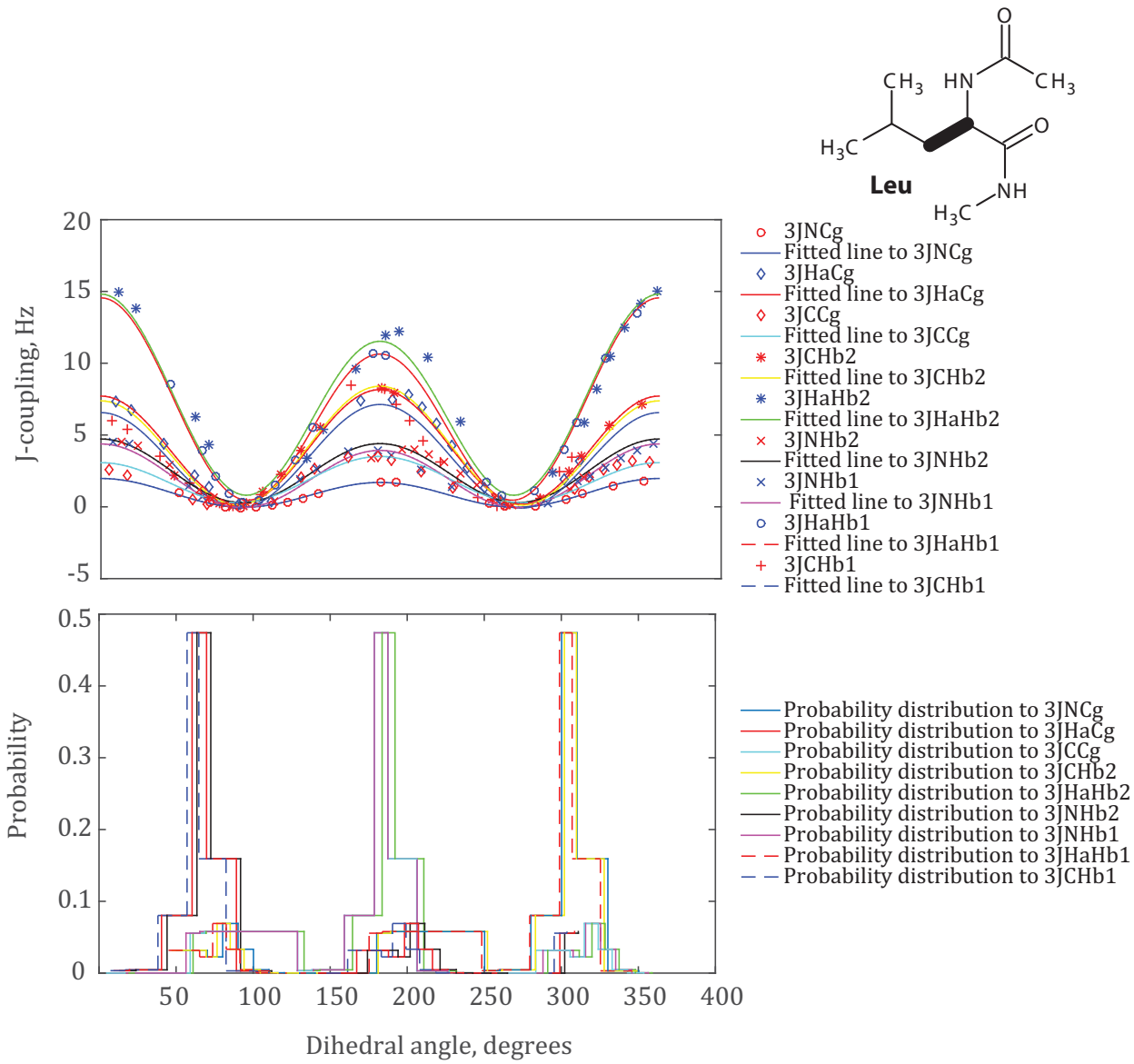

Figure 48 *LEU* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for torsion angles associated to the* χ_1 *dihedral angle and bottom panel probability distribution graphs for the respective torsion angles.*

The effect from the presence of branched sidechain hydrocarbon is more pronounced for the χ_1 dihedral angle. The amplitudes of the interaction energy are smaller compared to the unbranched amino acids, thus the appearance of small bumps in the probability distribution graphs and jumps in Karplus curves are mainly due to this effect and may be also due to tiny contribution from hydrogen bond effect.

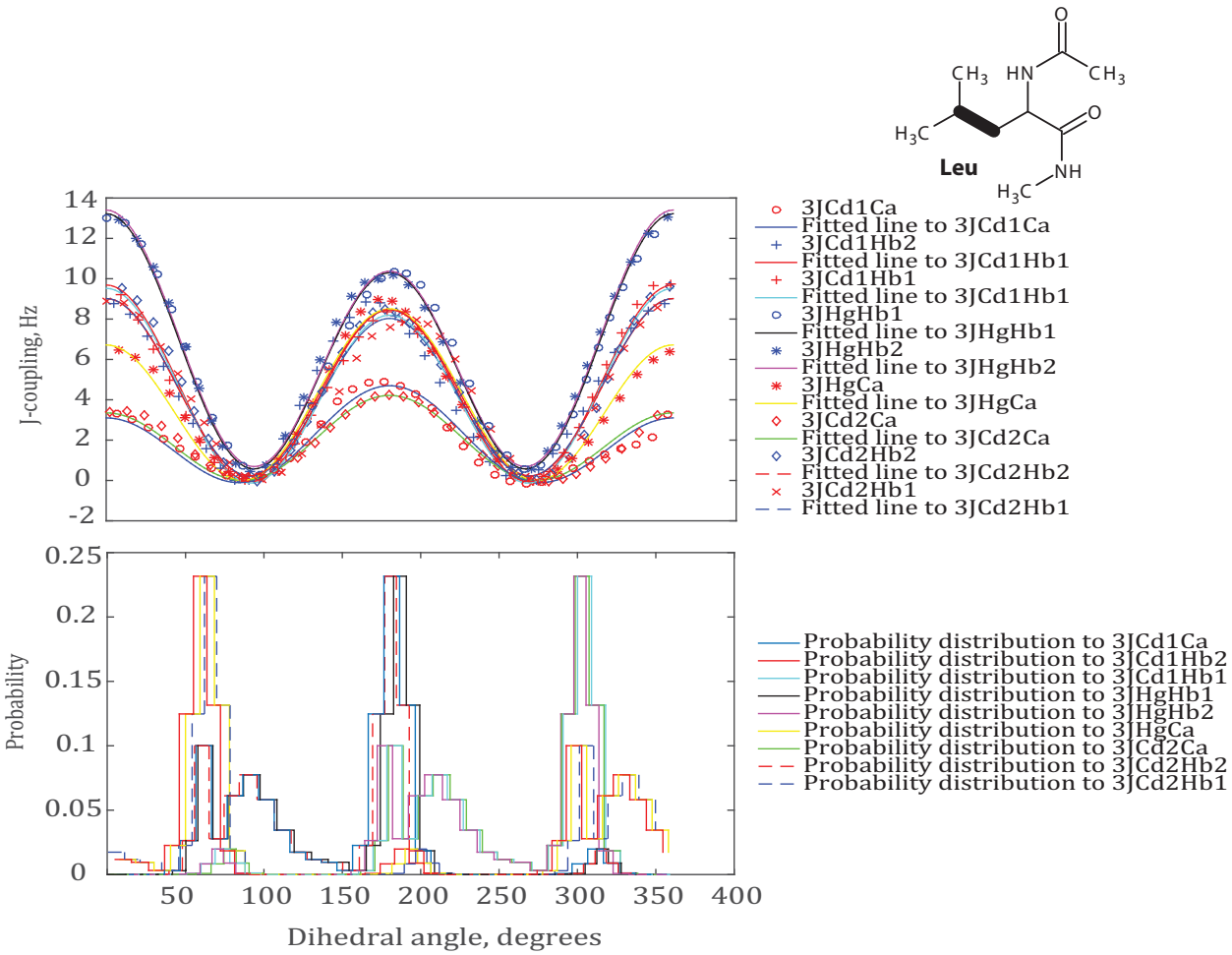

Figure 49 *LEU* χ_2 *side-chain dihedral angle scan. Top panel Karplus curves for torsion angles associated to the* χ_2 *dihedral angle and bottom panel probability distribution graphs for the respective torsion angles.*

3.9.13 Karplus coefficients, LYS

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	-0.01	0.31	-0.14	0.16	-0.03	0.14
HA	CA	C	N	0.86	0.32	-0.25	0.14	-0.37	0.17
CB	CA	C	N	-0.04	0.36	-0.23	0.14	-0.06	0.18
HA	CA	CB	CG	8.93	0.37	0.04	0.13	0.06	0.20
N	CA	CB	CG	1.46	0.28	-0.21	0.16	-0.10	0.15
C	CA	CB	CG	3.09	0.35	-0.67	0.15	-0.04	0.21

HA	CA	CB	HB1	13.30	0.32	1.97	0.17	0.73	0.20
N	CA	CB	HB1	3.48	0.35	-0.39	0.11	0.05	0.26
C	CA	CB	HB1	6.81	0.35	-0.73	0.13	0.37	0.19
HA	CA	CB	HB2	12.22	0.36	1.02	0.17	0.55	0.18
N	CA	CB	HB2	3.31	0.37	-0.25	0.17	0.01	0.25
C	CA	CB	HB2	7.44	0.36	-0.82	0.12	-0.30	0.23
H	N	CA	HA	10.36	0.30	0.48	0.11	-0.23	0.22
C	N	CA	HA	5.38	0.30	-1.73	0.15	0.34	0.16
H	N	CA	CB	6.26	0.27	-0.61	0.12	-0.55	0.18
C	N	CA	CB	2.22	0.29	-0.83	0.16	0.08	0.15
H	N	CA	C	6.48	0.44	-1.12	0.19	-0.45	0.23
C	N	CA	C	2.53	0.25	-0.86	0.12	0.09	0.16
CA	CB	CG	CD	5.12	0.34	-0.08	0.15	0.08	0.19
HB1	CB	CG	CD	11.52	0.32	0.27	0.15	-0.03	0.16
HB2	CB	CG	CD	10.72	0.43	-0.09	0.22	0.13	0.17
CA	CB	CG	HG1	8.23	0.34	-0.25	0.16	0.02	0.19
HB1	CB	CG	HG1	11.92	0.37	1.27	0.16	0.73	0.17
HB2	CB	CG	HG1	13.07	0.37	1.34	0.20	0.29	0.16
CA	CB	CG	HG2	7.67	0.31	0.08	0.16	0.18	0.17
HB1	CB	CG	HG2	12.67	0.30	1.51	0.11	0.45	0.21
HB2	CB	CG	HG2	12.27	0.34	0.78	0.13	0.68	0.19
CB	CG	CD	CE	7.70	0.30	0.80	0.14	0.01	0.17
HG1	CG	CD	CE	14.37	0.46	-0.19	0.32	0.12	0.28
HG2	CG	CD	CE	14.07	0.36	-0.03	0.15	0.03	0.18
CB	CG	CD	HD1	7.27	0.31	-0.43	0.16	0.06	0.17
HG1	CG	CD	HD1	13.17	0.43	1.12	0.22	0.30	0.21
HG2	CG	CD	HD1	12.13	0.31	1.14	0.13	0.71	0.17
CB	CG	CD	HD2	7.97	0.36	0.20	0.17	0.05	0.23
HG1	CG	CD	HD2	12.46	0.36	1.59	0.16	0.59	0.20
HG2	CG	CD	HD2	12.93	0.42	1.50	0.16	0.42	0.19
CG	CD	CE	NZ	2.32	0.34	0.12	0.16	-0.04	0.23
HD1	CD	CE	NZ	6.48	0.37	0.17	0.15	-0.02	0.19

HD2	CD	CE	NZ	6.83	0.39	0.11	0.17	-0.05	0.20
CG	CD	CE	HE1	7.45	0.36	0.04	0.14	-0.00	0.19
HD1	CD	CE	HE1	13.60	0.45	1.13	0.32	0.00	0.23
HD2	CD	CE	HE1	12.62	0.35	0.59	0.13	0.60	0.19
CG	CD	CE	HE2	7.33	0.39	-0.02	0.15	0.00	0.20
HD1	CD	CE	HE2	12.06	0.35	0.78	0.13	0.64	0.18
HD2	CD	CE	HE2	13.75	0.61	0.97	0.56	0.05	0.25

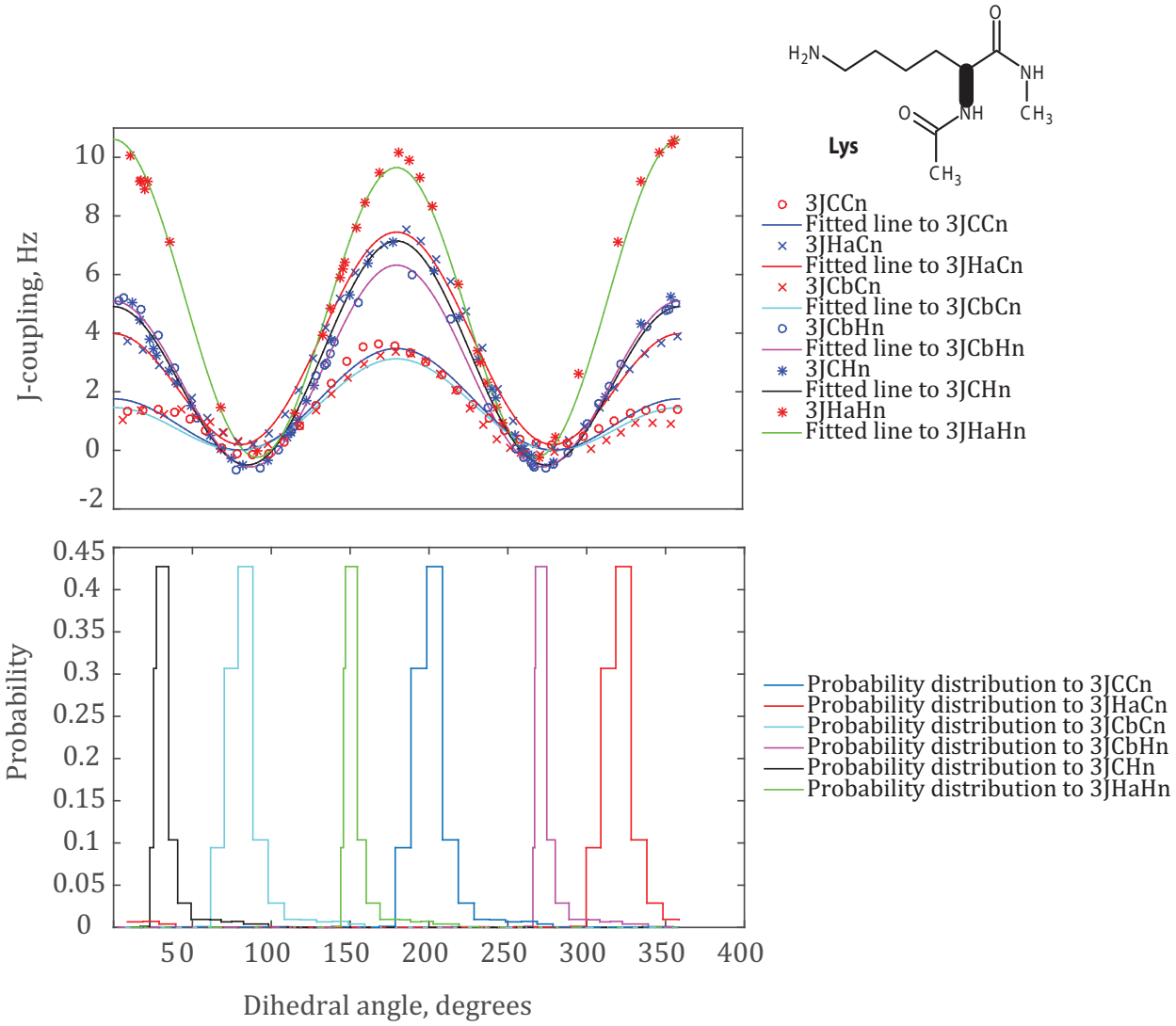

Figure 50 *LYS ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and probability distribution histograms for the torsion angles associated to the ϕ dihedral angle respectively.*

Hydrogen bond making and braking as the ϕ dihedral angle rotated is responsible for the jumps in the Karplus curves and the sharp appearance of the probability distribution graphs.

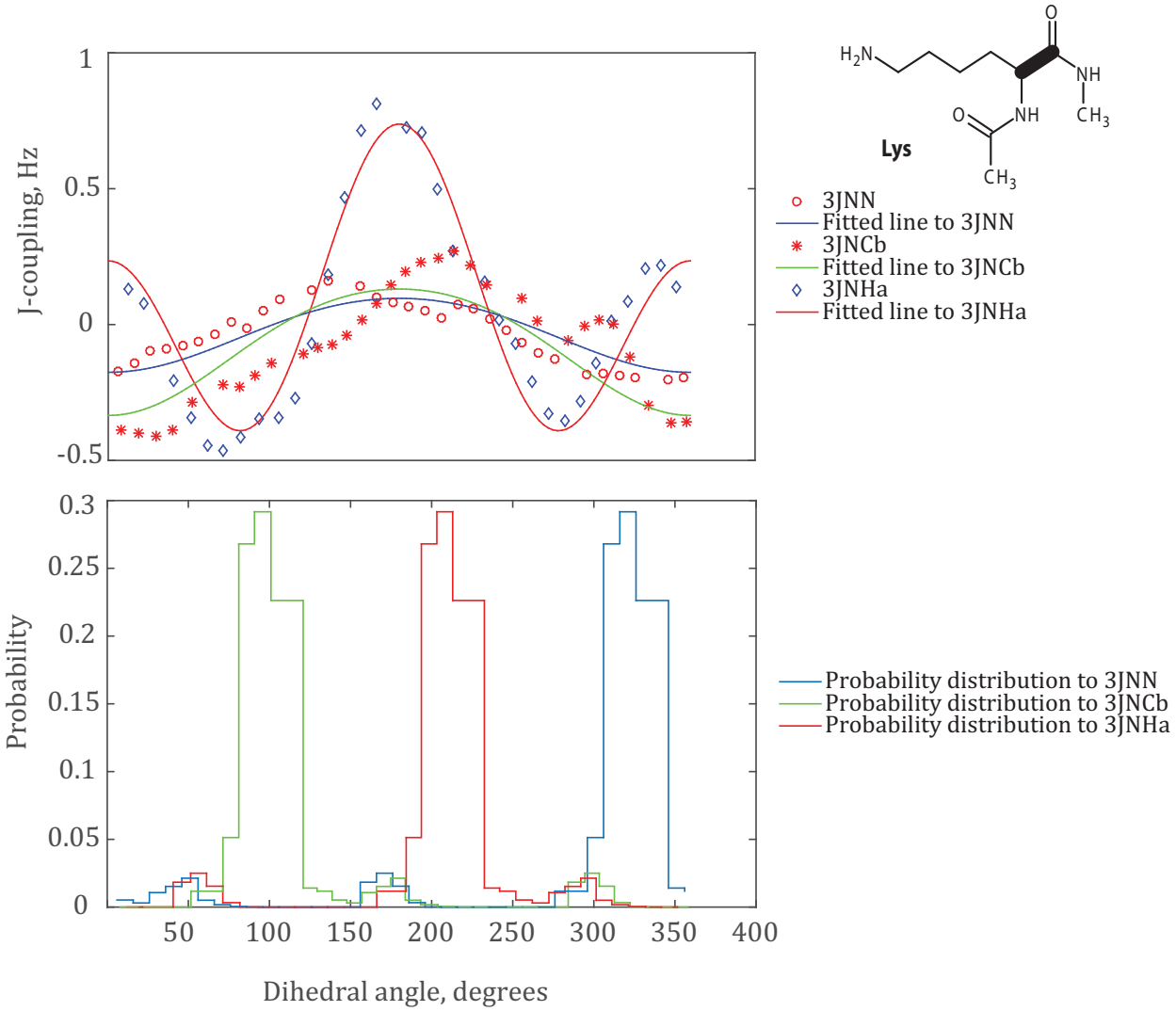

Figure 51 *LYS ψ backbone dihedral angle scan. Top and bottom are Karplus curves and Boltzmann probability distribution graphs for torsion angles linked to ψ dihedral angle scan respectively.*

Small amplitude vicinal coupling of nitrogen with nitrogen and carbon atoms of the ψ dihedral angle is responsible for the broken curves. Steric effect, that arise from the longer chain of hydrocarbons, coupled to hydrogen bond effect are responsible for the presence of the tiny bumps in the energy probability distribution graphs and breaks of the $^{3}J_{NHa}$ Karplus curve.

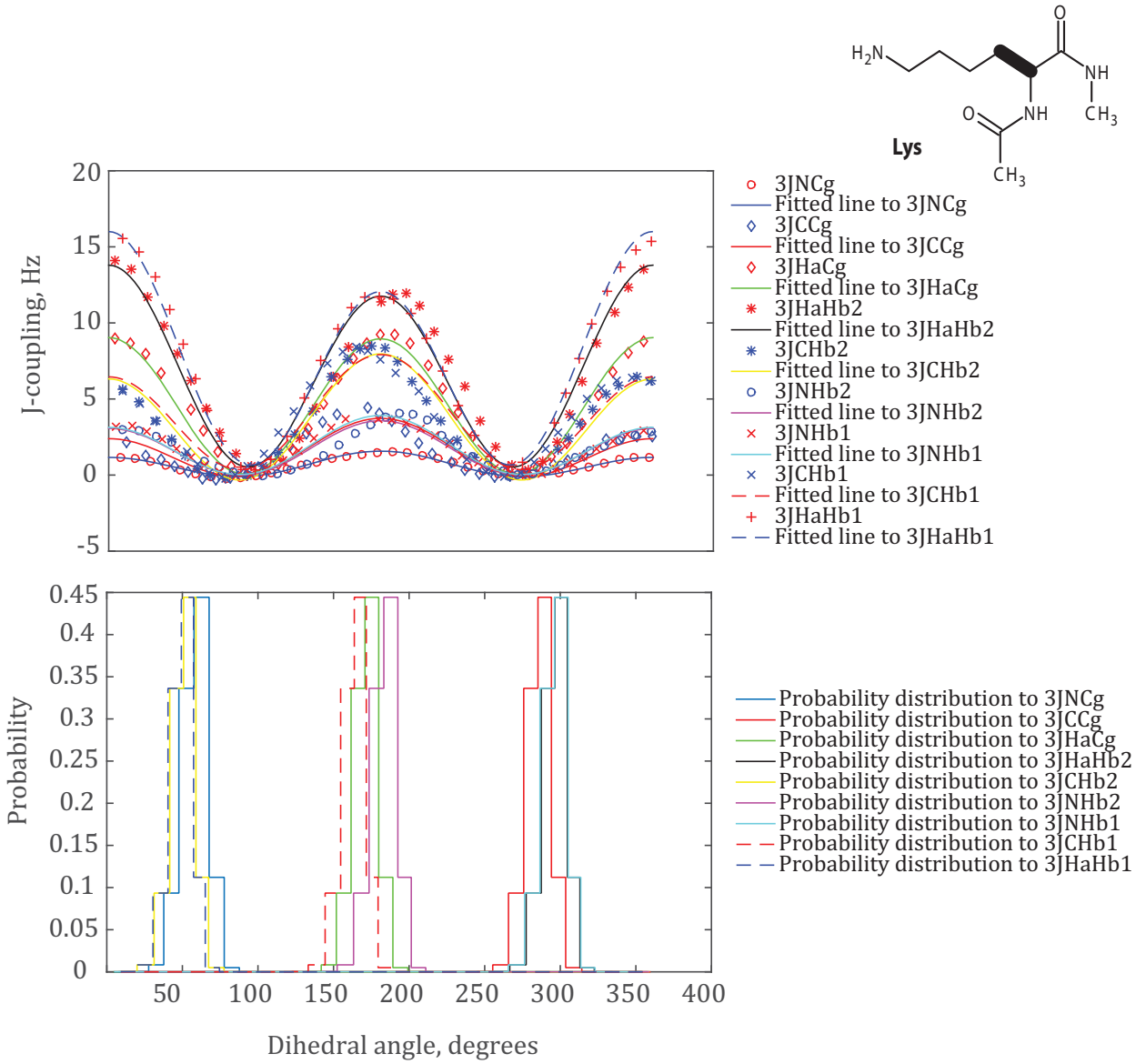

Figure 52 *LYS* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles linked to the* χ_1 *dihedral angle and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

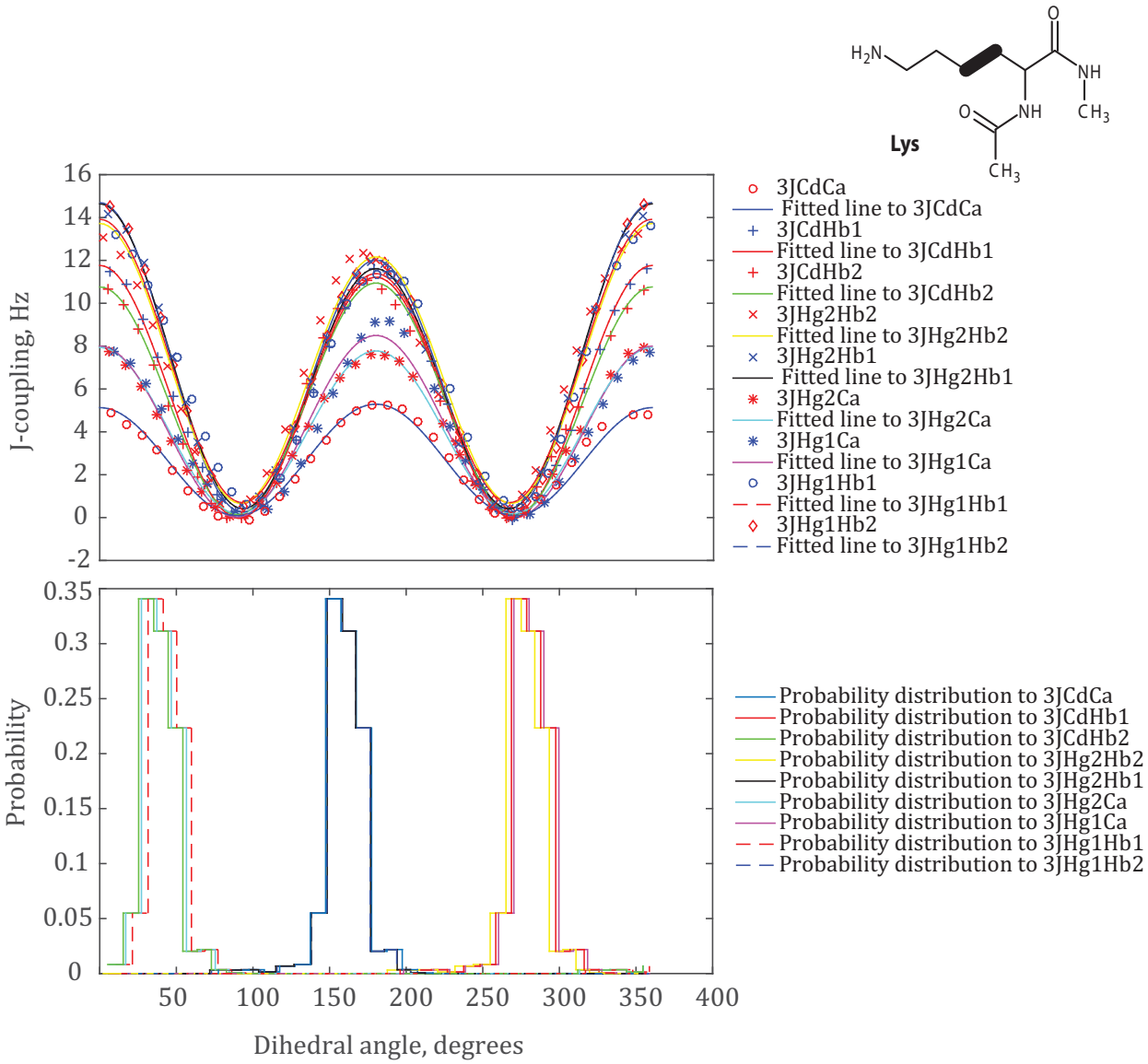

Figure 53 *LYS* χ_2 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles associated to the* χ_2 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

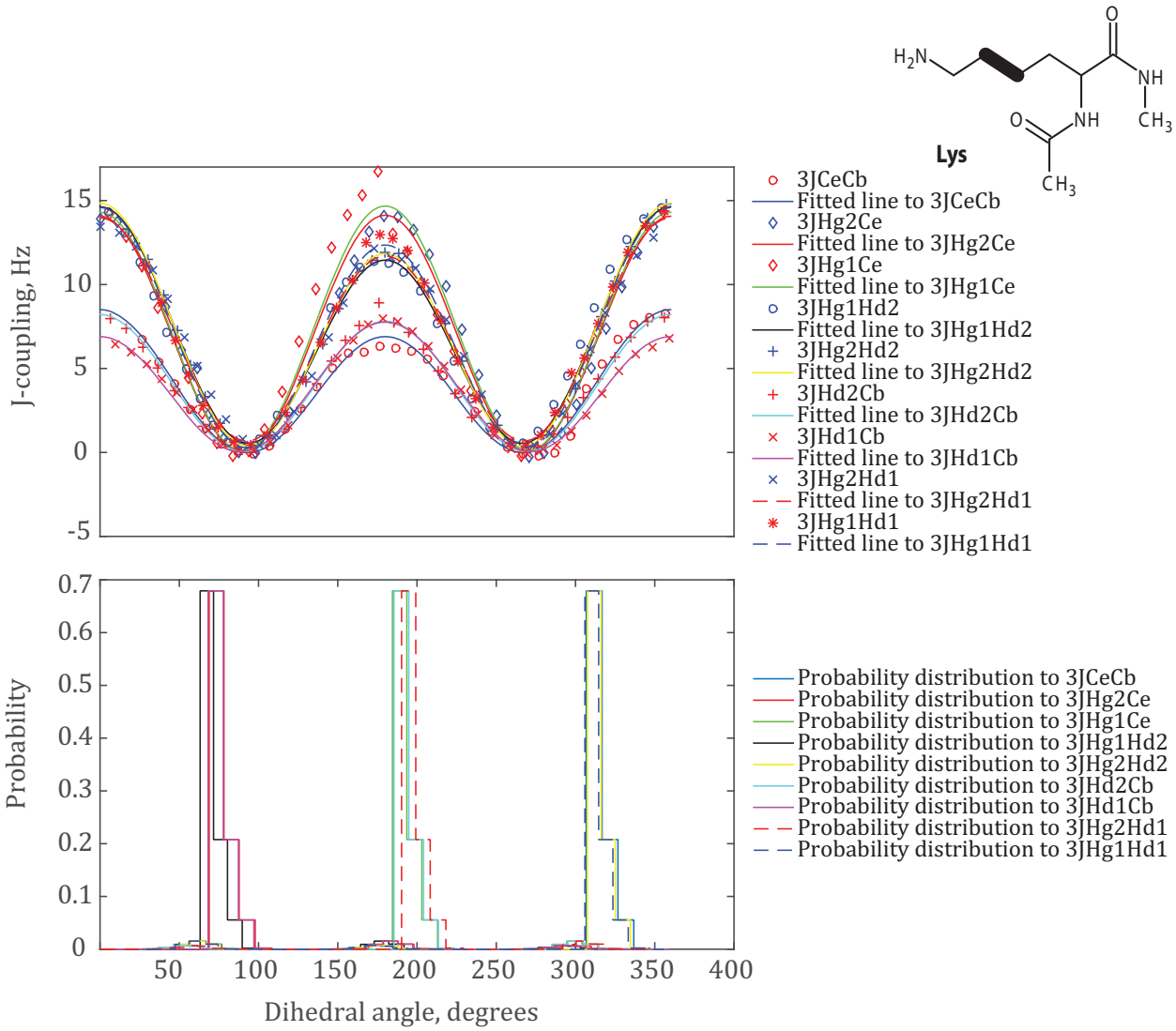

Figure 54 *LYS* χ_3 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles associated to the* χ_3 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

Three bond *J*-couplings that involved CE nuclei show an unusual character in which its Karplus coefficients have bigger values than the ones that arise from *J*-couplings between two proton atoms. This may be attributed to the extra amine present in the sidechain and the charge of the nitrogen atom of that amine group which is bonded to the CE atom. The two Karplus curves ($^3J_{CeHg2}$ and $^3J_{CeHg1}$) showed larger coefficients than the Karplus curves that involved vicinal couplings between proton atoms such as $^3J_{Hg1Hd1}$, $^3J_{Hg2Hd1}$, $^3J_{Hg1Hd2}$, and $^3J_{Hg2Hd2}$. In addition steric effect from the presence of large chain of methyl hydrocarbons in the amino acids sidechain might likely have a contribution to the jump in the Karplus curve of the $^3J_{CeHg1}$.

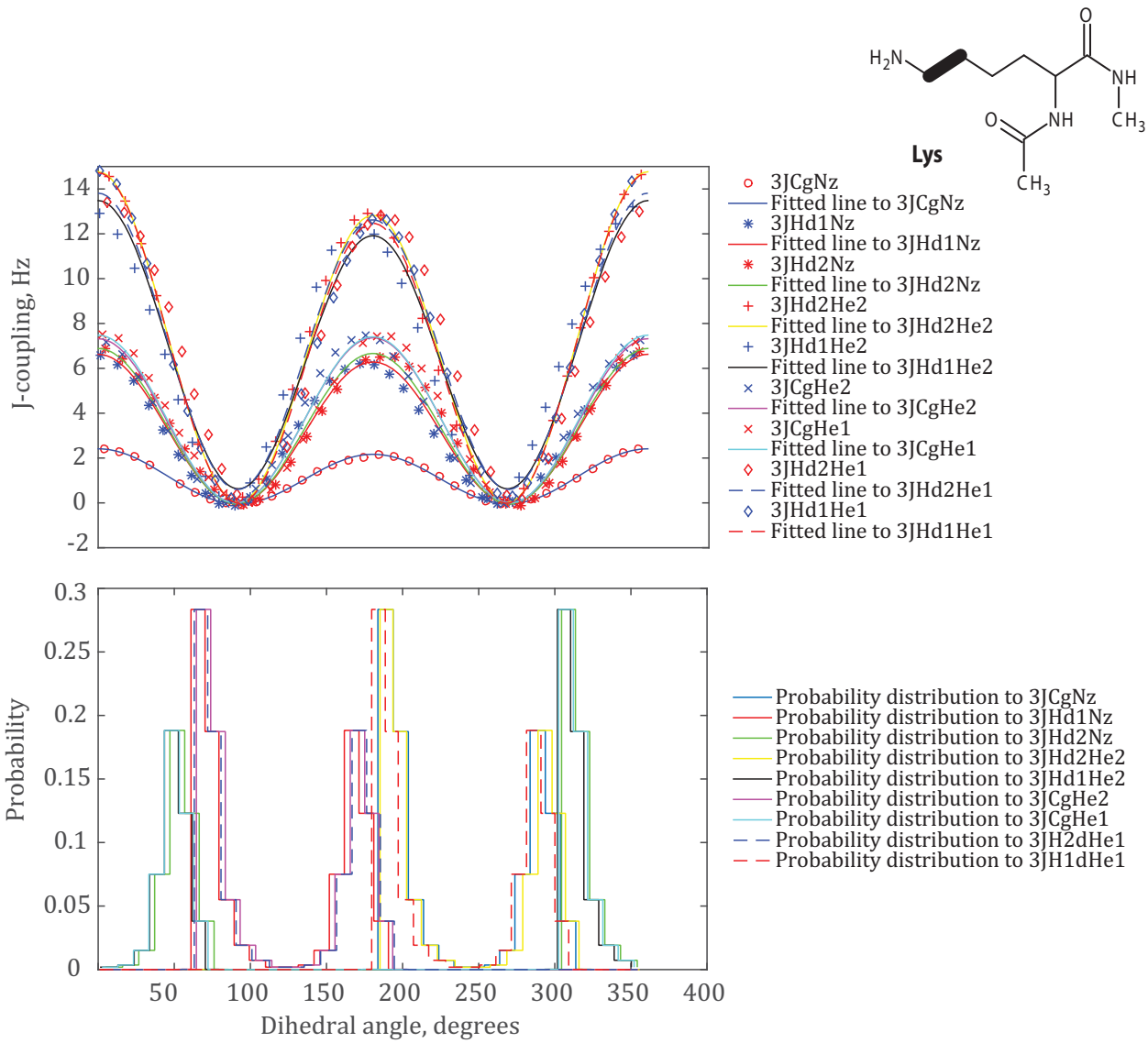

Figure 55 *LYS* χ_4 *side-chain dihedral angle scan. Top panel Karplus curves of the torsion angles associated to the* χ_4 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The Karplus curves of the dihedral angle appear to split into three groups depending on the magnitude of the amplitude of the spins involved in vicinal coupling. The presence of an extra amide and steric effect resulting from the larger hydrocarbon chain are likely to be responsible the staggered appearance of the energy probability distribution graphs.

3.9.14 Karplus coefficients, MET

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
N	C	CA	N	0.03	0.41	-0.10	0.18	0.00	0.25
HA	CA	C	N	1.17	0.35	-0.08	0.14	-0.20	0.22
CB	CA	C	N	0.20	0.32	0.00	0.18	0.03	0.20
HA	CA	CB	CG	9.24	0.34	1.02	0.14	0.18	0.19
N	CA	CB	CG	1.48	0.26	-0.12	0.14	-0.04	0.15
C	CA	CB	CG	3.05	0.36	-0.9	0.18	0.25	0.20
HA	CA	CB	HB1	12.39	0.34	1.77	0.14	0.69	0.19
N	CA	CB	HB1	3.39	0.37	-0.33	0.14	0.16	0.22
C	CA	CB	HB1	7.35	0.40	-0.16	0.19	-0.14	0.19
HA	CA	CB	HB2	12.68	0.30	1.09	0.15	0.53	0.17
N	CA	CB	HB2	3.44	0.47	-0.15	0.13	0.02	0.37
C	CA	CB	HB2	6.16	0.35	-0.89	0.13	0.42	0.21
H	N	CA	HA	11.70	0.31	-0.54	0.12	-0.11	0.21
C	N	CA	HA	5.65	0.31	-1.67	0.13	0.53	0.23
H	N	CA	CB	5.83	0.34	-0.33	0.20	-0.40	0.14
C	N	CA	CB	2.13	0.27	-0.92	0.14	0.13	0.15
H	N	CA	C	6.20	0.49	-1.32	0.29	-0.29	0.18
C	N	CA	C	2.79	0.29	-1.00	0.15	0.07	0.15
CA	CB	CG	SD	3.63	0.49	0.64	0.41	0.00	0.30
HB1	CB	CG	SD	7.08	0.45	-0.15	0.32	-0.26	0.29
HB2	CB	CG	SD	7.15	0.35	-0.09	0.17	-0.03	0.18
CA	CB	CG	HG1	8.87	0.34	-0.72	0.18	-0.09	0.19
HB1	CB	CG	HG1	13.25	0.32	0.89	0.16	0.51	0.20
HB2	CB	CG	HG1	13.02	0.29	0.58	0.15	0.84	0.15
CA	CB	CG	HG2	8.17	0.38	-0.24	0.18	0.18	0.21
HB1	CB	CG	HG2	12.46	0.34	0.48	0.12	0.81	0.21
HB2	CB	CG	HG2	13.91	0.25	1.00	0.13	0.42	0.16
CB	CG	SD	CE	5.59	0.37	1.33	0.26	-0.05	0.14
HG1	CG	SD	CE	8.52	0.38	0.52	0.15	-0.03	0.20

HG2	CG	SD	CE	8.50	0.37	0.45	0.13	0.03	0.22

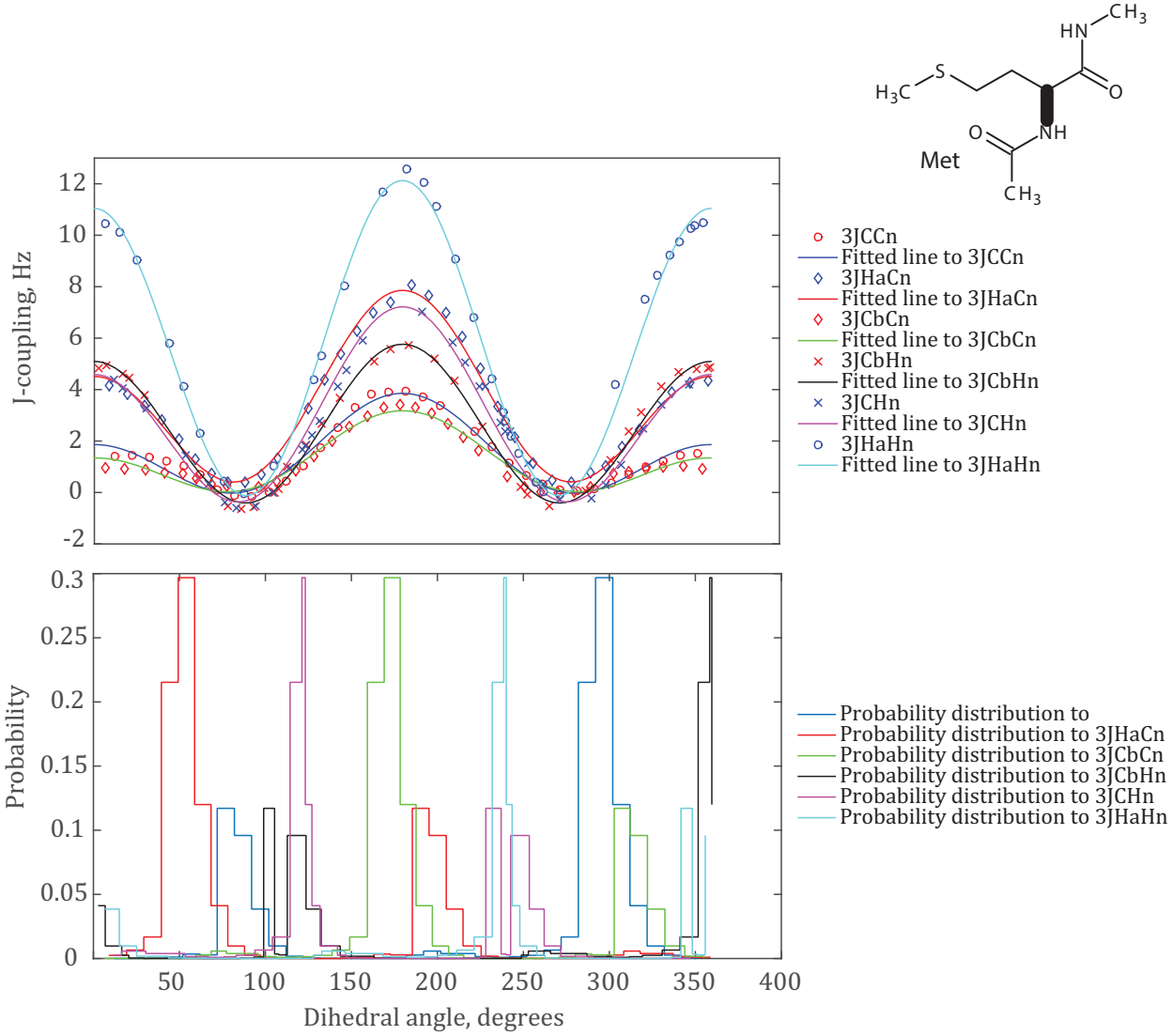

Figure 56 *MET ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles related to the ϕ dihedral angle scan.*

Though there exists a hydrogen bond that does not involve nuclei that participated in vicinal *J*-couplings associated to the scan of the ϕ dihedral angle, the Karplus curves looks fit the data nicely in a Karplus like pattern. Thus, hydrogen bond coupled with steric effect from longer chain of the sidechain may be likely to affect the appearance of the Boltzmann probability distribution graphs.

