Structure and dynamics of protein excited states with millisecond lifetimes

Pramodh Vallurupalli

Abstract | An in-depth understanding of biological processes often requires detailed atomic resolution structures of the molecules involved. However in solution where most of these processes occur the conformation of biomolecules like RNA, DNA and proteins is not static but fluctuates. Routinely used structural techniques like X-ray crystallography, NMR spectroscopy and cryo-electron microscopy have almost always been used to determine the structure of the dominant conformation or obtain an average structure of the biomolecule in solution with very little detailed information regarding the dynamics of these molecules in solution. Over the last few years, NMR based methods have been developed to study the dynamics of these biomolecules in solution in a site-specific manner with the aim of generating structures of the different conformations that these molecules can adopt in solution. One powerful technique is the Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiment, which can be used to detect and characterize protein excited states that are populated for as less as 0.5% of the time with ~0.5–10 millisecond lifetimes. Due to recent advances in NMR pulse sequences and labeling methodology, it is now possible to determine the structures of these transiently populated excited states with millisecond lifetimes by obtaining accurate chemical shifts, residual dipolar couplings (RDCs) and residual chemical shift anisotropies (RCSAs) of these excited states. In these excited states the dynamics of some methyl containing residues can also be studied.

NMR Research Centre, Indian Institute of Science, Bangalore 560012
pramodh@sif.iisc.ernet.in

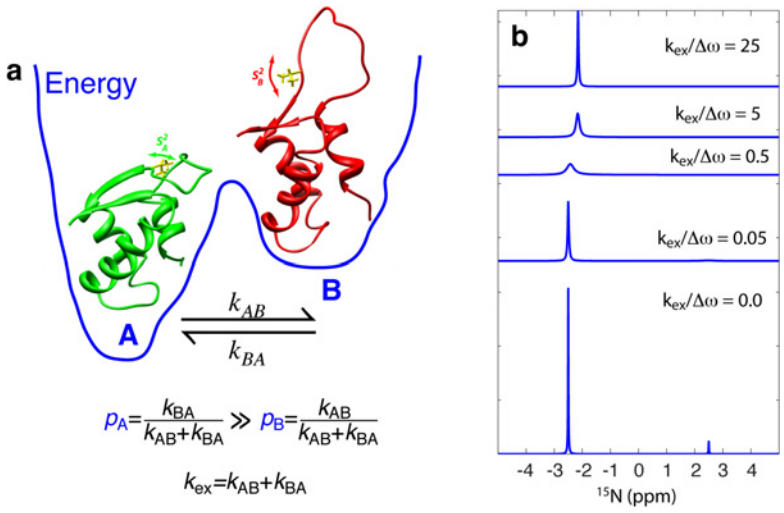
Keywords: Carr-Purcell-Meiboom-Gill (CPMG), Structure, Dynamics, Protein, Residual Dipolar Couplings (RDCs), Residual Chemical Shift Anisotropies (RCSAs), Chemical Shifts

Introduction

Biological macromolecules like proteins and nucleic acids are not rigid but populate a range of conformations [1–3]. These conformational fluctuations occur on a wide range of timescales ranging from picoseconds (ps) to several seconds. Several processes like ligand binding, enzymatic catalysis, protein folding etc occur on the microsecond (μs)-second (s) timescale [3,4]. Early NMR experiments performed on BPTI clearly indicated that the phenylalanine rings present in the molecule undergo rotations, which are coupled to motion in the rest of the protein [5]. Fast dynamics (in the ps-nanosecond (ns)) occurring in each of the states affects the thermodynamics of these processes by affecting the entropy. For example several intrinsically disordered proteins (undergoing fast conformational dynamics) become ordered and fold into a stable tertiary structure upon binding their target [6]. An example of this kind of coupled folding and binding is the pKID domain of the CREB transcription factor, which is disordered when

it is free in solution, but adopts a structure consisting of two helices when it binds the KIX domain of the CBP protein [7]. The entropic penalty during the disorder-order transition is compensated by favorable enthalpic interactions. Here the induced fit disorder-order transition reduces the affinity of pKID for KIX but leads to a complex with high specificity. Low affinity and high specificity means the protein has a low chance of associating with the wrong partner and can readily dissociate after the signaling event is complete [6]. This type of disorder-order induced-fit is not limited to protein-protein complexes but has been observed even in protein-RNA complexes where either the protein or the RNA or both are disordered when they are free and fold upon complex formation [8]. Rather than folding upon binding, the opposite unfolding for binding is observed when talin binds vinculin. Here part of the talin protein unfolds when it forms a complex with vinculin [9]. Any biomolecular conformational dynamics occurring on or slower than the timescale of the processes of interest affects the kinetics of the processes. As proteins and nucleic acids move on a complex energy surface [2] due to the large number of protein and solvent degrees of freedom, the molecule can populate conformations other than the ground state as illustrated in figure 1a using a simplified one-dimensional energy surface. When the energy difference between the ground (major) state and excited (minor) state is $5k_BT$ (k_B is the Boltzman constant and T is the absolute temperature) the population of the excited state will still be ~0.7% of the ground state population. In the case of the hammerhead ribozyme it is not the ground state but such a transiently populated excited state formed by significant conformational rearrangements of the ground state structure that is the active conformation. Thus the rate of exchange between the ground state and the active excited state affects the rate of catalysis [10]. Similarly NMR experiments have demonstrated that the SH3 domain from the Fyn tyrosine kinase (Fyn SH3 domain) folds via a transiently populated intermediate [11]. These examples show that it is not enough to think of proteins and nucleic acids as static structures, but as an ensemble of interconverting conformations. Sometimes as in the case of the hammerhead ribozyme it is such a transiently formed excited state with a low population that is important for function. Single molecule experiments have also clearly shown that the conformational dynamics of enzymes affects their rate of catalysis [12]. Just as the structures of the predominantly populated ground states of proteins and nucleic acids have aided in understanding their function, it is necessary to

Figure 1: (a) A protein undergoing exchange between two states A and B. A is the ground (major) state and B is the excited state (minor) state. In addition to the change in structure there can be a change in the local dynamics. This is illustrated for a valine side chain (shown in yellow), which becomes more flexible in the excited state ($S_A^2 > S_B^2$). The relationship between the populations of the two states and forward and reverse rate constants is also shown. (b) Calculated ^{15}N NMR spectrum (of a site in the protein) for various values of k_{ex}. Here the difference in chemical shifts between the two states $\Delta\varpi = \varpi_B - \varpi_A = 5$ ppm, the field strength is 800 MHz and the minor state population $p_b = 0.07$. The difference in the resonance frequency of the nucleus between the two states A and B $\Delta\omega$ (rad/s), scales with the field strength.



characterize the dynamics and obtain structures of these transiently formed excited states to understand processes like the ones mentioned above. NMR spectroscopy is a particularly powerful technique to characterize the dynamics of biomolecules, as it is sensitive to motion occurring in the entire picosecond-second timescale [4,13]. Further the dynamics can be monitored at potentially every NMR active nucleus in the molecule. Thus by suitable labeling dynamics can be monitored at various hydrogen (^{1}H or ^{2}H), nitrogen (^{15}N), carbon (^{13}C), phosphorus (^{31}P) sites in the molecule. Over the last few years there has been considerable progress in obtaining structures of excited states and ensembles of structures, which represent the conformations that biomolecules adopt in solution [14-19]. These include the use of order parameters to obtain an ensemble of structures that the molecule samples over the ps-ns timescale [14], residual dipolar couplings (RDCs) to obtain an ensemble of structures that the molecule adopts over the ps-μs timescale [15], paramagnetic relaxation enhancement (PRE) [16,17] and Carr-Purcell-Meiboom-Gill (CPMG) experiments to characterize transiently formed structures [18,19]. Some of the recently developed CPMG NMR experiments to obtain structures and characterize the dynamics of the excited states of proteins with millisecond lifetimes are surveyed in this article.