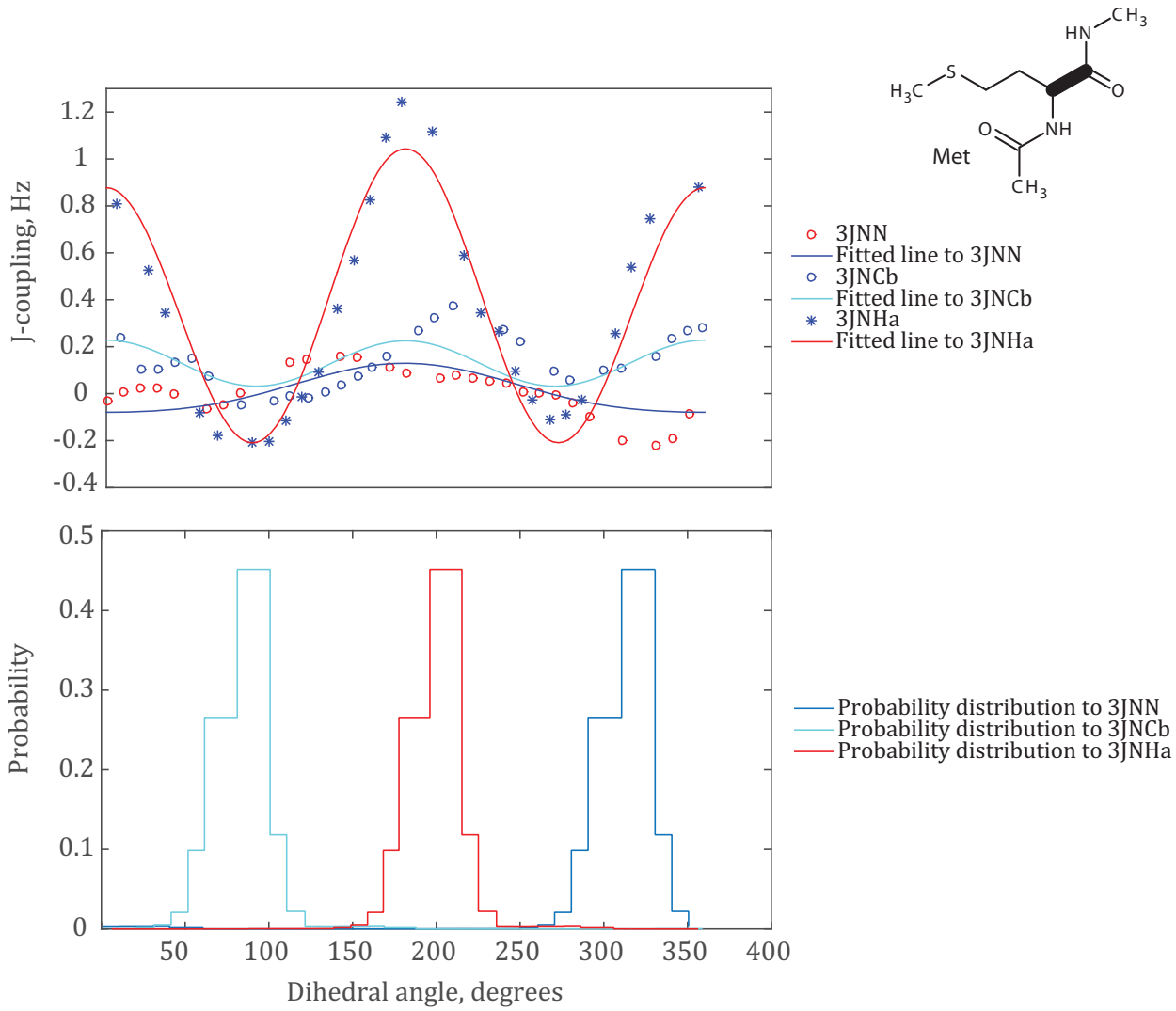

Figure 57 *MET ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ψ dihedral angle scan respectively.*

$^{3}J_{NN}$ and $^{3}J_{NCb}$ vicinal couplings have small amplitudes, making the fitting of the Karplus curve difficult. The other torsion angle Karplus curve *i.e.* $^{3}J_{NHa}$ shows a break in the fitting curve, which is likely down to hydrogen bond making and breaking as well as steric effect of the longer chains of methyl groups of the amino acid sidechain.

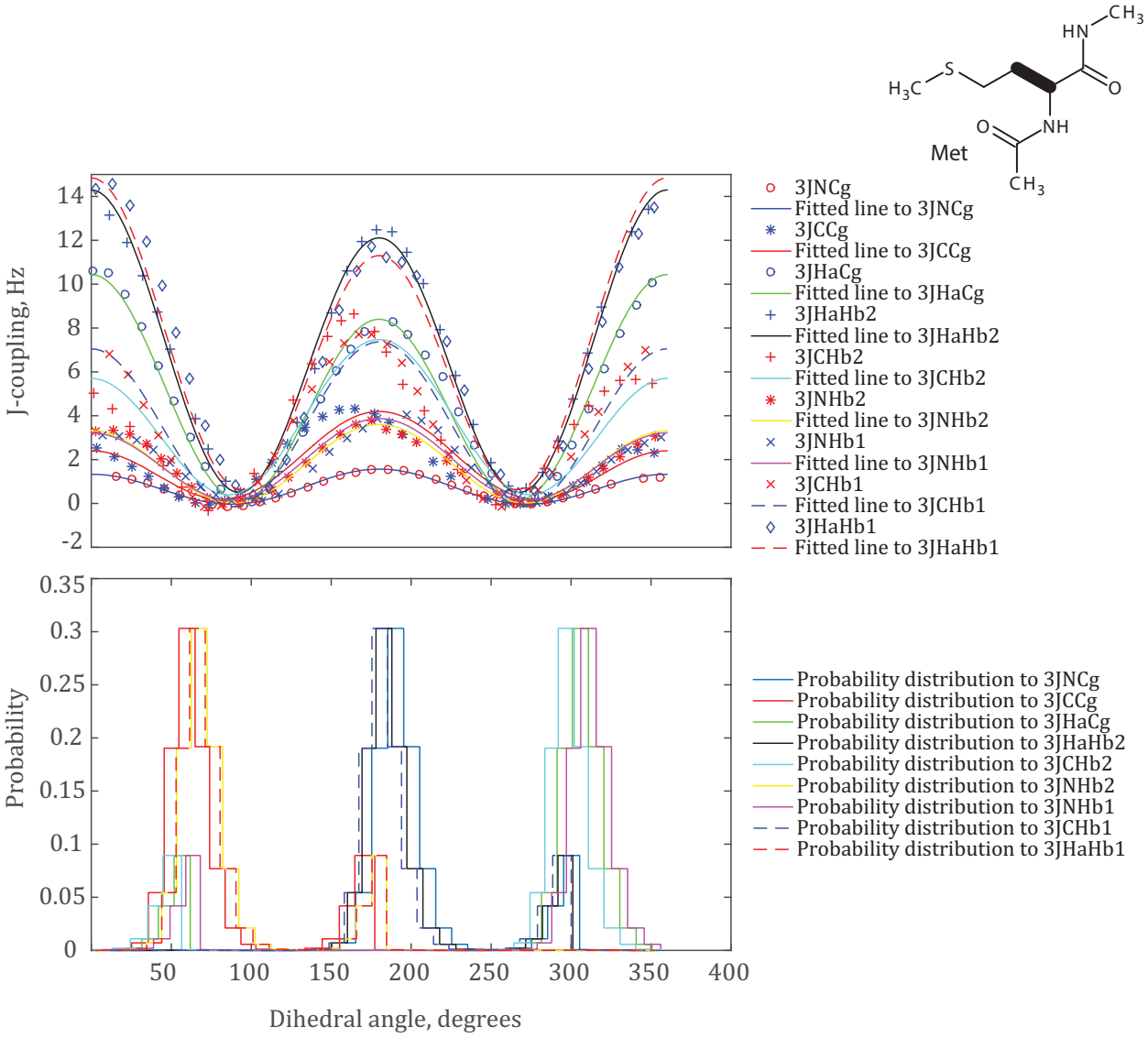

Figure 58 *MET* χ_1 *side-chain dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the* χ_1 *dihedral angle scan.*

All the Karplus curves in **Figure 58** (above) nicely fitted to the vicinal coupling data. This shows the absence of hydrogen interactions in the scan of the χ_1 dihedral angle.

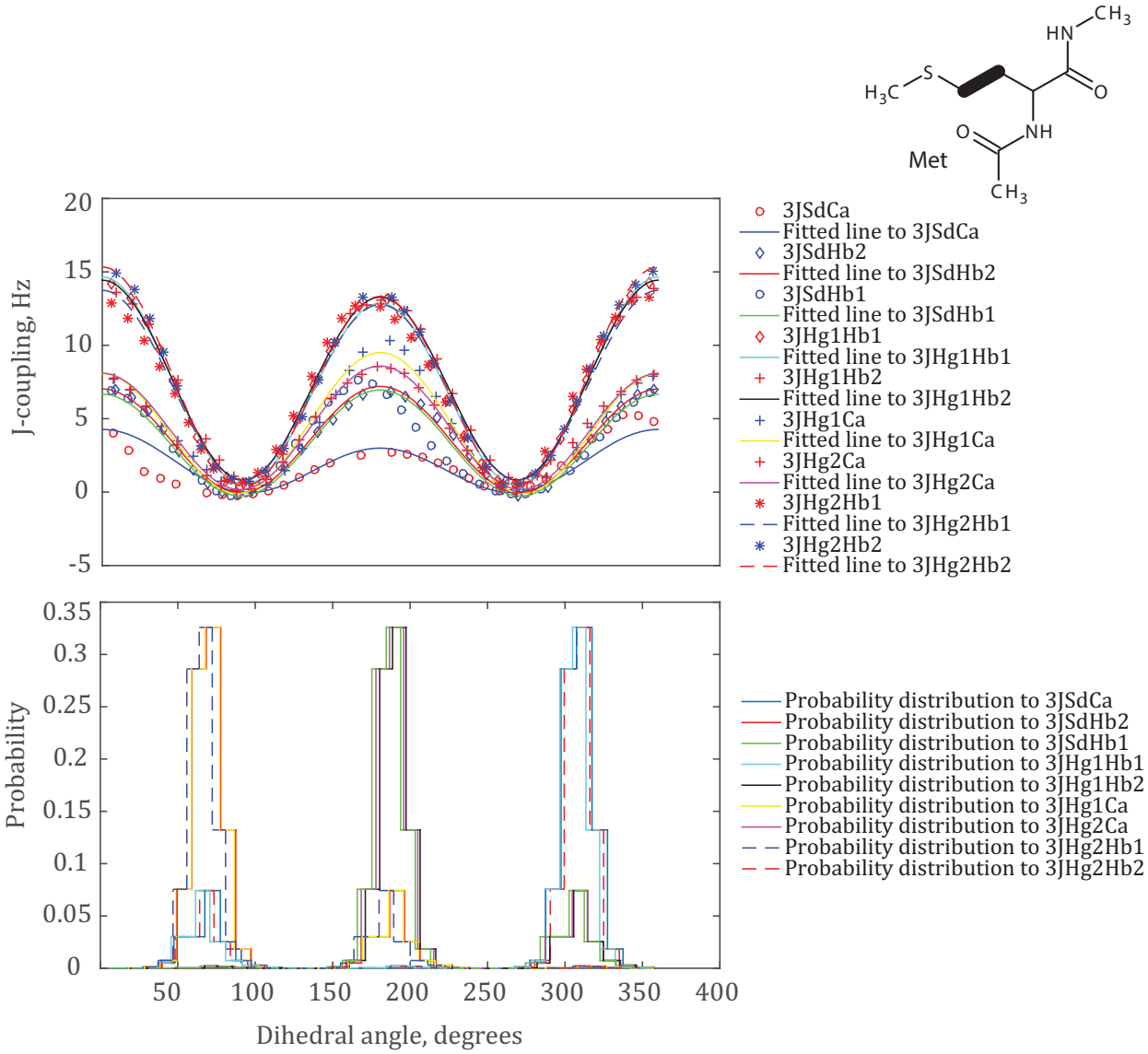

Figure 59 *MET* χ_2 *side-chain dihedral angle scan. Top panel Karplus curve of the torsion angles associated to the* χ_2 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs of the respective torsion angles.*

The scan of the χ_2 dihedral angle scan brings the sidechain sulphur atom closer the backbone amide hydrogen to form hydrogen bond. Sulphur is known to form noncovalent interactions such as hydrogen bond [111].

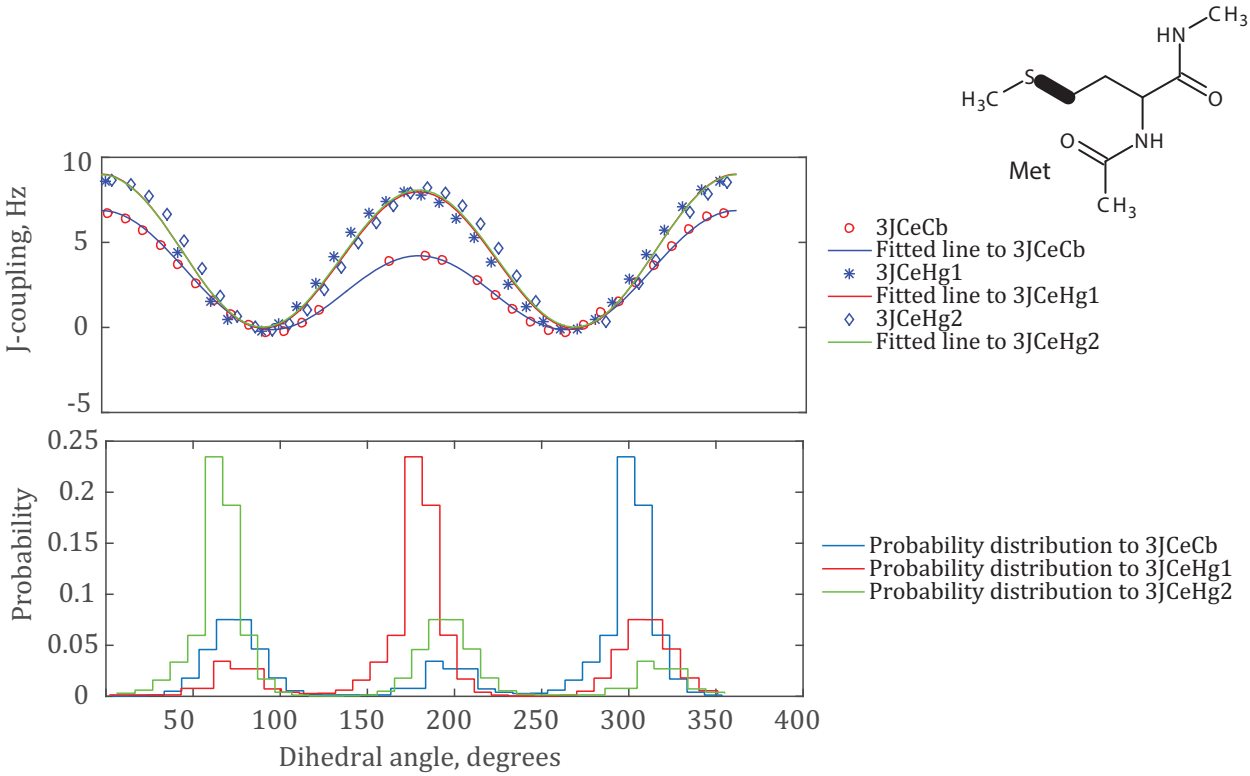

Figure 60 *MET* χ_3 *side-chain dihedral angle scan. Top panel Karplus curves of the torsion angles related to the* χ_3 *dihedral angle scan and bottom panel Boltzmann probability distribution for the respective torsion angles.*

3.9.15 Karplus coefficients, PHE

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	-0.03	0.53	-0.04	0.23	0.01	0.33
HA	CA	C	N	1.13	0.49	0.18	0.16	-0.22	0.35
CB	CA	C	N	0.08	0.45	-0.14	0.27	0.08	0.17
HA	CA	CB	CG	6.37	0.36	-0.43	0.14	0.31	0.20
N	CA	CB	CG	1.07	0.29	-0.23	0.12	0.02	0.18
C	CA	CB	CG	3.21	0.36	-0.46	0.13	0.06	0.22
HA	CA	CB	HB1	11.31	0.35	1.58	0.13	0.67	0.20
N	CA	CB	HB1	2.78	0.37	0.00	0.13	0.09	0.20
C	CA	CB	HB1	6.93	0.30	-0.26	0.15	0.19	0.17
HA	CA	CB	HB2	11.52	0.32	1.77	0.13	0.57	0.19

N	CA	CB	HB2	2.96	0.39	-0.11	0.14	0.03	0.27
C	CA	CB	HB2	6.96	0.35	-0.40	0.13	0.14	0.20
H	N	CA	HA	11.93	0.37	0.64	0.11	0.02	0.30
C	N	CA	HA	6.67	0.32	-2.19	0.14	-0.07	0.17
H	N	CA	CB	7.03	0.49	-0.40	0.28	-0.63	0.25
C	N	CA	CB	2.77	0.30	-0.74	0.18	-0.20	0.18
H	N	CA	C	7.27	0.44	-1.49	0.28	-0.31	0.18
C	N	CA	C	3.36	0.25	-1.01	0.12	-0.22	0.18

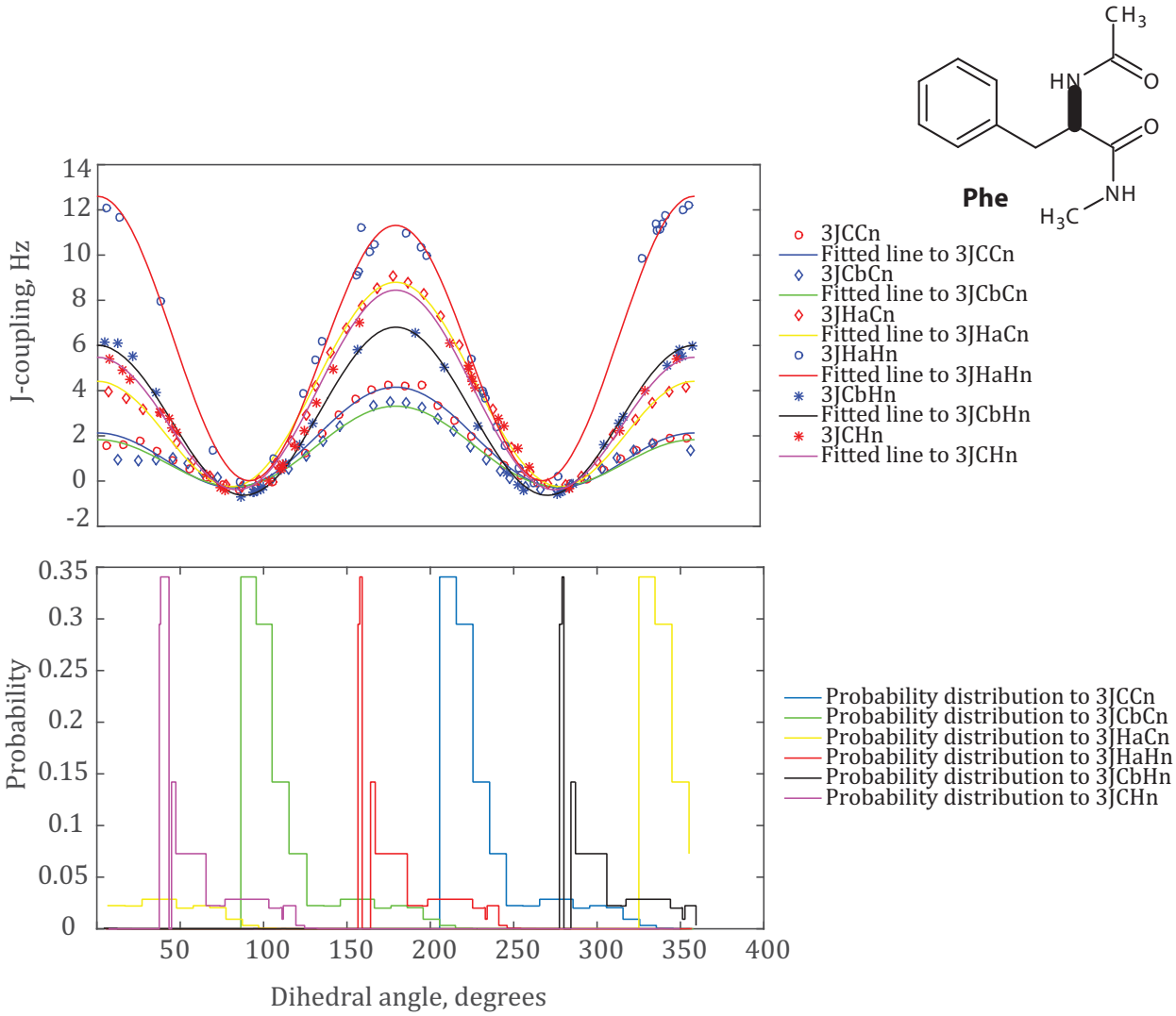

Figure 61 *PHE ϕ backbone dihedral angle scan. Top and bottom are Karplus curves and Boltzmann probability distribution graphs for torsion angles associated to the ϕ dihedral angle scan respectively.*

Karplus curves that involve backbone amide proton as one of the *J*-coupled nuclei, are broken. This is due to fact that the atom is involved in a hydrogen bond with oxygen atom attached to the carbon bonded to the C_α. Thus the scan of the ϕ dihedral angle breaks that bond hence that jumps and breaks are observed in the torsion angles that are associated to the above mentioned nuclei.

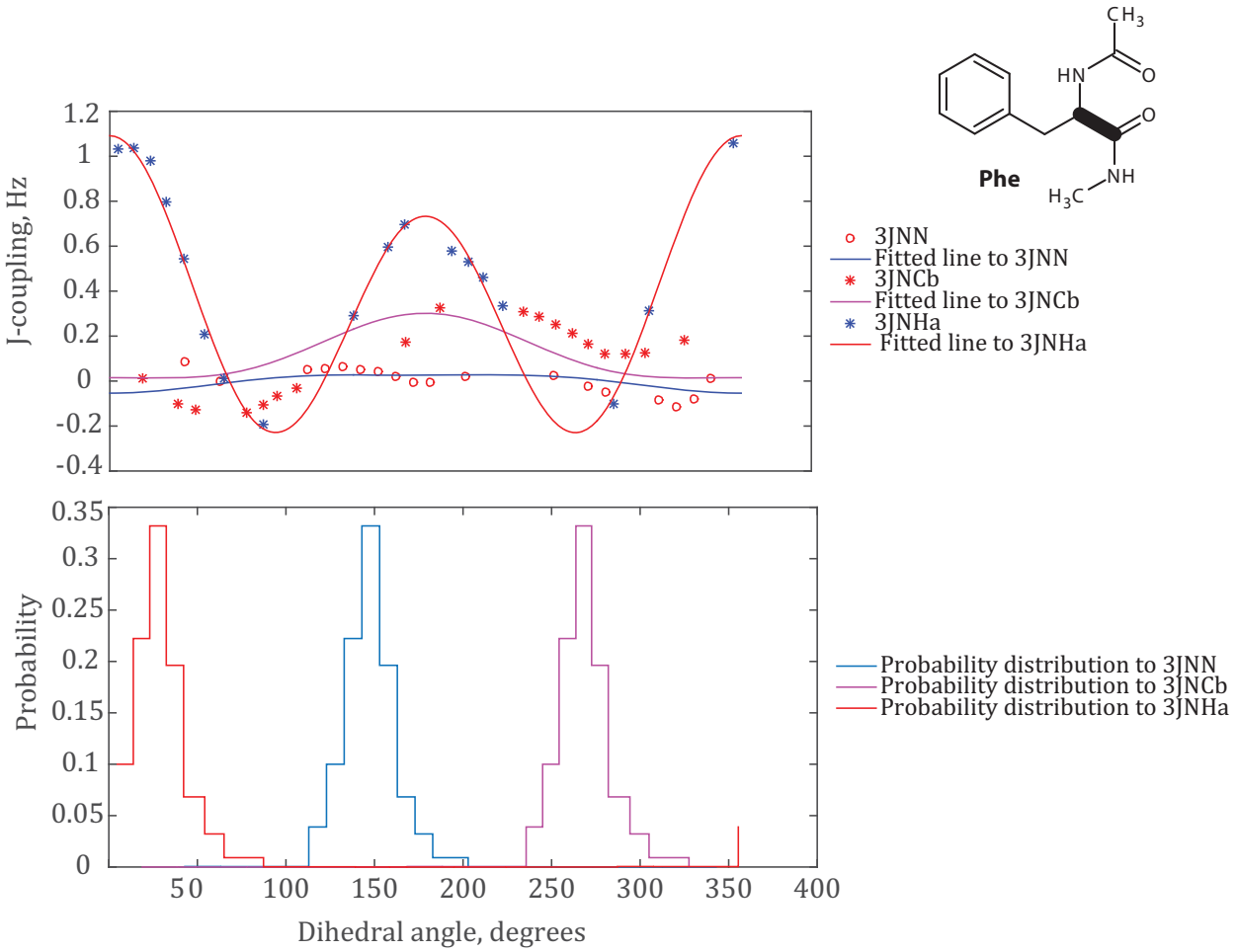

Figure 62 *PHE ψ backbone dihedral angle scan. Top and bottom are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ψ dihedral angle scan respectively.*

As mentioned above the scanning of the Karplus curve to the small amplitude *J*-coupling data of the $^{3}J_{NN}$ and $^{3}J_{NCb}$ torsion angles appears to produce random noise. The jumps in the other Karplus curve are likely due to the effect of hydrogen bond making and breaking together with the steric effect.

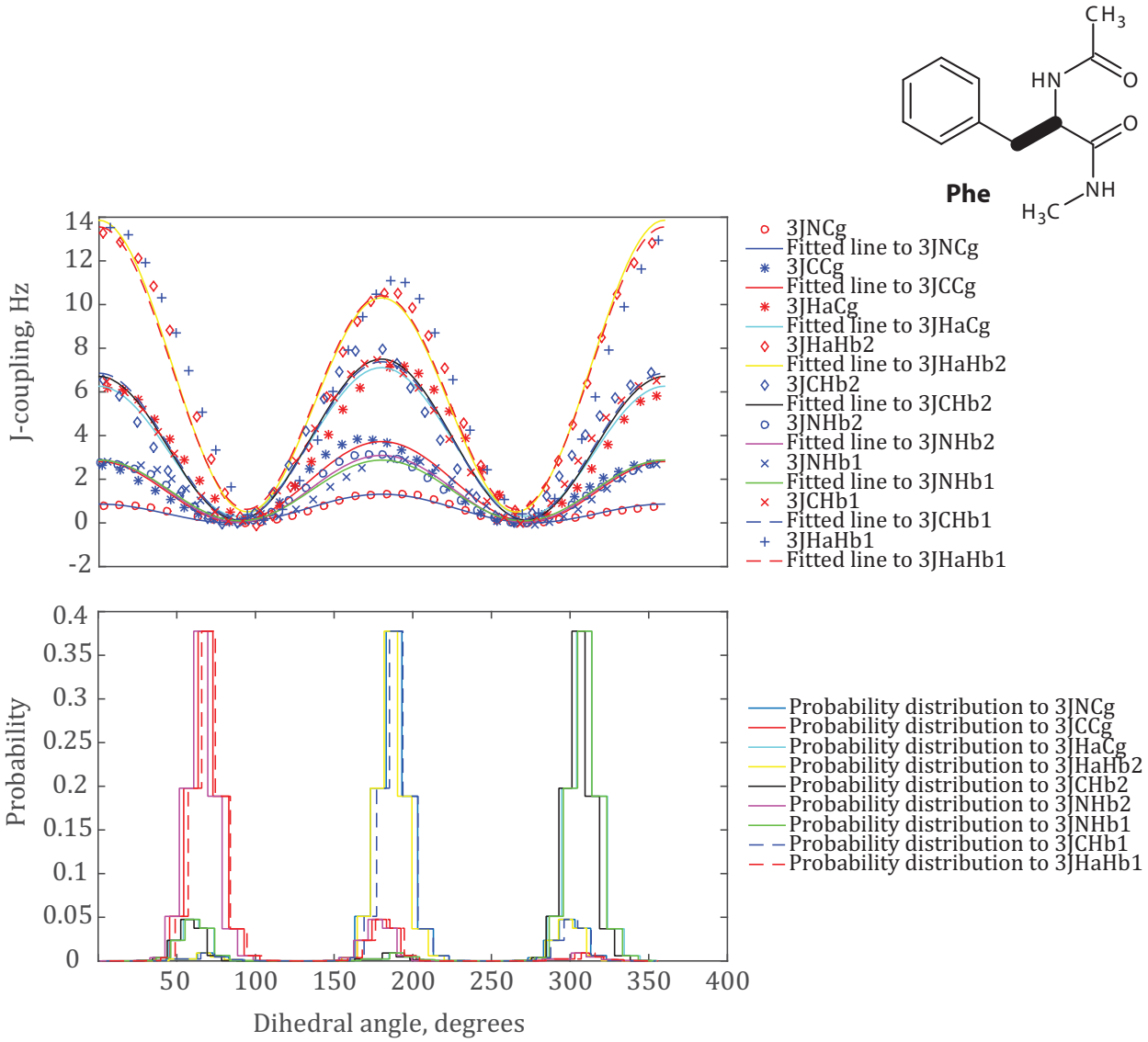

Figure 63 *PHE* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles related to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

3.9.16 Karplus coefficients, PRO

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	B, Hz	C, Hz
HA	CA	N	CD	0.00	0.00	0.00
CA	N	CD	HD3	0.00	0.00	2.50
HG2	CG	CD	N	0.00	0.00	0.00

C	N	CD	CG	0.00	0.00	2.50
C	CA	N	CD	0.00	0.00	1.10
C	N	CD	HD2	0.00	0.00	1.00
C	N	CD	HD3	0.00	0.00	1.00
CB	CA	N	CD	0.00	0.00	0.00
N	CD	CG	CB	0.00	0.00	1.10
CA	N	CD	HD2	0.00	0.00	2.50
CA	N	CD	CG	0.00	0.00	0.00
HG3	CG	CD	N	0.00	0.00	0.00

This amino acid deemed to have no parts that are flexible and thus not considered for a relaxed dihedral angle scan. However, proline does isomerise between cis and trans forms and plays a central role in protein folding and function.

3.9.17 Karplus coefficients, SER

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	**Atom 2**	**Atom 3**	**Atom 4**	**A, Hz**	**A_{RMSD}**	**B, Hz**	**B_{RMSD}**	**C, Hz**	**C_{RMSD}**
N	C	CA	N	0.02	0.33	-0.07	0.13	0.00	0.21
HA	CA	C	N	1.03	0.35	-0.26	0.14	-0.15	0.20
CB	CA	C	N	0.12	0.34	-0.30	0.13	0.01	0.20
HA	CA	CB	HB1	10.89	0.44	1.72	0.36	0.51	0.18
N	CA	CB	HB1	3.34	0.47	0.29	0.32	0.03	0.29
C	CA	CB	HB1	5.99	0.35	-0.64	0.13	0.43	0.20
HA	CA	CB	HB2	9.32	0.37	1.72	0.14	1.00	0.20
N	CA	CB	HB2	2.75	0.37	0.42	0.13	0.24	0.22
C	CA	CB	HB2	6.48	0.34	-0.21	0.12	0.05	0.21
H	N	CA	HA	10.81	0.45	0.31	0.14	-0.14	0.29
C	N	CA	HA	0.33	0.33	-1.20	0.20	2.76	0.13
H	N	CA	CB	7.00	0.39	-1.43	0.17	-0.48	0.22
C	N	CA	CB	-0.27	0.30	0.43	0.14	1.54	0.17
H	N	CA	C	7.02	0.35	-1.77	0.16	-0.44	0.21

C	N	CA	C	-0.76	0.34	0.65	0.17	1.76	0.23

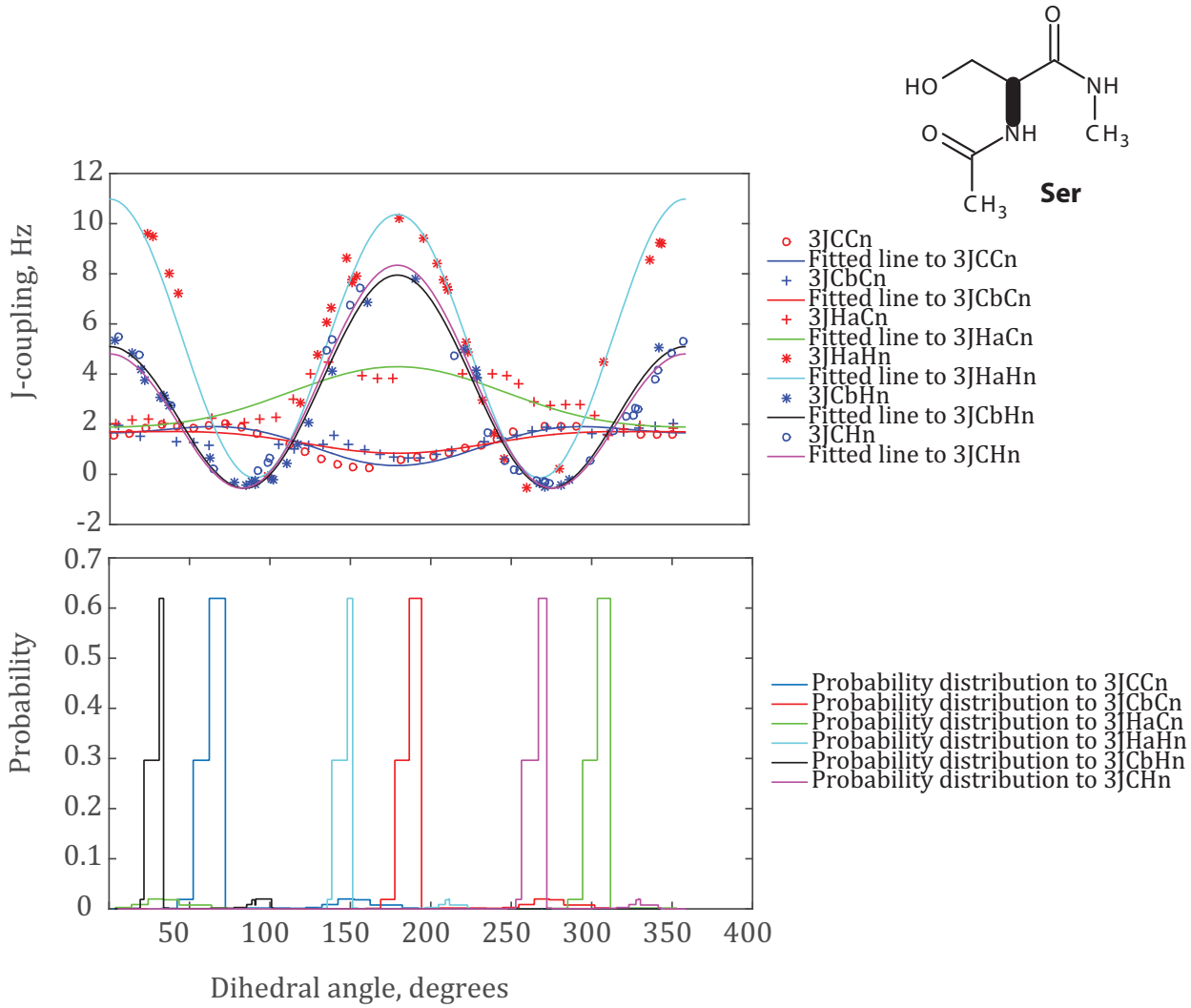

Figure 64 *SER ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ϕ dihedral angle scan respectively.*

The presence of hydroxyl ion attached to the carbon at the β position results in hydrogen bond formation. Thus, the rotation of the dihedral angle breaks the bond and is responsible for the scatter of the data. The fitting of Karplus curve to vicinal coupling data however looks fine. The hydrogen bond effect appears to affect all Karplus curves and the probability distribution graphs of the torsion angles related to the ϕ dihedral angle.

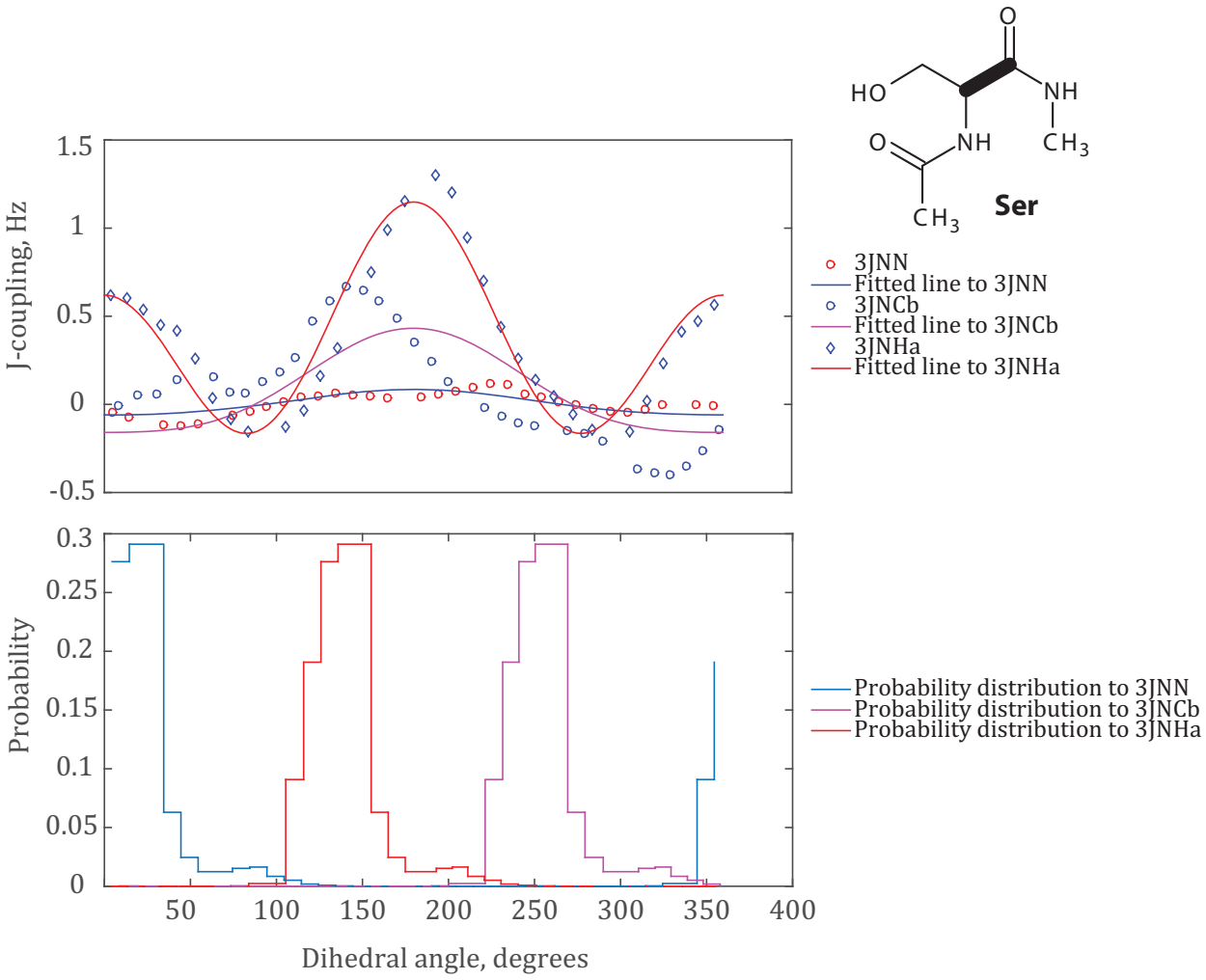

Figure 65 *SER ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the respective torsion angles respectively.*

Although the fitting of the $^{3}J_{NN}$ and $^{3}J_{NCb}$ Karplus curves appears random and noisy due to small amplitude of the coupling values, hydrogen bond effect is likely to be responsible for the broken Karplus curve of the $^{3}J_{NHa}$ torsion angle and the probability distribution graphs.

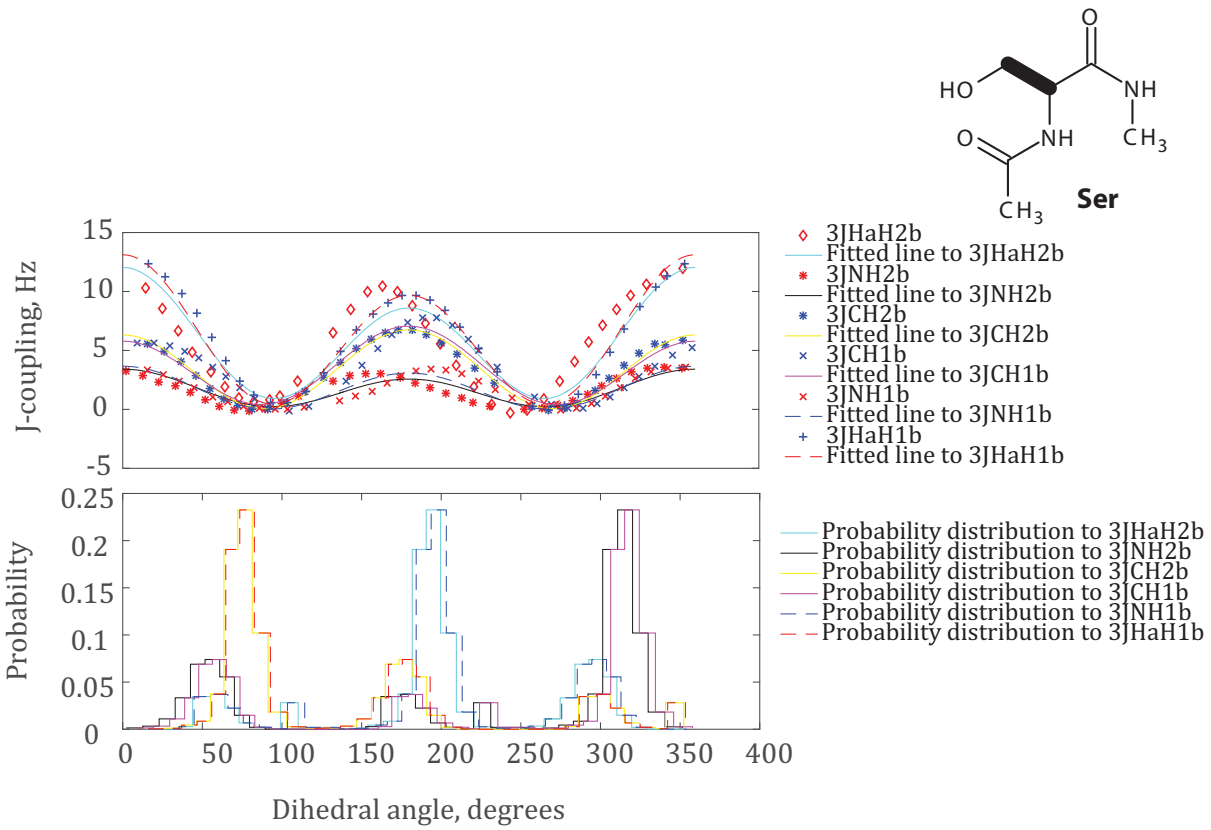

Figure 66 *SER* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles associated to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

All the Karplus curves are affected by hydrogen interaction as the dihedral angle scanned. The Karplus curves have either jumps or are broken as depicted in **Figure 66** (above).