Characterizing excited states using CPMG experiments

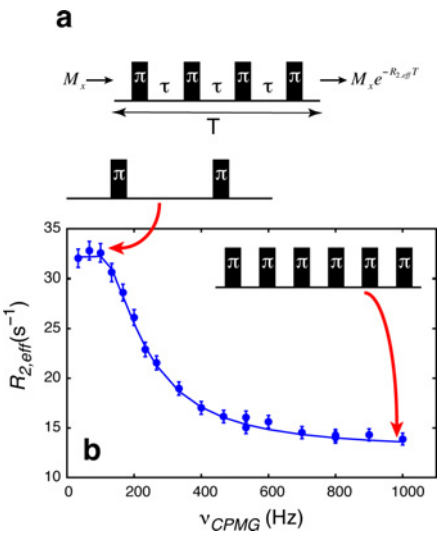
Since the excited states have very low populations they give rise to peaks with very low intensities in the NMR spectrum of the molecule. These weak peaks are further broadened because these excited states are transiently populated. This is illustrated in figure 1b for a ^{15}N nucleus when the minor state (state B) is populated to 7% of the ground state ($p_b = 0.07$). The Bloch-McConnell equations describe the evolution of magnetization under chemical exchange [20,21]. For two state exchange transverse magnetization evolves according to:

$$\frac{d}{dt}\begin{bmatrix} M_A^{\pm} \\ M_B^{\pm} \end{bmatrix} = \begin{bmatrix} -R_{2,A} \pm i\omega_A - k_{AB} & k_{BA} \\ k_{AB} & -R_{2,B} \pm i\omega_B - k_{BA} \end{bmatrix} \times \begin{bmatrix} M_A^{\pm} \\ M_B^{\pm} \end{bmatrix} \quad (1)$$

The spectrum can be obtained by carrying out a Fourier transformation of the total observed magnetization $M^{\pm} = M_x \pm iM_y = M_A^{\pm} + M_B^{\pm}$. k_{AB} and k_{BA} are the forward and reverse rate constants. $R_{2,A}$ and $R_{2,B}$ are the intrinsic relaxation rates for the states A and B. The line-shape is particularly dependent on the ratio $k_{ex}/|\Delta\omega|$, where $\Delta\omega = \omega_B - \omega_A$ (rad/s) is the difference in the resonance frequencies of states A and B. When exchange is slow ($k_{ex} \ll |\Delta\omega|$), we have two peaks at the chemical shifts of the two states ϖ_A and ϖ_B where the heights are proportional to the populations. However when the rate of exchange begins to increase we find that the minor peak broadens much more than the major peak and cannot be detected. Thus the minor state is sometimes called the 'invisible state'. The major state peak on the other hand can be seen even when the minor state peak cannot. Hence all information about the minor state has to be obtained indirectly form the visible major state resonances. When the rate of exchange is in the millisecond (ms) regime and the minor state is populated to 0.5% or higher, CPMG experiments can be used to obtain information about the minor state.

In a CPMG experiment the relaxation rate of transverse magnetization is monitored during the application of a series of π pulses [21, 22]. Transverse magnetization can be either due to single quantum (SQ) [23–25] or multi quantum (MQ) transitions [26–28]. In the constant time CPMG (CT-CPMG) experiment the effective relaxation rate $R_{2,\text{eff}}$ of the visible major state peak is monitored as function of the frequency (ν_{CPMG}) at which refocusing π pulses are applied during the constant time relaxation delay T (figure 2a) [29]. $R_{2,eff}$ is calculated by comparing intensity (I) of the peak of interest after the relaxation delay with that in a reference spectrum (I_0) which does not have the constant time delay T, $R_{2,eff} = -\frac{1}{T}\ln\left(\frac{I}{I_0}\right)$. Chemical exchange leads to a loss of magnetization (dephasing), which broadens the peaks (figure 1b). This loss of magnetization and subsequent broadening of the major state peak due to chemical exchange is quantified in terms of a relaxation rate R_{ex}. The π pulses refocus chemical shift evolution. When the protein jumps between states A and B the resonance frequency of the nucleus changes and refocusing is incomplete when the rate of pulsing is slow (figure 2b). Refocusing is nearly complete even in the presence of chemical exchange when the rate of pulsing is fast (figure 2b). Applying π pulses as shown in figure 2b effectively moves exchange from the slow to the fast regime making $R_{2,eff}$ depend on ν_{CPMG} (figure 2b). The effect of the CPMG sequence on the magnetization can easily be calculated by using the Bloch-McConnell equation (equation 1). The effect of the π pulse is to convert M^+ to M^- (which changes the sign of ω in equation 1).

Figure 2: Schematic description of the constant time CPMG experiment (a) The effective transverse relaxation rate $R_{2,eff}$ is measured as a function of the frequency ($\nu_{CPMG} = 1/(2\tau)$ where τ is the time between two π pulses) at which refocusing π pulses are applied during the constant time delay T. (b) In the presence of chemical exchange $R_{2,eff}$ decreases with increasing ν_{CPMG}.



Analysis of relaxation-dispersion curve using the Bloch-McConnell equations provides information on rate of exchange (k_{ex}), the population of the minor state (p_b) and the magnitude of the chemical shift difference between the major and minor state ($|\Delta\varpi|$). The parameters are obtained by minimizing the function

$$\chi^2(\zeta) = \sum \frac{\left(R^{\text{calc}}_{2,\text{eff}}(\xi) - R^{\text{exp}}_{2,\text{eff}}\right)^2}{\left(\Delta R^{\text{exp}}_{2,\text{eff}}\right)^2}.$$

Here $R^{\text{exp}}_{2,\text{eff}}$ and $\Delta R^{\text{exp}}_{2,\text{eff}}$ are the experimental effective relaxation rates and the associated errors. ξ refers to the various parameters like k_{ex} etc. The calculated relaxation rates $R^{\text{calc}}_{2,\text{eff}}$ are obtained by numerically propagating the Bloch-McConnell equations (11). The effects of the CPMG pulse train can be calculated more accurately by using a more complete basis for the Bloch-McConnell equations which can take the off-resonance effects of the pulses into effect (30).