3.9.18 Karplus coefficients, THR

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	-0.01	0.66	-0.12	0.36	0.03	0.45
HA	CA	C	N	0.97	0.50	-0.07	0.25	-0.12	0.34
CB	CA	C	N	-0.04	0.82	-0.29	0.73	0.01	0.23
HA	CA	CB	CG1	6.33	0.41	0.65	0.13	0.53	0.21
N	CA	CB	CG1	1.14	0.38	-0.10	0.21	0.01	0.22

N	CA	CB	CG1	2.01		0.21		-0.13 [110]	
C	CA	CB	CG1	3.02	0.29	-0.42	0.14	-0.15	0.16
C	CA	CB	CG1	2.76		-0.67		0.19 [110]	
HA	CA	CB	HB	11.23	0.33	2.26	0.14	0.41	0.23
N	CA	CB	HB	2.71	0.46	0.04	0.20	0.09	0.24
C	CA	CB	HB	5.29	0.36	-0.92	0.12	0.30	0.22
H	N	CA	HA	11.77	0.36	-0.14	0.10	-0.00	0.30
C	N	CA	HA	5.89	0.26	-1.54	0.13	0.41	0.15
H	N	CA	CB	6.45	0.36	-1.08	0.17	-0.40	0.15
C	N	CA	CB	1.88	0.27	-0.88	0.15	0.11	0.13
H	N	CA	C	6.41	0.39	-1.09	0.19	-0.45	0.15
C	N	CA	C	2.54	0.28	-0.97	0.14	0.09	0.16

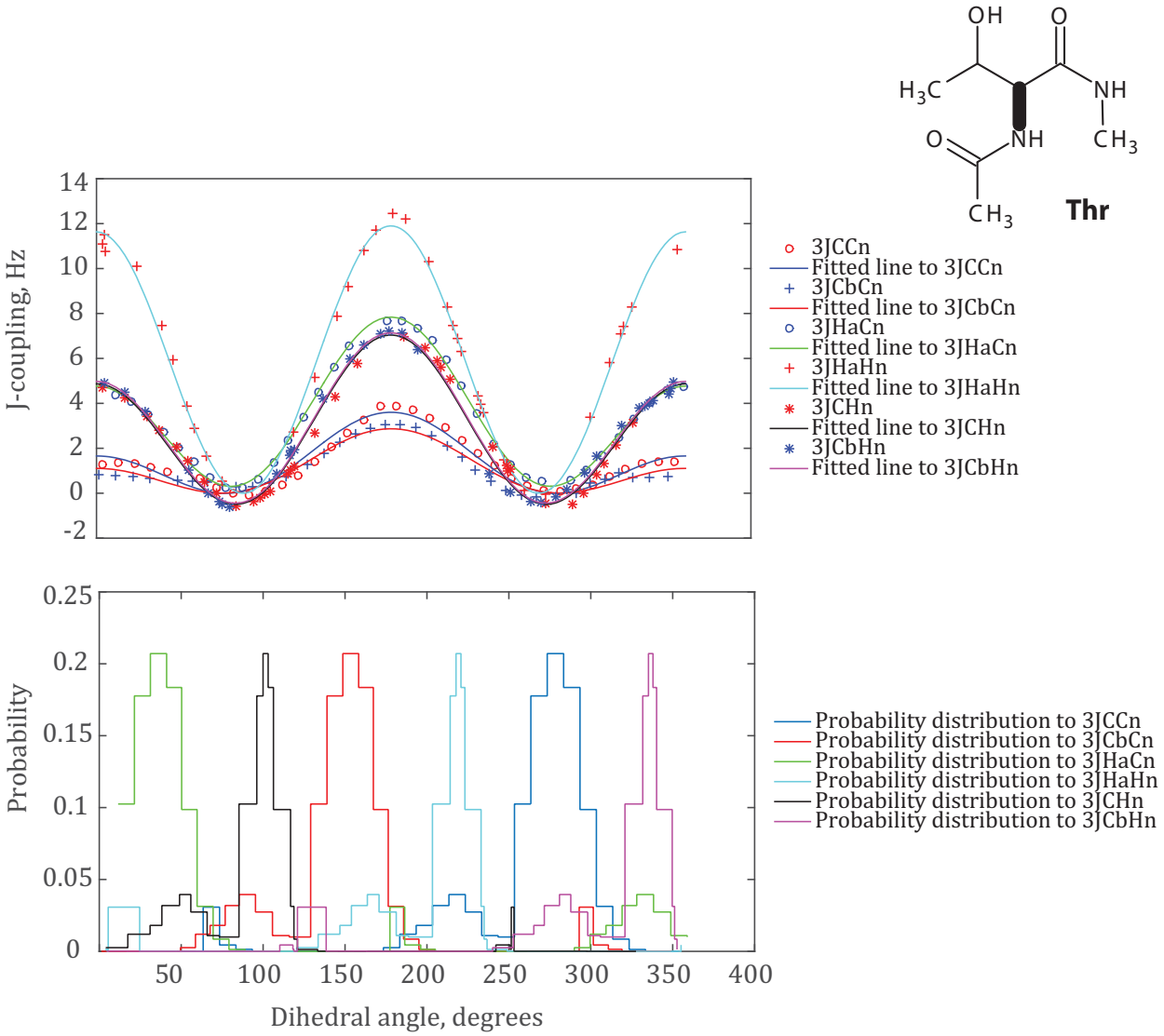

Figure 67 *THR ϕ backbone dihedral angle scan. Top panel Karplus curves for torsion angles associated to the ϕ dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

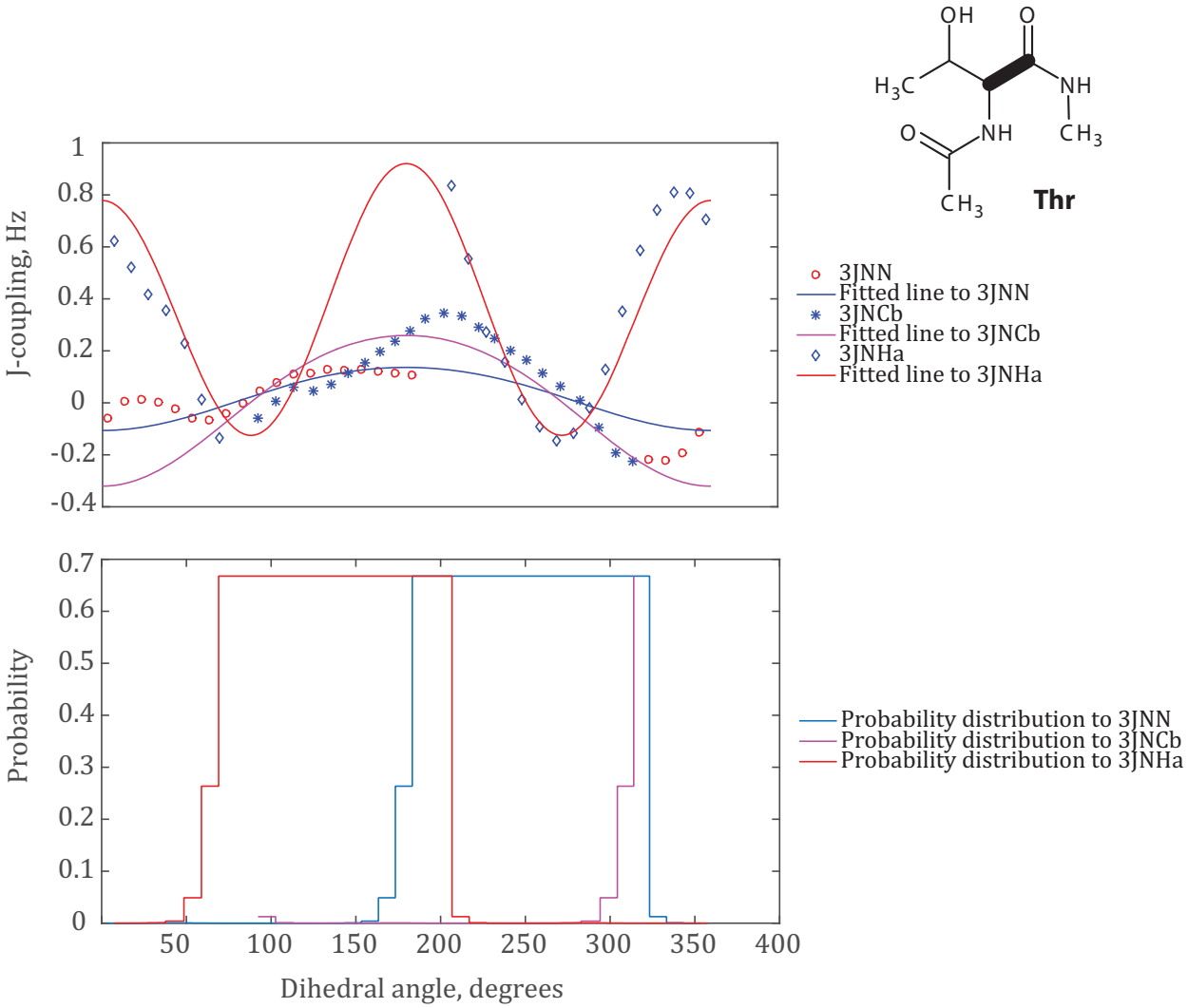

Figure 68 *THR ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ψ dihedral angle scan respectively.*

As normal, the $^{3}J_{NN}$ and $^{3}J_{NCb}$ *J*-coupling interaction amplitudes is small, hence the scan of the Karplus curves resembles random noise. However the presence of multiple hydrogen bond making and breaking is likely to be responsible for the breaking of the curve and broad appearance of the distribution histograms.

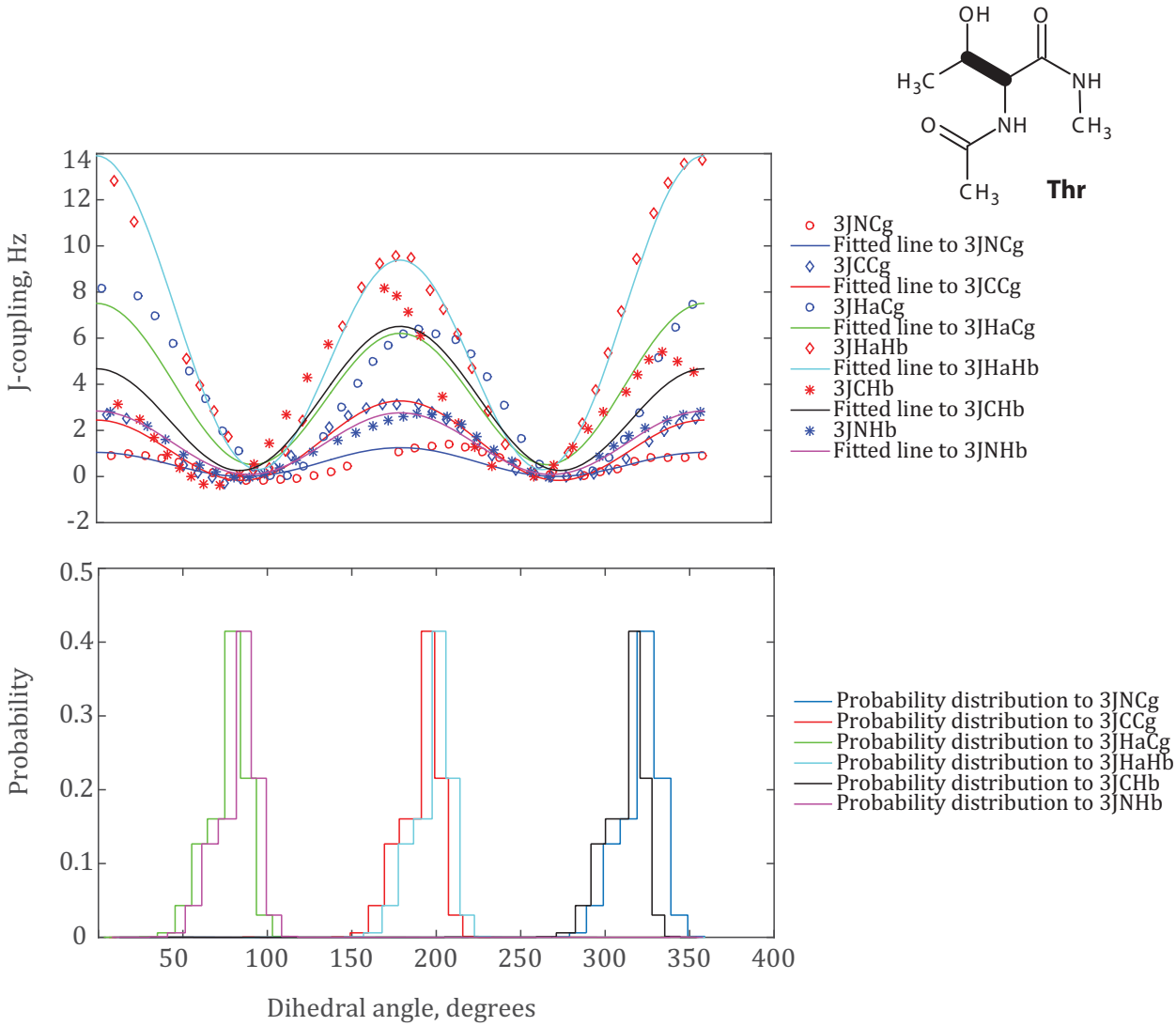

Figure 69 *THR* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles related to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

Though the atoms involved in the three bond separated *J*-couplings do not have any hydrogen bond, there are hydrogen bonds that involves an oxygen atom attached to the C_β and backbone amide proton. Thus the scan of the χ_1 dihedral angle breaks the hydrogen bond and causes the couplings of the spins to have jumps, hence the breaks of the Karplus curves. The effect is also likely to be responsible for the under and over prediction pattern evident in some of the curves.

3.9.19 Karplus coefficients, TRP

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	0.01	0.33	-0.15	0.13	0.02	0.20
HA	CA	C	N	1.30	0.30	-0.06	0.12	-0.21	0.18
CB	CA	C	N	0.22	0.29	-0.01	0.16	0.05	0.17
H	N	CA	HA	11.75	0.30	-0.58	0.13	-0.14	0.21
C	N	CA	HA	5.69	0.34	-1.70	0.15	0.52	0.24
H	N	CA	CB	5.93	0.30	-0.36	0.18	-0.39	0.12
C	N	CA	CB	2.26	0.28	-0.86	0.14	0.14	0.17
H	N	CA	C	6.14	0.56	-1.38	0.30	-0.21	0.20
C	N	CA	C	2.80	0.29	-0.98	0.15	0.04	0.15
HA	CA	CB	CG	7.71	0.36	0.06	0.20	0.26	0.21
N	CA	CB	CG	1.31	0.28	-0.29	0.15	0.01	0.15
C	CA	CB	CG	2.89	0.33	-0.80	0.15	0.21	0.19
HA	CA	CB	HB1	12.62	0.28	1.09	0.12	0.73	0.18
N	CA	CB	HB1	3.47	0.51	-0.08	0.12	-0.01	0.39
C	CA	CB	HB1	5.95	0.34	-0.73	0.13	0.49	0.19
HA	CA	CB	HB2	12.45	0.36	1.76	0.13	0.61	0.19
N	CA	CB	HB2	3.13	0.37	-0.32	0.14	0.15	0.22
C	CA	CB	HB2	7.05	0.35	-0.20	0.15	-0.14	0.18

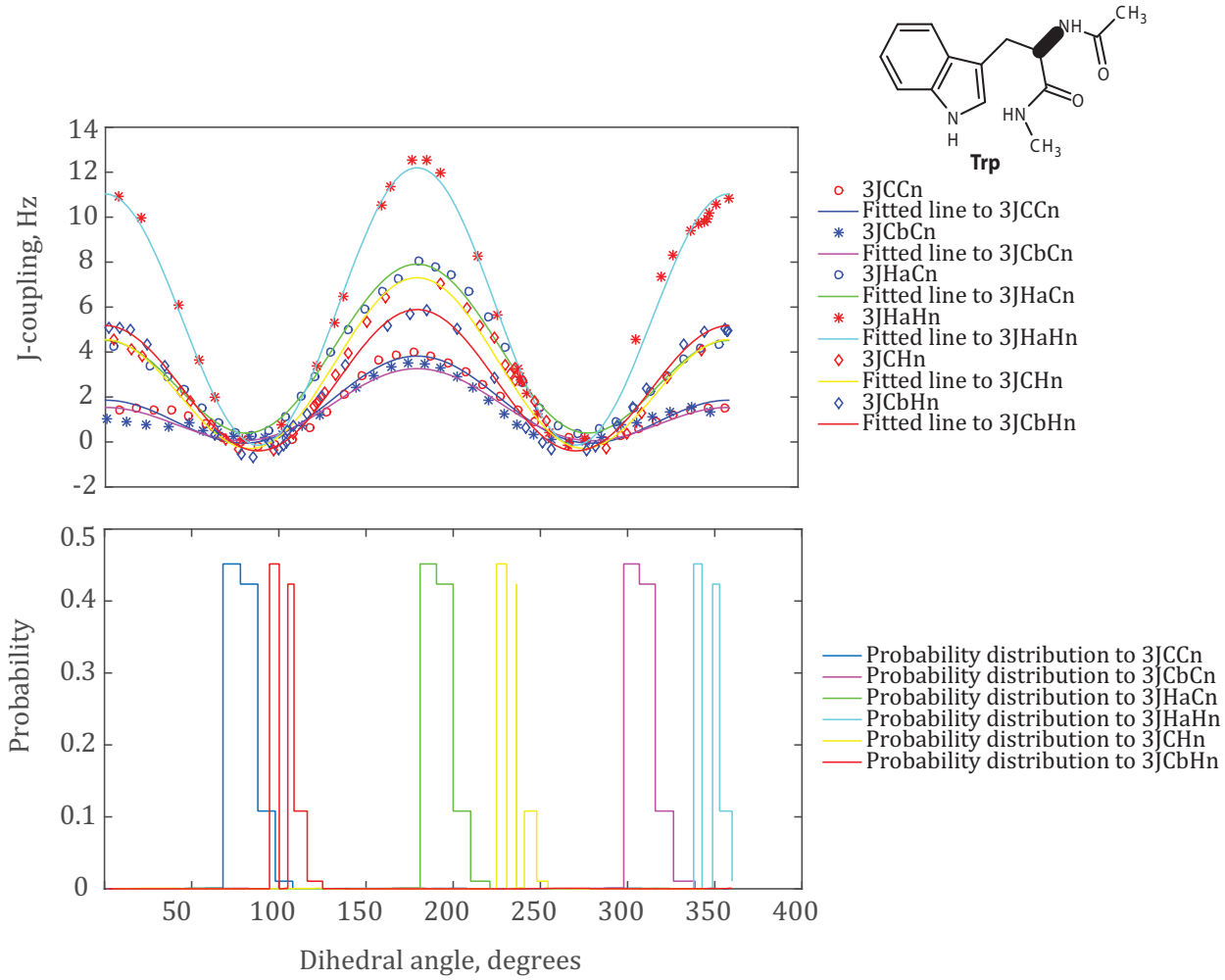

Figure 70 *TRP ϕ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ϕ dihedral angle scan respectively.*

The three bond *J*-coupled nuclei associated to the ϕ dihedral angle do not engage in hydrogen bond, however the jumps in some of the Karplus curves and the thin appearance of the probability distribution graphs might be due to steric effect that arise from the presence of large sidechain of the tryptophan amino acid.

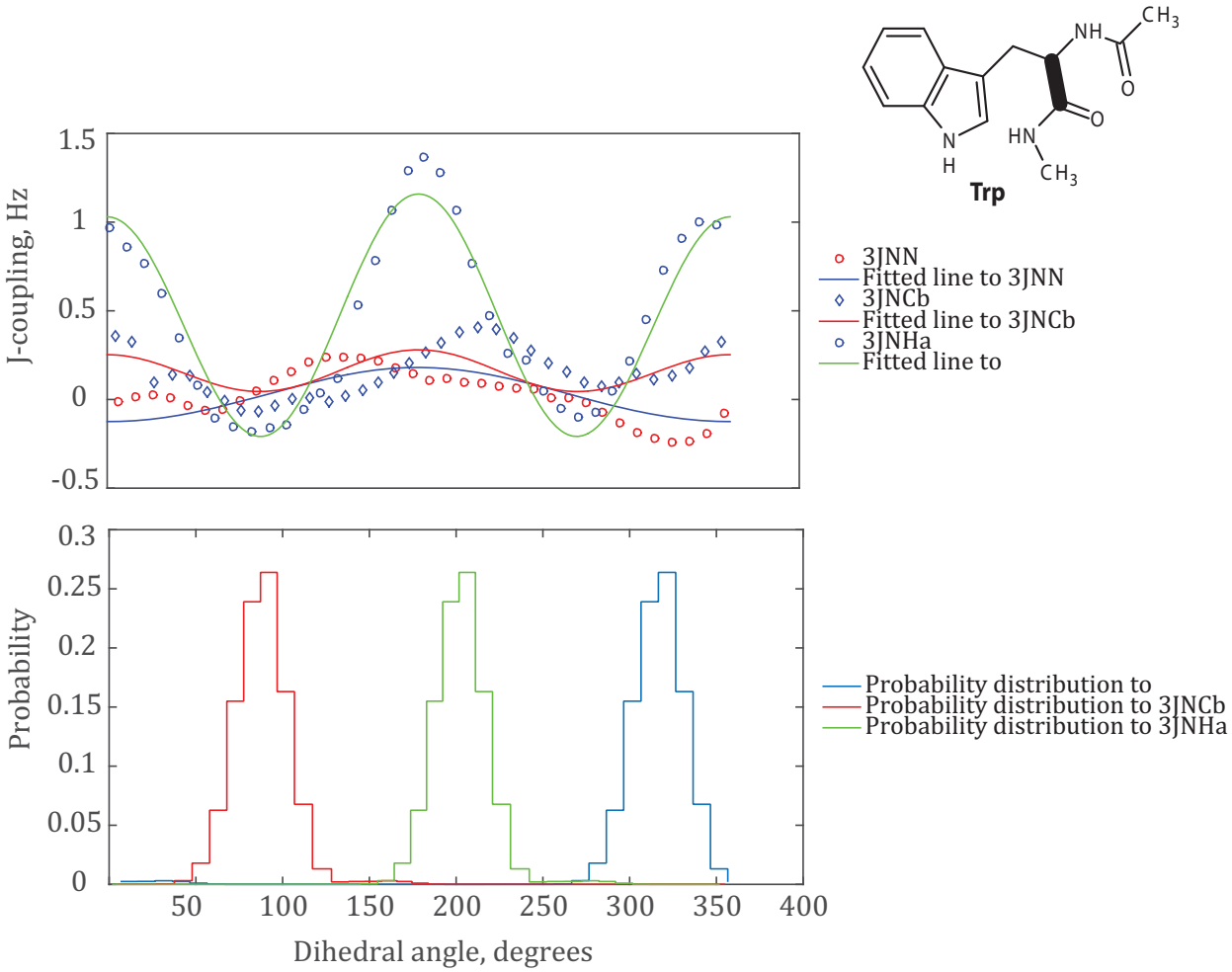

Figure 71 *TRP ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles related to the ψ dihedral angle scan respectively.*

Similar to the previous description of the Karplus curves related ψ dihedral angle, the amplitude of three bond *J*-couplings between nitrogen and another nitrogen and carbon atoms is small. The scan of the Karplus curves to these data appears to be random and noisy.

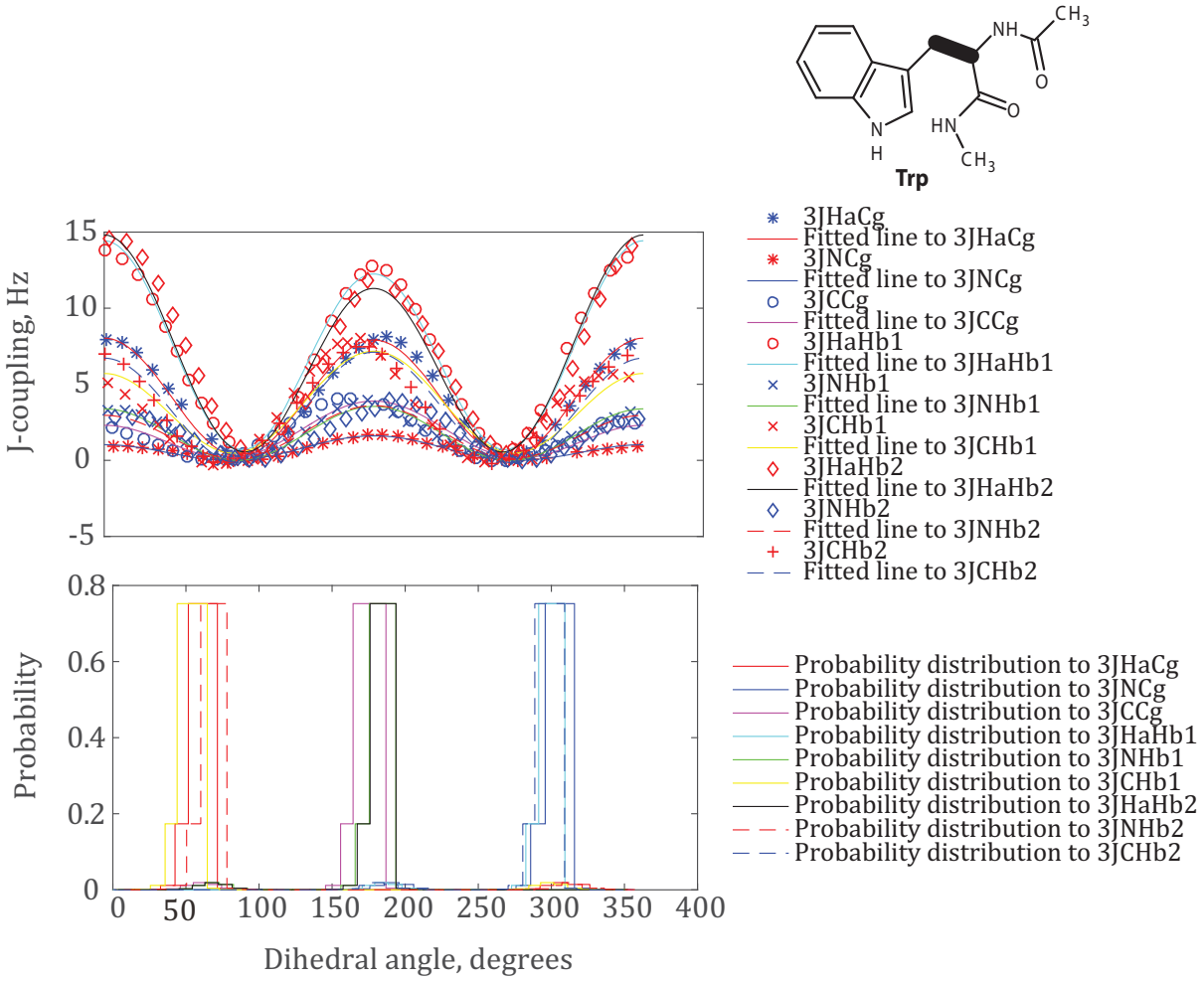

Figure 72 *TRP* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles related to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The Karplus curves that relates two proton spins appears to show higher amplitudes compared to curves constructed from dihedral angle that connects a proton and another spin (carbon and nitrogen). $^3J_{NCg}$ however has small intensity hence its Karplus curve appears to have the smallest amplitude of all curves in **Figure 72** (above).

3.9.20 Karplus coefficients, TYR

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	-0.18	0.74	-0.25	0.53	0.03	0.14
HA	CA	C	N	1.04	0.38	-0.20	0.20	-0.23	0.24

CB	CA	C	N	0.18	1.30	-0.15	1.42	0.09	0.42
HA	CA	CB	CG	7.74	0.40	-0.34	0.14	0.21	0.23
N	CA	CB	CG	1.78	0.57	-0.11	0.32	-0.00	0.33
C	CA	CB	CG	3.14	0.37	-0.67	0.14	0.20	0.20
HA	CA	CB	HB1	12.41	0.37	1.70	0.13	0.90	0.23
N	CA	CB	HB1	4.65	0.42	0.28	0.18	-0.09	0.29
C	CA	CB	HB1	6.47	0.36	-0.48	0.14	0.48	0.20
HA	CA	CB	HB2	12.64	0.36	2.13	0.14	0.88	0.20
N	CA	CB	HB2	3.98	0.33	0.38	0.14	0.12	0.19
C	CA	CB	HB2	7.70	0.41	0.01	0.13	-0.05	0.23
H	N	CA	HA	11.72	0.27	-0.48	0.12	-0.21	0.18
C	N	CA	HA	5.85	0.29	-1.74	0.12	0.45	0.18
H	N	CA	CB	5.47	0.37	-0.56	0.21	-0.31	0.18
C	N	CA	CB	2.03	0.26	-0.75	0.14	0.16	0.15
H	N	CA	C	6.44	0.41	-1.50	0.22	-0.34	0.15
C	N	CA	C	2.84	0.25	-0.98	0.13	0.03	0.14

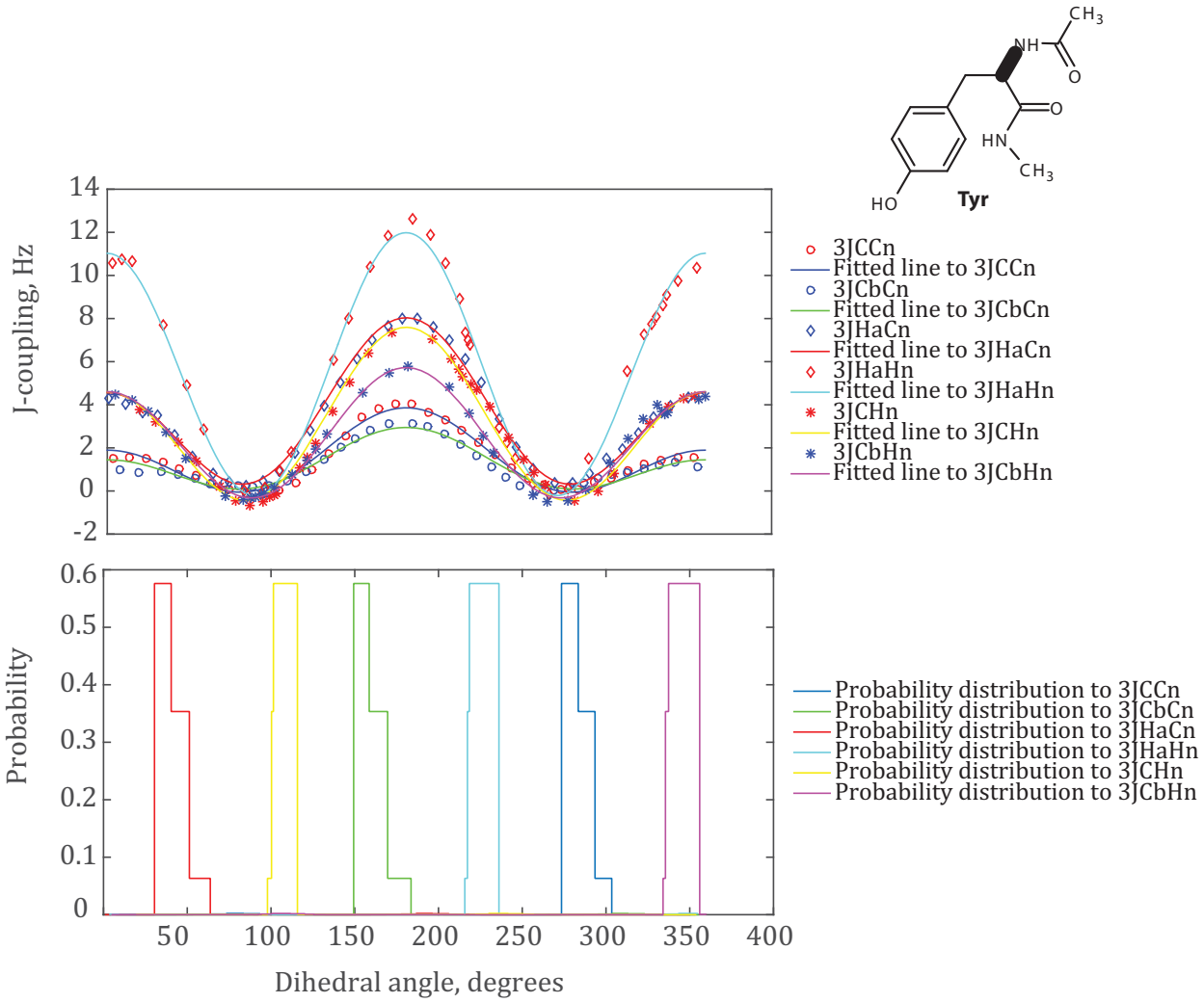

Figure 73 *TYR ϕ backbone dihedral angle scan. Top and bottom are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ϕ dihedral angle scan respectively.*

The nuclei involved in the *J*-couplings across the torsion angles associated to the ϕ dihedral angle do not engage in hydrogen bonds. However, there is hydrogen bond between oxygen atom attached to the backbone amide nitrogen and to the backbone amide proton of the other residue that might be responsible for the jumps in the Karplus curves.

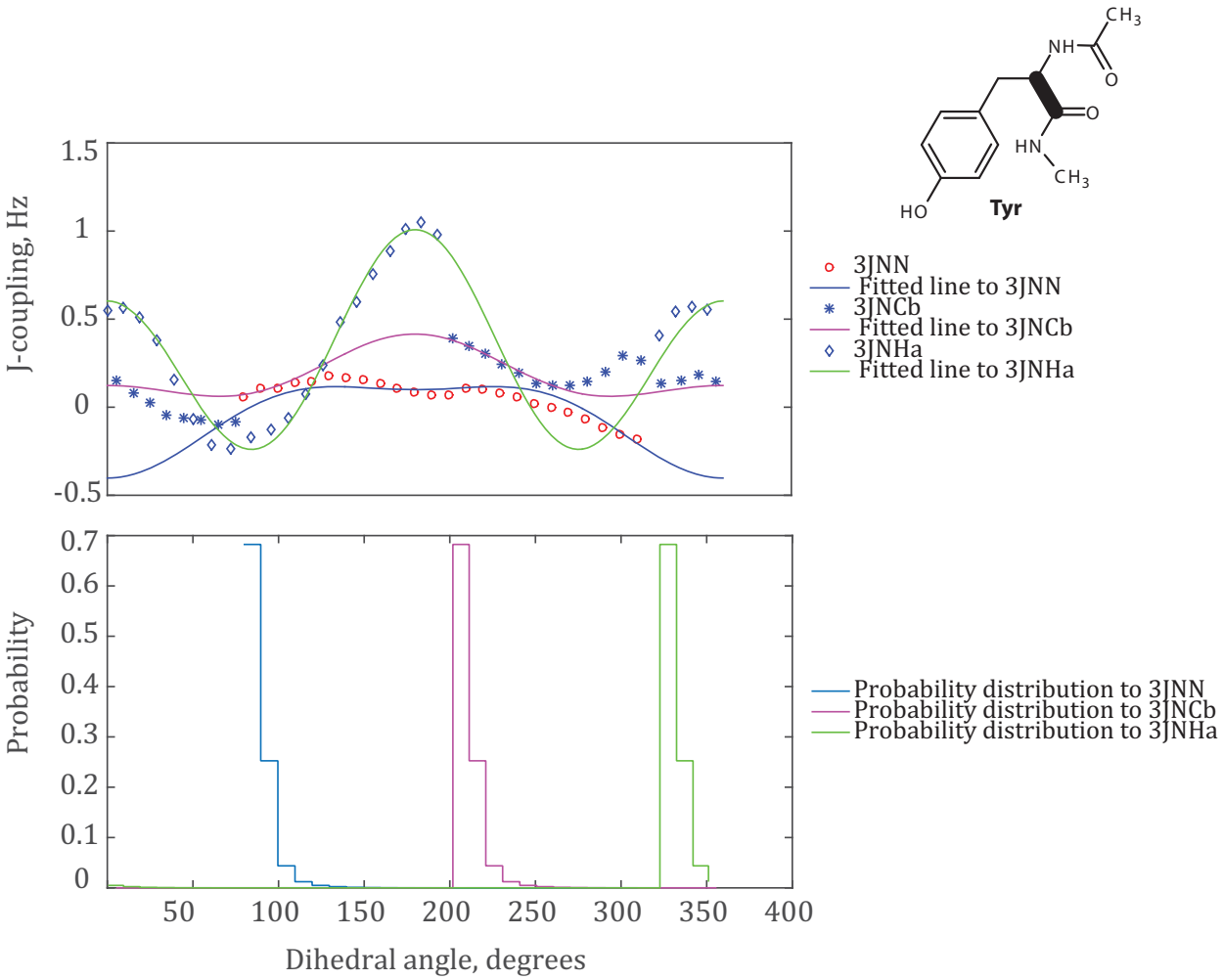

Figure 74 *TYR ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ψ dihedral angle scan respectively.*

As noted above, the amplitude of the vicinal coupling of the $^{3}J_{NN}$ and $^{3}J_{NCb}$ is small hence the scan of the Karplus curve appears as a random noise.

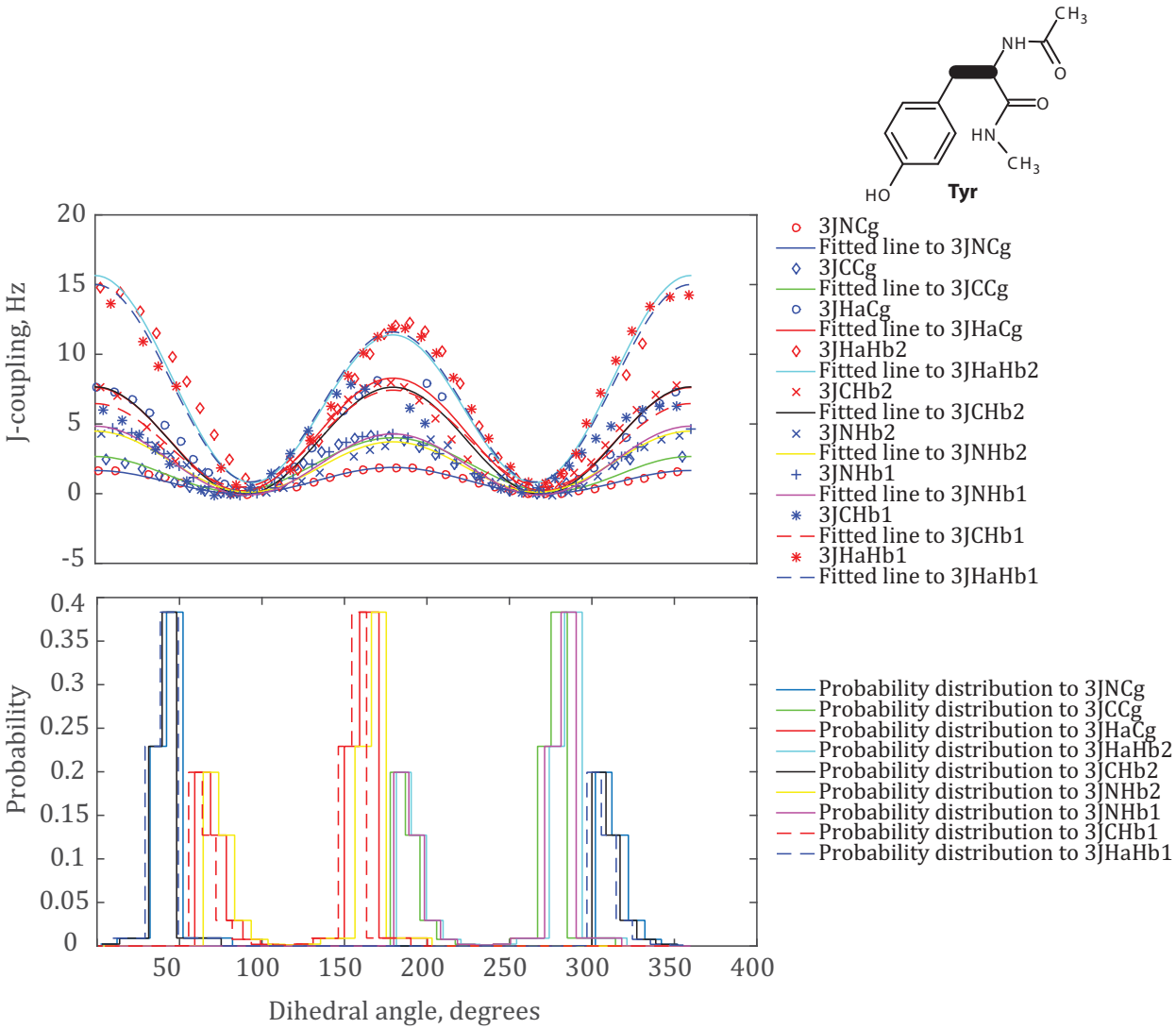

Figure 75 *TYR* χ_1 *side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles related to the* χ_1 *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

3.9.21 Karplus coefficients, VAL

Method: GIAO DFT B3LYP/M06/cc-pVTZ in SMD Water

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
N	C	CA	N	0.06	0.31	-0.06	0.20	-0.03	0.16
HA	CA	C	N	1.39	0.37	-0.03	0.16	-0.14	0.19
CB	CA	C	N	0.13	0.33	-0.19	0.14	0.03	0.16

HA	CA	CB	CG1	6.91	0.38	0.74	0.14	0.18	0.23
N	CA	CB	CG1	1.15	0.34	0.04	0.14	0.11	0.21
N	CA	CB	CG1	2.64		0.26		-0.22 [110]	
C	CA	CB	CG1	3.40	0.31	-0.13	0.13	0.11	0.21
C	CA	CB	CG1	3.42		-0.59		0.17 [110]	
HA	CA	CB	CG2	6.77	0.34	1.07	0.12	0.23	0.25
N	CA	CB	CG2	1.35	0.37	-0.00	0.17	-0.00	0.16
C	CA	CB	CG2	3.25	0.37	-0.39	0.13	0.12	0.24
HA	CA	CB	HB	10.69	0.43	1.93	0.20	0.62	0.17
N	CA	CB	HB	2.84	0.43	0.05	0.21	0.07	0.28
C	CA	CB	HB	6.94	0.35	-0.61	0.16	0.05	0.18
H	N	CA	HA	11.98	0.29	0.14	0.14	-0.14	0.21
C	N	CA	HA	5.94	0.26	-1.93	0.14	0.50	0.15
H	N	CA	CB	5.17	0.33	-0.72	0.17	-0.19	0.13
C	N	CA	CB	1.99	0.29	-0.73	0.14	0.05	0.15
H	N	CA	C	7.21	0.46	-1.49	0.32	-0.48	0.19
C	N	CA	C	3.06	0.25	-1.04	0.14	0.00	0.14

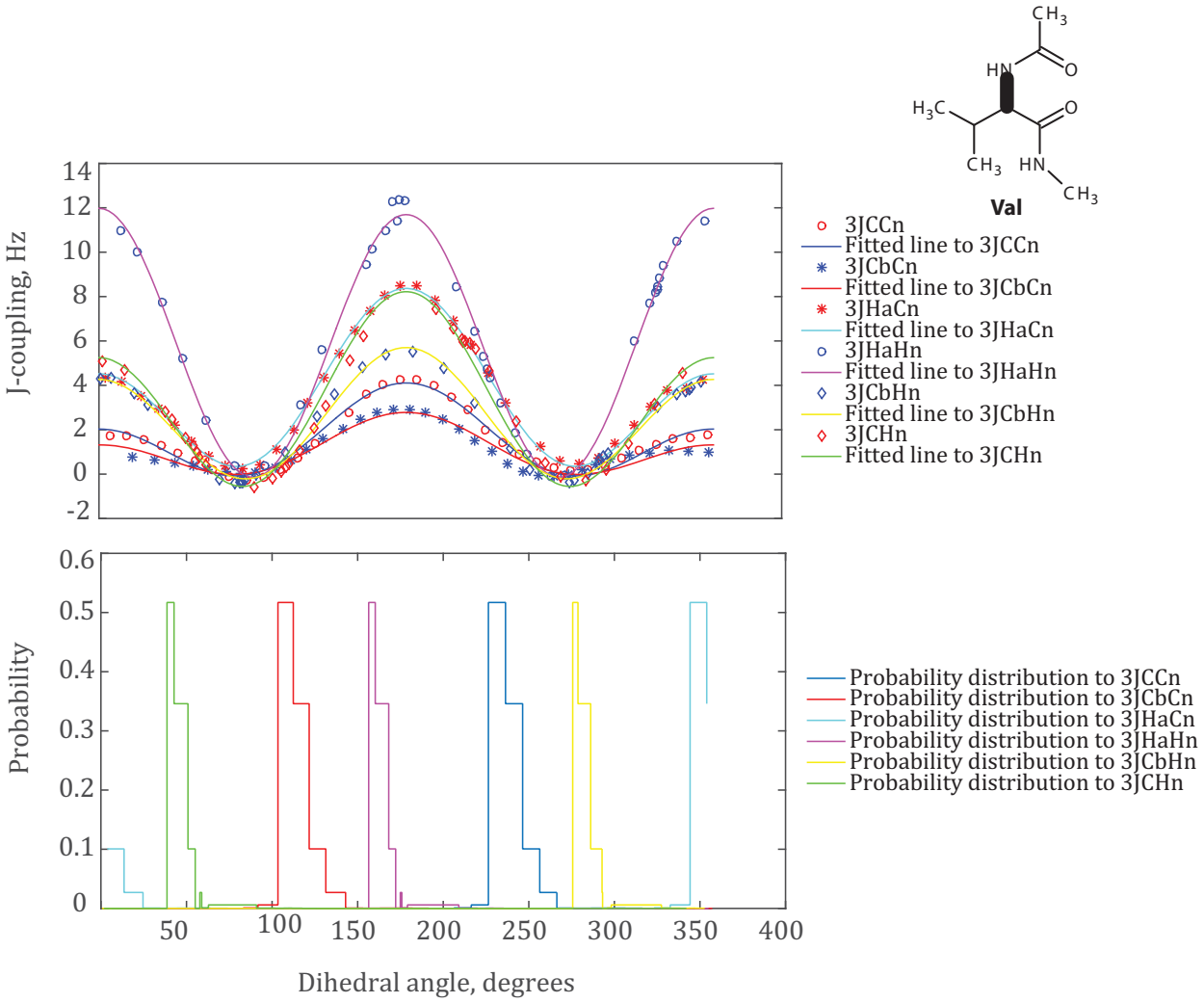

Figure 76 *VAL ϕ backbone dihedral angle scan. Top panel Karplus curve for the torsion angles linked to the ϕ dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The backbone amide proton is involved in hydrogen bond interactions, thus the rotation of the ϕ dihedral angle breaks the hydrogen bond and disturbs the coupling data. As a result, the fitting of the Karplus curves to the coupling data that participate the amide proton turns out to be just fine. The affected Karplus curves are $^{3}J_{\mathrm{HaHn}}$, $^{3}J_{\mathrm{CHn}}$ and $^{3}J_{\mathrm{CbHn}}$ coloured magenta, green, and yellow respectively. The Boltzmann probability distribution graphs for those torsion angles also appear sharp, which might be due to the hydrogen bond interaction.

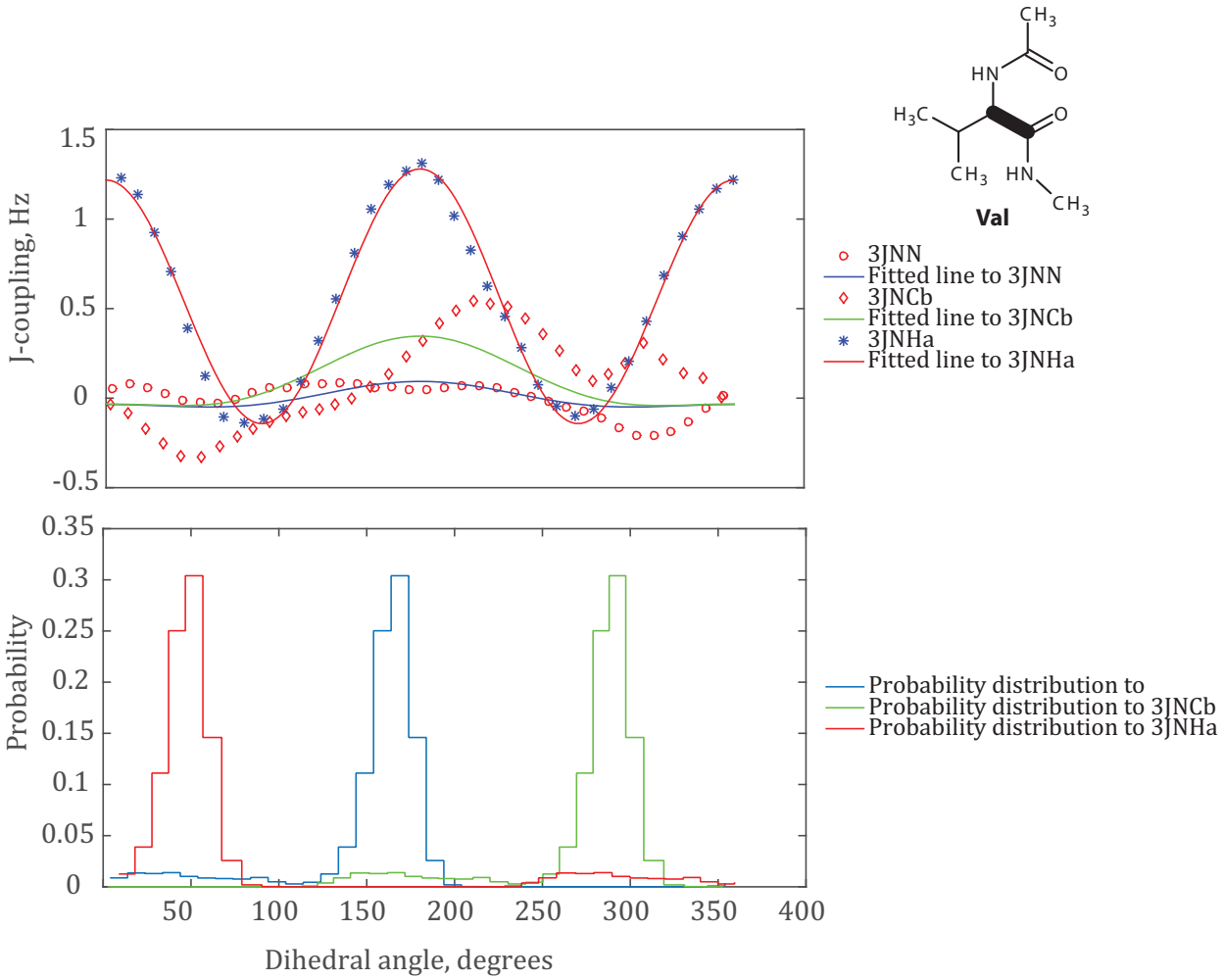

Figure 77 *VAL ψ backbone dihedral angle scan. Top and bottom panels are Karplus curves and Boltzmann probability distribution graphs for the torsion angles associated to the ψ dihedral angle scan respectively.*

As mentioned above, the $^{3}J_{NN}$ and $^{3}J_{NCb}$ have very small amplitudes. Thus the scan looks random and noisy. However the Karplus curve *i.e.* $^{3}J_{NHa}$ appears to fit the data from theory and preserve the Karplus like pattern where maximum values of vicinal coupling observed at 180°, 0° and 360° and minimum at 90° and 270°, while the values in between are related to the torsion angles between those angles.

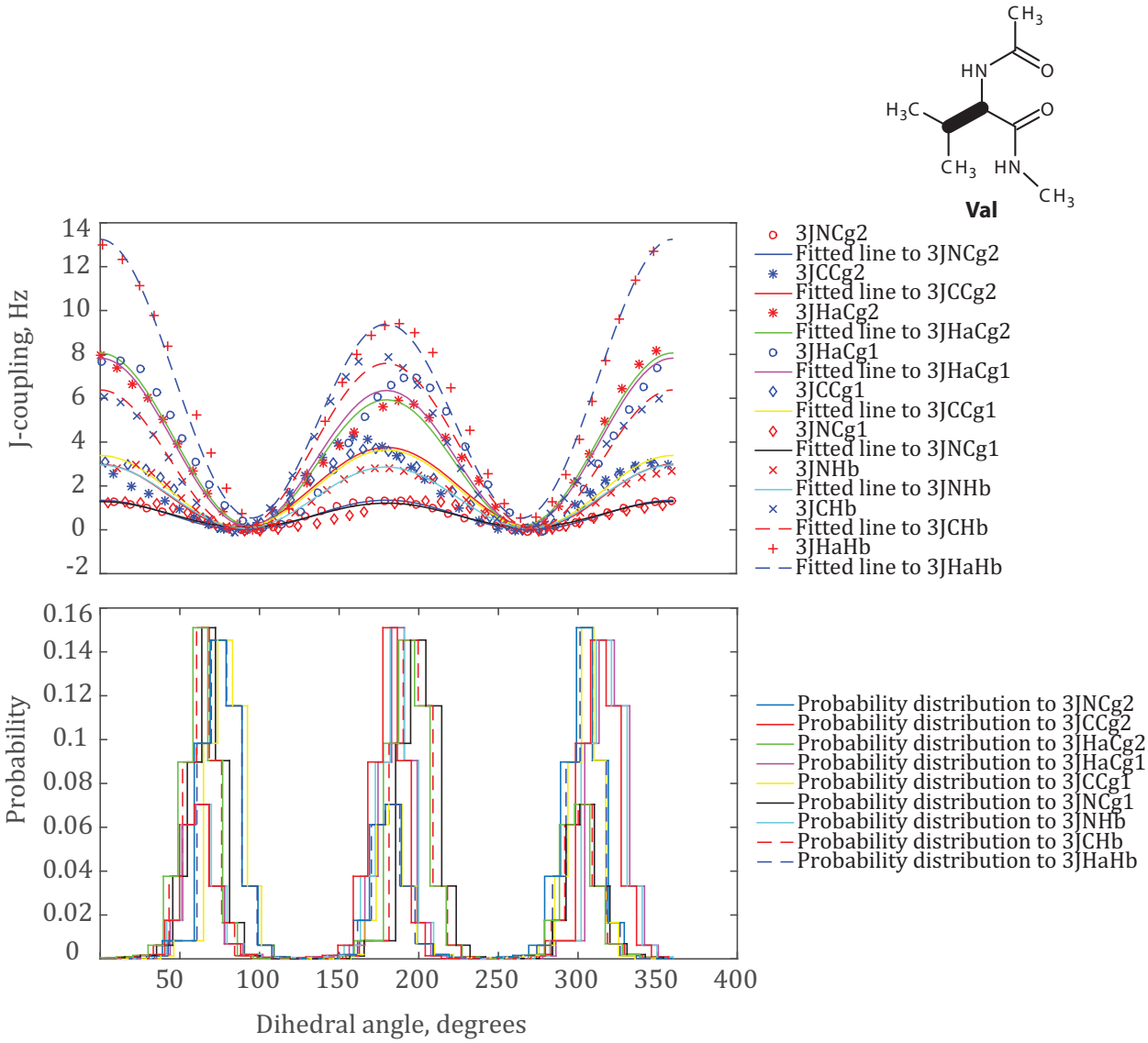

Figure 78 *VAL χ_1 side-chain dihedral angle scan. Top panel Karplus curves for the torsion angles associated to the χ_1 dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

There appears to be some hydrogen bond effect from the backbone groups that are not affected by the scan of the dihedral angle.

3.10 Comparison with experimental data

Overall, the *J*-coupling values obtained from DFT agree well to those measured experimentally. From extensive protein ^3J coupling data set that are available in litrature [24, 112-115], a recently published paper [116] for $^3J_{H\alpha H\beta 1}$ and [112, 113, 115] for the $^3J_{\alpha\beta}$ dihedral

angles are selected and the reported values are compared to the results from our DFT calculation. An example of the coupling data set is given in **Table 4** and **Table 5** (below) presented in **Figure 79** and **Figure 80** (below). The data shows the dependence of the spin-spin coupling of α proton to the amide proton for ϕ and the protons at the carbonyl positions α and β for χ_1.

For demonstration, GB3 protein 2N7J is used to show the relationship between theory and experiment for the ϕ dihedral angle. And the coupling values from theory for alpha and amide protons are within 95% prediction limit correlations to that from the experiment.

Most of the amino acids have two β protons. However, ILE, THR, VAL residues possesses only one β proton. An exception to both rules is GLY and PRO, which involve no β protons. Note that, as depicted in **Figure 80**, the $^3J_{H\alpha H\beta 1}$ of ARG and $^3J_{H\alpha H\beta 2}$ of HIS 2.00 and 0.95 Hz shows a considerable deviation from the experimental values of 5.7 and 2.6 Hz respectively. The dihedral angle χ_1 as of reference [24] is defined by the average dihedral angle and computed to be as 104 ± 35 for $^3J_{H\alpha H\beta 1}$ of ARG and 283 ± 27 for $^3J_{H\alpha H\beta 2}$ of HIS. While the respective theoretical dihedral angles are computed from atomic coordinates by scanning at the torsion angle. Chi1 dihedral angles where extracted from PDB. In this example hen lysozyme [112], FKBP bound ascomycin [113] and an unfolded urea denatured protein G [115] are used to demonstrate the correlation with the couplings from the theory.

Reported values from theory are for dihedral angles of phi and chi1 with ±47 degrees. This is because the calculations was performed by specifying the angles at a specific dihedral angles but the geometry optimization step refreshs the geometry to a ground state structure hence the disturbance of the specified angles by about ±47 degrees. The dihedral angles used for input setup were derived from PDB for the corresponding spin pairs.

J-coupling predictions better approximated using an ensemble structure [117], but the current DFT calculations are carried out on single amino acids due computational complexity.

Table 4 *An example that shows the comparison of predicted and experimentally determined [116] $^{3}J_{H\alpha Hn}$-coupling values for the main-chain dihedral angle phi.*

Amino acids and residue number	**Single letter amino acid code.**	**$^{3}J_{H\alpha Hn}$ coupling value in Hz from experiment**	**$^{3}J_{H\alpha Hn}$ coupling value in Hz from theory**	**Dihedral angle extracted from structure PDB (2N7J)**
GLN 2	Q2	7.55	6.82	0
LEU 5	L5	9.34	8.93	180
ILE 7	I7	9.46	11.19	0
GLY 9	G9	12.96	11.28	-0
GLY 9	G9	12.96	11.28	-0
LYS 10	K10	4.70	6.63	180
THR 11	T11	10.06	8.11	180
LEU 12	L12	7.84	8.93	180
LYS 13	K13	9.34	6.63	-180
GLU 15	E15	8.23	8.54	180
THR 16	T16	6.58	8.11	180
THR 17	T17	8.88	8.11	180
THR 18	T18	6.62	7.12	0
ALA 20	A20	7.41	9.58	0
ALA 26	A26	4.32	4.28	180
LYS 28	K28	3.78	4.81	0
ALA 29	A29	4.52	4.28	180
LYS 31	K31	3.96	4.81	180
GLN 32	Q32	4.68	6.82	0
ASN 37	N37	9.33	11.02	180
GLY 38	G38	12.05	11.28	0
GLY 41	G41	9.28	11.28	180
GLY 41	G41	9.28	6.85	180

VAL 42	V42	8.81	6.32	180
THR 44	T44	8.81	8.11	180
TYR 45	Y45	8.92	6.30	180
ASP 46	D46	9.91	11.79	180
ALA 48	A48	4.28	4.28	180
THR 49	T49	9.86	8.11	180
LYS 50	K50	7.17	4.81	180
THR 51	T51	10.07	8.11	180
PHE 52	F52	9.47	9.44	0
THR 53	T53	9.46	7.12	0
THR 55	T55	9.95	8.11	0
GLU 56	E56	8.49	8.73	0

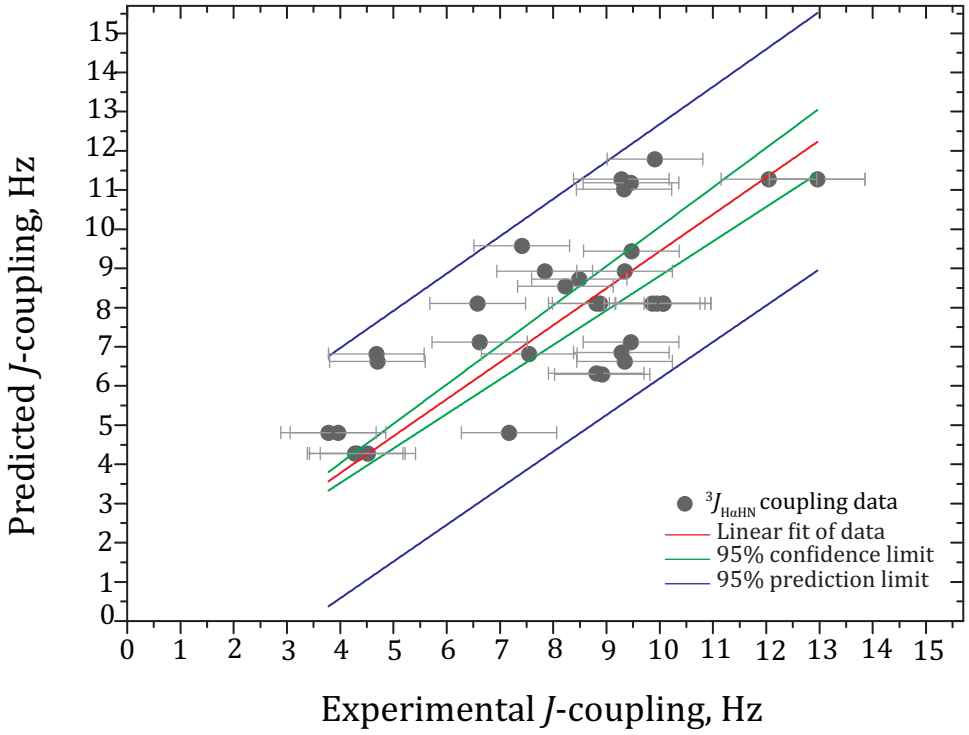

Figure 79 *Correlation of predicted and experimentally obtained coupling values for Hα and H_N for different ϕ dihedral angles.*

Table 5 *An example that shows the comparison of predicted and experimentally determined $^3J_{\alpha\beta}$-coupling values of the side-chain dihedral angle Chi1.*

Amino acid and residue number	**Single letter amino acid code**	**Scalar coupling over three bond**	**Alpha proton**	**Beta protons**	**$^3J_{H\alpha Hx\beta}$ coupling value in Hz from experiment**	**$^3J_{H\alpha Hx\beta}$ coupling value in Hz from theory**	**Dihedral angle from PDB; 5JXV, 1E8L, 1FKR and 9PCY**
TYR 3	Y3	^3J	HA	HB2	8.25 [115]	8.82	0
TYR 3	Y 3	^3J	HA	HB3	5.47 [115]	5.18	0
ASN 8	N 8	^3J	HA	HB3	5.50 [115]	5.15	0
PHE 52	F 52	^3J	HA	HB3	5.52 [115]	4.60	0
TYR 33	Y 33	^3J	HA	HB2	8.61 [115]	8.82	0
CYS 6	C6	^3J	HA	HB1	11.5 [112]	12.78	73.57
CYS 6	C6	^3J	HA	HB2	3.5 [112]	4.88	-50.16
PHE 3	F3	^3J	HA	HB1	10 [112]	10.39	76.38
PHE 3	F3	^3J	HA	HB2	3.0 [112]	5.01	161.21
HIS 15	H15	^3J	HA	HB1	11.2 [112]	10.09	70.18
HIS 15	H15	^3J	HA	HB2	2.6 [112]	0.95	12.41
ASP 18	D18	^3J	HA	HB1	4.2 [112]	5.04	48.14
ASP 18	D18	^3J	HA	HB2	11.0 [112]	13.18	-98.43
TYR 20	Y20	^3J	HA	HB1	2.3 [112]	5.19	48.03
TYR 20	Y20	^3J	HA	HB2	11.7 [112]	12.15	-73.52
ASN 27	N27	^3J	HA	HB1	10.3 [112]	13.10	75.40
ASN 27	N27	^3J	HA	HB2	2.4 [112]	5.29	-50.64
ARG 61	R61	^3J	HA	HB1	5.7 [112]	2.00	-176.79
ARG 61	R61	^3J	HA	HB2	10.8 [112]	10.53	49.65
TRP 123	W123	^3J	HA	HB1	10.6 [112]	11.30	69.73
TRP 123	W123	^3J	HA	HB2	2.9 [112]	5.20	-1.17
GLN 3	Q3	^3J	HA	HB1	9.2 [113]	11.31	45.91
GLN 20	Q20	^3J	HA	HB2	4.1 [113]	7.15	-88.61

SER 8	S8	^3J	HA	HB2	2.0 [113]	2.83	-66.53
SER 8	S8	^3J	HA	HB1	4.0 [113]	3.32	47
LEU 50	L50	^3J	HA	HB1	10.0 [113]	8.56	46.44
LEU 50	L50	^3J	HA	HB2	3.0 [113]	2.03	153.73
LYS 17	K17	^3J	HA	HB1	11.0 [113]	11.95	72.40
LYS 17	K17	^3J	HA	HB2	3.2 [113]	3.79	152.91
GLU 61	E61	^3J	HA	HB1	13.0 [113]	12.20	46.08
GLU 61	E61	^3J	HA	HB2	4.1 [113]	2.98	162.09
MET 49	M49	^3J	HA	HB1	11.0 [113]	11.36	44.34
MET 49	M49	^3J	HA	HB2	3.0 [113]	4.67	143.49
PRO 58	P58	^3J	HA	HB1	8.9 [114]	8.16	65.24
PRO 58	P58	^3J	HA	HB2	8.0 [114]	8.95	-13.64

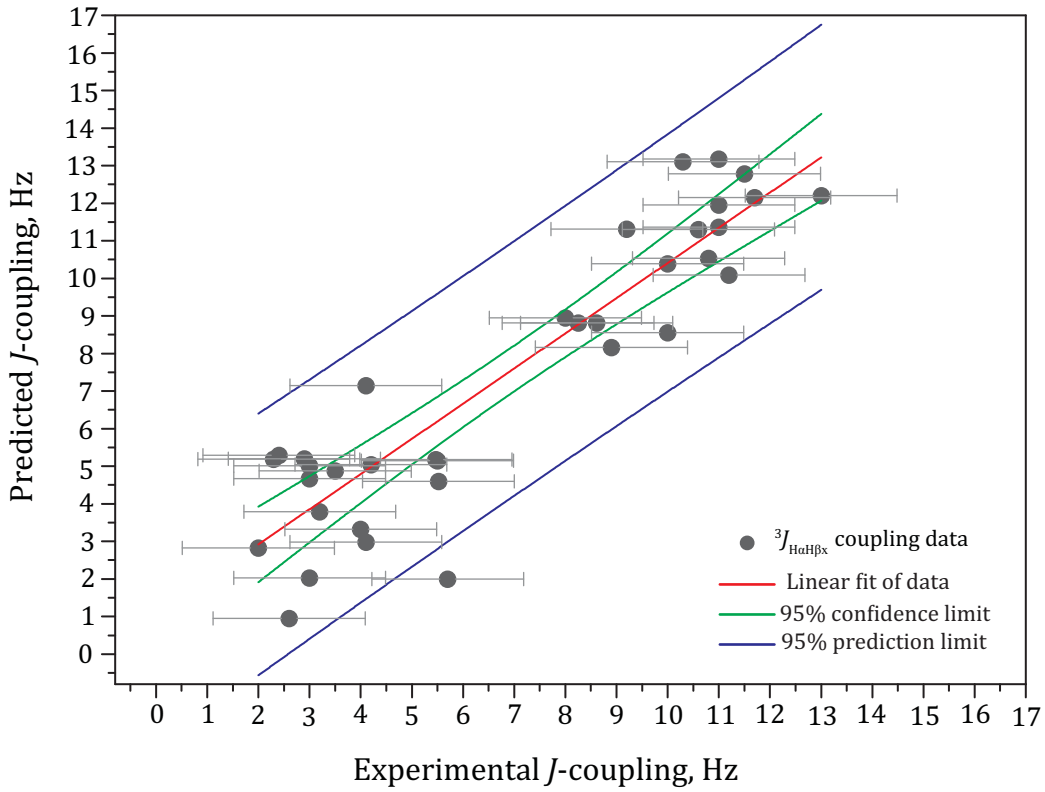

Figure 80 *Correlation of predicted and experimentally obtained coupling values for Hα and Hβx for different* χ_1 *dihedral angles, Hβx stands for Hβ1 or Hβ2.*

3.11 Concluding remarks on applicability and limitations

This work on spin-spin coupling constants has focused in NMR simulation work of protein systems. The work primarily involved the compilation of data to simulate NMR spectra of ubiquitin specifically, and biologically important molecules in general. The objective of the work was to theoretically compute spin-spin coupling values. To obtain the coupling values for a protein system, the spin-spin coupling constant of each of the 20 amino acids must be considered. This was done quantum mechanically using DFT with B3LYP/ cc-pVTZ method for structure optimization and GIAO M06/cc-pVTZ method for magnetic interactions computation. Practically, the computation required entering the amino acid structure using Gauss View software followed by a relaxed scan before running the spin-spin coupling job. The computationally predicted value of the spin-spin interactions involve isotropic contributions form Fermi contact (FC), spin-dipolar (SD), paramagnetic spin-orbit (PSO), and the diamagnetic spin-orbit (DSO) terms [94].

The focus of this research work was to determine the coefficients *A*, *B*, and *C* of the Karplus equation that aids in building an estimator function in *Spinach* software library [3], which estimates coupling constants in proteins during protein NMR simulation.

Most one and two-bond couplings are known already [14, 25]. Therefore, the focus turned into computing the ^{3}J-coupling values of all the possible conformations making a constraint at a particular dihedral angle. This was useful to construct the relationship between the *J*-coupling values and the associated dihedral angle. When the scan of the energy was completed, a script was used to extract the coordinates of each of the optimized conformations in to a Gaussian files to run spin-spin coupling job. This was done using Matlab. Another script was then used to extract the ^{3}J coupling constant values and draw a graph against the angles of rotation. The *A*, *B* and *C* parameters were extracted from the Karplus equation.

4 Quantum mechanical simulation of protein NOESY spectra

4.1 Nuclear Overhauser effect

NOE is one of the oldest and most researched NMR techniques that is widely used by structural biologists and organic chemists for structure determination and conformational studies [25, 29, 31] . The phenomenon is credited to Albert Overhauser [118] who initially proposed and derived detailed mathematical equations for nuclear spin polarization enhancement by polarizing conduction electron spins in metals. Carver and Slichter [119] experimentally verified the theory by observing lithium nuclear polarization enhancement as a result of electron saturation. Ronald Kaiser [120] has later extended the application of polarization enhancement technique to the analysis of NMR spectra. In his paper he showed the effect of saturating nuclear spin has on the spin polarization of the neighbouring nuclei.

NOE phenomenon arises from dipolar interaction between spins that are in close proximity, it is caused by population change that is brought about by cross-relaxation. The strength of NOE signal is proportional to the inverse sixth power of the distance between the interacting spins. The phenomenon is detectable for inter-nuclear distances up to about 6 Angstrom.

4.2 Basic NOE description using energy level diagrams

Consider a system with two spins I and S, which are neither magnetically equivalent nor *J*-coupled, but are dipolar coupled through space. The experimental manifestation of the nuclear Overhauser effect is that irradiation of the S spin would either increase, or decrease, or even invert the intensity of the signal from spin I. The mechanism involves cross-relaxation and spin-lattice relaxation pathways for the two spins. The effect of cross relaxation and signal intensity transfer between the spins can be understood from the energy level diagram shown in Figure 81 (below). The diagram shows that four relaxation pathways involve single spin with a rate W_1, while the other two pathways involve two spins either flipping in the same direction (rate W_2) or in opposite directions (rate W_0).

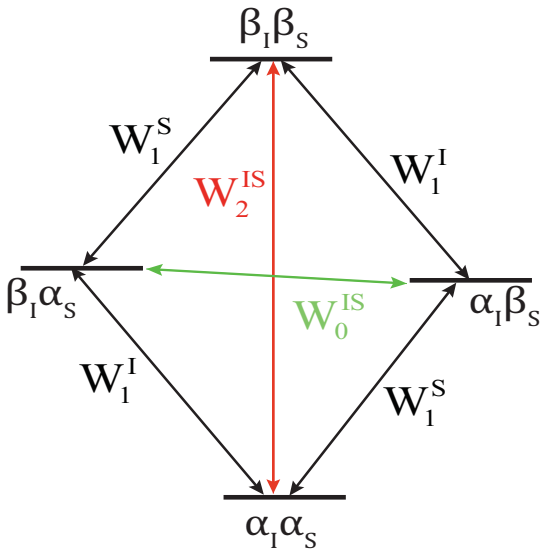

Figure 81 *Energy level diagram depicting relaxation pathways for two spins* I *and* S *in an NOE experiment. Black arrows indicate spin lattice relaxation pathways that involve single spin flips while red and green arrows show cross relaxation between the spins that involve two spin flipping either in the same or opposite direction.*

Rates W_2 and W_0 determine the sign of the enhancement. The equation of motion for the populations is an instance of Solomon equations, which were known before the Overhauser effect was discovered [121]. Solomon equations describe the evolution of energy level populations:

$$\begin{aligned}
\frac{dn_{\alpha\alpha}}{dt} &= -\left(W_{1I}+W_{1S}+W_{2IS}\right)n_{\alpha\alpha}+W_{1I}n_{\beta\alpha}+W_{1S}n_{\alpha\beta}+W_{2IS}n_{\beta\beta}\\
\frac{dn_{\alpha\beta}}{dt} &= -\left(W_{1I}+W_{1S}+W_{0IS}\right)n_{\alpha\beta}+W_{1I}n_{\beta\beta}+W_{1S}n_{\alpha\alpha}+W_{0IS}n_{\beta\alpha}\\
\frac{dn_{\beta\alpha}}{dt} &= -\left(W_{1I}+W_{1S}+W_{0IS}\right)n_{\beta\alpha}+W_{1S}n_{\beta\beta}+W_{1I}n_{\alpha\alpha}+W_{0IS}n_{\alpha\beta}\\
\frac{dn_{\beta\beta}}{dt} &= -\left(W_{1S}+W_{1I}+W_{2IS}\right)n_{\beta\beta}+W_{1I}n_{\alpha\beta}+W_{1S}n_{\beta\alpha}+W_{2IS}n_{\alpha\alpha}
\end{aligned} \tag{65}$$

In NOE experiments, the observable quantity is the longitudinal magnetization I_z and S_z. These are proportional to the difference in the number of spins between the α and β state and are given by

$$\begin{aligned}
S_z &= \left(n_{\alpha\alpha}+n_{\alpha\beta}\right)-\left(n_{\beta\alpha}-n_{\beta\beta}\right)\\
I_z &= \left(n_{\alpha\alpha}+n_{\beta\alpha}\right)-\left(n_{\alpha\beta}+n_{\beta\beta}\right)
\end{aligned} \tag{66}$$

Inserting Equation (66) in Equation (65) and rearranging the result produces a system of coupled differential equations for I_z and S_z :

$$\begin{aligned}\frac{dI_z}{dt} &= -\rho_I\left(I_z\right)-\sigma_{IS}\left(S_z\right)\\ \frac{dS_z}{dt} &= -\rho_S\left(S_z\right)-\sigma_{IS}\left(I_z\right)\end{aligned} \tag{67}$$

where

$$\begin{aligned}\rho_I &= W_{0IS}+2W_{1I}+W_{2IS}\\ \rho_S &= W_{0IS}+2W_{IS}+W_{2IS}\\ \sigma_{IS} &= W_{2IS}-W_{0IS}\end{aligned}$$

After correcting for the finite-temperature thermal equilibrium, we obtain:

$$\begin{aligned}\frac{dI_z}{dt} &= -\rho_I\left(I_z-I_z^0\right)-\sigma_{IS}\left(S_z-S_z^0\right)\\ \frac{dS_z}{dt} &= -\rho_S\left(S_z-S_z^0\right)-\sigma_{IS}\left(I_z-I_z^0\right)\end{aligned} \tag{68}$$

where I_z^0 and S_z^0 are the Boltzmann equilibrium values. NOE enhancement can be quantified as

$$\eta = \frac{I-I_0}{I_0} \tag{69}$$

where I_0 is the thermal equilibrium magnetization intensity and I is the perturbed magnetization. The pair of interacting spins may belong to different molecules (intermolecular NOE), or the same molecule (for example, in the amino acid chain of polypeptide) – this is called intramolecular NOE.

4.3 Advanced NOE description using Redfield theory

Relaxation theory, formulated by Bloch, Wangsness and Redfield (also known as Redfield theory) is one of the hardest and most important topics in the field of magnetic resonance [122, 123]. Relaxation theory is widely employed in structural biology and is useful for protein structure determination, conformational analysis, and ligand binding with a profound reliance in nuclear Overhauser effect, which is useful in determining inter-nuclear separation. The relevant relaxation theory in our case is the rotationally modulated relaxation theory in liquids.

A detailed derivation that follows broadly the argument of Ilya Kuprov's *Spin Dynamics* lecture course, and also keeps to the same notation, is presented in this section.

The Liouville - von Neumann equation for the density matrix $\hat{\rho}(t)$ is

$$\frac{\partial}{\partial t}\hat{\rho}(t) = -i\left[\hat{H}(t), \hat{\rho}(t)\right] = -i\hat{\hat{H}}(t)\hat{\rho}(t) \tag{70}$$

where $\hat{\hat{H}}$ is the Hamiltonian commutation superoperator. In solution state NMR, there are two types of contributions to the Hamiltonian superoperator, one that is static and predictable such as the chemical shift, *J*-coupling, and the other that arises due to the stochastic Brownian motion in liquid, which is assumed to be rotational diffusion. As a result, the Hamiltonian of the spin system is split into two – the static Hamiltonian $\hat{H}_0$, which acts on the spin system and the dynamic Hamiltonian that perturbs the system $\hat{H}_1(t)$:

$$\hat{H}(t) = \hat{H}_0 + \hat{H}_1(t) \tag{71}$$

The result is as follows:

$$\frac{\partial}{\partial t}\hat{\rho}(t) = -i\left(\hat{\hat{H}}_0 + \hat{\hat{H}}_1(t)\right)\hat{\rho}(t) \tag{72}$$

This equation has the large time-independent Hamiltonian, $\hat{H}_0$, and the small time-dependent stochastic Hamiltonian $\hat{H}_1(t)$.

Such systems can efficiently be treated in a formalism in which the noise is stationary, and its statistical properties are predictable. Our first assumption is that the ensemble average of the perturbing interaction vanishes *i.e.* $\left\langle \hat{\hat{H}}_1(t) \right\rangle = 0$ [124, 125]. If it is not zero, then that value is put back into the static Hamiltonian. The second assumption is that the norm of the noise is much smaller than that of the main Hamiltonian:

$$\left\| \hat{H}_1(t) \right\| \ll \left\| \hat{H}_0 \right\| \tag{73}$$

This is the perturbation theory approximation. The next step is to move into the interaction representation with respect to $\hat{H}_0$. This is achieved by introducing the following relations:

$$\hat{\sigma}(t) = e^{i\hat{\hat{H}}_0 t}\hat{\rho}(t) \qquad \hat{\hat{H}}_1^{\mathrm{R}}(t) = e^{i\hat{\hat{H}}_0 t}\hat{\hat{H}}_1(t)e^{-i\hat{\hat{H}}_0 t} \tag{74}$$

Plugging Equation (74) into the derivative of Equation (72) we obtain the following expression

$$\frac{\partial}{\partial t}\left(e^{-i\hat{\hat{H}}_0 t}\right)\hat{\sigma}(t)+e^{-i\hat{\hat{H}}_0 t}\frac{\partial}{\partial t}\left(\hat{\sigma}(t)\right)=-i\hat{\hat{H}}_0 e^{-i\hat{\hat{H}}_0 t}\hat{\sigma}(t)-ie^{-i\hat{\hat{H}}_0 t}\hat{\hat{H}}_1^{\mathrm{R}}(t)e^{i\hat{\hat{H}}_0 t}e^{-i\hat{\hat{H}}_0 t}\hat{\sigma}(t) \tag{75}$$

This simplifies into:

$$-i\hat{\hat{H}}_0 e^{-i\hat{\hat{H}}_0 t}\hat{\sigma}(t)+e^{-i\hat{\hat{H}}_0 t}\frac{\partial}{\partial t}\left(\hat{\sigma}(t)\right)=-i\hat{\hat{H}}_0 e^{-i\hat{\hat{H}}_0 t}\hat{\sigma}(t)-ie^{-i\hat{\hat{H}}_0 t}\hat{\hat{H}}_1^{\mathrm{R}}(t)e^{+i\hat{\hat{H}}_0 t}e^{-i\hat{\hat{H}}_0 t}\hat{\sigma}(t) \tag{76}$$

A lot of terms cancel and Equation (76) simplifies to

$$\frac{\partial}{\partial t}\hat{\sigma}(t)=-i\hat{\hat{H}}_1^{\mathrm{R}}(t)\hat{\sigma}(t) \tag{77}$$

where $\hat{H}_0$ has formally disappeared. It has gone into the definition of $\hat{\sigma}(t)$. This significantly accelerates the convergence under the time dependent perturbation theory (TDPT).

After formally integrating Equation (77), we obtain

$$\hat{\sigma}(t)-\hat{\sigma}(0)=-i\int_0^t \hat{\hat{H}}_1^{\mathrm{R}}(t_1)\hat{\sigma}(t_1)\,dt_1 \tag{78}$$

After moving $\hat{\sigma}(0)$ over to the right hand side and substituting it back into Equation (77) the following expression is obtained

$$\frac{\partial}{\partial t}\hat{\sigma}(t)=-i\hat{\hat{H}}_1^{\mathrm{R}}(t)\hat{\sigma}(0)-\int_0^t \hat{\hat{H}}_1^{\mathrm{R}}(t)\hat{\hat{H}}_1^{\mathrm{R}}(t_1)\hat{\sigma}(t_1)\,dt_1 \tag{79}$$

We will now apply ensemble averaging on both sides and denote it with angle brackets. Ensemble average is a linear operation, thus the average of a derivative is equivalent to the derivative of the average. Moreover $\hat{\sigma}(0)$ is a constant, therefore it can be taken out of the average.

$$\frac{\partial}{\partial t}\left\langle\hat{\sigma}(t)\right\rangle=-i\left\langle\hat{\hat{H}}_1^{\mathrm{R}}(t)\right\rangle\hat{\sigma}(0)-\int_0^t\left\langle\hat{\hat{H}}_1^{\mathrm{R}}(t)\hat{\hat{H}}_1^{\mathrm{R}}(t_1)\hat{\sigma}(t_1)\right\rangle dt_1 \tag{80}$$

As per our assumption at the beginning, the ensemble average of the noise is zero *i.e.* $\left\langle\hat{\hat{H}}_1^{\mathrm{R}}(t)\right\rangle=0$. Therefore the whole first term on the right vanishes, giving a simpler expression:

$$\frac{\partial}{\partial t}\left\langle\hat{\sigma}(t)\right\rangle=-\int_0^t\left\langle\hat{\hat{H}}_1^{\mathrm{R}}(t)\hat{\hat{H}}_1^{\mathrm{R}}(t_1)\hat{\sigma}(t_1)\right\rangle dt_1 \tag{81}$$

The time dependent rotating frame Hamiltonian operator can be factorized into time dependent coefficients with time independent operators. This is possible because time dependent matrices can be written as constant matrices with time dependent coefficients.

$$\hat{\hat{H}}_1(t) = \sum_k q_k(t)\hat{\hat{Q}}_k \;\; = \sum_m q_m^*(t)\hat{\hat{Q}}_m^\dagger \tag{82}$$

where $q_k(t)$ and $q_m^*(t)$ are the coefficients while $\hat{\hat{Q}}_k$ and $\hat{\hat{Q}}_m^\dagger$ are the time independent basis matrices. The second equality is the complex conjugate transpose. Thus, transforming Equation (82) into the interaction representation and plugging it into Equation (81) we obtain the following expression

$$\frac{\partial}{\partial t}\left\langle \hat{\sigma}(t)\right\rangle = -\sum_{km}\int_0^t \left\langle q_k(t)q_m^*(t_1)\hat{\hat{Q}}_1^{\mathrm{R}}(t)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}(t_1)\hat{\sigma}(t_1)\right\rangle dt_1 \tag{83}$$

It is necessary for later to paste one of the Hermitian superoperators $\hat{\hat{H}}_1^{\mathrm{R}}$ in a conjugated form. The perturbation is rotationally modulated and in such cases the coefficients $q_k(t)$ are Wigner functions and $\hat{\hat{Q}}_k$ are spherical tensor operators.

We will now remember our assumption at the very beginning. Equation (73) states that the amplitude of the perturbing Hamiltonian is much smaller than that of $\hat{H}_0$. Thus, the dynamics of the system is effectively uncorrelated with the noise in $\hat{H}_1(t)$. The average of a product can be separated into product of averages if the functions are uncorrelated *i.e.* the noise is so weak that it does not drive the system, but is strong enough to relax it. This allows taking the average separately for the state vector and the perturbing Hamiltonian:

$$\left\langle q_k(t)q_m^*(t_1)\hat{\hat{Q}}_1^{\mathrm{R}}(t)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}(t_1)\hat{\sigma}(t_1)\right\rangle = \left\langle q_k(t)q_m^*(t_1)\hat{\hat{Q}}_1^{\mathrm{R}}(t)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}(t_1)\right\rangle\left\langle \hat{\sigma}(t_1)\right\rangle \tag{84}$$

The time dependence in the $\hat{\hat{Q}}_1^{\mathrm{R}}(t)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}(t_1)$ term is not stochastic. They are therefore not affected by the ensemble average. Mathematically the ensemble average of the remaining term $q_k(t)q_m^*(t_1)$ is a correlation function:

$$\left\langle q_k(t)q_m^*(t_1)\right\rangle = g_{km}(t,t_1) \tag{85}$$

The noise in the system is assumed to be stationary, therefore the statistical properties of the noise do not depend on time. Therefore the correlation function can only depend on the difference of the time $\left(t-t_1\right)$ instead of the absolute times t and t_1:

$$g_{km}\left(t,t_1\right)=g_{km}\left(t-t_1\right) \tag{86}$$

After modifying Equation (83) using Equations (86) and (84) the following is obtained:

$$\frac{\partial}{\partial t}\left\langle\hat{\sigma}\left(t\right)\right\rangle=-\sum_{km}\int_0^t g_{km}\left(t-t_1\right)\hat{\hat{Q}}_1^{\mathrm{R}}\left(t\right)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}\left(t_1\right)\left\langle\hat{\sigma}\left(t_1\right)\right\rangle dt_1 \tag{87}$$

A further assumption is that the decay of the correlation function in small and non-viscous solvents is so fast that the spin dynamics of the system hardly occurs; hence the density matrix can be taken out of the integral to obtain

$$\frac{\partial}{\partial t}\left\langle\hat{\sigma}\left(t\right)\right\rangle=-\sum_{km}\int_0^t g_{km}\left(t-t_1\right)\hat{\hat{Q}}_1^{\mathrm{R}}\left(t\right)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}\left(t_1\right)dt_1\left\langle\hat{\sigma}\left(t_1\right)\right\rangle \tag{88}$$

Because the implicit coarse-graining schemes are known to be more stable than the explicit ones, we take $\hat{\sigma}\left(t\right)$ out of the integral, rather than $\hat{\sigma}\left(0\right)$. In the remaining integral we will perform a variable substitution $\tau=t-t_1$ and then turn the integration limits around to get:

$$\frac{\partial}{\partial t}\left\langle\hat{\sigma}\left(t\right)\right\rangle=-\sum_{km}\int_0^t g_{km}\left(\tau\right)\hat{\hat{Q}}_k^{\mathrm{R}}\left(t\right)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}\left(t-\tau\right)d\tau\left\langle\hat{\sigma}\left(t\right)\right\rangle \tag{89}$$

The second consequence of our assumption about the very rapid decay of the correlation functions is that it is permissible to extend the upper integration limit to infinity without affecting the value of the integral. We shall also drop the angular brackets on the density matrix for convenience:

$$\frac{\partial}{\partial t}\hat{\sigma}\left(t\right)=\left[-\sum_{km}\int_0^\infty g_{km}\left(\tau\right)\hat{\hat{Q}}_k^{\mathrm{R}}\left(t\right)\hat{\hat{Q}}_m^{\mathrm{R}\dagger}\left(t-\tau\right)d\tau\right]\hat{\sigma}\left(t\right) \tag{90}$$

After using Equations (74) to return back to the Schrödinger representation, we get:

$$\frac{\partial\hat{\rho}\left(t\right)}{\partial t}=-i\hat{\hat{H}}_0\hat{\rho}\left(t\right)-\sum_{km}\int_0^\infty g_{km}\left(\tau\right)\hat{\hat{Q}}_k e^{-i\hat{\hat{H}}_0\tau}\hat{\hat{Q}}_m^{\dagger}e^{i\hat{\hat{H}}_0\tau}d\tau\hat{\rho}\left(t\right) \tag{91}$$

The equation now has the following form:

$$\frac{\partial\hat{\rho}(t)}{\partial t} = -i\hat{\hat{H}}_0\hat{\rho}(t) + \hat{\hat{R}}\hat{\rho}(t) \tag{92}$$

In which the relaxation superoperator is:

$$\hat{\hat{R}} = -\sum_{km}\int_0^\infty g_{km}(\tau)\hat{\hat{Q}}_k e^{-i\hat{\hat{H}}_0\tau}\hat{\hat{Q}}_m^\dagger e^{i\hat{\hat{H}}_0\tau}d\tau \tag{93}$$

Equation (92) is the "master equation". In practice the master equation requires a correction to drive the system towards equilibrium instead to zero, which the currently derived equation does. The derivation of the correction term is beyond the scope of the project work.

An important bit of the relaxation superoperator that needs special attention is the correlation function. Mathematically, the correlation function is followed by the multiplication of two functions ensemble average. The Hamiltonian for a rotationally modulated interaction is:

$$\hat{H}(t) = \sum_{km=-2}^{2}\mathfrak{D}_{km}^{(2)}(t)\hat{Q}_{km} \tag{94}$$

where $\mathfrak{D}_{km}^{(2)}$ are second rank Wigner *D*-functions that depend on the angles of rotation which further depend on time and $\hat{Q}_{km}$ are the rotational basis set.

The correlation function between two Wigner functions is given by Equation (95) with the angle bracket denoting ensemble average:

$$G_{kmpq}(\tau) = \left\langle \mathfrak{D}_{km}^{(2)}(\tau)\mathfrak{D}_{pq}^{(2)*}(t+\tau)\right\rangle = \left\langle \mathfrak{D}_{km}^{(2)}(0)\mathfrak{D}_{pq}^{(2)*}(\tau)\right\rangle \tag{95}$$

It only depends on the time difference, hence the second equality in which the function is translated to the origin. Correlation function analysis is similar to the analysis of probability distributions thus Equation (95) can be expressed in probability distributions

$$G_{kmpq}(\tau) = \int\int \mathfrak{D}_{km}^{(2)}(\Omega_0)\mathfrak{D}_{pq}^{(2)*}(\Omega)P(\Omega_0)P(\Omega,\tau\,|\,\Omega_0,0)d\Omega_0 d\Omega \tag{96}$$

Our first assumption is that the sample is in equilibrium, the molecules are spherical and undergo isotropic rotational diffusion [25]. For systems that involve anisotropic rotational diffusion detailed description is given in [126, 127]. Thus $P(\Omega_0)$ is a constant and equal to $\frac{1}{8\pi^2}$ due to normalization. Thus, Equation (96) simplifies to

$$G_{kmpq}(\tau)=\frac{1}{8\pi^2}\int\int \mathfrak{D}^{(2)}_{km}(\Omega_0)\mathfrak{D}^{(2)*}_{pq}(\Omega)P(\Omega,\tau\,|\,\Omega_0,0)\,d\Omega_0 d\Omega \tag{97}$$

The term $P(\Omega,\tau\,|\,\Omega_0,0)$ is not known. Hence for the moment the focus would be to determine this term and substitute it into Equation (97). The system obeys the isotropic rotational diffusion equation in space. The equation is:

$$\frac{\partial}{\partial\tau}P(\Omega,\tau\,|\,\Omega_0,0)=-D\left(\hat{L}^2_{\mathrm{X}}+\hat{L}^2_{\mathrm{Y}}+\hat{L}^2_{\mathrm{Z}}\right)P(\Omega,\tau\,|\,\Omega_0,0)=-D\hat{L}^2P(\Omega,\tau\,|\,\Omega_0,0) \tag{98}$$

where D is the rotational diffusion constant and $\hat{L}^2_{\mathrm{X}}$, $\hat{L}^2_{\mathrm{Y}}$, and $\hat{L}^2_{\mathrm{Z}}$ are infinitesimal diffusion generators along the respective directions. $\left(\hat{L}^2_{\mathrm{X}}+\hat{L}^2_{\mathrm{Y}}+\hat{L}^2_{\mathrm{Z}}\right)$ is the angular momentum operator and its eigenfunctions are Wigner functions $\hat{L}^2\mathfrak{D}^{(l)}_{mk}=l(l+1)\mathfrak{D}^{(l)}_{mk}$.