High resolution structures of the invisible excited state

Solution structures of proteins and nucleic acids are calculated using distance and torsion angle restraints. Distance restraints are obtained from ^{1}H-^{1}H NOE data and the torsion angle restraints are obtained from ^{3}J couplings and in the case of proteins backbone chemical shifts [21, 31]. This data can be supplemented with residual dipolar couplings (RDCs) and residual chemical shift anisotropies (RCSAs), which can significantly improve the quality of the structures [32–34]. Although it is not yet possible to obtain distance restraints for the excited state, it is now possible to obtain backbone chemical shifts, RDCs and RCSAs of the excited state using CPMG experiments. The experiments to obtain this information and their use in the determination of an 'invisible state' structure are described below.

Chemical shifts of the excited states

As described in the previous section the difference in the chemical shift between the ground and excited state can be obtained by analyzing the relaxation dispersion data ($R_{2,eff}$ VS ν_{CPMG}) obtained from CPMG experiments. Ideally these experiments should be performed on an isolated spin to prevent magnetization transfer to other spins. As it is not practical to work with isolated spins experiments have been (and are being) developed to work with coupled spins. CPMG experiments can readily be performed on systems where the J couplings to the spin of interest are refocused by the π pulses (like an A-X spin system). The amide ^{15}N nucleus is not J coupled to other ^{15}N nuclei and the effect of the attached proton can be largely ignored by performing the experiment using the 'Palmer-element' which interconvert's inphase and antiphase ^{15}N magnetization components thus accounting for the differential relaxation effects of the two components [23]. ^{1}H CW decoupling of the amide proton can also be used to generate an isolated ^{15}N nucleus [35]. In the case of the ^{15}N TROSY-CPMG sequence the 'Palmer-element' suppresses cross-relaxation between the Trosy and Anti-Trosy components by inverting one with respect to the other in the middle of the CPMG period [36]. Labeling strategies have also been used to generate the desired spin systems [37]. Amide proton ^{1}HN CPMG experiments can be performed in protein samples where the non-exchangeable protons have been replaced by ^{2}H [25]. Overexpressing protein in *E. coli* grown on [2-^{13}C]-glucose as the sole carbon source generates isolated ^{13}Cα sites (without ^{13}C enrichment at carbonyl and Cβ positions) for seventeen amino acid types [38]. Isolated ^{13}CO groups can be generated using [1-^{13}C]-pyruvate and [^{13}C]-bicarbonate as the carbon sources [24]. Carbonyl CPMG experiments can also be performed on uniformly ^{13}C enriched proteins using selective pulses as the carbonyl resonance frequencies are well separated from that of other carbons [39]. It is not yet straightforward to generate isolated ^{1}Hα sites.

However it is possible to study the Hα sites using a protein sample expressed in *E. coli* grown in 50% D_2O with [$^2H_7$,$^{13}C_6$] glucose as the carbon source. This results in protein where 50–88% of Hβ protons are replaced by ^2H [40, 41]. The effects of the residual Hα-Hβ couplings are suppressed by a pulse sequence element placed in the middle of the CPMG sequence [40]. Proteins with isolated $^{13}C\beta$ sites have been generated by overexpressing the protein in an *E. coli* strain where the enzyme succinate dehydrogenase has been knocked out, using [1-^{13}C]-glucose/Na$H^{12}CO_3$ or [2-^{13}C]-glucose as the sole carbon sources [42]. ^{13}C CPMG experiments can be performed to study the dynamics of methyl groups in Ile, Leu and Val residues using suitably labeled proteins [43]. Although it is still not possible to study the dynamics at all the sites using a single sample, three appropriately labeled samples [37] are enough to study the dynamics at the amide ^1HN, ^{15}N, $^{13}C\alpha$, ^{13}CO and $^1H\alpha$ sites and study the sidechain dynamics in Asn and Gln residues [44]. Chemical shifts of these five backbone sites are useful to obtain information on the backbone torsion angles ϕ and ψ [45, 46].

To obtain accurate estimates of $\Delta\varpi$ the CPMG experiments described above are usually recorded at two well separated fields (for example 500 and 800 MHz) to deconvolute the $|\Delta\varpi|$ from other kinetic parameters like k_{AB} and k_{BA}, as none of the kinetic parameters change with field strength, but $|\Delta\omega|$ scales linearly. The sign of $|\Delta\varpi|$ can be obtained by comparing the peak positions of the major state peak in HSQC spectra recorded at two fields or in a pair of HSQC/HMQC spectra recorded at a single field [47]. The sign can be obtained by comparing HSQC spectra at two fields because the visible major state peak moves more towards the minor state peak at lower fields [47]. Under certain conditions that are usually satisfied the major state peak in a HSQC is shifted towards the minor state compared to its position in a HMQC [47]. Zero and double quantum CPMG dispersions can be used to obtain sign information of one nucleus relative to the other [48]. The sign of the amide proton ^1HN is usually obtained by performing ZQ/DQ dispersions as the sign of the amide ^{15}N can be obtained using HSQC/HMQC methods [48]. With $|\Delta\varpi|$ available from CPMG experiments and the sign information also available the excited state chemical shifts (ϖ_B) can be calculated ($\varpi_B = \varpi_A + \Delta\varpi$) as the ground state shift (ϖ_A) can be measured directly. Accurate estimates of the excited state chemical shifts are required as they are used to obtain restraints for structure calculation. The accuracy of the CPMG experiments in determining the excited state shifts has been tested using the SH3 domain from yeast Abp1p (P) and a 17 residue target peptide (L) from the Ark1p protein which associate in a simple two state manner ($P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$) where exchange

Figure 3: The spectrum of 6.8% bound Abp1p sample (a) is similar to the 0% bound (free) sample (b) and not the 100% bound sample (c). (d) Comparison of the change in ^{15}N chemicals shifts of backbone amide obtained using ^{15}N CPMG experiments and directly measured values in the case of the Abp1p-Ark1p system. (a–c) adapted from [30].