The solution to Equation (98) can be achieved by standard variable separation in terms of linear combination of Wigner functions. If we now partition the probability function into sum of products of space and time, we find a solution of the form

$$P(\Omega,\tau\,|\,\Omega_0,0)=\sum_{l=0}^{\infty}\sum_{k,m=-l}^{l}A^{(l)}_{km}(\tau\,|\,\Omega_0,0)\mathfrak{D}^{(l)}_{km}(\Omega) \tag{99}$$

After substituting Equation (99) into Equation (98) we get

$$\sum_{l=0}^{\infty}\sum_{k,m=-l}^{l}\frac{\partial}{\partial\tau}A^{(l)}_{km}(\tau\,|\,\Omega_0,0)\mathfrak{D}^{(l)}_{km}(\Omega)=\sum_{l=0}^{\infty}\sum_{k,m=-l}^{l}A^{(l)}_{km}(\tau\,|\,\Omega_0,0)\left(-Dl(l+1)\right)\mathfrak{D}^{(l)}_{km}(\Omega) \tag{100}$$

Wigner functions are orthogonal. Therefore, after multiplying both sides of Equation (100) with $\mathfrak{D}^{(l')}_{k'm'}(\Omega)$ and integrating, we get:

$$\frac{\partial}{\partial\tau}A^{(l)}_{km}(\tau\,|\,\Omega_0,0)=-Dl(l+1)A^{(l)}_{km}(\tau\,|\,\Omega_0,0) \tag{101}$$

This is a large set of equations but the equations are independent. They may be solved in the standard way:

$$A^{(l)}_{km}(\tau\,|\,\Omega_0,0)=e^{-Dl(l+1)\tau}A^{(l)}_{km}(0) \tag{102}$$

Replacing $A^{(l)}_{km}(0)$ with $C^{(l)}_{km}(0)$ to emphasize that they are constants, the general solution for Equation (99) would be

$$P\left(\Omega,\tau\mid\Omega_0,0\right)=\sum_{l=0}^{\infty}\sum_{k,m=-l}^{l}C_{km}^{(l)}\left(0\right)D_{km}^{(l)}\left(\Omega\right)e^{-\mathrm{D}l(l+1)\tau} \tag{103}$$

If we now apply the initial condition, the solution at time $\tau=0$ would be of the form

$$\sum_{l=0}^{\infty}\sum_{k,m=-l}^{l}C_{km}^{(l)}\left(0\right)\mathfrak{D}_{km}^{(l)}\left(\Omega\right)e^{-Dl(l+1)\tau}=\delta\left(\Omega-\Omega_0\right) \tag{104}$$

We now need to multiply again by the Wigner function $\mathfrak{D}_{k'm'}^{(l')}\left(\Omega\right)$ to extract $C_{km}^{(l)}$. Using the orthogonality condition of the Wigner functions from the literature:

$$C_{km}^{(l)}=\frac{2l+1}{8\pi^2}\mathfrak{D}_{km}^{(l)*}\left(\Omega_0\right) \tag{105}$$

and plugging Equation (105) into Equation (103) we obtain

$$P\left(\Omega,\tau\mid\Omega_0,0\right)=\sum_{l=0}^{\infty}\frac{2l+1}{8\pi^2}\sum_{k,m=-l}^{l}\mathfrak{D}_{km}^{(l)*}\left(\Omega_0\right)\mathfrak{D}_{km}^{(l)}\left(\Omega\right)e^{-D(l(l+1))\tau} \tag{106}$$

$$\int\mathfrak{D}_{k_1m_1}^{(l_1)*}\left(\Omega\right)\mathfrak{D}_{k_2m_2}^{(l_2)}\left(\Omega\right)d\Omega=\frac{8\pi^2}{2l+1}\delta_{l_1l_2}\delta_{k_1k_2}\delta_{m_1m_2} \tag{107}$$

So, to compute the correlation function, Equation (106) should be substituted into Equation (97) and, after taking the sum out of the integral for simplicity, we get

$$\begin{aligned}&G_{abcd}\left(\tau\right)=\\&\frac{1}{8\pi^2}\sum_{l=0}^{\infty}\frac{2l+1}{8\pi^2}e^{-D(l(l+1))\tau}\sum_{k,m=-l}^{l}\int\int\mathfrak{D}_{ab}^{(2)}\left(\Omega_0\right)\mathfrak{D}_{cd}^{(2)*}\left(\Omega\right)\mathfrak{D}_{km}^{(l)*}\left(\Omega_0\right)\mathfrak{D}_{km}^{(l)}\left(\Omega\right)d\Omega_0d\Omega\end{aligned} \tag{108}$$

The double integrals can now be partitioned into two single integrals with one integrating over Ω_0 and the other over Ω. The two integrals resemble Equation (107) which gives the orthogonality condition, and so the integrals can be replaced with the answers in Equation (107) :

$$G_{abcd}(\tau) =$$

$$\frac{1}{8\pi^2}\sum_{l=0}^{\infty}\frac{2l+1}{8\pi^2}e^{-D(l(l+1))\tau}\sum_{k,m=-l}^{l}\left[\int \mathfrak{D}_{ab}^{(2)}(\Omega_0)\mathfrak{D}_{km}^{(l)*}(\Omega_0)d\Omega_0\right]\left[\int \mathfrak{D}_{cd}^{(2)*}(\Omega)\mathfrak{D}_{km}^{(l)}(\Omega)d\Omega\right]$$

$$= \frac{1}{8\pi^2}\sum_{l=0}^{\infty}\frac{2l+1}{8\pi^2}e^{-D(l(l+1))\tau}\sum_{k,m=-l}^{l}\left[\frac{8\pi^2}{2l+1}\delta_{l,2}\delta_{k,a}\delta_{m,b}\right]\left[\frac{8\pi^2}{2l+1}\delta_{l,2}\delta_{k,c}\delta_{m,d}\right] \tag{109}$$

$$= \frac{\delta_{a,c}\delta_{b,d}}{5}e^{-6D\tau} = \frac{\delta_{a,c}\delta_{b,d}}{5}e^{-\frac{\tau}{\tau_c}}$$

where τ_c is the characteristic decay time of the correlation function:

$$\tau_{\mathrm{C}} = \frac{1}{6D} \tag{110}$$

and the rotational diffusion coefficient D is defined [128] from hydrodynamics using the following expression

$$D = \frac{k\mathrm{T}}{8\pi\eta r^3} \tag{111}$$

where T is the temperature, η viscosity of solvent, k and r are the Boltzmann's constant and molecular radius respectivley.

The rotational correlation time can be determined from the ratio of the longitudinal to transverse relaxation times [129, 130]. For small to medium sized biomolecules, such as proteins and nucleic acids, τ_c is obtained from the ratio of $^{15}\mathrm{N}$ longitudinal and transverse rates [130, 131]. However this approach cannot be employed for the analysis of larger proteins due to local motion contribution to the longitudinal relaxation [130]. Lee *et al.* have proposed a method for the determination of correlation time which is suitable for rapid estimation of τ_c based on the TROSY principle [132, 133]. Correlation time value varies depending on molecular size and shape, solvent viscosity, and temperature. τ_c is in picoseconds for small organic molecules and nanoseconds for biological macromolecules.

Our next step in the derivation is to define the spectral power density as the Fourier transform of the correlation function derived above. If we now calculate an element of the relaxation superoperator from state $\hat{\rho}_a$ to state $\hat{\rho}_b$

$$\left\langle \hat{\rho}_a \middle| \hat{\hat{R}}_{ab} \middle| \hat{\rho}_b \right\rangle = \sum_{kmpq}\int_0^{\infty} G_{km}(\tau)\,\mathrm{Tr}\left(\hat{\rho}_a^{\dagger}\left[\hat{\hat{Q}}_{km},\left[e^{i\hat{\hat{H}}_0\tau}\hat{\hat{Q}}_{pq}^{\dagger}e^{-i\hat{\hat{H}}_0\tau},\hat{\rho}_b\right]\right]\right)d\tau \tag{112}$$

Equation (112) involves commutators, which can be reordered by applying the cyclic permutation rule to facilitate the derivation

$$\begin{aligned}\left\langle \hat{\rho}_a \left| \hat{\hat{R}}_{ab} \right| \hat{\rho}_b \right\rangle &= -\sum_{ijkmpq} \left[\hat{Q}^{\dagger}_{pq} \right]_{ij} \left[\hat{P}^{(a,b)}_{km} \right]_{ji} \int_0^{\infty} G_{kmpq}(\tau) e^{i\omega_{ij}\tau} d\tau = \\ &= -\sum_{ijkmpq} \left[\hat{Q}^{\dagger}_{pq} \right]_{ij} \left[\hat{P}^{(a,b)}_{km} \right]_{ji} J_{kmpq}(\omega_{ij}) \end{aligned} \tag{113}$$

$$J_{kmpq}(\omega) = \int_0^{\infty} G_{kmpq}(\tau) e^{i\omega_{ij}\tau} d\tau, \qquad \omega_{ij} = \omega_i - \omega_j$$

Equation (113) involves a complex exponential that depends on ω_{ij} and the time τ, $\hat{Q}^{\dagger}_{pq}$ is a constant matrix, and $\hat{P}^{(a,b)}_{km}$ is also a constant matrix. Hence integration of Equation (113) clearly gives a number of Fourier transforms. For the approximation that we employed to derive the correlation function *i.e.* the isotropic tumbling approximation, the spectral density function would be given as

$$J(\omega) = \int_0^{\infty} e^{-\frac{\tau}{\tau_C}} e^{i\omega_{ij}\tau} d\tau = \int_0^{\infty} e^{\left(i\omega - \frac{1}{\tau_C}\right)\tau} d\tau = \frac{1}{\frac{1}{\tau_C} + i\omega} \tag{114}$$

For mathematical reasons it is not convenient to have a complex number in the denominator, thus the spectral density function is split into real and imaginary part to give

$$J(\omega) = \frac{\tau_C}{1+\omega^2\tau_C^2} + \frac{i\omega\tau_C^2}{1+\omega^2\tau_C^2} \tag{115}$$

The real part is a Lorentzian curve. The imaginary part is fairly small and is commonly ignored, reducing the spectral density function to

$$J(\omega) = \frac{\tau_C}{1+\omega^2\tau_C^2} \tag{116}$$

Physically, the spectral density function gives the energy density of stochastic perturbations at a given frequency. If the molecular motion is rapid, which means short correlation times, this gives rise to a broad spectral density function and samples a wide range of the frequencies. This limit is known as extreme narrowing limit. Mathematically expressed as $\omega^2\tau_C^2 \ll 1$ then $J(\omega) \approx J(0)$.

The final step in our derivation is to use the relaxation theory obtained above specifically for the dipolar interaction. Dipolar relaxation is an ever present interaction in magnetic resonance spectroscopy. The focus will be on two isolated spins I and S that are close enough in space. The amplitude of dipolar interaction depends on the inverse cube of the distance r between the spins, and the angle between that vector and the direction of the two spins [22]. The Hamiltonian to describe the interaction is [22, 28] given in chapter one Equation (23).

Running the Hamiltonian through the theory described above (a highly detailed derivation is available in Arthur Palmer's chapter in [25]) results in the following expressions for the self-relaxation and the cross-relaxation rates for the two spins:

$$\begin{aligned} R_{II} &= \frac{\gamma_I^2\gamma_S^2\hbar^2}{10}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\tau_C}{r_{IS}^6}\left(\frac{3}{1+\omega_I^2\tau_C^2}+\frac{6}{1+(\omega_I+\omega_S)^2\tau_C^2}+\frac{1}{1+(\omega_I-\omega_S)^2\tau_C^2}\right) \\ R_{SS} &= \frac{\gamma_I^2\gamma_S^2\hbar^2}{10}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\tau_C}{r_{IS}^6}\left(\frac{3}{1+\omega_S^2\tau_C^2}+\frac{6}{1+(\omega_I+\omega_S)^2\tau_C^2}+\frac{1}{1+(\omega_I-\omega_S)^2\tau_C^2}\right) \\ \sigma_{IS} &= \frac{\gamma_I^2\gamma_S^2\hbar^2}{10}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\tau_C}{r_{IS}^6}\left(\frac{6}{1+(\omega_I+\omega_S)^2\tau_C^2}-\frac{1}{1+(\omega_I-\omega_S)^2\tau_C^2}\right) \end{aligned} \quad (117)$$

Cross-relaxation is responsible for the observed cross peaks in NOESY spectrum. Thus, the intensity of the peaks is correlated to the distance between spins in space. The measurement of distance from NOESY cross peak intensity is nontrivial due to spin diffusion effect in dense networks of proton spins in proteins. Spin diffusion [134] arises from magnetization transfer between spins via multiple short steps and complicates distance determination [135, 136].

Note that relaxation theory, as given above, requires diagonalization of the Hamiltonian superoperator. This is not a problem for small systems but quickly becomes computationally complex as the system size increases [20]. In order to overcome this, specifically for the purposes of our work diagonalization of the Hamiltonian is avoided [125]. Along with the state space restriction method [20], this made the simulation of proteins feasible. To ease the constraint in the memory upon large system simulation the diagonalization step is replaced by numerical evaluations of the integral that is present in the BRW relaxation Hamiltonian [125].

4.4 NOE experiments and their applications

In its simplest form, the NOESY pulse sequence consists of three 90° pulses with an evolution time t_1 separating the first two pulses. This is followed by a fixed delay time τ_m, also known as mixing time, between the second and third pulses during which magnetization transfer

happens between the spins *via* NOE. The last step of the sequence is the acquisition of the oscillating and decaying magnetization, which is known in the jargon as FID during the last t_2 period. The sequence is repeated for a range of t_1 values to acquire FID, the time domain signal that are necessary for 2D experiment. Fourier transform of the FIDs with a suitable phase cycle generates frequency domain 2D NMR spectra [137, 138].

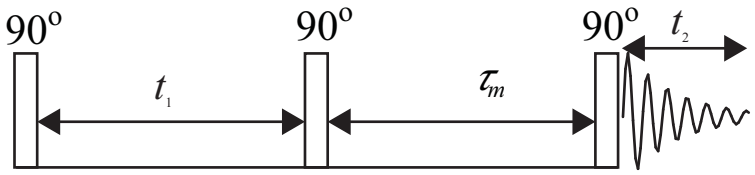

Figure 82 *Illustration of NOESY pulse sequence [138]. The pulse sequence involves application of a 90-degree rf pulse to the system, the system is then allowed to evolve for a period of* t_1*. This is followed by another 90-degree pulse and let it evolve for time* τ_m *in which the spins undergo cross relaxation. Finally another 90-degree pulse is applied followed by the record of the free induction decay.*

NOESY spectrum consists of diagonal and cross peaks, an example is given in **Figure 83** (below). Diagonal peaks are those that correspond to the 1D experiment, while cross peak signals are those that arise from cross-relaxation between spins that are in close proximity.

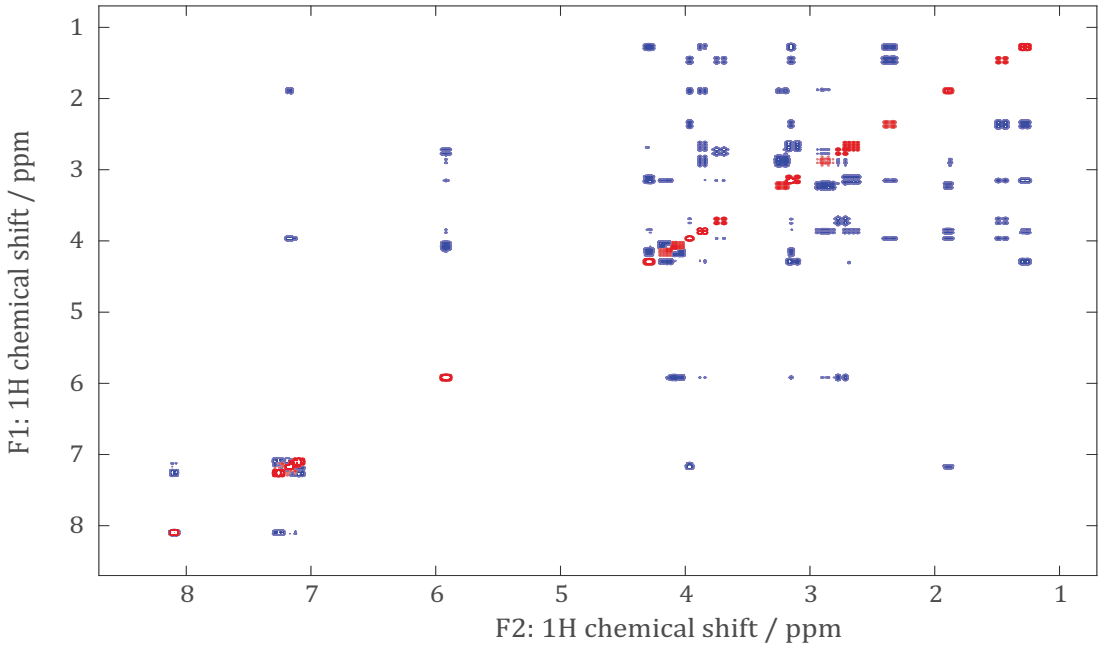

Figure 83 *An illustration of 2D NOESY spectrum showing the cross (blue) and diagonal (red) peaks. The cross peaks arise from the transfers of magnetization of one spin to another via Overhauser effect. This gives information on the separation between the nuclei and is useful in biomolecular NMR for structure determination and conformational analysis. The spin system used for the simulation of the spectrum is strychnine. The blue color of the cross-peaks was just chosen to highlight the cross-peaks that are observed from the NOE and does not indicate the negative NOE that are normally seen in NOESY of large biological molecules.*

1D NOE experiments may be split into two types depending on the area of their application. These are steady state and transient NOEs. Steady state experiment is a popular tool among organic chemists for structural and conformational studies. It relies on selective and continuous irradiation of the sample

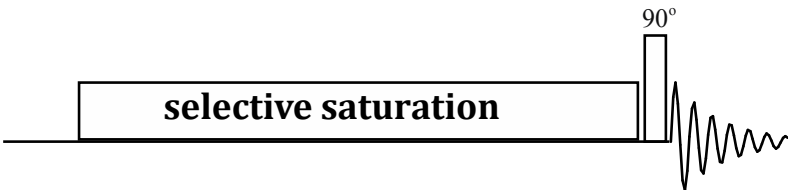

Figure 84 *Steady state NOE pulse sequence, involving selectively saturating signal of one of the spins and then application of 90º pulse to all spins followed by FID acquisition.*

The other is transient 1D NOE. It is a technique that relies in production of NOE signals brought about by population disturbance. The pulse sequence for transient NOE is depicted in **Figure 85** (below).

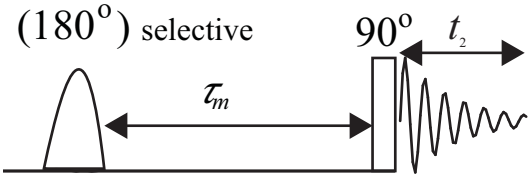

Figure 85 *Illustration of transient 1D NOE pulse sequence. In this type of experiment a single resonance is inverted selectively. It is a requirement that the resonance to be pulsed is well separated from other resonances.*

Due to its importance to structural biologists, NOESY spectroscopy is still one of the most regarded and popular technique widely used in structure determination of biological macromolecules, such as proteins, nucleic acids, enzymes, and membranes. Protein structure determination using NMR has seen a rise ever since the first NMR structure and is useful to determine structure at the potential physiological environment specifically using solution state NMR. Structure calculation can either be conducted by using signal assignment [134, 139] as a prerequisite for structure calculation or carried out using distance restrains without

requirement of signal assignment [140]. In both cases, NOESY cross-peaks are of paramount importance in determining the structure of proteins or biological macromolecules in general.

4.5 Our method for protein NOESY simulation

In the discussion of protein-scale NOESY simulation, I will be following closely our published work on quantum mechanical NMR simulation of protein systems [59]. The NMR simulation of proteins undoubtedly down to the work carried out at Oxford and then Southampton by IK *et al.* in the past 10 years [3, 20, 58, 125, 141, 142]. One of the many discoveries was the polynomially scaling simulation algorithm [20]. This became possible after the discovery of unpopulated density operator subspaces in spin systems, which normally are safe to ignore from analysis [20]. For reasons explored in [3, 20, 125], NMR simulation of arbitrary size system was made achievable.

In the simulation work reported here, the reduced operator algebra was built from populated spin product states in the basis that were mapped by analysing the topology of the spin interaction network. Spin interaction network in NMR could either be through chemical bonds, which is electron mediated and known as *J*-coupling network, or dipolar coupling network defined by through space magnetization transfer between dipole-dipole coupled spins. In solution state, NMR *J*-coupling is responsible for peak splitting while dipolar coupling is related to line width, and cross relaxation processes. In NMR experiments, magnetization flows across both networks [59].

The reduced basis set for the liquid state NOESY spectra simulation of our work was generated by using *J*-coupling graph (JCG) and dipolar coupling graph (DCG) that are generated from *J*-coupling data and Cartesian coordinates. User specified thresholds for minimum *J*-coupling value of 2 Hz and maximum distance of 5Å between spins were used. Complete list of connected overlapping subgraphs that involved user specified number of spins for both JCG and DCG was generated using the depth first search algorithm [143].

For each subgraph G_k, a complete description of the basis set of the corresponding spin subsystem is generated with the dimension D_k of the basis set is been equal to the product of the sequences of multiplicities of each spin in G_k. The basis operator represented as direct products of irreducible spherical tensors has the following from:

$$\hat{T}_{l_1,m_1} \otimes \hat{T}_{l_2,m_2} \otimes ... \otimes \hat{T}_{l_j,m_j} \otimes ... \otimes \hat{T}_{l_{|G_k|},m_{|G_k|}} \tag{118}$$

where $\hat{T}_{l_j,m_j}$ is an irreducible spherical tensor operator of rank l_j and projection m_j acting on spin j with $|G_k|$ is the number of spins in subgraph G_k. Equation (118) is useful to determine the structure of each basis operator based on the sequence index $\{l_j, m_j\}$ with a complete description of operator basis of a given subgraph G_k requiring $2|G_k|D_k$ integers of storage space. These operator basis descriptors will henceforth be referred to as the "state list" [59].

Low orders of spin correlation are included between spins proximate in the coupling graph after merging the state lists of all subgraphs and eliminating repetitions. The basis that results from this, describes the entire system without gaps or cuts. All these procedures result in a global list of spin operators that are engaged during spin system evolution. The accuracy of the basis set can systematically be varied by changing the subgraph size. Accuracy analysis is highly technical and beyond the scope of this work but it has been covered elsewhere [142]. The basis sets used for the simulation work are implemented in Spinach [3]. The description is given in **Table 6** (below).

Table 6 *Nomenclature of the reduced basis sets generated as described in the main text and implemented in Spinach library. This nomenclature is only applicable for liquid state NMR.*

IK-0	Includes all spin correlation up to user-specified level between all spins.
IK-1	All spin correlations up to user-specified level between directly *J*-coupled spins with coupling value above user specified threshold and up to user-specified level between spins that are proximate through space with distances below the user specified threshold.
IK-2	For every spin this includes all of its correlations with directly *J*-coupled spins and up to user-specified level between spins that are spatially proximate with distances below user specified threshold.

After mapping the populated space, the spin operators and state vectors should be represented as matrices. The Hamiltonian that describes spin interactions between two particles has a direct product structure [18] of the form

$$\hat{H}_n = \omega_n \left[\sigma_{n,1} \otimes \sigma_{n,2} \otimes ... \otimes \sigma_{n,N} \right] \tag{119}$$

where ω_n are spin interaction magnitudes, N is the total number of spins, and $\sigma_{n,k}$ are identity matrices, Pauli matrices or spherical tensor operators of dimension $2s_k+1$ with s_k is the spin quantum number of the *k*-th nucleus.

The commutation superoperator $\hat{\hat{H}}_n^{(\mathrm{C})}$ for the Hamiltonian in Equation (119) defined by its actions on the density operator $\hat{\rho}$, is represented as the difference between the left-side product superoperator $\hat{\hat{H}}_n^{(\mathrm{L})}$ and the right-side product superoperator $\hat{\hat{H}}_n^{(\mathrm{R})}$.

$$\begin{aligned}\hat{\hat{H}}^{(\mathrm{C})} &= \sum_n \hat{\hat{H}}_n^{(\mathrm{C})} = \sum_n \left(\hat{\hat{H}}_n^{(\mathrm{L})} - \hat{\hat{H}}_n^{(\mathrm{R})}\right) \\ \hat{\hat{H}}_n^{(\mathrm{C})}\hat{\rho} &= \left[\hat{H}_n, \hat{\rho}\right] = \hat{H}_n\hat{\rho} - \hat{\rho}\hat{H}_n \\ \hat{\hat{H}}_n^{(\mathrm{L})}\hat{\rho} &= \hat{H}_n\hat{\rho}, \quad \hat{\hat{H}}_n^{(\mathrm{R})}\hat{\rho} = \hat{\rho}\hat{H}_n\end{aligned} \tag{120}$$

If we now represent the Hamiltonian superoperator in the low correlation order basis set as discussed above, the complexity for the computation is reduced [141]. In a given operator basis $\{\hat{o}_k\}$ the left-side product superoperator would be

$$\begin{aligned}\left[\hat{\hat{H}}_n^{(\mathrm{L})}\right]_{jk} &= \left\langle \hat{o}_j \middle| \hat{\hat{H}}_n^{(\mathrm{L})} \middle| \hat{o}_k \right\rangle = \mathrm{Tr}\left[\hat{o}_j^{\dagger}\hat{H}_n\hat{o}_k\right] \\ &= \mathrm{Tr}\left[\left(\bigotimes_{m=1}^{N}\hat{\sigma}_{j,m}^{\dagger}\right)\left(\bigotimes_{m=1}^{N}\hat{\sigma}_{n,m}\right)\left(\bigotimes_{m=1}^{N}\hat{\sigma}_{k,m}\right)\right]\end{aligned} \tag{121}$$

As dot products commute with direct products, Equation (121) may be rearranged to obtain

$$\left[\hat{\hat{H}}_n^{(\mathrm{L})}\right]_{jk} = \mathrm{Tr}\left[\bigotimes_{m=1}^{N}\left(\hat{\sigma}_{j,m}^{\dagger}\hat{\sigma}_{n,m}\hat{\sigma}_{k,m}\right)\right] = \prod_{m=1}^{N}\mathrm{Tr}\left[\hat{\sigma}_{j,m}^{\dagger}\hat{\sigma}_{n,m}\hat{\sigma}_{k,m}\right] \tag{122}$$

Due to the tiny size of the Pauli matrices, the complexity of computing the trace in Equation (122) is $O(1)$ while the complexity of computing one matrix element is $O(N)$ multiplications. With $O\left(N^2\right)$ interactions in spin system, this makes the building of the Hamiltonian $O\left(N^3D^2\right)$ expensive. D is the dimension of the reduced basis set. However the sparsity of spin Hamiltonians [144] and sparse spin interaction networks of protein systems reduce the complexity to around $O\left(N^2D\right)$. Computational complexity is further reduced by the discovery of unpopulated states even in the low correlation order subspace, presence of multiple

independently evolving subspaces, and because not all populated states evolve into the detection state [145].

Spinach implementation of the NOESY pulse sequence itself is straightforward, annotated source code, coded as part of Luke Edwards and Ilya Kuprov's work, appears below.

```
function fid=noesy(spin_system,parameters,H,R,K)

% Consistency check
grumble(spin_system,parameters,H,R,K);

% Compose Liouvillian
L=H+1i*R+1i*K;

% Coherent evolution timestep
timestep=1./parameters.sweep;

% Detection state
coil=state(spin_system,'L+',parameters.spins {1},'cheap');

% Pulse operators
Lp=operator(spin_system,'L+',parameters.spins {1});
Lx=(Lp+Lp')/2; Ly=(Lp-Lp')/2i;

% First pulse
rho=step(spin_system,Lx,parameters.rho0,pi/2);

% F1 evolution
rho_stack=evolution(spin_system,L,[],rho,timestep(1),...
                    parameters.npoints(1)-1,'trajectory');

% Second pulse
rho_stack_cos=step(spin_system,Lx,rho_stack,pi/2);
rho_stack_sin=step(spin_system,Ly,rho_stack,pi/2);

% Homospoil
rho_stack_cos=homospoil(spin_system,rho_stack_cos,'destroy');
rho_stack_sin=homospoil(spin_system,rho_stack_sin,'destroy');

% Mixing time
rho_stack_cos=evolution(spin_system,1i*R+1i*K,[],rho_stack_cos,...
```

```
                        parameters.tmix,1,'final');
rho_stack_sin=evolution(spin_system,1i*R+1i*K,[],rho_stack_sin,...
                        parameters.tmix,1,'final');

% Homospoil
rho_stack_cos=homospoil(spin_system,rho_stack_cos,'destroy');
rho_stack_sin=homospoil(spin_system,rho_stack_sin,'destroy');

% Third pulse
rho_stack_cos=step(spin_system,Ly,rho_stack_cos,pi/2);
rho_stack_sin=step(spin_system,Ly,rho_stack_sin,pi/2);

% F2 evolution
fid.cos=evolution(spin_system,L,coil,rho_stack_cos,...
                  timestep(2),parameters.npoints(2)-1,'observable');
fid.sin=evolution(spin_system,L,coil,rho_stack_sin,...
                  timestep(2),parameters.npoints(2)-1,'observable');


end
```

The simulation, however, practically required the following parameters: coordinates of each magnetic nucleus to extract dipolar couplings, chemical shielding tensors, scalar couplings, quadrupolar couplings for any quadrupolar nuclei present (*e.g.* for ^{14}N), rotational correlation times – Redfield theory with rigid body rotational diffusion, and the maximum level of spin correlations expected depending on the pulse sequence. In the present work the coordinates came from the PDB database and chemical shifts including the side chains from the BMRB database. For the backbone shielding anisotropies and scalar couplings typical literature values were assumed. Sidechain scalar couplings and shielding anisotropies were estimated using GIAO DFT M06/cc-pVTZ method in Gaussian09 for the appropriately capped individual amino acids in vacuum. Detailed description on *J*-couplings generation is given in chapter three.

After the data is collated, the following procedure was used to simulate the NOESY spectrum:

1. Generate matrices for the spin quantum numbers that the system contains; that is the Pauli matrices, which are the matrix representations for the spin operators.

2. Generate spin operators using Kronecker products in the reduced state space and then build the Hamiltonian.

3. Take the initial condition, detection state, and the time step with the best sampling rate using Nyquist condition for optimal sampling, and then building the step propagator to record the free induction decay.

4. Finally the signal is multiplied by a window function, Fourier transformed and then the spectrum is plotted.

The specification of the system is done by specifying the magnet field in units of Tesla along with the coordinates, and isotropic chemical shift of the system read in from PDB and a file of chemical shift data supplied by Christian Griesinger and Donghan Lee for ubiquitin spin system and from Imai Shunsuke, Gerhard Wagner, and Scott Robson of Wagner Lab at Harvard Medical School for the stem loop RNA system.

The cut-off tolerances for dipolar coupling and *J*-couplings are set using the following commands

```
sys.tols.inter_cutoff=1.0
sys.tols.prox_cutoff=4.0
```

Cut-off tolerance for proximity is specified in Angstrom and cut-off for *J*-coupling is specified in Hz. In the example above, dipolar couplings between spins that are further than 4.0 Angstro apart would not be taken into account and any *J*-couplings smaller than 1.0 Hz would be ignored. The next thing to be specified is the relaxation theory. Redfield theory is used in the simulation.

```
inter.relaxation='redfield';
inter.rlx_keep='kite';
inter.tau_c=5e-9;
```

The first command requests full Redfield relaxation theory: DD, CSA, nuclear quadrupolar interaction (NQI) and all cross-correlations that are available for the system. As mentioned before, dipolar relaxation rates are computed from atomic coordinates, CSAs and NQIs must be provided – for those involved in simulations that require the two terms visit the manual available on http://spindynamics.org on how to add the data for CSA and NQI to *Spinach* input data structures. The second line in the specification requests the Redfield "kite". This takes into account cross-relaxation between longitudinal spin states only. For those who require the treatment of all cross-relaxation processes, need to specify "secular" rather than "kite". The inclusion of cross-relaxation terms in the simulation, would significantly increase the CPU

time. The last line of the commands specifies the rotational correlation time (τ_C) in seconds. The relaxation rates depend on τ_C.

The next step in the simulation processes is the option to specify the basis set. This is a complicated topic by itself and the reader is referred to the original papers [3, 59]. However the minimal basis set required for accurate results in solution state is given below:

```
bas.formalism='sphten-liouv';
bas.approximation='IK-1';
bas.connectivity='scalar_couplings';
bas.level=5;
bas.space_level=3;
```

The IK-1(5,3) requests connectivity-adaptive basis set that includes local correlations of up to five spins on the *J*-coupling graph and local correlations of up to three spins on the spatial proximity graph. IK-(5,4) should be used for accurate simulations that require the analysis of NOESY peak volumes. In the next stage the function calls the Spinach constructor functions using the command

```
spin_system=create(sys,inter)
```

And generates the `spin_system` data structure – the fundamental structure that contains all information about the spin system. The spin composition of the system was read from the file of chemical shift data. All the protons present in the system were used in the simulation, as the interest was to simulate 2D NOESY NMR spectrum of the molecule supplied, other spins present in the system are excluded from the simulation using the following commands.

```
spin_system=kill_spin(spin_system,strcmp('13C',spin_system.comp.isotopes));
spin_system=kill_spin(spin_system,strcmp('15N',spin_system.comp.isotopes));
```

These spins are either present in ^{12}C and ^{14}N forms, or decoupled. For experiments that require the presence of these spins: HSQC, HCP, HNCO and other sequences, the commands should be altered depending on which spins should be included and/or excluded in the simulation.

The next step is to specify experiment parameters. For the purposes of the 2D ^{1}H-^{1}H NOESY calculations, the following sets of parameters were used:

```
parameters.tmix=0.065;
parameters.offset=3473;
```

```
parameters.sweep=[7500 7500];
parameters.npoints=[512 1024];
parameters.zerofill=[1024 4096];
parameters.spins= {'1H'};
parameters.axis_units='ppm';
parameters.rho0=state(spin_system,'Lz','1H','cheap');
```

Finally, after the parameters are specified, two FIDs: one cosine modulated and the other sine modulated for the pulse sequence is simulated using a Spinach NOESY function and multiplied both by a decaying Gaussian function, which then Fourier transformed both with respect to t_2 using the following commands

```
f1_cos=real(fftshift(fft(fid.cos,parameters.zerofill(2),1),1);
f1_sin=real(fftshift(fft(fid.sin,parameters.zerofill(2),1),1));
```

The imaginary parts are then ignored and the sine modulated data is multiplied by -*i* and added to the cosine modulated data using

```
f1_states=f1_cos-1i*f1_sin.
```

This is then Fourier transformed with respect to t_1 with the command

```
spec=fftshift(fft(f1_states,parameters.zerofill(1),2),2).
```

And then plotted using the `plot_2d` function.

The result from the simulation work is given in **Figure 86** (below); with the spin system where the quantum mechanical simulations were carried out on is shown in **Figure 87** (below).

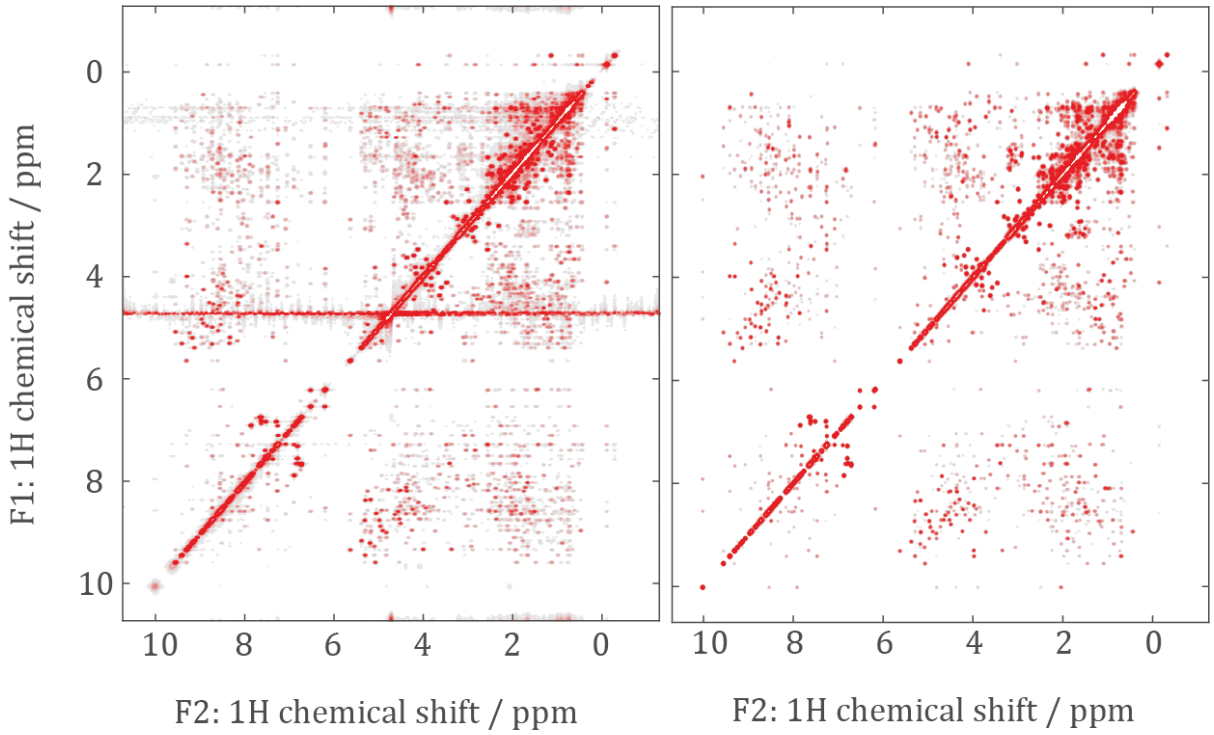

Figure 86 *Ubiquitin NOESY spectra at 900 MHz proton frequency – (left panel) experimental spectrum, data obtained from Christian Griesinger and Donghan Lee. Right panel theoretically simulated spectrum. Simulated spectrum is computed with through space distance cut-off of 4.0 Angstrom for dipolar interactions. Mixing time of 0.065 s was used in the simulation process. BRW relaxation theory with a single global rotational diffusion time of 5 ns was used for relaxation superoperator.*

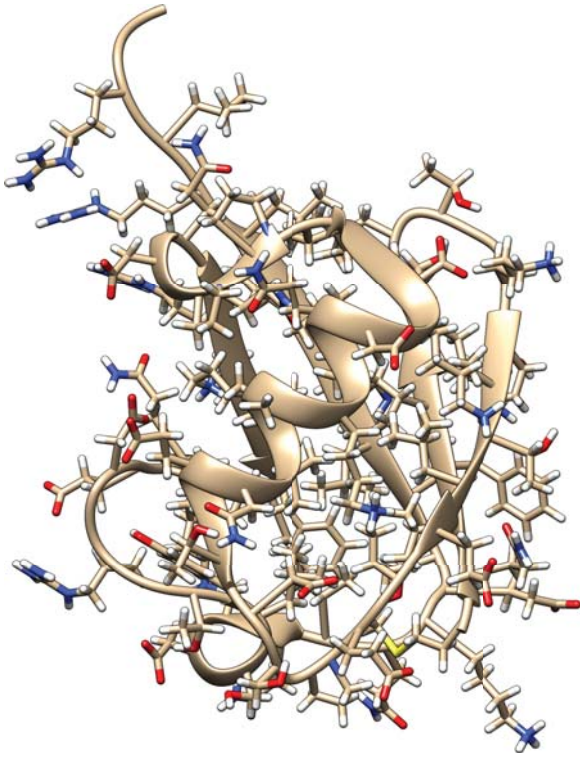

Figure 87 *Ubiquitin (PDB 1D3Z): 76 amino acids, 1060 magnetic nuclei, about 60,000 spin-spin couplings.*

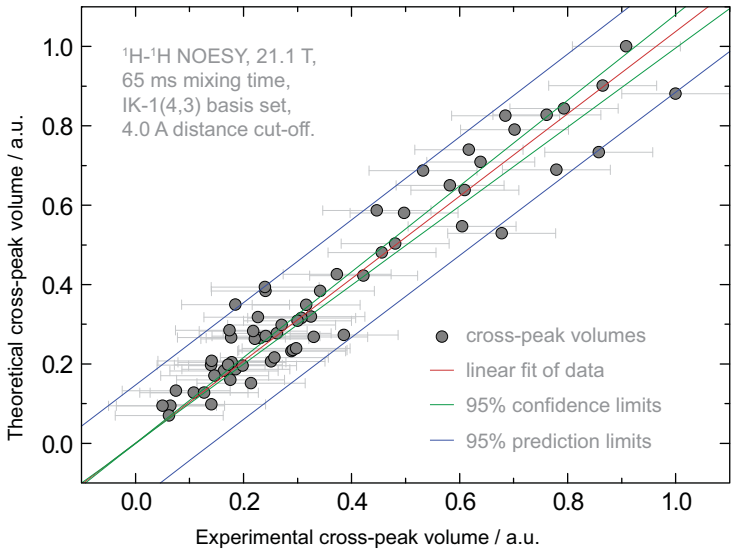

Figure 88 *Correlation between experimental and theoretical ^{1}H-^{1}H cross-peak volume for ubiquitin at 900 MHz proton frequency for the spectra depicted in Figure 86 (above). The intensity between the experimentally and theoretically obtained cross peaks were compared. Simpson's method [146] was used to integrate the volume of the cross peak taking care of the baseline. The cross peak volumes obtained from the experiment were then plotted against those from the theory to observe the quantitative relationship between the spectra from experiment and theory. This figure is copied from our published paper [59].*

Discussion and result analysis

A large number almost hundreds of NOEs are observed in the simulated spectrum **Figure 86** (above). These cross peaks are due to the mutual cross relaxations between two proton spins which happens during the mixing time period of the NOESY pulse sequence **Figure 82 (above)**.

Mixing time is an important parameter for NOE hence one mixing time –a parameter normally required for an NOE experiment, was specified for the simulation *i.e.* the same mixing time used for generating the experimental data. This is normally found to be small due to the large molecular weight of proteins and the value should also be kept as small as possible to linearly approximate NOE build-up and avoid spin diffusion. When the mixing time is too short only diagonal peaks are observed and useful coupling information about the interacting spins start to appear when there is exchange of magnetization between spins via cross relaxation. In a similar fashion the effect of a long mixing time is disappearance of cross peaks from a NOESY spectrum due to the establishment of equilibrium and the loss of frequency labelling.

Overall rotational diffusion and internal motion of protein are the factors that normally modulate spin relaxation in proteins. The term rotational diffusion of biological macromolecules however includes the overall tumbling times and the anisotropic motion. Internal motions are not efficient at driving relaxation but are included at occasions when accurate simulations are required. Overall tumbling time depends on molecular shape and size. However these are not the only factors that affect the tumbling rates. Temperature and solvent viscosities are also important factors in determining the rotation correlation time in which the NOE is dependent on. Thus, larger molecules such as proteins in viscosities solvents have larger correlation times hence negative NOE's, while small organic molecules in non-viscous solvents have positive NOEs. Correlation time along with the distance between the interacting protons determines the position of the maximum NOE and the rate at which the NOE builds up.

The NOE is inversely proportional to the sixth power of the internuclear distance, thus small changes in spin coordinates can bring in significant variations in the observed cross peak intensities. Nevertheless signals are observed when the distance between the spins in not more than 5 Angstroms.

Variation in NOE data were observed from the dynamic property of sidechain when an ensemble averaging was performed. The variability showed little or no impact on the cross

peak NOE simulation we conducted. Note, however that this is a conclusion from our observation from a single project study carried out by Zihan Fang.

Cut-off variation from 4 to 5 or 6 Angstrom increases the volume (and therefore the number of active spins) used for simulation and this effect goes cubically with the radius. In a similar way the complexity increases cubically with the number of active spins used to approximate the simulation, causing the computational effort to rise very steeply.

When *J*-couplings are ignored from the simulation process the quality of the simulated spectra would be affected as the multiplicity pattern vanishes and also all transfer routes across zero-quantum coherences disappears. It is likely that the simulation would probably be a passable approximation, however it would not deem to be quantitative.