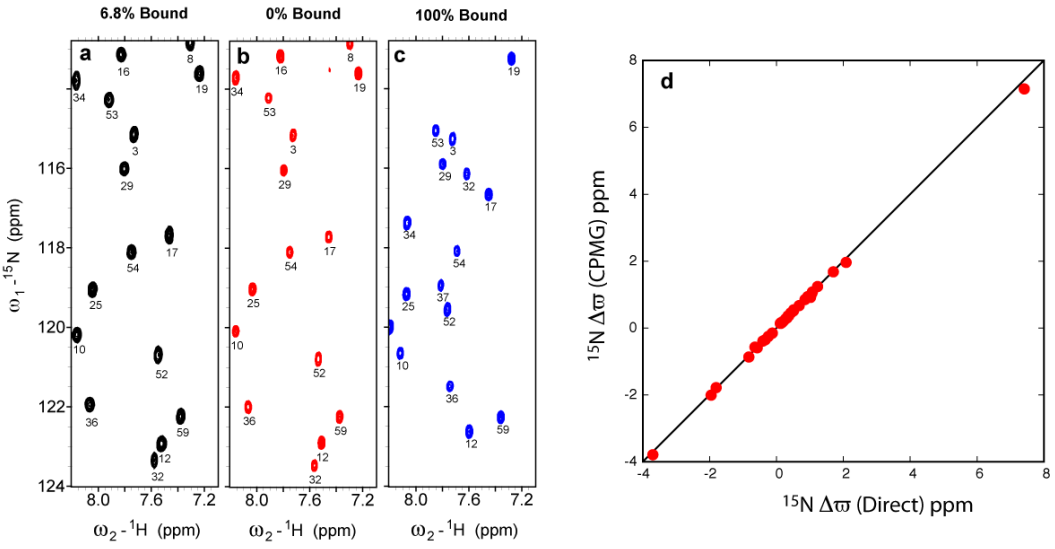
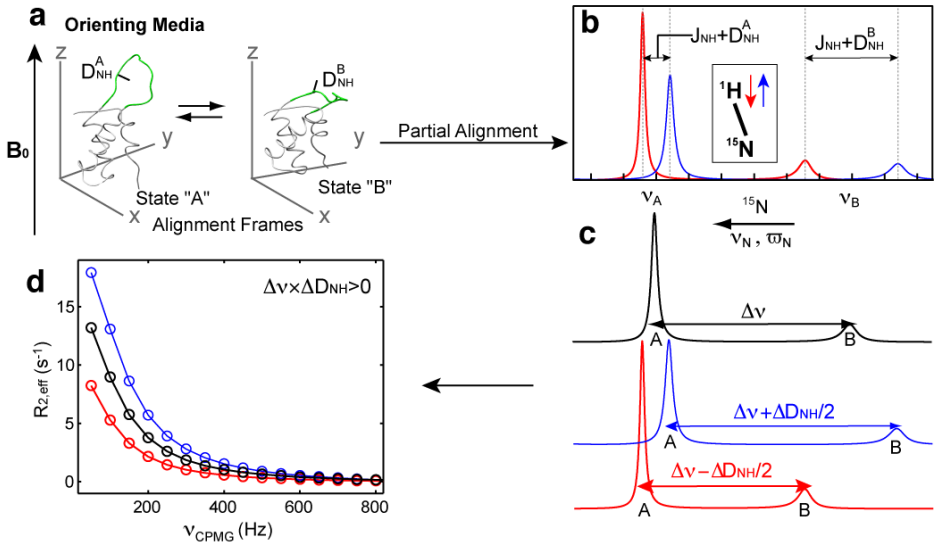


Figure 4: Measurement of ^{1}H-^{15}N RDCs in the 'invisible' excited states of proteins [30]. (a) Under conditions of weak alignment conformational exchange can lead to a change in the orientation of the bond-vector with respect to the alignment tensor, resulting in a change in the dipolar coupling between states A and B. (b) The splitting between the major and minor state peaks is now different. The minor state is shown for the purpose of illustration only. (c) The frequency difference between the major and minor state resonances in the case of ^{1}H decoupling (Black), Trosy (Red) and anti-Trosy (Blue) components are not the same. (d) This gives rise to different ^{15}N CPMG dispersion curves, which can be analyzed to extract the change in the dipolar coupling between the minor and major states ΔD_{NH} and the change in chemical shift between $\Delta\nu$ the two states. Figure adapted from [30].



is in the millisecond time scale. When a small amount of peptide (~7%) is added to the protein the protein:peptide complex is the 'invisible state' (Figure 3a–c). The chemical shifts of various sites in the bound form (complex) can be determined using CPMG experiments. These values can then be compared to values directly measured in a second sample with saturating amounts of peptide added so that almost all the protein bound to the peptide and the bound state is visible. As illustrated in the case of the amide ^{15}N sites (Figure 3d) the chemical shifts of the excited state can be determined extremely accurately (RMSD of 0.03 ppm for ^{15}N, 0.02 ppm for ^{13}CO, 0.04 ppm for ^{13}Cα and 0.01 ppm for ^{1}HN) using the CPMG based methods.

Residual dipolar couplings (RDCs) of the excited state

Under the secular approximation the dipolar coupling Hamiltonian H_{IS}^{D} between two nuclei I and S is given by [21, 32, 33]:

$$H_{IS}^{D} = -\left(\frac{\mu_o}{4\pi}\right)\frac{\gamma_I\gamma_S h}{2\pi^2} I_z S_z \frac{P_2(\cos\theta)}{r_{IS}^3}$$

where γ_I and γ_S are the gyromagnetic ratios of the nuclei I and S respectively. r_{IS} is the distance between the two nuclei and θ is the angle between the internuclear vector and the magnetic field. In the cases considered here the internuclear distance r_{IS} is fixed. In solution θ is continuously varying due to molecular tumbling and only the average effect of H_{IS}^{D} is observed. In isotropic solution H_{IS}^{D} averages to zero and the dipolar Hamiltonian does not affect the positions of peaks in the spectrum to a first approximation. In anisotropic media $\langle P_2(\cos\theta)\rangle \neq 0$, and H_{IS}^{D} adds a term just like the J coupling between the I and S nuclei. The magnitude of the dipolar coupling D_{IS} depends both on the degree of alignment of the protein D^A, and the orientation of the I-S vector in this global alignment frame, with $D_{IS} = D_{IS}^{A}\left[\left(3\cos^2\theta - 1\right) + \frac{3}{2}R\sin^2\theta\cos 2\varphi\right]$ [32, 33]. Here R is the rombicity of the alignment tensor and θ and φ describe the orientation of the I-S vector in the global alignment frame. Thus measuring dipolar couplings between various nuclei provides detailed structural information [32, 33].

The idea behind measuring one-bond ^{1}H-^{15}N dipolar couplings in the excited state is illustrated in figure 4 [30, 49]. When the ^{15}N CPMG experiment is performed in a (^{1}H) spin-state selective manner the difference in frequency between the ground

Figure 5: Measurement of the 'invisible' bound state RDCs in the Abp1p-Ark system [30]. (a,b) Disperson curves recorded at 25°C, 800 MHz and alignment was obtained using Pf1 phage [52]. (a) For D35 ΔD_{NH} is small and the Trosy and anti-Trosy curves are nearly identical. (b) In the case of Y8 ΔD_{NH} is substantial and the Trosy and anti-Trosy curves differ. (c) A comparison of CPMG derived bound form RDCs and directly measured bound form RDCs shows that CPMG derived RDCs are accurate. Figure adapted from [30].

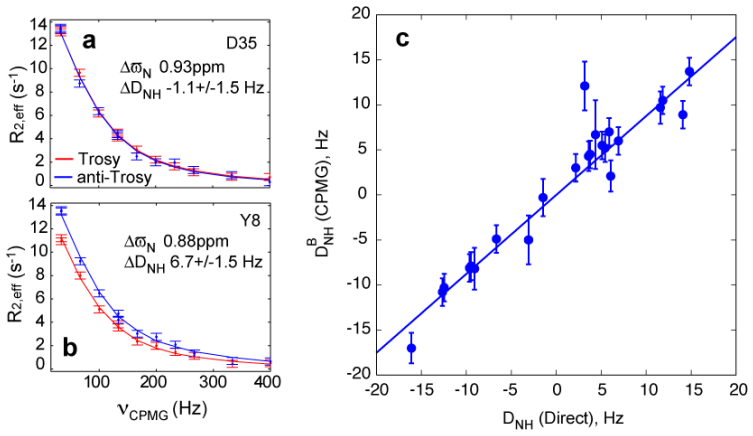
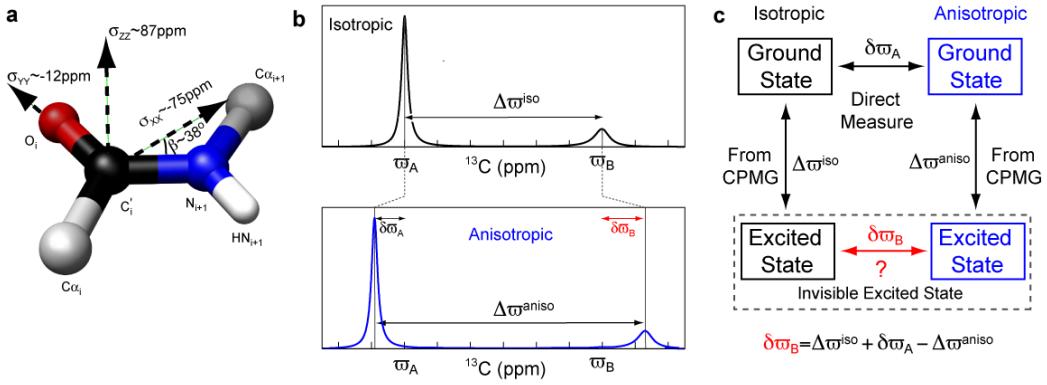


Figure 6: Measurement of excited state RCSA shifts using CPMG experiments [53]. (a) The orientation of the carbonyl CSA tensor is shown with respect to the protein backbone [54]. (b) The positions of the carbonyl peaks change upon alignment. RCSA shift $\delta\varpi_B$ contains information about the excited state conformation. The minor state peaks are usually not visible. (c) The 'chemical shift cycle' to obtain $\delta\varpi_B$. Two CPMG experiments are performed one with alignment and one without alignment. Figure adapted from [53].



and excited state for the TROSY and anti-TROSY components is $\left|\Delta\nu - \frac{\Delta D}{2}\right|$ and $\left|\Delta\nu + \frac{\Delta D}{2}\right|$. Here $\Delta\nu$ is the change in chemical shift between the ground and excited state in hertz and ΔD is the change in the RDC between the ground and excited state. As the ground state RDCs can be measured directly, the excited state RDCs can be determined, so long as ΔD can be accurately determined using CPMG experiments. For the RDCs to be useful for structure determination the errors must be small (∼10%) compared to the range of the observed RDC values. Since the excited state RDCs are going to be obtained from CPMG experiments and not measured directly it is advantageous to increase the range of the RDCs by obtaining the highest possible degree of alignment. This has to be done without compromising the quality of the spectrum. Long-range proton–roton RDCs reduce the quality of the spectrum. ^{2}H labeling of the protein samples significantly reduces the effects of proton–proton RDCs and the samples can be aligned so that ^{1}H-^{15}N one-bond RDCs lie in the ±25Hz range [30]. Proton

'spin-flips' interconvert the TROSY and anti-TROSY components, which reduces the accuracy of the experiments, ^{2}H labeling also increases the accuracy by reducing the ^{1}H 'spin-flip' rates. In addition to performing the two spin state selective CPMG experiments; it is also advantageous to perform the ^{15}N CPMG experiment with ^{1}H decoupling [35]. As this experiment is sensitive only to $\Delta\nu$ and not to ΔD, it complements the spin-selective experiments and helps in obtaining more accurate ΔD values [30]. Using data from all three experiments recorded at two field strengths extremely accurate ($\sim \pm 1.5$ Hz) ^{1}H-^{15}N RDC values could be obtained for the 'invisible state' (bound state) in the Abp1p SH3-Ark1p system (Figure 5), even though the minor state peaks are more than 100Hz wide! Using appropriate labeling schemes it is now possible to accurately measure $^{1}H\alpha - ^{13}C\alpha$ [50], two bond ^{1}HN-^{13}CO [50] and methyl ^{1}H-^{13}C [51] RDCs in the excited states of proteins.

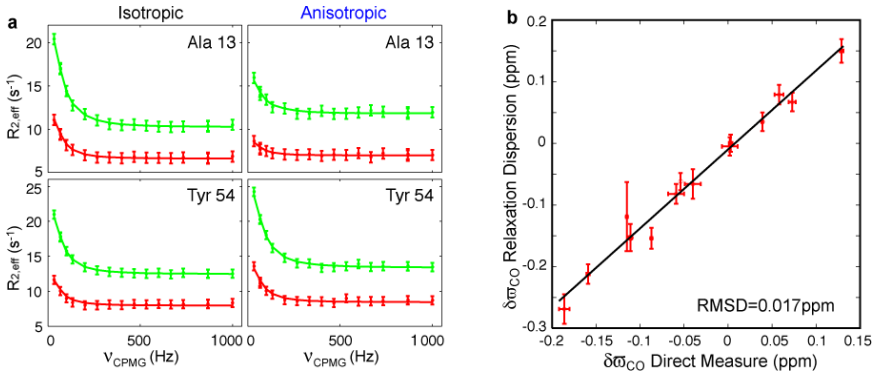
Residual chemical shift anisotropies (RCSAs) of the excited state

The chemical shift of any nucleus depends on the electronic environment around the nucleus. The electronic environment around the nucleus in turn depends on the neighboring nuclei and is not isotropic giving rise to the chemical shift anisotropy (CSA). Thus the chemical shift of the nucleus depends on the orientation of the CSA tensor with respect to the external magnetic field. The orientation of the CSA tensor for a backbone carbonyl group shown in figure 6a is largely independent of the backbone conformation. Due to fast molecular tumbling in isotropic solution only the average chemical shift is observed (figure 6b). Under conditions of anisotropic alignment, the averaging is not complete and the observed chemical shift is different from the one observed in the absence of alignment (figure 6b). The change in chemical shift due to incomplete averaging of the CSA tensor is called the residual chemical shift anisotropy (RCSA). As in the case of the RDCs, the RCSAs depends on the orientation of the carbonyl group with respect to the alignment frame of the molecule, it contains information regarding the orientation of the carbonyl group and consequently structural information about the backbone conformation [34]. The RCSAs have to be determined very accurately as under conditions of weak alignment typically used for protein samples carbonyl RCSAs are quite small (± 0.2 ppm). The RCSA shift of the excited state can be determined using the 'chemical shift cycle' shown in figure 6c [53]. The change in the position of the ground state peak can be measured directly by recording spectra in isotropic and anisotropic media to obtain $\delta\varpi_A$. Carbonyl ^{13}C CPMG experiments are performed in isotropic and anisotropic media to obtain $\Delta\varpi^{iso}$ and $\Delta\varpi^{aniso}$. Since three out of the four shifts in the 'chemical shift cycle' are known the RCSAs of the excited state can be obtained as $\delta\varpi_B = \Delta\varpi^{iso} + \delta\varpi_A - \Delta\varpi^{aniso}$. Examples of ^{13}CO dispersions from the protein-ligand system are show in figure 7a. There is a change in the size of the dispersions between the data recorded under isotropic and anisotropic conditions (figure 7a) for both Ala 13 and Tyr 54 as $\Delta\varpi^{iso} \neq \Delta\varpi^{aniso}$. Very accurate RCSAs for the excited state (± 0.02 ppm) could be obtained for the Abp1p-Ark1p system (figure 7b). The errors in excited state RCSA shifts are small (± 0.02 ppm) compared to the range of the RCSAs ($\sim \quad \pm 0.2$ ppm) and are accurate enough to be useful for structure determination.