An ensemble of structures for ubiquitin are available from the PDB (1D3Z), however a single structure was used in the NOESY simulation conducted as part of the project. The cross-peak volume was analysed and fitted to those from experiment and showed linear correlation **Figure 88**. For comprehensive peak volume analysis every off-diagonal peak should have been included however, this was impossible with the techniques currently available to us. Thus the cross-peaks of the spectrum were cut into small squares with the volume underneath each square is computed and ignoring squares with low integration value from the analysis.

To integrate the overall motion of ubiquitin, cross-peak volume from a single structure and that from an ensemble average was compared and showed a negligible multi-structural effect, on a separate project study where the results are not published.

Based on the results from the simulation the 4 Å dipolar distance cut-off used is sufficient for ubiquitin NOESY simulation. However, in cases where the computational resources are available 5 Å threshold is recommended for better accuracy.

4.6 Conclusions

An NMR spectrum of protein spin system has been simulated using the reduced state space approximation method. Ubiquitin that possesses 76 amino acids, 1060 magnetic nuclei, about 60,000 spin-spin couplings is used as an example. In the simulation processes the spin system was treated as a whole in one go - without cutting it into fragments. The simulation procedure discussed above has massively improved the simulation time compared to the exact ones. Redfield relaxation superoperator with a single global rotational diffusion tensor including cross-correlations was used in the calculation and was computed using the diagonalization-free algorithm [125]. In practice the simulation required shielding tensors, *J*-couplings, quadrupolar tensors and Cartesian coordinates for each participating spin. The simulation given above was carried out on a Dell PowerEdge 820 system with 32 Intel Sandy Bridge cores and 1024 GB of memory. In the latest versions of Spinach the memory requirement is reduced to 128 GB.

5 Large-scale simulations and fitting of chemical exchange NMR

5.1 Introduction

NMR spectroscopy is useful for the study of exchange process [147-149] at atomic level for biologically important molecules, such as proteins and nucleic acids. Exchange behaviour depends on the conformations of the molecules which can be captured by NMR due to the chemical shift differences between conformations.

NMR spectroscopy can also be used to quantify chemical exchange processes in the time scale from microseconds to several hours. C. D. Kroenke *et al* [148] has quantitatively determined chemical exchange effects in the timescale micro to milliseconds, which was demonstrated in the study of *Escherichia coli* ribonuclease H. Exchange processes are classified into three types [23, 35, 150], depending on the time scale.

1. "Slow" time scale characterizes motions in the micro to millisecond range [148, 151]. This is known to affect the line shape and chemical shift of NMR signal. The slow time scale motion contribution to the transverse relaxation can be studied using different techniques depending in the type of the motion. Motions in the microsecond to millisecond can be investigated using CPMG (Carr-Purcell-Meiboom-Gill) [39, 46] based experiments. For processes in the millisecond to second, especially for ^{15}N nuclei, the ZZ-exchange spectroscopy is suitable to measure the rate exchange.

2. "The intermediate" regime which can be detected in the nanosecond to microsecond is known for broad and coalesced peak, can be described using the evolution of the density matrix over time. Motions which are in the range of fast microsecond region can be studied using dynamic line shape analysis or $R_{1\rho}$ [42] experiments.

3. "Fast" exchange processes characterize motions in the picosecond to nanosecond time scale. The fast regime is usually studied using relaxation methods. In a fast exchange process the contribution of exchange is determined by T_2 experiments.

Exchange processes alter T_2 in NMR experiments and this effect may be similar to the normal relaxation processes [150]. The former is distinguishable in the property that they increase the transverse relaxation rate R_2 while the longitudinal relaxation normally remains unaffected by the exchange. Generally, there are two ways to analyse exchange effects, namely the lineshape analysis and magnetization transfer [35]. However the analysis depends on the region of exchange and the degree of the order (first or second) of the exchange process. NMR spectra and relaxation studies are useful for the analysis of conformational and chemical exchange in biomolecules. Chemical shift and spin-spin couplings change the appearance of spectra in the presence of exchange processes. Relaxation however senses chemical and conformational processes depending upon line width. These NMR parameters are very useful to determine the kinetics and thermodynamic processes on the microsecond to second time scale, which is characteristic for the dissociation and association of protein complexes. **Table 7** (below) shows the classification of exchange processes depending on different NMR parameters. In NMR time scale, a slow or fast classification of rate constant is categorized depending on how fast or slow is the exchange rate in relation to the difference of the resonance frequency of the two states [23, 35, 152].

Table 7 *Showing the classification of exchange processes depending in the NMR timescale. It is defined as the difference of magnetic signal in the two environments relative to the exchange rate. Exchange rate is affected experimentally by an increase in the temperature of the system. Increasing the temperature drives the system into the fast exchange regime.*

Exchange processes			
NMR Time scale	**Slow exchange**	**Intermediate exchange**	**Fast exchange**
Chemical shift	$k \ll \delta_A - \delta_B$	$k = \delta_A - \delta_B$	$k \gg \delta_A - \delta_B$
Coupling constant	$k \ll J_A - J_B$	$k = J_A - J_B$	$k \gg J_A - J_B$
Relaxation time	$k \ll \frac{1}{T_{2,A}} - \frac{1}{T_{2,B}}$	$k = \frac{1}{T_{2,A}} - \frac{1}{T_{2,B}}$	$k \gg \frac{1}{T_{2,A}} - \frac{1}{T_{2,B}}$

Anything in between the two extremes is considered as intermediate. There is a simplified mathematical expression for the intermediate regime processes and detailed description of this

can be found in reference [23, 35]. Now back to the description of the spectra in **Figure 90** (below). When the exchange rate is slow, two separate signals are observed at the resonance frequency of the nuclei in the two sites. In contrast in the fast exchange regime a single resonance at the mean frequency of the two sites is observed. The computed spectra in **Figure 90** are a special case of equal populations of the most common two site exchange processes. A variety to this is when the population of the two states are different. An example of this case can be found in many NMR text books. This is the case when concentration of the two states are different hence their rate constant [23]. The first step in the analysis of exchange processes is to identify to which region of the exchange the spectra belongs. EXSY [38] is a technique normally employed to establish the presence of slow exchange. A selective inversion experiment is useful for quantitative analysis. Exchange process affects NMR spectra in variety of different ways, however all follow the same basic mechanisms.

5.1.1 Ligand Binding

Ligand binding is the study of binding of a ligand to a macromolecule such as proteins and nucleic acids. Ligand binding may induce structural changes. The structural rearrangements are large at binding sites compared to other parts of the macromolecule-ligand complex. The interaction of ligand to a macromolecule partner is due to the presence of binding sites that are charged and/or hydrophobic. An example of ligand binding study with a positively charged binding pocket is given in the paper by Anderson *et al*. [153]. The binding of ligand to macromolecule happens by Van der Waals forces, ionic bonds, and hydrogen bonds. The rate of ligand binding is known as affinity and is determined by K_d. It is used to characterize the strength of binding effect. Protein complexation occurs in the region of slow exchange processes, typically around 10^{-6} to 10^{-1} seconds. In dynamic NMR, this may be assumed in the simplest case as a two-site exchange processes, observed at two distinct chemical environments. These can be studied by chemical shift line shape analysis using Bloch-McConnell equation given in Equation (127).

Ligand binding can be mapped using chemical shift changes in the case where the resonance difference associated with the binding is on the fast time scale. However the slow time scale binding kinetics of the bound state should be established using alternative or additional experiments. One of the application of ligand binding is in drug design. Thus it is important to understand ligand binding mechanism at molecular level.

5.1.2 Chemical shift mapping

Chemical shift perturbation mapping method is used to map the interface of protein complexes [154]. In such experiments, a ^{15}N-^{1}H HSQC spectrum of the protein is monitored under stepwise addition of another protein or ligand. The interaction affects the resonance frequency of the NMR active nuclei in that area.

5.1.3 Visualisation

Chemical shift changes due to titration are mapped to the protein structure using a weighted average of the chemical shift difference of the amide ^{1}H and ^{15}N backbone resonances [155] using the formula $d = \sqrt{\frac{1}{2}\left[\delta_{\mathrm{H}}^2 + \alpha \cdot \delta_{\mathrm{N}}^2\right]}$. The value for α used to map the residue shift in calmodulin (PDB: 1CLL) depicted in **Figure 89** (below) was 0.14. Another application of resonance perturbation is to map side chain interactions upon protein complexation [156]. Side chains paly a dominant role in molecular recognition. Chemical shift perturbation mapping provide the location of the binding site, however it doesn't explain or describe how the partners interact at atomic level. Shift perturbation experiments can be used to estimate the affinity, specificity, and kinetics of the binding.

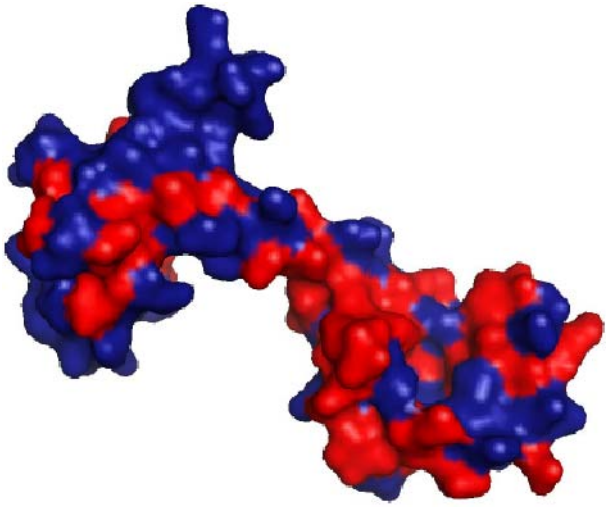

Figure 89 *PyMOL surface representation of calmodulin titrated using a ligand with the geometrical distance moved higher that 0.2 ppm by the peak. Red illustrates hetronuclear N-H spin interactions of the residues with a shift greater than 0.2 ppm. The NMR data of the spin system were obtained from Joern Werner and Kelly Hooper.*

5.2 Classical theory: Bloch-McConnell equation

Below is the derivation of the Bloch-McConnell equation for a simple system with two sites of different chemical environment and a single spin exchanging between two sites A and B *i.e.* for a reaction scheme given in Equation (131).

The Bloch equations for the transverse magnetization without chemical exchange are

$$\frac{d}{d\mathrm{t}}\begin{bmatrix} M_{\mathrm{A}}^{+} \\ M_{\mathrm{B}}^{+} \end{bmatrix} = \begin{bmatrix} i\omega_{\mathrm{A}} - R_{2\mathrm{A}} & 0 \\ 0 & i\omega_{\mathrm{B}} - R_{2\mathrm{B}} \end{bmatrix} \begin{bmatrix} M_{\mathrm{A}}^{+} \\ M_{\mathrm{B}}^{+} \end{bmatrix} \qquad (123)$$
$$M^{+} = M_{\mathrm{x}} + iM_{\mathrm{y}}$$

where $\begin{pmatrix} M_{\mathrm{A}}^{+} \\ M_{\mathrm{B}}^{+} \end{pmatrix}$ are complex magnetization amplitudes. If we now take the case when the spin interchanges between the two states which are of different frequency.

The kinetics of the system that interconverts between two states can be described by the rate equations below

$$\begin{aligned} \frac{d[\mathrm{A}]}{dt} &= -k_{1}[\mathrm{A}] + k_{-1}[\mathrm{B}] \\ \frac{d[\mathrm{B}]}{dt} &= k_{1}[\mathrm{A}] - k_{-1}[\mathrm{B}] \end{aligned} \qquad (124)$$

The kinetic rate matrix is

$$\begin{bmatrix} -k_{1} & k_{-1} \\ k_{1} & -k_{-1} \end{bmatrix} \qquad (125)$$

This kinetic equation can be incorporated into the Bloch equation to give the equation which is known as the Bloch-McConnell equation, which is given as:

$$\frac{d}{d\mathrm{t}}\begin{bmatrix} M_{\mathrm{A}}^{+} \\ M_{\mathrm{B}}^{+} \end{bmatrix} = \left\{ \begin{bmatrix} \mathrm{i}\omega_{\mathrm{A}} - R_{2\mathrm{A}} & 0 \\ 0 & \mathrm{i}\omega_{\mathrm{B}} - R_{2\mathrm{B}} \end{bmatrix} + \begin{bmatrix} -k_{1} & k_{-1} \\ k_{1} & -k_{-1} \end{bmatrix} \right\} \begin{bmatrix} M_{\mathrm{A}}^{+} \\ M_{\mathrm{B}}^{+} \end{bmatrix} \qquad (126)$$

This can be rewritten in a matrix form:

$$\frac{d}{d\mathrm{t}}\begin{pmatrix} M_{\mathrm{A}}^{+} \\ M_{\mathrm{B}}^{+} \end{pmatrix} = \left[i\boldsymbol{\Omega} + \mathbf{R} + \mathbf{K} \right] \begin{pmatrix} M_{\mathrm{A}}^{+} \\ M_{\mathrm{B}}^{+} \end{pmatrix} \qquad (127)$$

where $\boldsymbol{\Omega}$ is the frequency matrix, $\mathbf{R}$ is relaxation matrix, and $\mathbf{K}$ – the exchange matrix are given by

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_{\mathrm{A}} & 0 \\ 0 & \omega_{\mathrm{B}} \end{pmatrix};\ \mathbf{R} = \begin{pmatrix} -\frac{1}{T_{2}} & 0 \\ 0 & -\frac{1}{T_{2}} \end{pmatrix};\ \mathbf{K} = \begin{bmatrix} -k_{1} & k_{-1} \\ k_{1} & -k_{-1} \end{bmatrix}. \qquad (128)$$

Equation (127) is a first order differential equation that can be solved as:

$$\begin{pmatrix} M_{\mathrm{A}}^{+}(t) \\ M_{\mathrm{B}}^{+}(t) \end{pmatrix} = e^{[i\boldsymbol{\Omega}+\mathbf{R}+\mathbf{K}]t} \begin{pmatrix} M_{\mathrm{A}}^{+}(0) \\ M_{\mathrm{B}}^{+}(0) \end{pmatrix} \tag{129}$$

The longitudinal and transverse magnetization components evolve independently and are given by Equations (130) and (129) respectively.

$$\begin{pmatrix} M_{\mathrm{z}}^{(\mathrm{A})}(t) \\ M_{\mathrm{z}}^{(\mathrm{B})}(t) \end{pmatrix} = e^{[i\boldsymbol{\Omega}+-\mathbf{R}_1+\mathbf{K}]t} \left(\begin{bmatrix} M_{\mathrm{z}}^{(\mathrm{A})}(t) \\ M_{\mathrm{z}}^{(\mathrm{B})}(t) \end{bmatrix} - \begin{bmatrix} M_0^{(\mathrm{A})} \\ M_0^{(\mathrm{B})} \end{bmatrix} \right) + \begin{bmatrix} -k_1 & k_{-1} \\ k_1 & -k_{-1} \end{bmatrix} \cdot \begin{pmatrix} M_0^{(\mathrm{A})} \\ M_0^{(\mathrm{B})} \end{pmatrix} \tag{130}$$

where $\mathbf{R}_1$ is the longitudinal relaxation matrix. This can further be adapted to study transient phenomena [18], in which a system starts from non-equilibrium and monitor the return to equilibrium overtime.

5.3 Simple examples

A classical example is the first order isomerization reaction with different forward and backward rate [23, 151, 157].

$$\mathrm{A} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{B} \tag{131}$$

where A and B are the states of the system and k is the rate of exchange between the states.

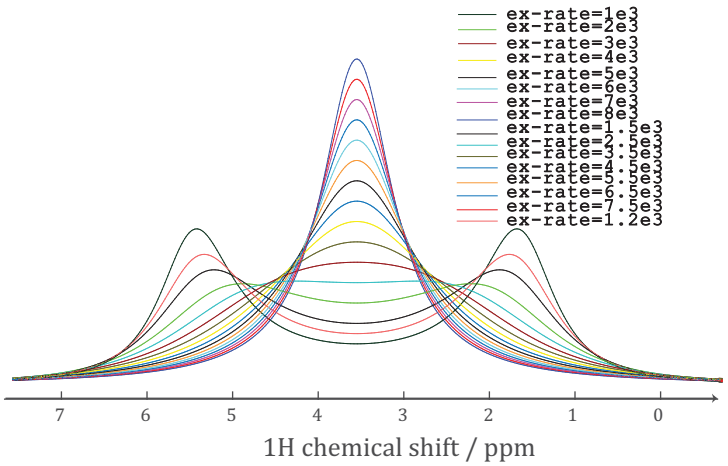

Figure 90 *Simulation of two site exchange process depicting change in line shape as k_{ex}, the rate constant abrivated as ex-rate in the legend moves on from slow to fast in NMR time scale. These NMR spectra are*

calculated using Spinach [3]. NMR frequency of 600 MHz with 900 Hz offset and sweep width of 5000 Hz were used as parameters in the simulation. The exchange rate is in Hertz.

The spectra in **Figure 90** (above) are calculated using different values of the rate constants, k. NMR parameters that are distinct for states A and B are the chemical shift (δ_A, δ_B), longitudinal ($R_{1,A}$, $R_{1,B}$) and transverse ($R_{2,A}$, $R_{2,B}$) relaxation rates.

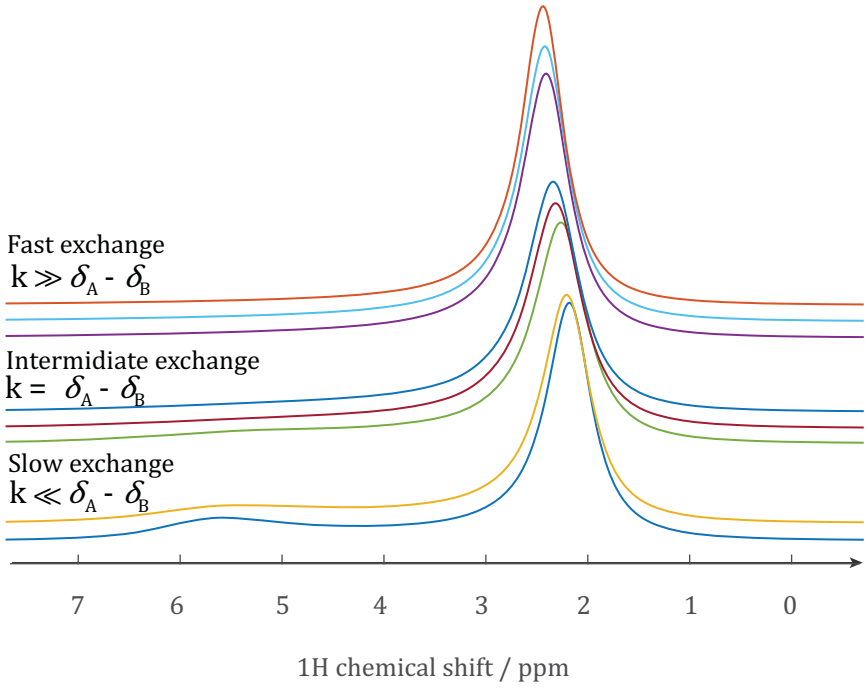

Figure 91 *Computed NMR spectra for a pair of nuclei in two states that possess different populations using NMR proton Larmor frequency of 600 MHz with an offset and sweep width of 900 Hz and 5000 Hz respectively. Where k indicates the exchange rate constant.*

The population of the two states should be equal to unity *i.e.* $(p_A + p_B = 1)$. The lifetimes τ_A and τ_B for the two states can be given by the equation below:

$$\tau_A = \frac{1}{k_A}, \qquad \tau_B = \frac{1}{k_B}, \tag{132}$$

where k_A and k_B are the exchange rates for the states.

In the slow regime, the effect of the exchange at the two sites can be given by the equation below.

$$\Delta\omega_{\mathrm{A}} = \frac{1}{\pi\tau_{\mathrm{A}}}, \qquad \Delta\omega_{\mathrm{B}} = \frac{1}{\pi\tau_{\mathrm{B}}}, \tag{133}$$

where $\Delta\omega_{\mathrm{A}}$ and $\Delta\omega_{\mathrm{B}}$ are the change in line width due to exchange effect at states A and B respectively. For the fast regime a single peak at the weighted average of the resonance frequency of the two sites is observed. Its line width is given by

$$\Delta\omega = \frac{4\pi \mathrm{p}_{\mathrm{A}}\mathrm{p}_{\mathrm{B}}(\delta\omega)^2}{k_{\mathrm{A}} + k_{\mathrm{B}}}, \qquad \text{where} \quad \Delta\omega = \omega_{\mathrm{A}} + \omega_{\mathrm{B}}. \tag{134}$$

5.4 Why Bloch-McConnell equation can't describe most NMR pulse sequences

The Bloch-McConnell equation used in the study of exchange processes assumes to describe the process in the absence of scalar coupling. That is, the equations are derived to illustrate the evolution of magnetization for isolated spins. HSQC, however, is a two dimensional experiment that transfers magnetization between ^{1}H and heteronuclei via scalar coupling.

5.5 Quantum mechanical description of chemical exchange

The most convenient way to approach exchange processes for complicated systems that involve two or more spins that are coupled, is the density matrix.

The equation of motion of ρ includes the frequency matrix, $\mathbf{L}$, the relaxation matrix, $\mathbf{R}$, and the exchange rate matrix, $\mathbf{K}$, to provide the following equation:

$$\frac{\partial}{\partial t}\rho(t) = -\left[i\mathbf{L}+\mathbf{R}+\mathbf{K}\right]\rho(t) \tag{135}$$

It can be solved using a matrix exponential:

$$\rho(t) = \exp\left(-\left[i\mathbf{L}+\mathbf{R}+\mathbf{K}\right]t\right)\rho(0) \tag{136}$$

In NMR, the spectrometers usually detect magnetic signals along the transverse x- or y- axis. $\hat{I}_{\mathrm{x}}$, is the operator that is used to represent the magnetization. This operator is a vector in the Liouville space. Thus the measured signal $\mathrm{F}(t)$ can be described by

$$\mathrm{F}(t) = \left\langle \hat{I}_{\mathrm{x}} \mid \rho(t) \right\rangle = \left\langle \hat{I}_{\mathrm{x}} \mid \exp\left(-\left[i\mathbf{L}+\mathbf{R}+\mathbf{K}\right]t\right)\rho(0) \right\rangle \tag{137}$$

Density operator equation that describes two-site exchange processes has been implemented into SPINACH and is given in Equation (138).

$$\frac{d}{dt}\begin{pmatrix}\hat{\rho}_{\mathrm{A}}\\ \hat{\rho}_{\mathrm{B}}\end{pmatrix}=-i\begin{pmatrix}\hat{\hat{H}}_{\mathrm{A}}+i\hat{\hat{R}}_{\mathrm{A}} & 0\\ 0 & \hat{\hat{H}}_{\mathrm{B}}+i\hat{\hat{R}}_{\mathrm{B}}\end{pmatrix}\begin{pmatrix}\hat{\rho}_{\mathrm{A}}\\ \hat{\rho}_{\mathrm{B}}\end{pmatrix}+\begin{pmatrix}-k_{+}\hat{\hat{E}} & k_{-}\hat{\hat{E}}\\ -k_{-}\hat{\hat{E}} & k_{+}\hat{\hat{E}}\end{pmatrix}\begin{pmatrix}\hat{\rho}_{\mathrm{A}}\\ \hat{\rho}_{\mathrm{B}}\end{pmatrix} \tag{138}$$

where $\hat{\rho}_{\mathrm{A,B}}$ are density matrices of the spin systems of chemical species A and B, $\hat{\hat{H}}_{\mathrm{A,B}}$ are their spin Hamiltonian commutation superoperators, $\hat{\hat{R}}_{\mathrm{A,B}}$ are their relaxation superoperators, $\hat{\hat{E}}$ is a unit superoperator and $k_{\pm}$ are the reaction rates.

5.6 Fitting method

Here in, the project work ^{1}H-^{15}N HSQC spectra of ligand titrated calmodulin protein were fitted to investigate kinetics and extract exchange parameters for the protein system. And this was achieved using least squares method [158, 159].

Least squares is a mathematical approach to find optimum parameters that best fit a function, $f(x)$, of data points $x_1, x_2, x_3 \ldots x_n$, with the purpose to minimize the sum of the squares of the deviations between the observed and calculated values of the data. This is measured using "error functional".

For a given set of data points $\{y_i, x_i\}$ and any function $f(x;\vec{\alpha})$, the "error functional" can be described using Equation (139):

$$s(\vec{\alpha})=\sum_{i=1}^{m}\left[y_i-f(x;\vec{\alpha})\right]^2 \tag{139}$$

where $s(\vec{\alpha})$ is the residual of the sum of squares of all the elements of the data points, *i.e.* the difference between the actual data acquired from experiment and the model function with respect to the parameter $\vec{\alpha}$.

The objective function was minimized using Nelder–Mead simplex method [160, 161]. In numerical analysis the Nelder-Mead simplex belongs to the direct search method category. The method is used to minimize a real valued function. In addition it is a technique which is used

to handle functions which are discontinuous, noisy, nonlinear, and non-differentiable. The method minimizes a function of several variables by computing a simplex from an initial estimate. This is summarized as given in Equation (140)

$$f(x),\ x \in \Re^n \tag{140}$$

This is followed by the evaluation of the function at the points and reflection operation is performed on the worst value of the function through the centroid of the other points. Depending on the value of the function at the reflected point a contraction or expansion step is taken. These steps are represented by β and γ variables. The expansion coefficient is usually greater than unity while that of the contraction step is between 0 and 1. The Nelder-Mead generation of simplex and steps to minimum is depicted in **Figure 92** (below).

Figure 92 *Simplex, which is a polygon of three vertices in two dimensional spaces, followed by reflection, expansion, contraction, and shrinkage. The figure is taken and modified from Ilya Kuprov's Spin Dynamics lectures.*

The objective function, which is the least square function, computes the square of the difference between the theory and experimental spectra as given in Equation (141).

$$\chi^2(a,b) = \sum_{n=1}^{N}\left(\left(\text{spec\_theo}(\text{parameters})\right)_n - \left(\text{spec\_exp}\right)_n\right)^2 \tag{141}$$

where $\chi^2(a,b)$ is the value of the function at the parameters, spec_exp - the experimental spectra, $\text{spec\_theo}(\text{parameters})$ is the theoretical spectra with the parameters to be optimized and N is the number of iterations.

5.7 Fitting of exchange processes applied to ^{1}H-^{15}N HSQC spectra of calmodulin

Joern Werner and Kelly Hooper at the centre for biological science of the University of Southampton were interested in a protein simulation task collaboration, and requested us to run simulations of chemical exchange between different protein conformations from an NMR titration experiment.

The system investigated was a protein calmodulin (PDB: 1CLL) titrated by a ligand with the reaction monitored using HSQC, a 2D nuclear magnetic resonance technique that stands for heteronuclear single quantum coherence.

Calmodulin abbreviated as CaM for calcium-modulated protein is a ubiquitous calcium binding protein that plays an important role in Ca^{+2} signal transduction within the cell [162]. Calmodulin consists of 148 amino acid residues and has a dumbbell shaped topological structure. It contains four calcium binding cites with two on each of the two globular domains connected by a flexible antiparallel β-sheet linker [163]. Calmodulin is a versatile molecule that interacts with generic proteins involved in a wide variety of functions including cell signalling and ion transport by regulating the activities of kinases and phosphatases [164].

HSQC spectra of Ca^{+2} calmodulin and that of the samples with subsequent addition of ligand were recorded and analysed. As part of our collaboration we quantum mechanically simulated a candidate spectrum and fitted the experimental data to the computed once to extract exchange parameters such as the k_{on} and k_{off}.

HSQC spectra of calmodulin in the presence of chemical exchange were calculated theoretically using Spinach v1.5 [3] for a collection of standalone N–H pairs. To model spin dynamics in the presence of a unimolecular chemical reaction Equation (131), Spinach uses Equation (138) as the equation of motion. The formalism given for the case of two chemical species is a special case of a more general implementation of chemical kinetics in Spinach, which operates in the direct product of spin state space and chemical state space.

The spin part of the problem contains 254 N–H pairs and was therefore handled using the restricted state space approximation [20]. Because a collection of standalone N–H pairs is assumed in this case, state space restriction to two-spin orders connecting directly coupled spins is in this case exact. A single spin pair in Liouville space yields a 16-dimensional problem and therefore the total dimension of the matrices in Equation (138) is 254 x 16 x 2 = 8128 –

small enough to be handled comfortably by the sparse matrix functionality implemented in Spinach.

However, the current project work involves a second order reaction with a chemical reaction structure of $\mathrm{CaM} + \mathrm{Pep} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{CaM{:}Pep}$.

Spinach NMR exchange simulation work requires the following items for that reaction:

I. $^{1}\mathrm{H}$-$^{15}\mathrm{N}$ chemical shifts of Ca^{+2} CaM

II. $^{1}\mathrm{H}$-$^{15}\mathrm{N}$ chemical shifts of Ca^{+2} CaM:Pep complex

III. Forward and backward rates in the $\mathrm{CaM} + \mathrm{Pep} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{CaM{:}Pep}$ reaction, which are assumed to be first order in both directions were extracted as described below.

The parameters supplied to `create.m` function of Spinach are a cell array of the type { [1 2] [3 4]} assigned to `inter.chem.parts`. This specifies a vector of integers *i.e.* spins 1 and 2 belongs to the first chemical subsystem while spins 3 and 4 to the second chemical subsystem. The next parameter is `inter.chem.rates`. This is a matrix of real numbers that indicates the reaction matrix rate between the subsystems specified in the inter.chem.parts variable. Finally the initial concentrations of the subsystems should be supplied to the assigned command variable `inter.chem.concs`.

To proceed with the simulation of the exchanging spectra we required a kinetic binding model. K_{d} the binding affinity for the titrant peptide was measured using ITC and found to be around 200nM. Values obtained from ITC measurements provide useful insights into the binding mechanism. ITC an acronym for isothermal titration calorimetry is an experimental technique used to directly measure the heat generated or absorbed when two biomolecules interact.

Assuming that this K_{d} (200nM) value corresponds to the slow exchange part of the binding our collaborators Joern Warner and Kelly Hooper simplified the multistep binding to a simple and most generic binding model or mechanism. The binding mechanism involved a first order reaction in the absence of scalar coupling that computes equilibrium concentrations in the reaction: $\mathrm{CaM} + \mathrm{Pep} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{CaM{:}Pep}$ given the initial concentrations of CaM and Pep as well as the overall binding affinity K_{d}. With K_{d} is assumed to be equivalent to $\dfrac{[\mathrm{CaM}]\cdot[\mathrm{Pep}]}{[\mathrm{CaM{:}Pep}]}$.

For simple equilibrium systems like this, $K_d = \frac{k_{off}}{k_{on}}$. The diffusion limited k_{on} for the peptide involved from literature shows the values in the range of 0.1M $M^{-1}s^{-1}$ to 0.1G$M^{-1}s^{-1}$ giving a k_{off} in the range of 0.02 to 20 along with a life time $\left(\tau = \frac{1}{k_{off}}\right)$ for the complex formed around 50ms to 50s. However, the very slow and fast k_{on} rate are extreme values, hence a k_{off} of 0.2 to $2s^{-1}$ *i.e.* a life times of 500ms to 5s for the complex formed are considered realistic. This estimate came from Joern Warner and Kelly Hooper our collaborators in the project work.

The titration experiment with a reaction scheme $\mathrm{CaM + Pep} \rightleftharpoons \mathrm{CaM{:}Pep}$ is a second-order reaction where only the effective first order is actually observed. Therefore the following ordinary differential equations would describe the rate of the forward and reverse reaction:

$$\begin{aligned} \frac{d[\mathrm{CaM}]}{dt} &= -k_{off} \cdot [\mathrm{CaM}] \cdot [\mathrm{Pep}] + k_{on} \cdot [\mathrm{CaM{:}Pep}] \\ \frac{d[\mathrm{CaM{:}Pep}]}{dt} &= +k_{off} \cdot [\mathrm{CaM}] \cdot [\mathrm{Pep}] - k_{on} \cdot [\mathrm{CaM{:}Pep}] \end{aligned} \tag{142}$$

This makes the calculation to require the concentration of the peptide in which its solution involves solving a quadratic equation. The law of conservation of matter in an isolated system states that the total quantity of reactants must be equal to that of the products. Thus the concentrations of $[\mathrm{Cam}]$ and $[\mathrm{Pep}]$ can be given by the following equations:

$$\begin{aligned} [\mathrm{CaM}] &= [\mathrm{CaM}]_{tot} - [\mathrm{CaM:Pep}] \\ [\mathrm{Pep}] &= [\mathrm{Pep}]_{tot} - [\mathrm{CaM:Pep}] \end{aligned} \tag{143}$$

where $[\mathrm{CaM}]_{tot}$ and $[\mathrm{Pep}]_{tot}$ are the total concentrations of the protein calmodulin and the peptide. Thus, if we know $[\mathrm{CaM:Pep}]$, we can easily determine the concentration of $[\mathrm{CaM}]$ and $[\mathrm{Pep}]$. This sorts out equilibrium concentrations, as a consequence, the nonlinearity problem of Equation (142). After equilibrium is established, the concentration of the peptide $([\mathrm{Pep}])$ becomes stationary and the quadratic term becomes linear.

To calculate the concentration of $[\mathrm{CaM:Pep}]$ we can substitute Equation (143) into the equation of the overall binding affinity to obtain a quadratic equation, given below in Equation (144)

$$\left([\mathrm{CaM:Pep}]\right)^2 - \left([\mathrm{CaM}]_{\mathrm{tot}} + [\mathrm{Pep}]_{\mathrm{tot}} + K_{\mathrm{d}}\right) \cdot \left([\mathrm{CaM:Pep}]\right) + \left([\mathrm{CaM}]_{\mathrm{tot}} \cdot [\mathrm{Pep}]_{\mathrm{tot}}\right) = 0 \quad (144)$$

As we are working on concentrations, only positive solutions for Equation (144) are considered.

The coded Matlab program for the kinetic model is available under the function name "`kin_model`".

With the state A data came from the naked protein calmodulin with a starting concentration of 100μM and zero for the peptide (practically this was given as a finitely small number), and 20μM for the binding constant (K_{d}). While for state B the values are 100μM for the protein and the peptide and 20μM for K_{d}.

These input data were used for the simulation of the exchanging HSQC spectra of the titration experiment with an example shown in **Figure 93** (below), and were calculated with respect to the rate constants of the forward and backward reactions, where both rates presumed "slow". The parameters for the theoretical data were matched to that of the experiment including the point spacing, timing, and the apodization function. The chemical shifts of the theory and experiment were also aligned to make sure that the signals are overlapping, this consequently makes the data quantitatively fittable.

The only adjustable parameter was the kinetic dissociation rate (k_{dis}) and association rate (k_{ass}) was computed from the equilibrium condition, *i.e.* the equilibrium condition from the concentrations and K_{d} that were provided.

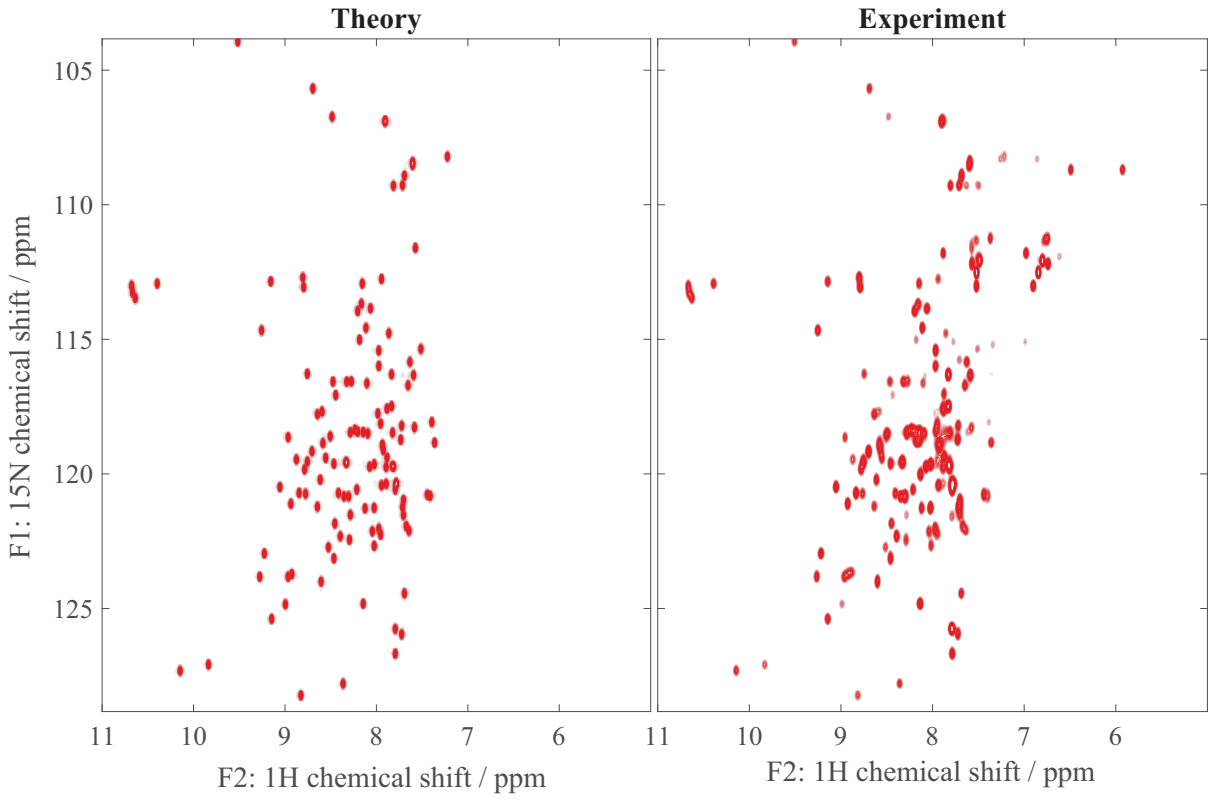

Figure 93 *^{1}H-^{15}N HSQC NMR spectra of calmodulin experimentally obtained right panel and left panel theoretically computed. Two of all the spectra from the titration of calmodulin upon addition of a ligand to show data match. Some missing peaks in the simulated spectrum.*

It is evident from the simulation that the features of titration are reproduced using the simulation work with an example given in **Figure 96**. An example of two residues that are well resolved and in an isolated regions of the spectrum that show exchange characteristics are Asp 64 and Ile 130 at H=8.75-8.95ppm and N=127.5-129ppm, and H=8.25-8.45ppm and N=127.5-129ppm coordinate axis for the two residues respectively. Both the residues show the characteristics of fast exchange processes in the F1 *i.e.* direct and F2 the indirect axis. The exchange phenomenon is also accompanied by a change in direction of peak movement.

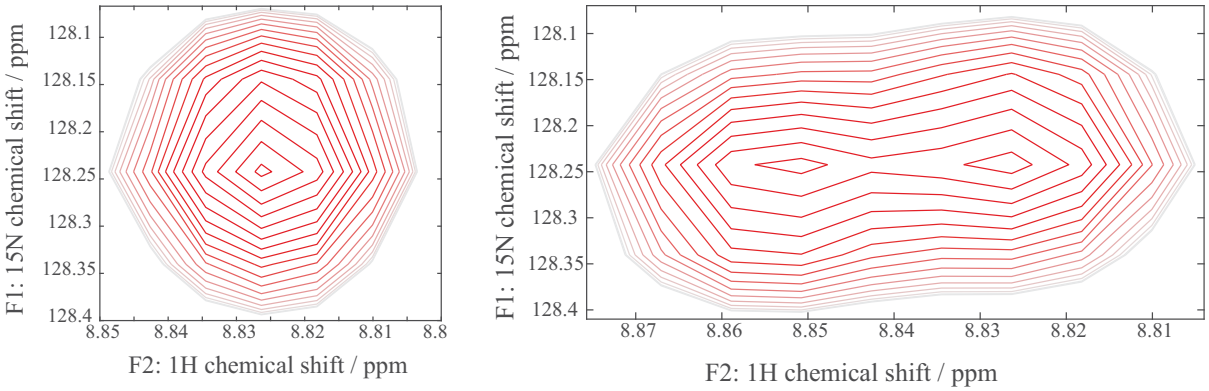

Figure 94 *An illustration of chemical exchange phenomenon from ^{1}H-^{15}N HSQC spectra for the residue Asp 64. The two unresolved peaks on the right panel merge and form a single peak as the concentration of the peptide increased in a titration experiment. The ratio of the concentration of the protein to the peptide is 1:0.05 and 1:1 for the right and left panel, with kinetics matrix of [-0.098, 2; 0.098, -2] and [-0.033, 2; 0.033, -2] respectively. The kinetics matrix consists chemical reaction exchange rate matrix between subsystems as its elements. Due to the law of conservation of matter, the diagonal elements of the kinetic matrix must balance out those that are off diagonal.*

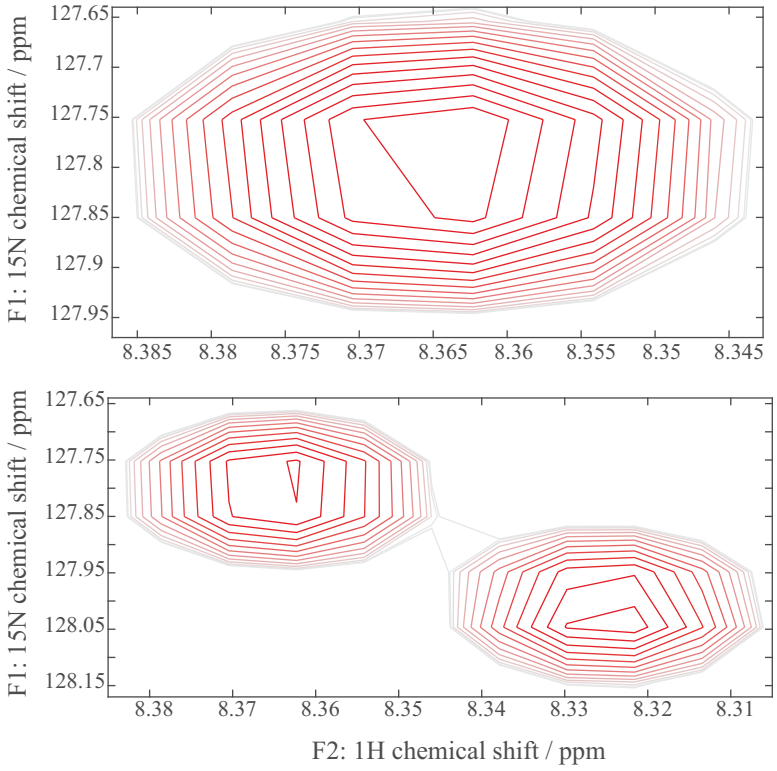

Figure 95 *Illustration of fast exchange phenomenon of a well resolved peak from Ile 130. Top panel – a single peak for complex of 1:1 ratio of protein and ligand, while bottom panel is a peak from 1:0.05 ratio.*

After accomplishing this, *i.e.* simulation of the HSQC spectrum which is almost identical (line shape and position) to that of the experiment– the work proceeded in minimizing the difference of the volumes of the peaks between the experimental and simulated data as a function of the exchange rate. This is conducted using global optimization of the peak to peak fitting as depicted in **Figure 96** of those signals that are well resolved.

The simulated spectra were fitted to the experimental spectra using the least squares method [165] to obtain chi squared (χ^2) – the value of the function evaluated as the parameters passed over to the objective function. The least squares method minimizes the difference between the experiment and the theory. The optimized parameters are K_d– the overall binding affinity, k_{dis}– the dissociation constant, proton and ^{15}N relaxation constants, and intensity multiplier.

In the fitting process K_d was initially fixed and the fitting was run for the other four parameters, and then K_d was varied from result obtained from previous fitting run. An example of the results from the fitting are depicted below in **Figure 96**.

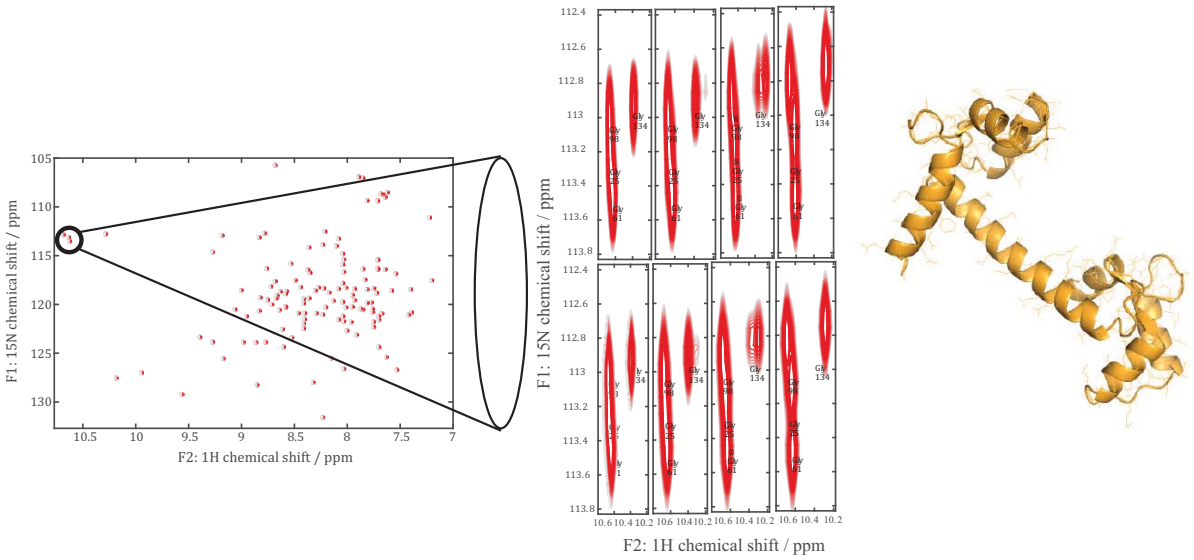

Figure 96 *Fitting of ^1H-^{15}N 2D HSQC spectra with respect to ligand binding kinetic parameters. Left panel ^1H-^{15}N HSQC spectrum of protein ligand complex Figure 93 (above). Middle panel – a fragment of the peak by peak fit undertaken in this work – (Top row) experimental peak and (Bottom row) theoretically simulated peak at three different concentrations of ligand. Good agreement with peak position and line width is apparent between theory and experiment. Right panel Calmodulin (PDB: 1CLL): 148 amino acids, 1134 magnetic nuclei.*

The simulation has reproduced the line shape as it is apparent in **Figure 96**. The values obtained for the fitted parameters from the fitting procedure described above are 9.62, 1.81, -8.09, 79.86, 28.54, 0.58 for the χ^2, the logarithm of k_{dis} and K_d, proton and ^{15}N relaxation constants, and intensity multiplier respectively.

A subtraction plot that is the difference between experimental and simulated peaks was also conducted to investigate the extent of matching between the two data types. An example, a fragment of the peaks studied is depicted in **Figure 97**.

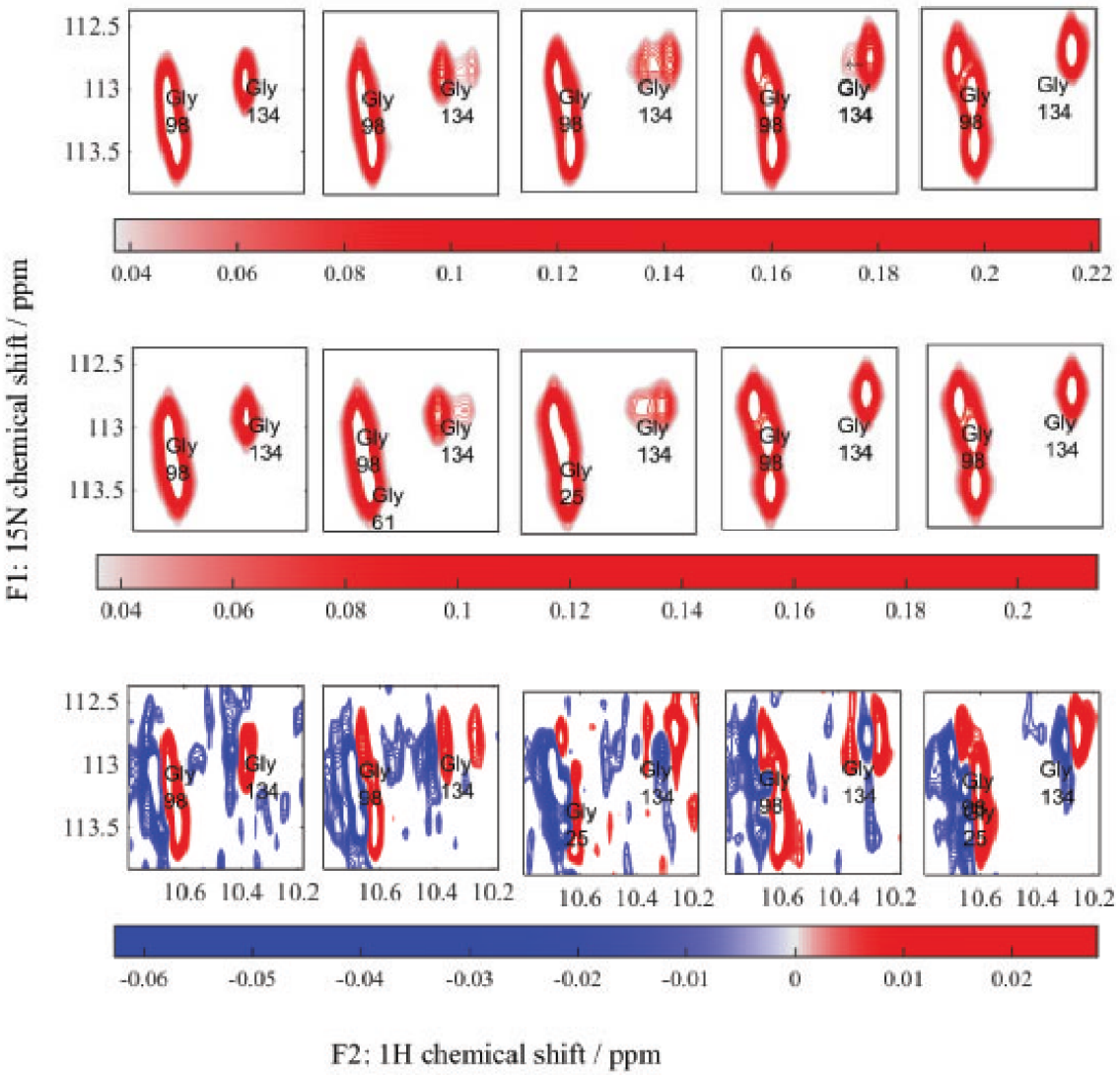

Figure 97 *A fragment from the 2D fitting of ^{1}H-^{15}N HSQC spectra of calmodulin (PDB: 1CLL). The top panel of 5 peaks is from experiment, the middle row is the theoretically simulated peaks while the bottom panel is the difference plot between the experiment and theory.*

It is apparent from **Figure 97** (above) that the difference plot depicts an almost flat peak with a lot of noise. Beneath all the peaks is a colour bar that illustrates the intensity of the corresponding signals. A closer look to the colour bar shows that the intensity of the peaks spans from 0 to ~0.25, and 0 to ~0.20 for the experimental and theoretically simulated peaks respectively. But for the difference plot (bottom panel) it is an order of magnitude smaller than

that of the theory and experiment. In addition the difference plot depicts a considerable amount of noise (apparent blue colour) around the entire area. In this research work, we are yet to make a conclusion on the conformation of the protein calmodulin upon step wise addition of a ligand using line shape analysis as a function of exchange processes. This is as far as we can go for this fitting work and to bring it to a logical end, there needs another simulation work to be conducted using another model that involves two site binding of the ligand to the protein.

Table 8 *Objective function and fitting parameters value extracted from the fitting work conducted as part of this research work. The left most column represents an arbitrary number given to 19 peaks randomly picked from the titration spectra recorded at eleven different concentrations of the peptide to the fixed concentration of the protein calmodulin. The Log_K_d and Log_k_{diss} are the natural logarithm of the binding affinity and dissociation constants respectively. χ is the local minimizer of the error functional i.e. its values are the local minimizers of the objective function (the minimized function for fitting the data from experiment to that obtained from theory).*

Arbitrary number of peak function	**χ**	**Log_k_{diss}**	**Log_K_d**	**R_2 for ^{1}H**	**R_2 for ^{15}N**	**Intensity multiplier – A**
0	9.62	1.81	-8.09	79.86	28.54	0.58
1	0.00	1.24	-8.02	25.56	14.79	
2	3.94	1.79	-5.86	68.71	31.12	0.74
3	0.00	1.41	-7.07	66.24	35.75	0.90
4	0.00	1.48	-6.78	72.63	24.25	0.77
5	4.78	1.58	-6.02	69.10	43.67	0.61
6	5.06	1.77	-6.96	70.32	28.32	0.60
7	2.81	1.89	-7.69	67.44	29.17	0.56
8	7.27	1.79	-6.08	62.76	27.40	0.75
9	1.41	1.70	-6.32	68.66	31.60	0.49
10	0.49	-2.94	-6.60	95.69	41.85	0.55
11	6.21	2.49	-5.65	57.03	28.07	0.86
12	7.06	-0.13	-4.56	44.23	25.11	0.79

13	2.48	-2.74	-6.04	76.02	34.46	0.84
14	3.40	2.68	-5.63	76.02	34.23	0.73
15	1.85	1.94	-6.41	89.98	37.47	0.55
16	1.31	1.20	-11.08	82.39	3.32	0.57
17	0.41	1.77	-5.99	82.12	40.97	0.48
18	16.93	1.92	-6.15	69.29	33.34	0.64

5.8 Concluding remark

This project work involved: the fitting of CaM protein ^{1}H-^{15}N HSQC spectra. The simulation of the spin systems was run quantum mechanically using the reduced state space approximation [20]. The spectra were computed using the standard Spinach functionality with pulse sequence and acquisition parameters matched to the experimental values. As it is discussed in detail in the main text, the spectra from theory were fitted to the data from the experiment using the least squares method with the error functional minimization performed by the Nelder-Mead Simplex algorithm. In practice, the fitting was performed by minimizing the discrepancies between the theoretical model and the data from the experiment. The minimization technique makes sure that the optimizing parameters are optimized to best match the theory to the experiment. The optimized parameters are K_{d} – the binding affinity, K_{dis} – the dissociation constant, proton and ^{15}N relaxation constants, and intensity multiplier.

A peak by peak fitting was conducted for the extraction of parameters used in the evaluation of the conformation of calmodulin using exchange process as the system investigated was treated as a two site exchanging system.

The residual of the sum of squares function employed in the project work was the method of choice for the fitting purposes due to its practicality and firm establishment in statistics. The objective function *i.e.* the residual function is minimized using the Nelder-Mead simplex algorithm implemented in Matlab. The Nelder-Mead simplex algorithm is a direct search method that does not rely on gradient and Hessian of the function. It is the method of choice to handle functions which are discontinuous, noisy, nonlinear, and non-differentiable. The algorithm minimizes a function of several variables by computing a simplex from an initial

estimate and evaluates the function at the points, sorts the values of the function to perform a reflection, expansion and contraction operation and depending on the value of the function at the reflected point a change in shape or size of the simplex is taken as require

6 Quantum mechanical simulation and fitting of nucleic acid NMR spectra

6.1 Introduction

Before delving into the discussion of the quantum mechanical NMR NOESY simulation and fitting of RNA; here in briefly introduced on where RNA comes from and why is it an interesting system to study.

Cell is a basic building unit of structural and functional biology of all living organisms [166, 167]. The cell consists of a compartmentalized organelles, and one of these structures, enclosed at the centre is “the nucleus”. The nucleus contains hereditary genetic particles known as nucleic acids (DNA & RNA). It is the place were DNA transcription and RNA synthesis occurs. DNA is transcribed into special type RNA (mRNA), which is transported out of the nucleus where it gets translated into another important biomolecule “the protein”. Proteins are known for biochemical catalysis, and structure were its specific property and function is determined from the translated hereditary gene [167].

RNA plays an important role in living cells as biocatalysts [168, 169] of chemical reaction (a property discovered in early 1980), gene expression and gene regulation. The route for the transmission of genetic messages is the encoding of the DNA *i.e.* the three letter word sequence (CAG ACT GCC) that contain the genetic code to RNA. The sequence of the code (three letter word) in turn determines the amount and combination of the amino acids in protein synthesis. This convey of information from DNA to proteins carried out by RNA is the base for different structural and functional variations in living organisms. RNA also transfers specific amino acids to a polypeptide chain during translation processes.

Although RNAs are single strand, the property to self-complement makes it to base-pair and fold to form complex three dimensional structure. The shape of the RNA is determined by the amount and combination of the nitrogenous bases. A change in the bases would cause a change in shape or structure, hence function. This versatility in structure makes RNA such an important

molecule in biology and interesting system to investigate its structure and dynamics using NMR and crystallographic technique.

6.2 Chemical structure and NMR spectroscopy of nucleic acids

Nucleic acids are composed of fundamental building blocks called nucleotides. A nucleotide unit consists of three components: the base, the ribose also known as sugar, and the phosphate section. The nucleotides are connected together through phosphodiester bonds attached at both ends to the C5' and hydroxyl at C3' to form a chain of nucleotides. There are two types of nucleic acids, DNA and RNA.

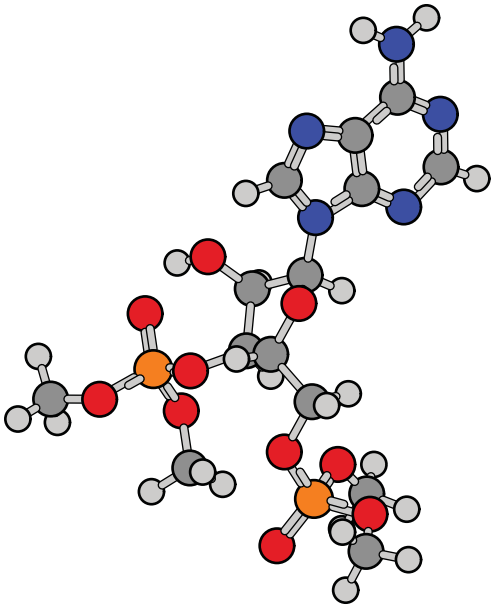

Figure 98 *An example – chemical structure of the adenine nucleotide.*

The difference between DNA and RNA is the elimination of oxygen atom from the sugar component of DNA nucleotides to yield deoxyribose. There are two types of nucleotide bases composed of two groups of nitrogen heterocycle bases, pyrimidine and purine rings.

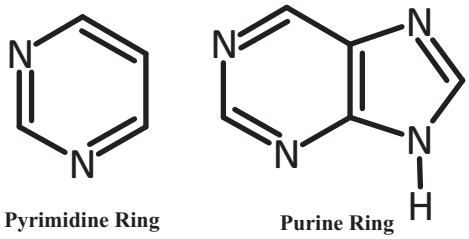

Figure 99 *Chemical structure of the bases for nucleic acids.*

The local conformation of nucleic acids depends on backbone dihedral angles α, β, γ, δ, ε [170], ζ, and the glycosidic bond dihedral angle χ [171]. Changes in these parameters are observed using different NMR experiments. Some examples of these experimental techniques are: HSQC for through bond scalar couplings that records correlations over one bond (^1J), and NOESY that makes use of dipolar coupling between spins to estimate distance. Another experimental technique that is useful in determining backbone dihedral angle is the ^{31}P-^{1}H correlation experiment [172].

The conformation of the sugar component can be identified from the ^3J (H,H) coupling values of proton spins attached at C1' and C2' together with coupling value of the protons at positions C3' and C4' of the sugar ring [173].

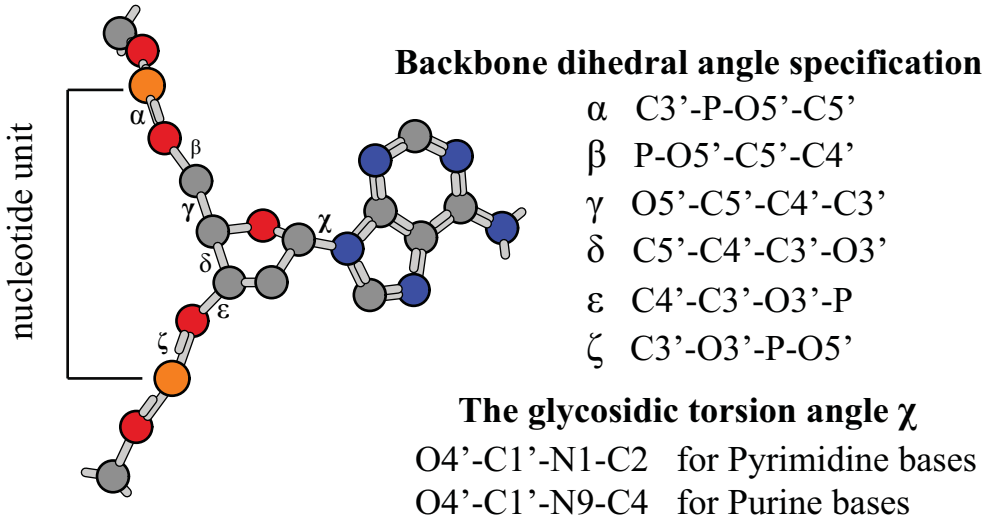

Figure 100 *Graphical illustration of the backbone and glycosidic dihedral angles scanned for a nucleotide unit using Gaussian 09 package at DFT theory level with basis set M06/6-31G**.*

One of the applications of NOESY in nucleic acids NMR is to establish base pairing pattern. There are two type of imino protons in nucleic acid nucleobases. These are the H3 imino proton of uracil and the H1 proton of guanine. The resonance signals of these protons provide information about nucleic acid pairing and are only observed when the spins are involved in hydrogen bonding. Therefore the number of resonance signals from these proton spins gives the number of bases that are paired; 2D NOESY cross peak correlates spins that are a few Angstroms apart [137, 174].

Observation of cross peaks in NOESY experiment is an indication that there are dipolar couplings between spins. These interactions are fundamental for assignment of base-pairing

resonances. In simple models, NOESY cross peaks are related to distance qualitatively and hence grouped into "strong", "medium", and "weak" for the purposes of structure determination depending on the intensity of the observed signals.

Amino protons (example are those of cytosine nucleobase) can be identified by doing NMR with ^{13}C and ^{15}N labelled nucleic acid molecules, which are useful for further correlating to H5 and H6 of the unsaturated ring protons using correlation experiments. 2D (^{15}N, ^{1}H) HSQC experiment in particular, using ^{15}N labelled systems, provides information for the elucidation of base pairing – giving distinct signals for uracil and guanine imino proton resonance signals – where the imino nitrogen resonance frequency has around 10 ppm difference for the two bases. HNN-COSY is also another technique used to determine scalar coupling across hydrogen bonds that connects base pairs to elucidate secondary and tertiary structures.

The complete assignment of nucleic acid resonances can be accomplished by the identification of the spin resonance frequencies of the nucleobase and assignment of nuclei such as H^N, NH_2, H_2, H5, H6 and H8. The assignment of sugar proton spins is then followed. Once these are established, the connection between ribose and nucleobase spin system can be established using NOE enhancements observed for H1 proton of sugar and aromatic spins of the nucleobase. HCN type techniques can also correlate the same proton of the sugar to aromatic spins [175]. The ribose spin system is then completed by using HCCH – TOCSY [176, 177] experiment.

Finally, sequential NOE experiments establishes the sequential connection between adjacent nucleotides. Other applicable experiments include HCP and the HCP – TOCSY techniques.

A variety of nucleic acid structure calculation strategies are described in the literature [178-180]. However all have one thing in common, *i.e.* all require some of the following NMR parameters for structure calculation [181]:

- NOESY cross-peak for distance constraint
- Chemical shifts – to identify the type of each spin
- Homonuclear and heteronuclear coupling constants – identifies nearest neighbour and next nearest neighbour spins. For nuclei that are three bonds apart, the scalar coupling value depends on the corresponding dihedral angle.
- Cross correlated relaxation rates – most commonly used parameters are the DD-CSA or the DD-DD relaxation rates.

- NOE from hydroxyl (OH) groups.
- Residual dipolar couplings (RDC)

6.3 *J*-coupling predictor function for RNA

We need all *J*-couplings in RNA for our simulation work. Hence, we looked for scalar coupling values from literature, as many of the *J*-couplings are not available in literature, we did DFT estimation of *J*-coupling values for values that could not be located in literature.

Nucleic acid *J*-coupling predictor function was built and implemented into Spinach [3].

Vicinal coupling data from DFT method GIAO M06/cc-pVTZ in SMD Water (a detailed description of the methods involved is given in chapter three) was fitted to Karplus equation, with an example result is given below in **Figure 101** (below). The equation relates local dihedral angle to three bond scalar spin-spin couplings.

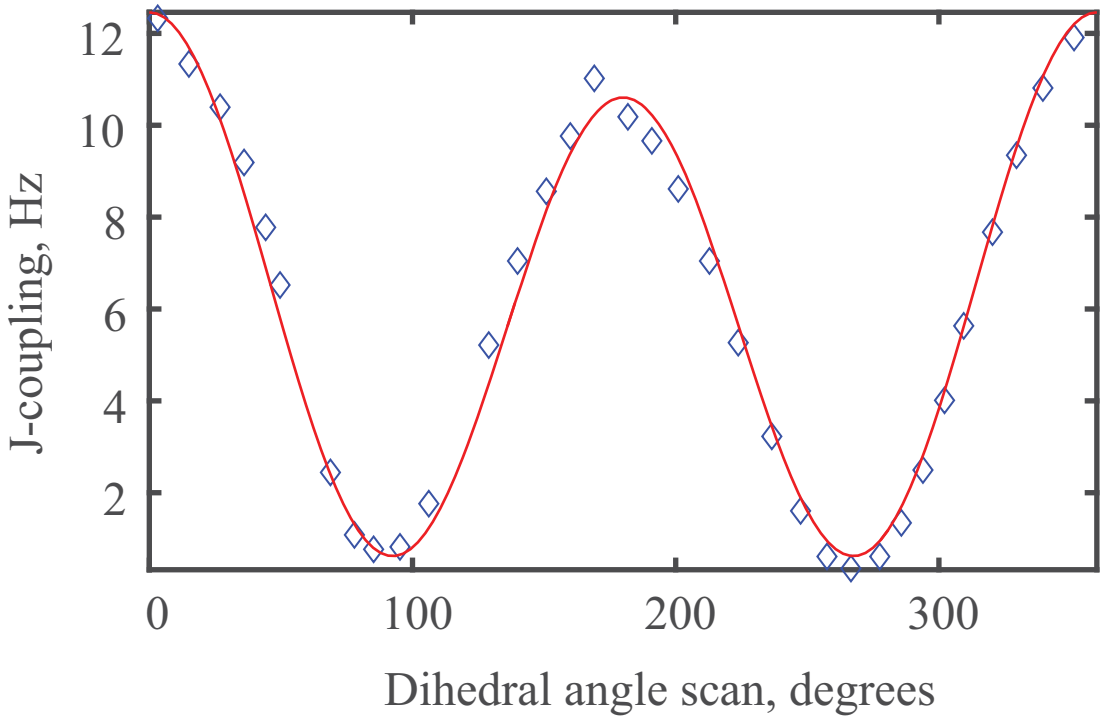

Figure 101 *Fitted curve of* ${}^{3}J_{H4H5pp}$ *coupling constant at the* γ *dihedral angle. Scanned coordinates at position* $C^{i-1}NC\alpha C^{i+1}$ *of the* γ *dihedral angle of uracil nucleotide. The fitting curve is achieved using Karplus equation* $\left({}^{3}J = A \cdot cos^{2}\left(\phi\right) + B \cdot cos\left(\phi\right) + C\right)$. *The Karplus coefficients derived from the fitting have values of 10.89, 0.93, and 0.64 for A, B, and C of the Karplus equation in hertz. The data points "blue diamond" are J-coupling values from electronic structure calculations with DFT method GIAO M06/6-31G** in [SMD] [Water] and the curve is the model applied Karplus equation.*

The method proceeds by splitting the molecular bonding graph into clusters of sequentially connected four atoms (*i.e.* connected by three bonds). The dihedral angle is computed from atomic coordinates, and the three bond *J*-couplings from a database of Karplus curves.

The work primarily involved the compilation of data that would allow to simulate the NMR spectra of RNA specifically, and nucleic acid molecules in general.

To obtain the coupling between all spins in nucleic acids, the calculation of *J*-coupling constants for each of the five nucleotides has been done using DFT theory level with M06 exchange-correlation functional for geometry optimization and for *J*-couplings along with cc-pVTZ basis set. Practically the computation required the construction of the nucleotides structure using Gauss View 5.0.9 package, then optimization into a minimum energy molecular structure before running the spin-spin coupling calculation. Atomic coordinates in Gaussian 09 can normally be specified in two ways: the Cartesian coordinate system and the internal coordinate system also known as Z-matrix. By default geometry optimization in Gaussian is performed in internal coordinates. Pulay et al. [182, 183] has demonstrated that internal coordinates are preferred for optimizing large flexible systems to the Cartesian coordinate, which is efficient for handling rigid ring systems.

Gaussian uses quasi-Newton BFGS algorithm [184] for optimizing molecular structure, which requires the gradient of the energy to be supplied at each iterate with no requirement to compute the Hessian. The Gaussian jobs were submitted to Iridis4 cluster.

The coupling constants obtained from the computation include one bond– ^{1}J, two bond– ^{2}J, and three bond– ^{3}J. Longer range couplings are of vanishingly small size and are not part of the research work. Electrostatic environment was ignored because its effect on the *J*-coupling is minimal [185, 186].

However, care was taken for nucleotides that contain heterocyclic bases, where long range couplings would provide useful structural information – those ^{4}J couplings fall in the range of 2 to 3 Hz.

The computationally predicted value of the spin-spin interactions involves isotropic contributions form Fermi contact (FC), spin-dipolar (SD), paramagnetic spin-orbit (PSO), and the diamagnetic spin-orbit (DSO) terms. With the Fermi contact contribution being the most prominent contributor of all the other terms to the calculated scalar coupling values.

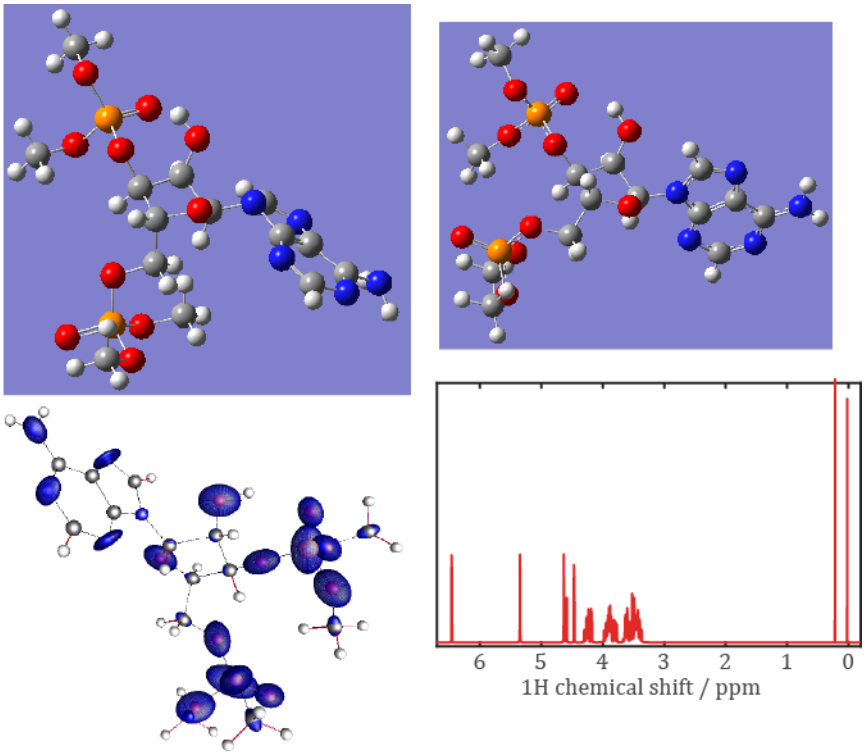

Figure 102 *Top panels are the constructed molecular structure of adenine before and after geometry optimization, left bottom panel is SPINACH representation of magnetic shielding tensors in adenine, and right bottom panel is proton NMR spectrum of adenine computed using SPINACH at proton resonance frequency of 250 MHz.*

The contribution for one bond ^{1}J coupling mostly comes from the Fermi contact term [187, 188], while that for two and three bond couplings, there are sufficient contribution from the other terms. Literature review was conducted to supplement the remaining coupling values. The DFT computation of the coupling constant for the four nucleotides was done considering the gas phase environment.

6.4 Results and scan plots for RNA

6.4.1 Karplus coefficients, C

*Method: GIAO DFT M06/6-31G** in SMD Water*

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
P5’	O5’	C5’	H5’’	20.17	0.28	-4.83	0.15	2.87	0.16
P5’	O5’	C5’	H5’	17.33	0.34	-8.08	0.14	2.71	0.18
P5’	O5’	C5’	C4’	21.90	0.34	-0.50	0.19	-7.95	0.16

H5''	C5'	C4'	C3'	6.80	0.51	-0.62	0.13	-0.02	0.39
H5''	C5'	C4'	H4'	9.31	0.47	1.29	0.17	1.70	0.26
H5'	C5'	C4'	C3'	4.38	0.42	-0.28	0.15	1.13	0.22
H5'	C5'	C4'	H4'	11.85	0.35	1.51	0.14	-0.87	0.23

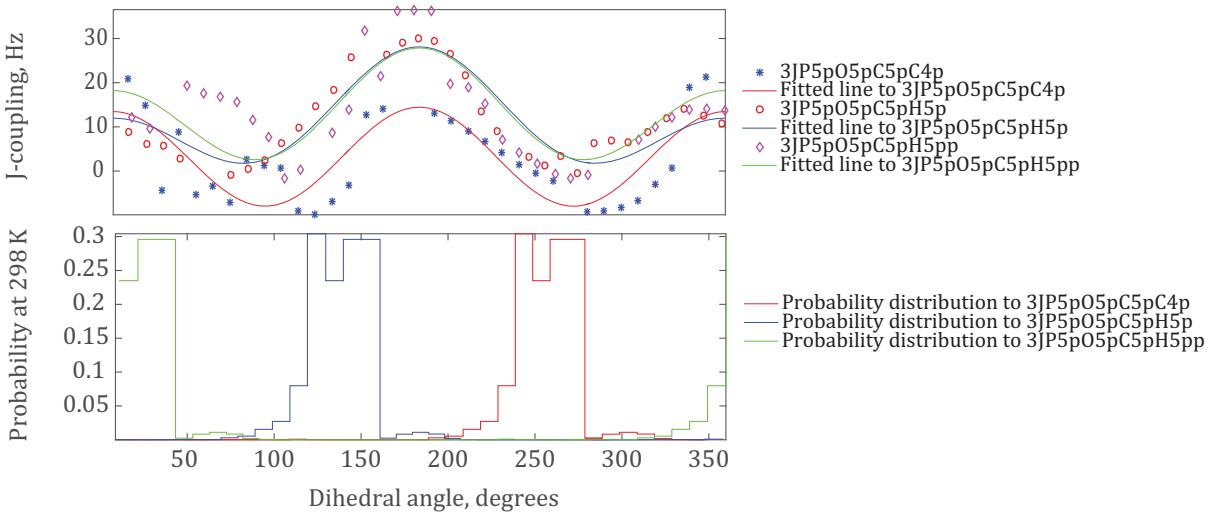

Figure 103 *Cytosine* β *main-chain dihedral angle scan. Top panel fitting of Karplus curves associated to the* β *dihedral angle to theoretical NMR vicinal coupling data and bottom panel probability distribution histograms for the respective torsion angles.*

As it is depicted in **Figure 103** the Karplus curves for the $^{3}J_{PC}$, $^{3}J_{PH5'}$, and $^{3}J_{PH5''}$ coupling data shows a pattern of under and over prediction which might be due to the effect that steric hindrance brings to the scan of the β dihedral angle.

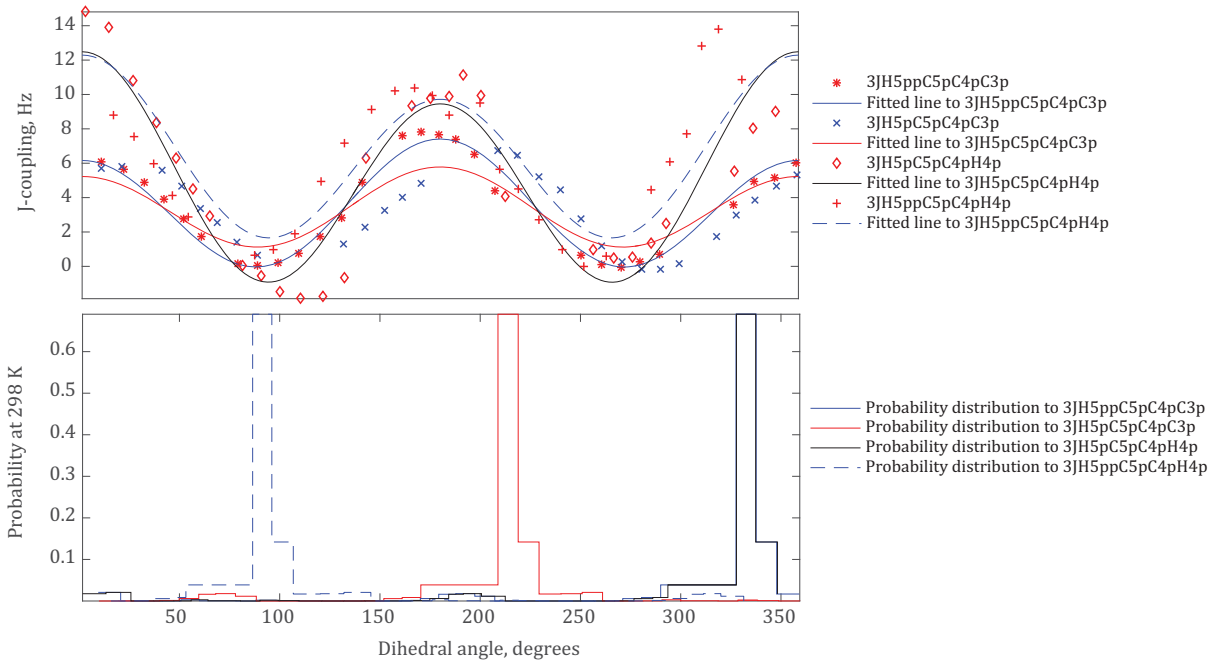

Figure 104 *Cytosine γ main-chain dihedral angle scan. Top panel Karplus curves for torsion angles related to the γ dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

Similarly the steric effect is also likely to be responsible for the under and over prediction pattern evident in some of the curves and for the broken Karplus curves.

6.4.2 Karplus coefficients, U

*Method: GIAO DFT M06/6-31G** in SMD Water*

Atom 1	Atom 2	Atom 3	Atom 4	A, Hz	A_{RMSD}	B, Hz	B_{RMSD}	C, Hz	C_{RMSD}
P5’	O5’	C5’	H5’’	26.59	0.37	-13.58	0.20	4.66	0.21
P5’	O5’	C5’	H5’	28.60	0.42	-7.92	0.16	0.87	0.26
P5’	O5’	C5’	C4’	0.35	0.62	-3.61	0.37	-1.87	0.27
H5’’	C5’	C4’	C3’	5.14	0.38	0.53	0.17	0.66	0.19
H5’’	C5’	C4’	H4’	9.37	0.36	0.20	0.14	1.44	0.20
H5’	C5’	C4’	C3’	6.61	0.37	-0.31	0.18	-0.09	0.23
H5’	C5’	C4’	H4’	10.89	0.32	0.93	0.13	0.64	0.20

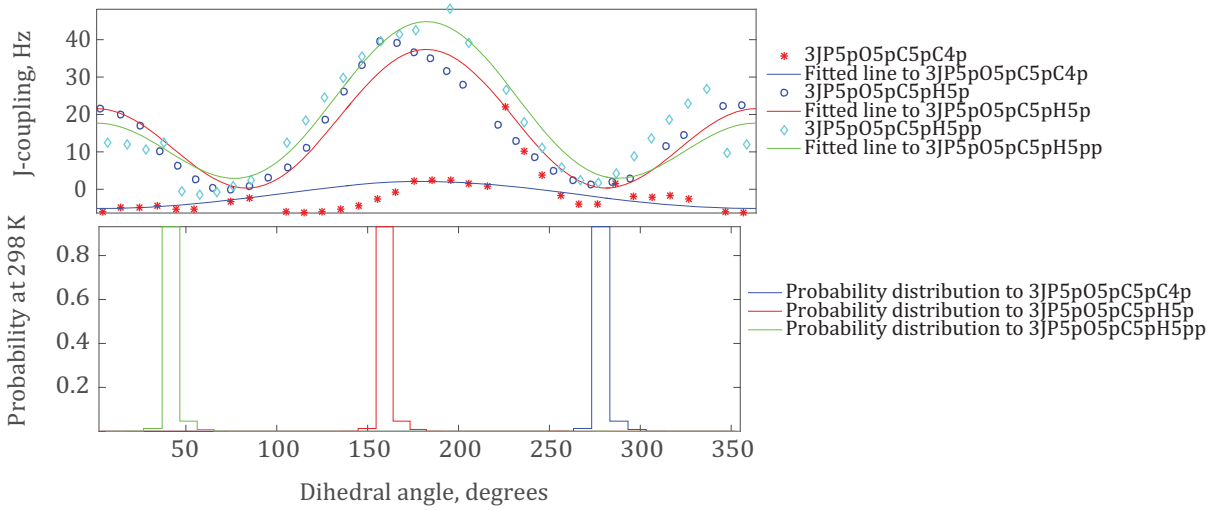

Figure 105 *Uracil* β *main-chain dihedral angle scan. Top panel Karplus curves for the torsion angles related to the* β *dihedral angle scan and bottom panel Boltzmann probability distribution graphs for the respective torsion angles.*

The result of the fitting of ^3J coupling data to Karplus curves for the three torsion angles is given in **Figure 105** and shows the dependence of the coupling values on the Karplus relation. Top panel shows the dihedral angle scan data with under and over predicting the function for two of the Karplus curves ($^3J_{PH5'}$, and $^3J_{PH5''}$). However, the $^3J_{PC}$ data for the broken curve is due to the predicted small amplitudes.

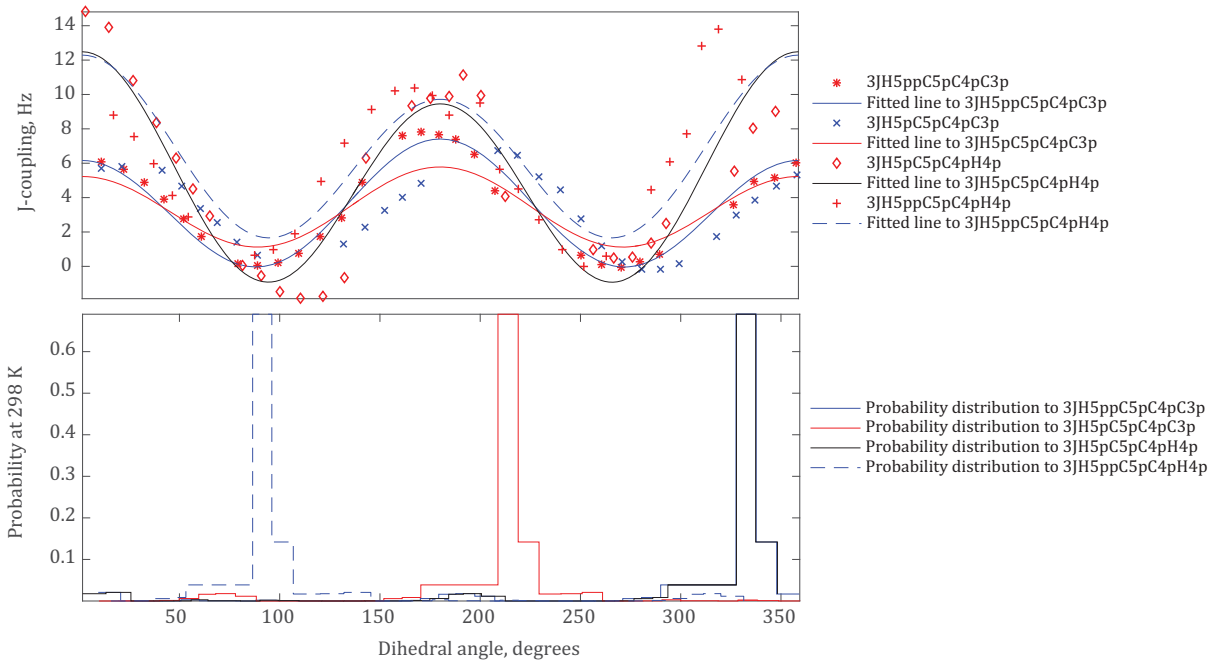

Figure 106 *Uracil γ main-chain dihedral angle scan. Top panel Karplus curves for torsion angles related to the γ dihedral angle scan and bottom panel Boltzmann probability distribution histograms for the respective torsion angles.*

All the scan of the γ dihedral angle are affected by steric effect as the fitting of the coupling data from the scan to the Karplus curves are either under and over predicted or have jumps as depicted in **Figure 106**.

6.5 RNA NOESY spectrum simulation and fitting with respect to correlation time (τ_C) and frequency offset

The detailed description of the theoretical background of NOESY NMR simulation technique along with the quantum mechanical treatment of the interaction Hamiltonians is given in chapter four. The simulation work for the NOESY RNA followed the procedure outlined in the chapter.

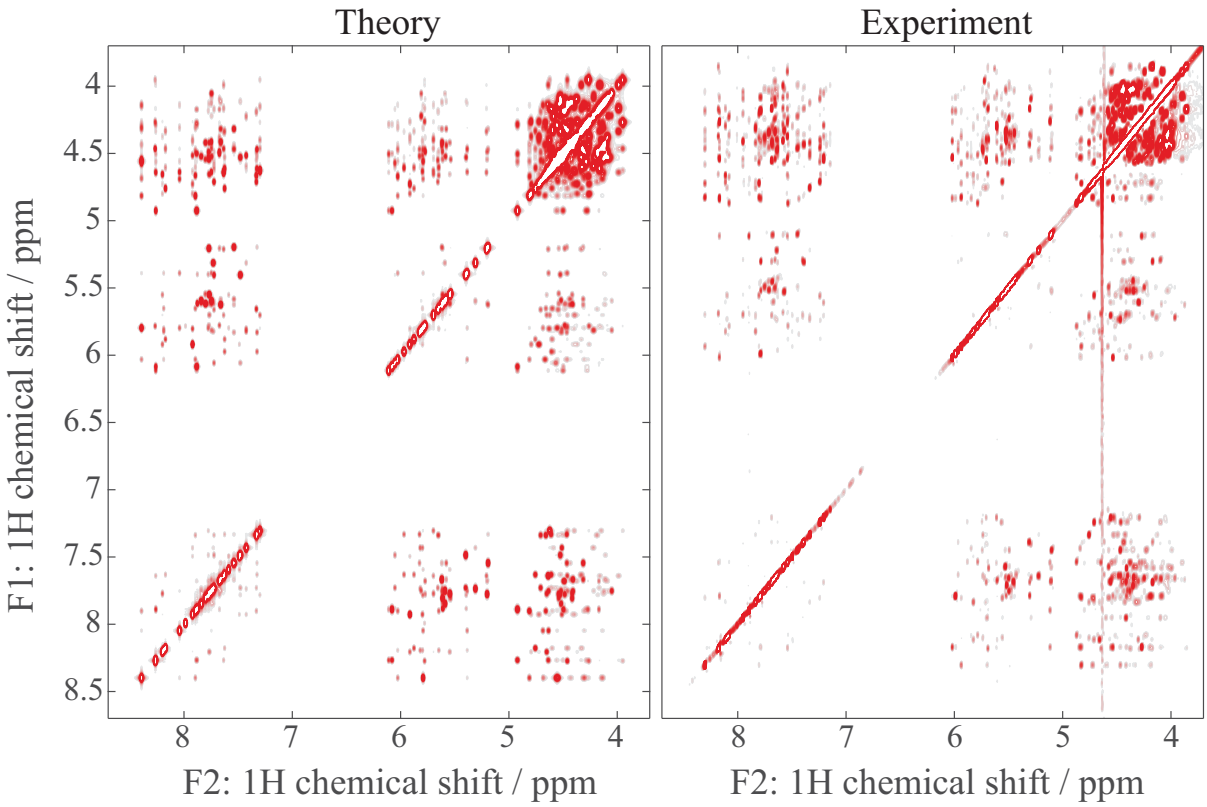

Figure 107 *Spinach simulated ^{1}H-^{1}H 2D NOESY NMR spectrum of RNA (28 residues and around 910 spins) calculated with a cut-off of 5Å and mixing time of 200ms, recorded at a 750MHz proton Larmor precession spectrometer. This illustrates that Spinach can deal with a large number of coupled spin system, while any conventional software would crash.*

The simulation included one-bond, two-bond and three-bond *J*-couplings, correct magnet field and mixing time. Only important Karplus curves that involved two proton spins are included in the simulation. Three bond couplings between spins that consist other nuclei are assigned zero.

The next step after the simulation work is to fit a simulated 2D NOESY spectrum of the RNA molecule supplied by Shunsuke *et al* of Wagner lab of the Harvard Medical School to a candidate structure from PDB.

The experimental spectrum provided by Imai Shunsuke and his colleagues was acquired by using the following parameters: mixing time of 200ms, sweep width of 7500Hz for the direct and 7496.25Hz for the indirect dimension, and offset of 3473Hz in an NMR spectrometer with a proton Larmor frequency of 750MHz. The rotational correlation time of the system is estimated to be around 3ns depending on previously reported rotational correlation time for the other RNAs and the molecular weight of the RNA molecule *i.e.* 9.2kDa. The same parameters

were used to simulate a candidate theoretical spectrum that matches the one from the experiment and would be employed as a decent initial guess in fitting to the data from experiment. The enclosed spectrum depicted in **Figure 107** (above) is the result of matching the simulation parameters to the experimental data. The extra peaks that appear in the simulation probably belong to the protons that got exchanged for deuterium in D_2O and the missing peaks may be attributed to the protons that have not been assigned.