Structures of the excited states

CPMG experiments have typically been used to characterize the thermodynamic and kinetic parameters of the processes of interest and qualitative structural information about the excited state has been obtained indirectly [55–59]. Chemical shifts of the excited states were interpreted in a qualitative manner, to identify parts of the protein that were becoming disordered etc. However it is necessary to obtain atomic resolution three dimensional structures of the excited states to understand how the protein functions. This has been demonstrated recently with Abp1p SH3-Ark1p system [18]. CPMG derived chemical shifts, RDCs and RCSAs (recorded with Pf1 phage [52] used as an alignment media) shifts were used to obtain the structure of the 'invisible' bound form in the Abp1p-Ark1p system. The CPMG derived $\Delta\varpi$ values for the backbone ^{1}HN, ^{15}N, $^{13}C\alpha$ and ^{13}CO sites were used to identify regions of the protein undergoing a structural change upon binding the peptide. Based upon large changes in $\Delta\varpi$, three segments of the protein that were possibly undergoing a change in structure upon binding the peptide were identified. Random structures for these three regions with the rest of the molecule fixed to the crystal structure of the free form were generated using high temperature MD simulations (Figure 8a). The 'invisible excited state' structures were then calculated using a restrained MD protocol with RDC, RCSA, torsion angle restraints derived using TALOS [45] from the backbone chemical shifts of the excited state (Figure 8b). Since these structures have been calculated without using any direct structural information (like NOEs, ^{3}J couplings etc) the structures were cross-validated to test their accuracy. The structures were cross-validated against

Figure 7: (a) Carbonyl ^{13}C dispersion profiles for Abp1p SH3 Ala13 and Tyr54 from the ligand binding system described in the text. Data recorded at 500 and 800 MHz are shown in red and green respectively. (b) Correlation between the carbonyl RCSA shifts $\delta\varpi_{CO}(=\delta\varpi_B)$ measured directly on the 100% bound sample and measured using the CPMG experiments described on the ∼7% bound sample. Figure adapted from [53].



a second set of ^{1}H-^{15}N RDCs measured using CPMG experiments in the PEG/hexanol media [60] (Figure 8c,d). The final structure calculated with all available CPMG data has an RMSD of 0.43±0.4 Å to the structure of the bound form determined using conventional NMR methods clearly demonstrating that it is possible to determine the structures of protein excited states with millisecond lifetimes.

Dynamics of the excited states

As the excited states studied using the CPMG experiments have millisecond lifetimes, motion occurring on the picosecond-millisecond timescale can affect the properties of these states. For example, is the Fyn SH3 folding intermediate flexible or rigid? Is a part of the intermediate structured while another part is disordered? New experiments developed to address these type questions [61, 62] are described here.

Pico-nanosecond timescale dynamics of methyl containing side chains

The intrinsic transverse relaxation rate $R_{2,\mathrm{int}}$, of a state contains information regarding the dynamics at that site [4, 13]. R_1, R_2 and NOE are routinely used to study ps-ns dynamics of the ground state [4, 13, 21, 63, 64]. In the macromolecular limit when a site becomes more flexible in the ps-ns timescale $R_{2,\mathrm{int}}$

Figure 8: Three dimensional structure of the 'invisible' state derived from CPMG data [18]. (a) Starting structures with random structures for regions undergoing chemical exchange. (b) Ten final structures calculated using TALOS derived torsion angle restraints using the ^{1}HN, ^{15}N, $^{13}C\alpha$ and ^{13}CO shifts and ^{1}H-^{15}N, $^{1}H\alpha-^{13}C\alpha$, ^{1}H-^{13}CO RDCs and ^{13}CO RCSAs measured using Pf1 alignment media. (c) The calculated structures (b) cross-validated against ^{1}H-^{15}N RDCs recorded using CPMG experiments in PEG/hexanol media. (d) The ^{1}H-^{15}N RDCs measured in the PEG/hexanol media differ considerably from the RDCs measured in the Pf1 media. Figure adapted from [18].

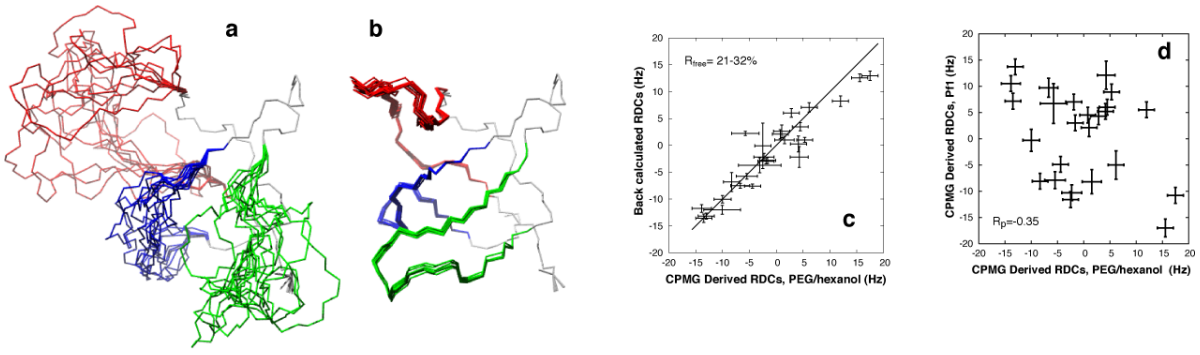
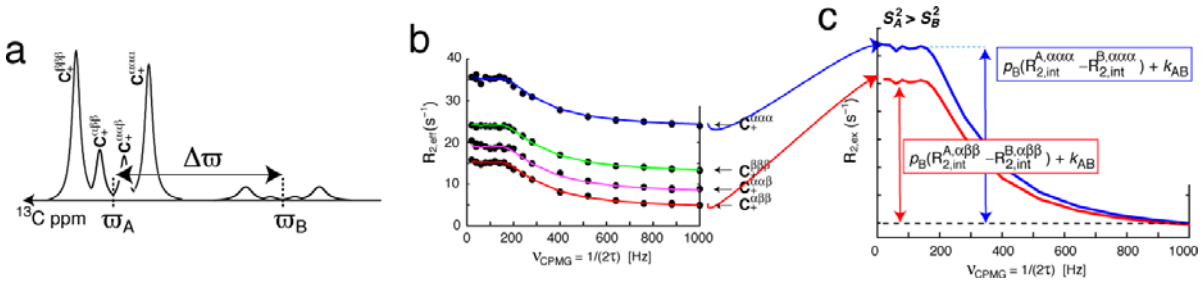


Figure 9: Studying the ps-ns timescale motions of methyl containing sidechains in the excited states of proteins [61]. (a) ^{13}C NMR spectrum of a methyl group undergoing exchange. $\alpha\alpha\alpha$, $\alpha\alpha\beta$, $\alpha\beta\beta$ and $\beta\beta\beta$ refer to the spin-states of the three attached protons. The minor state is shown for the purpose of illustration only. (b) Spin-state selective CPMG dispersions are recorded for the four lines. The CPMG curves are displaced from one another because the intrinsic relaxation rate for each of the components is different. (c) Differences between the (four) dispersion profiles provide information regarding the ps-ns dynamics of the excited state. Figure adapted from [61].