The proton isotropic chemical shift tensor assignment file have fewer protons assigned than the PDB file has protons. And it is the more protons of chemical shift known available the better the result from the simulation would be. The assignment input file contains a peak list of 208 protons out of 308 protons in the PDB file. The difference stems from the exchangeable protons such as HO2' for all residues; H1, H21, H22 of all guanine residues; H41, H42 of cytosine residues; H3 for uridine residues; H61 and H62 for adenine residues and the terminal residues protons, which are H5T and H3T that are not observable in D_2O. To get the simulation setup closer to the experimental conditions we have come to the idea of deuteriating those exchangeable protons that are available in the PDB file. If this is not accounted for, the line widths of the non-exchangeable protons in the simulated spectra would include the effect of dipolar interactions with the exchangeable protons – apparently huge for some of them – that are not present in the experimental data.

A script is then written that would deuterate the list of protons in the structure of the RNA from PDB that are exchangeable in D_2O. Hence Spinach would do the calculation having the exchangeable protons deuterated.

For the purposes of the fitting work a comparison between a simulated and experimental spectra is necessary, therefore this requires a better resolved spectra. For this reason a NOESY spectrum in D_2O is simulated and was directly compared to experimental spectrum. After deuterating the exchangeable protons the two things apparent from the calculation are a lot of missing signals and a few variations in intensities in the simulation. The former would raise an important questions such as:

Is it possible to use the simulation to get the rest of the assignments?

When all missing assignments were dumped at zero ppm in the calculation, they generate cross-peaks that are correct with respect to other coupled partners. This is useful to take the dipolar interactions of the missing signals into account so that we can compare the line widths of the

peaks with the experimental ones and would be helpful to identify missing protons by sliding down until they match a missing peak.

The observation of variation in intensities between the simulated and experimental spectra is an indication of mismatch between the PDB geometry and the real geometry. Fitting would iron this out but it is required that as many assigned protons as possible to have to start the fitting, unless the fitting might end in astray.

Eliminating unassigned peaks from the theoretical calculation would produce sharp simulated peaks for the sugar protons H1' and H2', which could be another difficult during the fitting runs of the computed spectra against the experimental data. An example, sugar protons such as H4' and H5' are difficult to assign as a result of the severe signal overlaps, but the line widths of H1' and H2' should be affected by the presence and or absence of the other two protons.

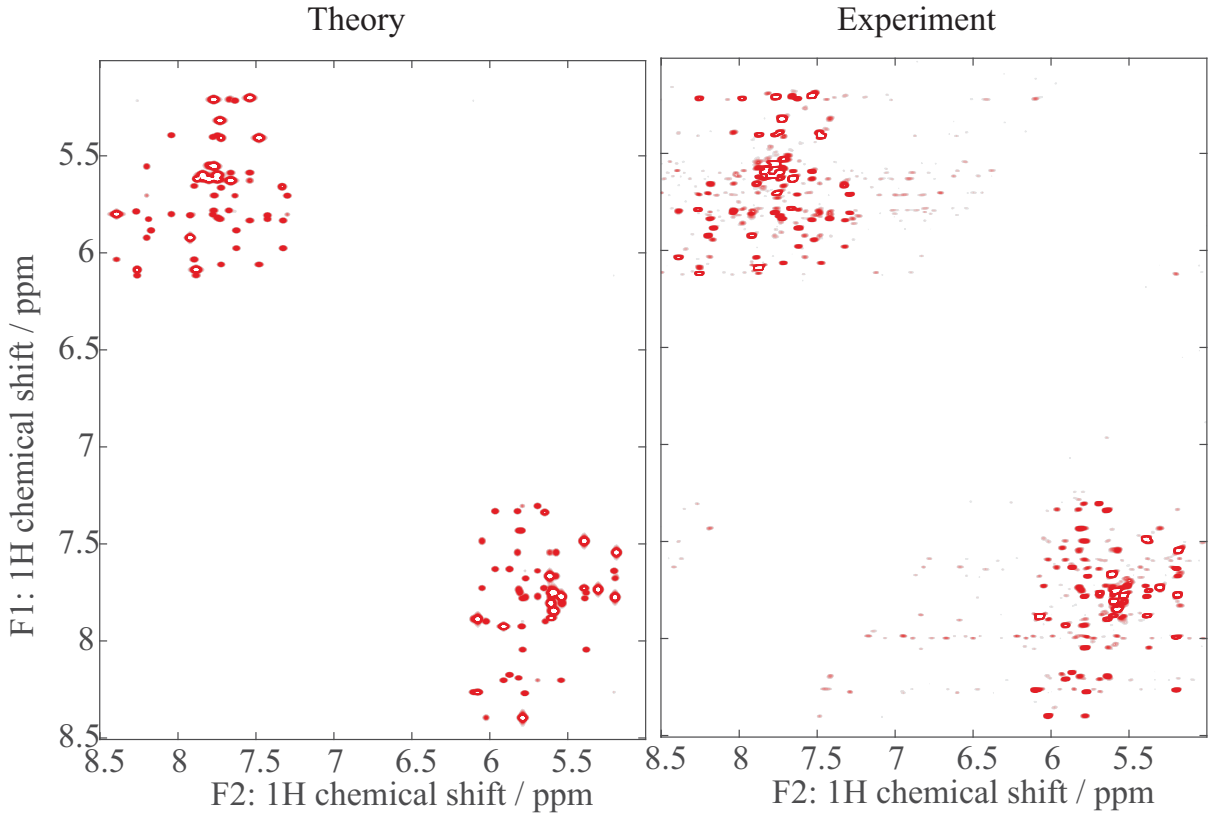

Figure 108 *RNA fitting work undertaken as part of this project. The simulation was run under the same conditions of the simulated and experimental spectra depicted in Figure 107.*

More assignments can be obtained from simulation using computational programs such as ARIA [189, 190] to refine protein structures by iteratively minimizing the deference between the theoretical and experimental spectra to optimize protein structures, or alternative methods

described in detail in [191] that do not require the NOE cross peak assigned unequivocally to proton pairs to use as initial guess to compute three dimensional protein structures by loosely describing the term ambiguous restrain using assigned isotropic chemical shift tensor and NOE cross peak that are not assigned to coupled proton spins. Though the aforementioned techniques enjoys huge success in solution protein NMR spectroscopy, this has not been the case for nucleic acids and specifically with RNAs, and similar programs for RNA have long been awaited in the field. In the simulated spectrum from the cross peaks it can qualitatively be observed that some variation in intensities *i.e.* some of the NOE's are more broadened than it is expected, and this effect might be down to the dynamics of the bulge region that depends on temperature.

To avoid having many unassigned cross-peaks and get around the problem of ambiguities that this creates, we have made the focus on the "fingerprint" region *i.e.* on the 5 to 6.5 ppm in F2 and 7 to 9 ppm in F1 where most of the signals are assigned along with the diagonal peaks.

The proposed region doesn't have too many unassigned peaks and usually contains much more structural information than the excluded region. If the structural refinement as describe above and another side project on extending assignments are established, both of these can be applied iteratively to refine the structure further.

In practice the fitting was conducted by zeroing the diagonal to keep the optimization function from getting stuck in a local minimum as the diagonals dominate the intensities if they are left in place during optimization process. The first runs of fitting of the project fits with respect to the rotational correlation time, experimental offset variable and overall scaling multiplier. The results from the fit is depicted in **Figure 108** (above). To fit with respect to coordinates, the Hessians are required.

The NOESY spectra were computed using standard Spinach functionality with pulse sequence and acquisition parameters matched to the experimental values. As usual, the spectra from theory were fitted to the data from the experiment using the least squares method with the error functional minimization performed by the Nelder-Mead Simplex algorithm. In practice, the fitting was performed by optimizing the parameters: correlation time $\left(\tau_{\mathrm{C}}\right)$ and frequency offset, in a similar fashion to that of the fitting performed in chapter five.

6.6 Concluding summary and outlook

Nucleic acid *J*-coupling predictor function was built and implemented into Spinach. This is a function, which is necessary in the simulation of biologically important molecules such as DNA and RNA.

After completion of the estimator function, a 2D NOESY RNA spectrum was simulated for the candidate data provided by the group from Wagner lab at Harvard Medical School. During the simulation processes, the estimator function as a wrapper function is used to estimate *J*-coupling values as mentioned above in the main text when called by another Spinach function to simulate nucleic acid systems.

The integration of the predictor function not only allowed us to carry the simulation of the system but also opened the way for the fitting of the system. As it is delineated and shown in section 6.5, we have managed to fit the NOESY experimental spectrum against the NOESY simulated spectrum with respect to the rotational correlation time, frequency offset variable and overall intensity scaling multiplier.

Once simulation stage is over, one possible area of exploration is, the fitting with respect to structure from PDB. For that purpose, one future direction is to refine the structure based on minimizing the NOE difference between the theoretically simulated spectrum and that from the experiment. The main structure of the RNA stem loops is a Watson-Crick double helix, where the loop is not well defined. The optimization should focus on the global fold of the structure rather than on the atomic coordinates. This will alleviate the computational complexity as the dimensions are significantly reduced. For that, the Hessian of the Energy is required, which would be generated using the molecular dynamics package GROMACS and diagonalise for the acquisition of the slow vibrational modes.

The slow normal mode stage of the Hessian matrix ensures that we should not worry about the coordinates of each participating atom and instead focus on optimizing the global fold. Thus at each iteration of the fitting procedure, the resulting Hessian from the previous iteration is used to update the coordinate search direction for the next step in the fitting process.

7 Reference

1. Senthamarai, R.R., I. Kuprov, and K. Pervushin, *Benchmarking NMR experiments: A relational database of protein pulse sequences.* Journal of Magnetic Resonance, 2010. **203**(1): p. 129-137.

2. McConnell, H.M., *Reaction Rates by Nuclear Magnetic Resonance.* The Journal of Chemical Physics, 1958. **28**(3): p. 430.

3. Hogben, H.J., et al., *Spinach – A software library for simulation of spin dynamics in large spin systems.* Journal of Magnetic Resonance, 2011. **208**(2): p. 179-194.

4. Veshtort, M. and R.G. Griffin, *SPINEVOLUTION: a powerful tool for the simulation of solid and liquid state NMR experiments.* J Magn Reson, 2006. **178**(2): p. 248-82.

5. Bak, M., J.T. Rasmussen, and N.C. Nielsen, *SIMPSON: A General Simulation Program for Solid-State NMR Spectroscopy.* Journal of Magnetic Resonance, 2000. **147**(2): p. 296-330.

6. Schmidt, J.M., et al., *Variation in protein Cα-related one-bond J couplings.* Magnetic Resonance in Chemistry, 2009. **47**(1): p. 16-30.

7. Sahakyan, A.B., et al., *Dielectric permittivity and temperature effects on spin-spin couplings studied on acetonitrile.* Magn Reson Chem, 2008. **46**(1): p. 63-8.

8. Peralta, J.E., et al., *Basis set dependence of NMR spin–spin couplings in density functional theory calculations: first row and hydrogen atoms.* Chemical Physics Letters, 2003. **375**(5–6): p. 452-458.

9. Delaglio, F., D.A. Torchia, and A. Bax, *Measurement of15N-13C J couplings in staphylococcal nuclease.* Journal of Biomolecular NMR, 1991. **1**(4): p. 439-446.

10. Karplus, M., *Vicinal Proton Coupling in Nuclear Magnetic Resonance.* Journal of the American Chemical Society, 1963. **85**(18): p. 2870-2871.

11. Kohn, W. and L.J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects.* Physical Review, 1965. **140**(4A): p. A1133-A1138.

12. Jensen, F., *Introduction to Computational Chemistry.* 2007. **Second Edition**.

13. Hehre, W.J., *A Guide to Molecular Mechanics and Quantum Chemical Calculations.* Wavefunction, Inc., 2003.

14. M. Kaupp, M.B., V. G. Malkin *Calculation of NMR and EPR Parameters: Theory and Applications.* 2004.

15. P. J. Hore, J.A.J., S. Wimperis, , *NMR: the toolkit*. Vol. 92. 2000: Oxford University Press Oxford.

16. Fano, U., *Description of States in Quantum mechanics by Density Matrix and Operator Techniques* REVIEWS Of MODERN PHYSICS, 1957. **29**: p. 74-93.

17. Munowtiz, M., *Coherence and NMR* John Wiley, Chichester, 1988.

18. Ernst, R.R., G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. 1987: Clarendon Press.

19. Kosloff, M.B.a.R., *Time-dependent solution of the Liouville-von Neumann equation: non-dissipative evolution.* Computer Physics Communications, 1991. **63**: p. 1-20.

20. Kuprov, I.W.-R., Nicola; Hore, P. J., *Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction.* Journal of Magnetic Resonance, 2007. **189**(2): p. 241-250.

21. Ramsey, N.F., *Magnetic Shielding of Nuclei in Molecules.* Physical Review, 1950. **78**(6): p. 699-703.

22. Atkins, P.W.a.R.S.F., *Molecular quantum mechanics.* Oxford: Oxford University Press, 2011. **Fifth ed**.

23. Hore, P.J., *Nuclear magnetic resonance*. 2015: Oxford University Press, USA.

24. Steiner, D., et al., *On the calculation of 3 J αβ-coupling constants for side chains in proteins.* Journal of Biomolecular NMR, 2012. **53**(3): p. 223-246.

25. Cavanagh, J., et al., *Protein NMR Spectroscopy: Principles and Practice*. 2010: Elsevier Science.

26. Levitt, M.H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. 2008.

27. Keeler, J., *Understanding NMR Spectroscopy*. 2011.

28. Kowalewski, J.a.M., L., *Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications*. 2006: Taylor & Francis.

29. Wüthrich, K., *NMR in Biological Research: Peptides and Proteins*. 1976: North-Holland Publishing Company.

30. Abragam, A., *The Principles of Nuclear Magnetism*. 1961: Clarendon Press.

31. Schirmer., J.H.N.R.E., *The nuclear Overhauser effect : chemical applications.* 1971.

32. Noggle, J.H. and R.E. Schirmer, *The nuclear Overhauser effect: chemical applications*. 1971: Academic Press.

33. Bakhmutov, V.I., *Practical Nuclear Magnetic Resonance Relaxation for Chemists*. 2005.

34. Neuhaus, D. and M.P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. 1989: Wiley.

35. Roberts, G.C.K., *NMR of Macromolecules; A Practical Approach.* The Practical approach series, 1993. **134**.

36. J. R. ALGER, R.G.S., *NMR studies of enzymatic rates in vitro and in vivo by magnetization transfer.* Quarterly Review of Biophysics, 1984. **17**(1): p. 83-124.

37. S. FORSEN, R.A.H., *Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance* THE JOURNAL OF CHEMICAL PHYSICS 1963. **39**(11): p. 2892-2901

38. J. Jeener, B.H.M., P. Bachmann, and R. R. Ernst, *Investigation of Exchange processes by two-dimensional NMR spectroscopy.* Journal of Chemical Physics 1979. **71**: p. 4546-4553.

39. S. Meiboom, D.G., *Modified SpinEcho Method for Measuring Nuclear Relaxation Times.* Review of Scientific Instruments, 1958. **29**(8): p. 688-691.

40. A. Allerhand, H.S.G., *Spin-Echo Studies of Chemical Exchange. II. Closed Formulas for Two Sites.* Journal of Chemical Physics, 1965. **42**: p. 1587-1599.

41. A. K. Mittermaier, L.E.K., *Observing biological dynamics at atomic resolution using NMR.* Cell Press, 2009. **07**(004): p. 601-611.

42. III, A.G.P., *NMR Characterization of the Dynamics of Biomacromolecules.* Chemical Reviews, 2004. **104**: p. 3623-3640.

43. I. R. Kleckner, M.P.F., *An introduction to NMR-based approaches for measuring protein dynamics.* Biochimica et Biophysica Acta, 2011. **1814**: p. 942–968.

44. Sandstrom, J., *Dynamic Nuclear Magnetic Resonance Spectroscopy.* 1982.

45. S. Braun, H.-O.K., S. Berger, *100 and More Basic NMR Experiments, A Practical Course.* 1996.

46. Baldwin, A.J., *An exact solution for R2,eff in CPMG experiments in the case of two site chemical exchange.* Journal of Magnetic Resonance, 2014. **244**: p. 114-124.

47. Fejzo, J., et al., *Elimination of cross-relaxation effects from two-dimensional chemical-exchange spectra of macromolecules.* Journal of the American Chemical Society, 1990. **112**(7): p. 2574-2577.

48. Fejzo, J., et al., *Strategies for eliminating unwanted cross-relaxation and coherence-transfer effects from two-dimensional chemical-exchange spectra.* Journal of Magnetic Resonance (1969), 1991. **92**(1): p. 20-29.

49. J. P. Loria, R.B.B., E. D. Watt, *Characterization of Enzyme Motions by Solution NMR Relaxation Dispersion.* ACCOUNTS OF CHEMICAL RESEARCH, 2008. **41**(02): p. 214-221.

50. Stoll, S. and A. Schweiger, *EasySpin, a comprehensive software package for spectral simulation and analysis in EPR.* Journal of Magnetic Resonance, 2006. **178**(1): p. 42-55.

51. Bak, M., J.T. Rasmussen, and N.C. Nielsen, *SIMPSON: A general simulation program for solid-state NMR spectroscopy.* Journal of Magnetic Resonance, 2011. **213**(2): p. 366-400.

52. Veshtort, M. and R.G. Griffin, *SPINEVOLUTION: A powerful tool for the simulation of solid and liquid state NMR experiments.* Journal of Magnetic Resonance, 2006. **178**(2): p. 248-282.

53. Dumez, J.-N.B., Mark C.; Emsley, Lyndon, *Numerical simulation of free evolution in solid-state nuclear magnetic resonance using low-order correlations in Liouville space.* The Journal of Chemical Physics, 2010. **133**(22): p. 224501.

54. Butler, M.C.D., Jean-Nicolas; Emsley, Lyndon, *Dynamics of large nuclear-spin systems from low-order correlations in Liouville space.* Chemical Physics Letters, 2009. **477**(4–6): p. 377-381.

55. Suzuki, M.M., Seiji; Kuroda, Akira, *Monte Carlo Simulation of Quantum Spin Systems. I.* Progress of Theoretical Physics, 1977. **58**(5): p. 1377-1387.

56. Lee, S., et al., *Quantum treatment of the effects of dipole–dipole interactions in liquid nuclear magnetic resonance.* The Journal of chemical physics, 1996. **105**(3): p. 874-900.

57. Bai, Z., *Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems.* Applied Numerical Mathematics, 2002. **43**(1–2): p. 9-44.

58. Kuprov, I., *Polynomially scaling spin dynamics II: Further state-space compression using Krylov subspace techniques and zero track elimination.* Journal of Magnetic Resonance, 2008. **195**(1): p. 45-51.

59. Edwards, L.J., et al., *Quantum mechanical NMR simulation algorithm for protein-size spin systems.* Journal of Magnetic Resonance, 2014. **243**: p. 107-113.

60. Kuprov, I.W.-R., Nicola; Hore, P. J., *Bloch-Redfield-Wangsness theory engine implementation using symbolic processing software.* Journal of Magnetic Resonance, 2007. **184**(2): p. 196-206.

61. Wijmenga, S.S. and B.N.M. van Buuren, *The use of NMR methods for conformational studies of nucleic acids.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1998. **32**(4): p. 287-387.

62. Ottiger, M., F. Delaglio, and A. Bax, *Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra.* Journal of Magnetic Resonance, 1998. **131**(2): p. 373-378.

63. Griesinger, C., O.W. Soerensen, and R.R. Ernst, *Two-dimensional correlation of connected NMR transitions.* Journal of the American Chemical Society, 1985. **107**(22): p. 6394-6396.

64. Bax, A., et al., *Measurement of homo- and heteronuclear J couplings from quantitative J correlation*, in *Methods in Enzymology*. 1994. p. 79-105.

65. Morris, G.A. and R. Freeman, *Enhancement of nuclear magnetic resonance signals by polarization transfer.* Journal of the American Chemical Society, 1979. **101**(3): p. 760-762.

66. Griesinger, C., O.W. Sorensen, and R.R. Ernst, *Correlation of connected transitions by two-dimensional NMR spectroscopy.* The Journal of Chemical Physics, 1986. **85**(12): p. 6837-6852.

67. Tjandra, N., S. Grzesiek, and A. Bax, *Magnetic Field Dependence of Nitrogen−Proton J Splittings in 15N-Enriched Human Ubiquitin Resulting from Relaxation Interference and Residual Dipolar Coupling.* Journal of the American Chemical Society, 1996. **118**(26): p. 6264-6272.

68. Vuister, G.W. and A. Bax, *Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins.* Journal of the American Chemical Society, 1993. **115**(17): p. 7772-7777.

69. Cobas, J.C. and F.J. Sardina, *Nuclear magnetic resonance data processing. MestRe-C: A software package for desktop computers.* Concepts in Magnetic Resonance Part A, 2003. **19**(2): p. 80-96.

70. Elyashberg, M.E., A.J. Williams, and G.E. Martin, *Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2008. **53**(1–2): p. 1-104.

71. Laatikainen, R., et al., *A computational strategy for the deconvolution of NMR spectra with multiplet structures and constraints: analysis of overlapping 13C-2H multiplets of 13C enriched metabolites from cell suspensions incubated in deuterated media.* Magn Reson Med, 1996. **36**(3): p. 359-65.

72. Laatikainen, R., et al., *General Strategies for Total-Lineshape-Type Spectral Analysis of NMR Spectra Using Integral-Transform Iterator.* Journal of Magnetic Resonance, Series A, 1996. **120**(1): p. 1-10.

73. Lehtivarjo, J., M. Niemitz, and S.-P. Korhonen, *Universal J-Coupling Prediction.* Journal of Chemical Information and Modeling, 2014. **54**(3): p. 810-817.

74. Binev, Y., M.M.B. Marques, and J. Aires-de-Sousa, *Prediction of 1H NMR Coupling Constants with Associative Neural Networks Trained for Chemical Shifts.* Journal of Chemical Information and Modeling, 2007. **47**(6): p. 2089-2097.

75. Tetko, I.V., *Neural network studies. 4. Introduction to associative neural networks.* J Chem Inf Comput Sci, 2002. **42**(3): p. 717-28.

76. Wang, B., X. He, and K.M. Merz, *Quantum Mechanical Study of Vicinal J Spin–Spin Coupling Constants for the Protein Backbone.* Journal of Chemical Theory and Computation, 2013. **9**(10): p. 4653-4659.

77. Xia, J. and C.J. Margulis, *Computational Study of the Conformational Structures of Saccharides in Solution Based on J Couplings and the "Fast Sugar Structure Prediction Software".* Biomacromolecules, 2009. **10**(11): p. 3081-3088.

78. Xia, J., et al., *Sugar Folding: A Novel Structural Prediction Tool for Oligosaccharides and Polysaccharides 1.* J Chem Theory Comput, 2007. **3**(4): p. 1620-8.

79. Islam, S.M., et al., *Conformational analysis of oligoarabinofuranosides: Overcoming torsional barriers with umbrella sampling.* 2011. **7**(9): p. 2989-3000.

80. Hohenberg, P. and W. Kohn, *Inhomogeneous Electron Gas.* Physical Review, 1964. **136**(3B): p. B864-B871.

81. W. Koch, a.M.C.H., *A Chemist's Guide to Density Functional Theory.* 2001. **Second Edition**.

82. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* The Journal of Chemical Physics, 1993. **98**(7): p. 5648-5652.

83. P. J. Stephens, F.J.D., C. F. Chabalowski, M. J. Frisch, *Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields.* The Journal of Physical Chemistry, 1994. **98**: p. 11623-11627.

84. Dunning, T.H., *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen.* The Journal of Chemical Physics, 1989. **90**(2): p. 1007.

85. Wilson, A.K., T. van Mourik, and T.H. Dunning Jr, *Gaussian basis sets for use in correlated molecular calculations. VI. Sextuple zeta correlation consistent basis sets*

for boron through neon. Journal of Molecular Structure: THEOCHEM, 1996. **388**: p. 339-349.

86. Leszczynski, J., *Computational Chemistry: Reviews of Current Trends*. 2004: World Scientific.

87. T. Helgaker, M.J., K. Ruud, *Ab Initio Methods for the Calculation of NMR Shielding and Indirect Spin-Spin Coupling Constants.* Chemical Reviews, 1999. **99**: p. 293-352.

88. Cheeseman, J.R.T., Gary W. Keith, Todd A. Frisch, Michael J., *A comparison of models for calculating nuclear magnetic resonance shielding tensors.* The Journal of Chemical Physics, 1996. **104**(14): p. 5497.

89. Ditchfield, R., *Self-consistent perturbation theory of diamagnetism. 1. Gauge-invariant LCAO method for N.M.R. chemical shifts.* Molecular Physics, 1974. **27** p. 789-807.

90. W. Deng, J.R.C., and M. J. Frisch, *Calculation of Nuclear Spin-Spin Coupling Constants of Molecules with First and Second Row Atoms in Study of Basis Set Dependence.* Journal of Chemical Theory and Computation., 2006. **2**: p. 1028-1037.

91. Peralta, J.E., Scuseria, Gustavo E., Cheeseman, James R., Frisch, Michael J., *Basis set dependence of NMR spin–spin couplings in density functional theory calculations: first row and hydrogen atoms.* Chemical Physics Letters, 2003. **375**(5-6): p. 452-458.

92. Aleksandr V. Marenich, C.J.C., and Donald G. Truhlar, *Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions.* J. Phys. Chem. B, 2009. **113**: p. 6378–6396.

93. R. E. Skyner, J.L.M., C. R. Groom, T. van Mourika and J. B. O. Mitchell, *A review of methods for the calculation of solution free energies and the modelling of systems in solution.* Phys.Chem.Chem.Phys., 2015. **15**: p. 6174-6191.

94. T. Helgaker, M.W., N. C. Handy, *Analytical calculation of nuclear magnetic resonance indirect spin–spin coupling constants at the generalized gradient approximation and hybrid levels of density-functional theory.* JOURNAL OF CHEMICAL PHYSICS, 2000. **113**: p. 9402-9409.

95. RAMSEY, N.F., *Electron Coupled Interactions between Nuclear Spins in Molecules.* PH YSI CAL REVI EW, 1953. **91**: p. 303-307.

96. T. Helgaker, M.J., M. Pecul *The quantum-chemical calculation of NMR indirect spin–spin coupling constants.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2008. **53**: p. 249–268.

97. Blondel, A.K., Martin, *New formulation for derivatives of torsion angles and improper torsion angles in molecular mechanics: Elimination of singularities.* Journal of Computational Chemistry, 1996. **17**(9): p. 1132-1141.

98. Haasnoot CAG, d.L.F., de Leeuw HPM, Altona C, *Interpretation of vicinal proton-proton coupling constants by a generalised Karplus relation. Conformational analysis of the exocyclic C4'-C5' bond in nucleosides and nucleotides.* Rec J Roy Neth Chem Soc, 1979. **98**: p. 576–577.

99. Haasnoot CAG, d.L.F., de Leeuw HPM, Altona C, *Relationship between proton-proton NMR coupling constants and substituent electronegativities. III. Conformational analysis of proline rings in solution using a generalized Karplus equation.* Bio polymer, 1981(20): p. 1211–1245.

100. Coxon, B., *DEVELOPMENTS IN THE KARPLUS EQUATION AS THEY RELATE TO THE NMR COUPLING CONSTANTS OF CARBOHYDRATES.* Adv Carbohydr Chem Biochem., 2009. **62**: p. 17-82.

101. SHAM, W.K.A.L.J., *Self-Consistent Equations Including Exchange and Correlation Effects.* PHYSICAL REVIEW, 1965. **140**: p. 1133-1138.

102. Daniel Zaccari, V.B., Juan E. Peralta, Rubén H. Contreras, Oscar E. Taurian, Ernesto Díez and Angel Esteban, *Solvent Effects on Nuclear Magnetic Resonance 2J(C,Hf) and 1J(C,Hf) Spin–Spin Coupling Constants in Acetaldehyde.* International Journal of Molecular Sciences, 2003. **4**: p. 93-106.

103. Markovsky, I. and S. Van Huffel, *Overview of total least-squares methods.* Signal Processing, 2007. **87**(10): p. 2283-2302.

104. Wang, A.C. and A. Bax, *Reparametrization of the Karplus Relation for (3)J(H-Alpha-N) and (3)J(H-N-C') in Peptides from Uniformly C-13/N-15-Enriched Human Ubiquitin.* Journal of the American Chemical Society, 1995. **117**(6): p. 1810-1813.

105. Vuister, G.W.a.A.B., *Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins.* Journal of the American Chemical Society, 1993. **115**: p. 7772-7777.

106. Bax, J.-S.a.A., *Determination of ohi and chi1 Angles in proteins from ^{13}C-^{13}C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond.* Joural of American Chemical Society, 1997. **119**: p. 6360-6368.

107. Bax, J.H.a.A., *Measurment of Three-bond 13C-13C J couplings between carbonyl and carbonyl/carbonyl carbons in Isotopically enriched proteins.* JACS, 1996. **118**: p. 8170-8171.

108. West, L.J.S.a.N.J., *Side-chains in Native and Random Coil Protein Conformations. Analysis of NMR Coupling Constants and X1 Torsion Angle Preferences.* Jornal of Molecular Biology, 1998. **280**: p. 867-877.

109. Fischman, A.J., et al., *Torsion Angles in the Cystine Bridge of Oxytocin in Aqueous-Solution - Measurements of Circumjacent Vicinal Couplings between H-1, C-13, and N-15(1a).* Journal of the American Chemical Society, 1980. **102**(8): p. 2533-2539.

110. Chou, J.J., D.A. Case, and A. Bax, *Insights into the mobility of methyl-bearing side chains in proteins from (3)J(CC) and (3)J(CN) couplings.* Journal of the American Chemical Society, 2003. **125**(29): p. 8959-8966.

111. Zhou, P., et al., *Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins.* Proteins, 2009. **76**(1): p. 151-63.

112. Smith L. J., S.M.J., Redfield C., Dobson C. M.,, *Analysis of phi and chi1 torsion angles from hen Lysozyme in solution from 1H NMR spin-spin coupling constants.* Biochemistry, 1991. **30**: p. 986–996.

113. Xu R. X., O.E.T., Fesik S. W., , *Stereospecific assignment and Chi1 rotamers for FKBP when bound to Ascomycin from 3JHa Hb and 3JNHb coupling constants.* FEBS Lett, 1992. **305**: p. 137–143.

114. Moore J. M., L.C.A., Gippler G. P., Chazin W. J., Case D. A., Wright P. E., *High-resolution solution structure of reduced french bean Plastocyanin and comparison with the crystal structure of poplar Plastocyanin.* J Mol Biol, 1991. **221**: p. 533–555.

115. Vajpai, N., et al., *Side-Chain χ1 Conformations in Urea-Denatured Ubiquitin and Protein G from 3J Coupling Constants and Residual Dipolar Couplings.* Journal of the American Chemical Society, 2010. **132**(9): p. 3196-3203.

116. Roche, J., J. Ying, and A. Bax, *Accurate measurement of 3JHNHα couplings in small or disordered proteins from WATERGATE-optimized TROSY spectra.* Journal of biomolecular NMR, 2016. **64**(1): p. 1-7.

117. Lindorff-Larsen, K., R.B. Best, and M. Vendruscolo, *Interpreting Dynamically-Averaged Scalar Couplings in Proteins.* Journal of Biomolecular NMR, 2005. **32**(4): p. 273-280.

118. Overhauser, A.W., *Polarization of Nuclei in Metals.* Physical Review, 1953. **92**(2): p. 411-415.

119. Carver, T.R. and C.P. Slichter, *Experimental verification of the Overhauser nuclear polarization effect.* Physical Review, 1956. **102**(4): p. 975.

120. Kaiser, R., *Use of the Nuclear Overhauser Effect in the Analysis of High-Resolution Nuclear Magnetic Resonance Spectra.* THE JOURNAL OF CHEMICAL PHYSICS, 1963. **39**(10): p. 2435.

121. Solomon, I., *Relaxation Processes in a System of Two Spins.* Physical Review, 1955. **99**(2): p. 559-565.

122. Redfield, A.G., *Relaxation Theory: Density Matrix Formulation*, in *eMagRes*. 2007, John Wiley & Sons, Ltd.

123. F., W.R.K.a.B., *The dynamical theory of nuclear induction.* Physical Review, 1953. **89**(4): p. 728.

124. M. W. F. Fischer, A.M., E. R. P. Zuiderweg, *Protein NMR relaxation: theory, applications and outlook.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1998. **33**: p. 207-272.

125. Kuprov, I., *Diagonalization-free implementation of spin relaxation theory for large spin systems.* J Magn Reson, 2011. **209**(1): p. 31-8.

126. Kitao, K.T.a.A., *Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density.* J. Phys. Chem. B, 2012. **116**: p. 6279−6287.

127. Woessner, D.E., *Nuclear Spin Relaxation in Ellipsoids Undergoing Rotational Brownian Motion.* THE JOURNAL OF CHEMICAL PHYSICS, 1962. **37**: p. 647-654.

128. Edward, J.T., *Molecular Volumes and the Stokes-Einstein Equation.* Journal of Chemical Education, 1970. **47**: p. 261-270.

129. Keller, C.W.R.a.C.E., *Direct Determination of NMR Correlation Times from Spin-Lattice and Spin-Spin Relaxation Times.* American Chemical Society, 1997. **101**: p. 3246-3250.

130. Donghan Lee, C.H., Gerhard Wider, and Kurt Wuthrich, *Effective rotational correlation times of proteins from NMR relaxation interference.* Journal of Magnetic Resonance, 2006. **178**: p. 72–76.

131. Wuthrich, P.L.a.K., *Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules.* Progress In Nuclear Magnetic Resonance Spectroscopy, 2002. **40**: p. 199–247.

132. Rance, M., J.P. Loria, and A.G. Palmer Iii, *Sensitivity Improvement of Transverse Relaxation-Optimized Spectroscopy.* Journal of Magnetic Resonance, 1999. **136**(1): p. 92-101.

133. Pervushin, K., et al., *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.* Proceedings of the National Academy of Sciences, 1997. **94**(23): p. 12366-12371.

134. Williamson, M.P., *Chapter 3 Applications of the NOE in Molecular Biology.* 2009. **65**: p. 77-109.

135. Kalk, A. and H.J.C. Berendsen, *Proton magnetic relaxation and spin diffusion in proteins.* Journal of Magnetic Resonance (1969), 1976. **24**(3): p. 343-366.

136. Gordon, S.L. and K. Wuethrich, *Transient proton-proton Overhauser effects in horse ferrocytochrome c.* Journal of the American Chemical Society, 1978. **100**(22): p. 7094-7096.

137. Kumar, A., R.R. Ernst, and K. Wüthrich, *A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton*

cross-relaxation networks in biological macromolecules. Biochemical and Biophysical Research Communications, 1980. **95**(1): p. 1-6.

138. Kumar, A., *Two dimensional nuclear Overhauser effect in biomolecules* Journal of Chemical Sciences - Indian Academy of Sciences 1985. **95**: p. 1-8.

139. Wagner, G., *NMR investigations of protein structure.* Progress In Nuclear Magnetic Resonance Spectroscopy, 1990. **22**: p. 101-139.

140. Saudek, R.A.A.a.V., *The direct determination of protein structure by NMR without assignment.* FEBS Letters, 2002. **510**: p. 1-4.

141. Hogben, H.J., P.J. Hore, and I. Kuprov, *Strategies for state space restriction in densely coupled spin systems with applications to spin chemistry.* J Chem Phys, 2010. **132**(17): p. 174101.

142. Karabanov, A., et al., *On the accuracy of the state space restriction approximation for spin dynamics simulations.* J Chem Phys, 2011. **135**(8): p. 084106.

143. Even, S., *Graph algorithms*. 2011: Cambridge University Press.

144. Edwards, L.J. and I. Kuprov, *Parallel density matrix propagation in spin dynamics simulations.* J Chem Phys, 2012. **136**(4): p. 044108.

145. Krzystyniak, M., L.J. Edwards, and I. Kuprov, *Destination state screening of active spaces in spin dynamics simulations.* J Magn Reson, 2011. **210**(2): p. 228-32.

146. Atkinson, K., *An Introduction to Numerical Analysis*. 1989: Wiley.

147. Gutowsky, H.S. and C.H. Holm, *Rate Processes and Nuclear Magnetic Resonance Spectra. II. Hindered Internal Rotation of Amides.* The Journal of Chemical Physics, 1956. **25**(6): p. 1228.

148. C. D. Kroenke, J.P.L., L. K. Lee, M. Rance, A. G. Palmer, III, *Longitudinal and Transverse 1H-15N Dipolar/15N Chemical Shift Anisotropy Relaxation Interference: Unambiguous Determination of Rotational Diffusion Tensors and Chemical Exchange Effects in Biological Macromolecules.* J. Am. Chem. Soc., 1998. **120**: p. 7905-7915.

149. J. Patrick Loriaa, M.R., Arthur G. Palmer, III, *A TROSY CPMG sequence for characterizing chemical exchange in large proteins.* Journal of Biomolecular NMR, 1999. **15**: p. 151–155.

150. Thuduppathy, G.R. and R.B. Hill, *Applications of NMR Spin Relaxation Methods for Measuring Biological Motions.* 2004. **384**: p. 243-264.

151. Bain, A.D., *Chemical exchange in NMR.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2003. **43**(3-4): p. 63-103.

152. Vallurupalli, P., *Chemical Exchange.* 2009.

153. Jablonka, W., et al., *Structure and Ligand-Binding Mechanism of a Cysteinyl Leukotriene-Binding Protein from a Blood-Feeding Disease Vector.* ACS Chemical Biology, 2016. **11**(7): p. 1934-1944.

154. Zuiderweg, E.R.P., *Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy.* Biochemistry, 2002. **41**: p. 1-7.

155. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Prog Nucl Magn Reson Spectrosc, 2013. **73**: p. 1-16.

156. Weiping Shao, S.-C.I., Erik R. P. Zuiderweg, Lucy Waskell, *Mapping the Binding Interface of the Cytochrome b5-Cytochrome c Complex by Nuclear Magnetic Resonance†.* Biochemistry, 2003. **42**: p. 14774-14784.

157. Abergel, D. and A.G. Palmer, *Approximate solutions of the Bloch-McConnell equations for two-site chemical exchange.* Chemphyschem, 2004. **5**(6): p. 787-93.

158. Motulsky, H.J. and L.A. Ransnas, *Fitting curves to data using nonlinear regression: a practical and nonmathematical review.* The FASEB Journal, 1987. **1**(5): p. 365-74.

159. Nocedal, J. and S. Wright, *Numerical Optimization*. 2000: Springer New York.

160. Nelder, J.A. and R. Mead, *A Simplex Method for Function Minimization.* The Computer Journal, 1965. **7**(4): p. 308-313.

161. JEFFREY C. LAGARIAS, J.A.R., MARGARET H. WRIGHT, AND PAUL E. WRIGHT, *CONVERGENCE PROPERTIES OF THE NELDER-MEAD SIMPLEX METHOD IN LOW DIMENSIONS.* Society for Industrial and Applied Mathematics J. OPTIM., 1998. **9**: p. 112-147.

162. Stevens, F.C., *Calmodulin: an introduction.* Canadian Journal of Biochemistry and Cell Biology, 1983. **61**(8): p. 906-910.

163. Chattopadhyaya, R.M., William E.; Means, Anthony R.; Quiocho, Florante A., *Calmodulin structure refined at 1.7 Å resolution.* Journal of Molecular Biology, 1992. **228**(4): p. 1177-1192.

164. Chou, J.J.L., S.; Klee, C. B.; Bax, A., *Solution structure of Ca(2+)-calmodulin reveals flexible hand-like properties of its domains.* Nat Struct Biol, 2001. **8**(11): p. 990-7.

165. MARQUARDT, D.W., *AN ALGORITHM FOR LEAST-SQUARES ESTIMATION OF NONLINEAR PARAMETERS.* Journal of the Society for Industrial and Applied Mathematics, 1963. **11**(2): p. 431-441.

166. Alberts, B.J., A.; Lewis, J.; Walter, P.; Raff, M.; Roberts, K., *Molecular Biology of the Cell 4th Edition: International Student Edition*. 2002: Routledge.

167. Clark, D.P., *Molecular Biology : Academic Cell Update Edition*. 2014: Academic Cell.

168. Guerrier-Takada, C., et al., *The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme.* Cell, 1983. **35**(3, Part 2): p. 849-857.

169. Bass, B.L. and T.R. Cech, *Ribozyme inhibitors: deoxyguanosine and dideoxyguanosine are competitive inhibitors of self-splicing of the Tetrahymena ribosomal ribonucleic acid precursor.* Biochemistry, 1986. **25**(16): p. 4473-7.

170. Clore, G.M., et al., *Determination of Three-Bond 1 H3'–31 P Couplings in Nucleic Acids and Protein–Nucleic Acid Complexes by QuantitativeJCorrelation Spectroscopy.* Journal of Magnetic Resonance, 1998. **134**(1): p. 164-167.

171. Trantírek, L., et al., *Determination of the glycosidic torsion angles in uniformly 13C-labeled nucleic acids from vicinal coupling constants 3JC2/4−H1' and 3JC6/8−H1'.* Journal of Biomolecular NMR. **23**(1): p. 1-12.

172. Sklenar, V. and A. Bax, *Measurement of proton-phosphorus-31 NMR coupling constants in double-stranded DNA fragments.* Journal of the American Chemical Society, 1987. **109**(24): p. 7525-7526.

173. Hosur, R.V., et al., *Solution structure of d-GAATTCGAATTC by 2D NMR. A new approach to determination of sugar geometries in DNA segments.* FEBS Lett, 1986. **205**(1): p. 71-6.

174. Furtig, B., et al., *NMR spectroscopy of RNA.* Chembiochem, 2003. **4**(10): p. 936-62.

175. Sklenář, V., et al., *Two-and three-dimensional HCN experiments for correlating base and sugar resonances in 15N, 13C-labeled RNA oligonucleotides.* Journal of Biomolecular NMR, 1993. **3**(6): p. 721-727.

176. Marino, J.P., J.H. Prestegard, and D.M. Crothers, *Correlation of Adenine H2/H8 Resonances in Uniformly 13C Labeled RNAs by 2D HCCH-TOCSY: A New Tool for 1H Assignment.* Journal of the American Chemical Society, 1994. **116**(5): p. 2205-2206.

177. Bax, A., G.M. Clore, and A.M. Gronenborn, *1H・1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins.* Journal of Magnetic Resonance (1969), 1990. **88**(2): p. 425-431.

178. Güntert, P., C. Mumenthaler, and K. Wüthrich, *Torsion angle dynamics for NMR structure calculation with the new program Dyana11Edited by P. E. Wright.* Journal of Molecular Biology, 1997. **273**(1): p. 283-298.

179. Herrmann, T., P. Güntert, and K. Wüthrich, *Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA.* Journal of Molecular Biology, 2002. **319**(1): p. 209-227.

180. Güntert, P., *Automated NMR Structure Calculation With CYANA*. 2004. p. 353-378.

181. Flinders, J. and T. Dieckmann, *NMR spectroscopy of ribonucleic acids.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2006. **48**(2): p. 137-159.

182. Pulay, P. and G. Fogarasi, *Geometry optimization in redundant internal coordinates.* The Journal of Chemical Physics, 1992. **96**(4): p. 2856-2860.

183. Fogarasi, G., et al., *The calculation of ab initio molecular geometries: efficient optimization by natural internal coordinates and empirical correction by offset forces.* Journal of the American Chemical Society, 1992. **114**(21): p. 8191-8201.

184. Nocedal, J., *Updating Quasi-Newton Matrices with Limited Storage.* Mathematics of Computation, 1980. **35**(151): p. 773-782.

185. Sahakyan, A.B., et al., *Dielectric permittivity and temperature effects on spin–spin couplings studied on acetonitrile.* Magnetic Resonance in Chemistry, 2008. **46**(1): p. 63-68.

186. Zaccari, D., et al., *Solvent Effects on Nuclear Magnetic Resonance 2J(C,Hf) and 1J(C,Hf) Spin–Spin Coupling Constants in Acetaldehyde.* International Journal of Molecular Sciences, 2003. **4**(3): p. 93.

187. Manfred, B., *The electron inside the nucleus: an almost classical derivation of the isotropic hyperfine interaction.* European Journal of Physics, 2000. **21**(1): p. 19.

188. Fermi, E., *Über die magnetischen Momente der Atomkerne.* Zeitschrift für Physik. **60**(5): p. 320-333.

189. Nilges, M., et al., *Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin.* Journal of molecular biology, 1997. **269**(3): p. 408-422.

190. Güntert, P., K.D. Berndt, and K. Wüthrich, *The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination.* Journal of Biomolecular NMR. **3**(5): p. 601-606.

191. Nilges, M., *Calculation of Protein Structures with Ambiguous Distance Restraints. Automated Assignment of Ambiguous NOE Crosspeaks and Disulphide Connectivities.* Journal of Molecular Biology, 1995. **245**(5): p. 645-660.