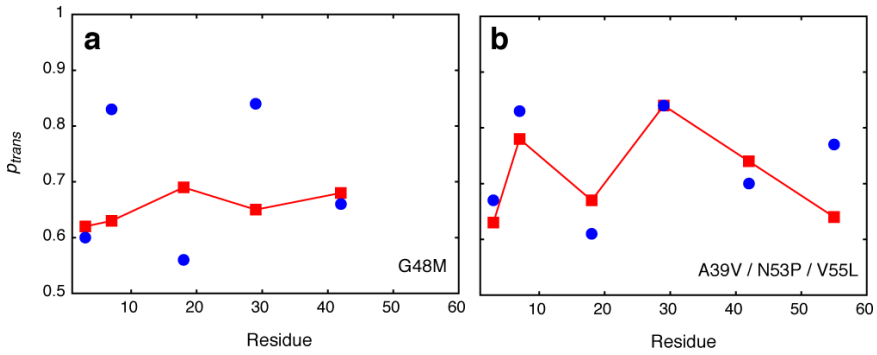


decreases. For a site in slow exchange the major peak relaxation rate $R_2^A = R_{2,\text{int}}^A + k_{AB}$, where $R_{2,\text{int}}^A$ is the intrinsic transverse relaxation rate of state A and it contains no information regarding state B. However when the site is in fast exchange the average peak relaxation rate $R_2 = p_A R_{2,\text{int}}^A + p_B R_{2,\text{int}}^B$ contains information regarding the minor state. Now when the relaxation rate for a site in the slow exchange limit is monitored in a CPMG experiment, exchange starts in the slow exchange regime when pulsing is slow and moves to the fast exchange regime under rapid pulsing. Thus the size of the total dispersion is $k_{AB} + p_B\left(R_{2,\text{int}}^A - R_{2,\text{int}}^B\right)$ not just k_{AB}. Thus the CPMG relaxation dispersion curve contains information regarding the change in the intrinsic relaxation rate between states A and B but it is not straightforward to deconvolute this information from all the other parameters (particularly k_{AB} from $R_{2,\text{int}}^A - R_{2,\text{int}}^B$) using a single CPMG curve [65]. This problem can be solved by performing spin state selective CPMG experiments when the different components have different $R_{2,\text{int}}$ values [61]. The scheme to extract the change in $R_{2,\text{int}}$ is illustrated in figure 9 for the case of a methyl group which becomes more flexible in state B (as in figure 1a). Although $R_{2,\text{int}}$ rates of all four components depends on the same spectral density functions, the $R_{2,\text{int}}$ values for the four components differ due to dipole-dipole and dipole-CSA, cross-correlated relaxation effects [66,67]. The four curves are reporting on the same exchange processes, but when there is a change in dynamics between the ground and excited state, the four curves will differ from each other because $R_{2,\text{int}}^A - R_{2,\text{int}}^B$ is different for each of the curves. This is illustrated for methyl groups when a ^{13}C CPMG experiment is performed on the coherences $C_+^{\alpha\alpha\alpha}$ and $C_+^{\alpha\beta\beta}$ where the protons are in the $\alpha\alpha\alpha$ and $\alpha\beta\beta$ state respectively with $R_{2,\text{int}}^{\alpha\alpha\alpha} > R_{2,\text{int}}^{\alpha\beta\beta}$ (figure 9b,c). Hence by comparing the small differences between the four curves it is possible to deconvolute k_{AB} from $R_{2,\text{int}}^A - R_{2,\text{int}}^B$. This experiment has been used to study the Ile, Leu and Val residues in the 'invisible' unfolded form of the Fyn SH3 G48M protein. As expected methyls in the unfolded state are more flexible by a factor of aproximately two [61].

Pico-millisecond timescale dynamics of Leucine sidechain

The leucine sidechain χ_2 dihedral angle samples two conformations gauche + and trans. Interestingly the difference in ^{13}C chemical shifts of the two methyl groups ($\delta 1$ and $\delta 2$) is related to populations of the gauche+($p_{\text{gauche+}}$) and trans (p_{trans}) conformations with $\Delta\delta(^{13}\text{C}) \approx \delta(^{13}\text{C}_{\delta 1}) - \delta(^{13}\text{C}_{\delta 2}) = -5\text{ ppm} + 10\text{ ppm} \times p_{\text{trans}}$ and $p_{\text{gacuhe+}} + p_{\text{trans}} = 1$ [68]. Since these chemical shifts of the excited states can be obtained accurately for the Leu δ sites, the conformational preferences of Leu residues around the χ_2 dihedral angle can easily be estimated [62]. The method has been applied to study the folding of the Fyn SH3 domain. In the case of the G48M mutant the unfolded state which is populated to ~3.5% can be studied by the CPMG methods. In the A39V/N53P/V55L triple mutant a folding intermediate populated to ~1.9% is studied using CPMG experiments [69]. In both the cases the ground state is the folded form of the protein. In the case of the G48M mutant the unfolded form has a uniform $p_{\text{trans}} \sim 65\%$ although p_{trans} varies from 59 to 84% for the folded from (Figure 10a). This is not surprising as the protein environment surrounding the different methyl groups in the folded molecule is different, while it is expected to be relatively uniform for the unfolded form. In the folding intermediate

Figure 10: Probing the conformational preferences of the Leucine χ_2 dihedral angle in the excited states of proteins using CPMG experiments. The gauche- conformation is almost never populated. (a,b) Conformational preferences of Leu residues in Fyn SH3 G48M and the triple mutant obtained from the Cδ1 and Cδ2 shifts. The conformational preferences of the excited states obtained by analyzing CPMG data is shown in red and the ground state preferences are shown in blue. The excited state in the case of the triple mutant is a folding intermediate.



however there is a significant change in p_{trans} only for Leu 55 (Figure 10b), suggesting that the protein environment around only Leu 55 is different in the intermediate when compared to the ground state.

Conclusions and future prospects

Due the advances in both labeling schemes and pulse sequences over the last ten years it is now possible to use CPMG experiments to study the μs-ms dynamics in proteins at almost all the backbone sites [23–25, 36, 37, 40, 42, 70] and the sidechains of methyl containing [43], aromatic [71], Asn and Gln residues [44]. The spectrum of the excited state can be reconstructed very accurately [24]. In addition anisotropic parameters like RDCs [30, 50, 51] and RCSA shifts [53] of the excited states can also be accurately measured using CPMG experiments. In the case of a test system the chemical shift, RDC and RCSA information obtained from CPMG experiments was used to obtain the structure of an 'invisible state' of a small SH3 domain [18].

There is still a significant need for the development of new (CPMG) experiments to study the 'invisible' excited states of biomolecules. The accuracy of anisotropic parameters (RDCs and RCSAs) extracted from the CPMG experiments described here has not been tested for larger proteins and three state (or higher order) exchange. However it is now possible to obtain structures of proteins (∼15 kDa) [72–74] and backbone dynamics [75] solely from chemical shifts which opens up the possibility of obtaining structures of the excited states for larger proteins (upto ∼15 kDa) and characterizing the backbone dynamics of excited states even in the presence of three state (or higher order) exchange. So far most of the μs-ms timescale dynamics studies of biomolecular dynamics using NMR have been limited to proteins. New CPMG experiments and labeling methods [76] have to be developed to study the μs-ms dynamics of nucleic acids. Unlike proteins where the correlation between backbone chemical shifts and structure is well established, chemical shifts have only been correlated to sugar pucker in the case of RNA [77]. Several excited states of proteins are expected to be highly dynamic [58]. Hence it is necessary to combine all the CPMG derived structural (chemical shifts, RDCs, RCSA) and dynamics information to generate ensembles of structures for excited state, which captures the ps-ms dynamics of the excited state with a millisecond lifetime. This will require combining CPMG derived experimental data with computational techniques to generate structures of the excited state as has been done recently for the 'visible' state in solution [14–16].

Despite the challenges mentioned above systems involving folding intermediates [69, 78], binding intermediates [7], excited states involved in enzyme catalysis [56,59,79] can be investigated with the available methods. Although there has to be considerable methodological developments before it becomes routine practice to determine structures of the 'invisible' excited states of biomolecules, structures of interesting excited states should be available in the next few years.

Acknowledgements

The author would like to thank D Flemming Hansen (University of Toronto) for critically reading the manuscript and providing some of the figures. All

the experiments were performed in the laboratory of Prof Lewis E. Kay (University of Toronto).

Received 13 November 2009; accepted 23 December 2009.

References

1. M. Karplus, *Journal of Physical Chemistry B* **104**, 11 (2000).
2. H. Frauenfelder, S. G. Sligar and P. G. Wolynes, *Science* **254**, 1598 (1991).
3. K. Henzler-Wildman and D. Kern, *Nature* **450**, 964 (2007).
4. A. G. Palmer, *Chemical Reviews* **104**, 3623 (2004).
5. G. Wagner, A. DeMarco and K. Wuthrich, *Biophys Struct Mech* **2**, 139 (1976).
6. H. J. Dyson and P. E. Wright, *Nat Rev Mol Cell Biol* **6**, 197 (2005).
7. K. Sugase, H. J. Dyson and P. E. Wright, *Nature* **447**, 1021 (2007).
8. J. R. Williamson, *Nat Struct Biol* **7**, 834 (2000).
9. G. C. Roberts and D. R. Critchley, *Biophys Rev* **1**, 61 (2009).
10. K. F. Blount and O. C. Uhlenbeck, *Annu Rev Biophys Biomol Struct* **34**, 415 (2005).
11. D. M. Korzhnev, et al., *Nature* **430**, 586 (2004).
12. W. Min, et al., *Acc Chem Res* **38**, 923 (2005).
13. A. Mittermaier and L. E. Kay, *Science* **312**, 224 (2006).
14. K. Lindorff-Larsen, R. B. Best, M. A. Depristo, C. M. Dobson and M. Vendruscolo, *Nature* **433**, 128 (2005).
15. O. F. Lange, et al., *Science* **320**, 1471 (2008).
16. G. M. Clore and J. Iwahara, *Chem Rev* **109**, 4108 (2009).
17. C. Tang, J. M. Louis, A. Aniana, J. Y. Suh and G. M. Clore, *Nature* **455**, 693 (2008).
18. P. Vallurupalli, D. F. Hansen and L. E. Kay, *Proc Natl Acad Sci U S A* **105**, 11766 (2008).
19. D. F. Hansen, P. Vallurupalli and L. E. Kay, *J Biomol NMR* **41**, 113 (2008).
20. H. M. McConnell, *Journal of Chemical Physics* **28**, 430 (1958).
21. J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance and N. J. Skelton, "Protein NMR Spectroscopy, Second Edition: Principles and Practice" Academic Press, 2006.
22. S. Meiboom and D. Gill, *Review of Scientific Instruments* **29**, 688 (1958).
23. J. P. Loria, M. Rance and A. G. Palmer, *Journal of the American Chemical Society* **121**, 2331 (1999).
24. D. F. Hansen, P. Vallurupalli, P. Lundstrom, P. Neudecker and L. E. Kay, *J Am Chem Soc* **130**, 2667 (2008).
25. R. Ishima, P. T. Wingfield, S. J. Stahl, J. D. Kaufman and D. A. Torchia, *Journal of the American Chemical Society* **120**, 10534 (1998).
26. D. M. Korzhnev, K. Kloiber and L. E. Kay, *J Am Chem Soc* **126**, 7320 (2004).
27. K. Kloiber and R. Konrat, *J Biomol NMR* **18**, 33 (2000).
28. J. Dittmer and G. Bodenhausen, *J Am Chem Soc* **126**, 1314 (2004).
29. M. Tollinger, N. R. Skrynnikov, F. A. Mulder, J. D. Forman-Kay and L. E. Kay, *J Am Chem Soc* **123**, 11341 (2001).
30. P. Vallurupalli, D. F. Hansen, E. Stollar, E. Meirovitch and L. E. Kay, *Proc Natl Acad Sci U S A* **104**, 18473 (2007).
31. K. Wuthrich, "NMR of Proteins and Nucleic Acids" Wiley-Interscience, 1986.
32. A. Bax, G. Kontaxis and N. Tjandra, *Methods Enzymol* **339**, 127 (2001).
33. J. H. Prestegard, H. M. al-Hashimi and J. R. Tolman, *Q Rev Biophys* **33**, 371 (2000).
34. R. S. Lipsitz and N. Tjandra, *J Am Chem Soc* **123**, 11065 (2001).
35. D. F. Hansen, P. Vallurupalli and L. E. Kay, *J Phys Chem B* **112**, 5898 (2008).
36. J. P. Loria, M. Rance and A. G. Palmer, *Journal of Biomolecular Nmr* **15**, 151 (1999).
37. P. Lundstrom, P. Vallurupalli, D. F. Hansen and L. E. Kay, *Nat Protoc* **4**, 1641 (2009).
38. P. Lundstrom, et al., *J Biomol NMR* **38**, 199 (2007).
39. P. Lundstrom, D. F. Hansen and L. E. Kay, *J Biomol NMR* **42**, 35 (2008).
40. P. Lundstrom, D. F. Hansen, P. Vallurupalli and L. E. Kay, *J Am Chem Soc* **131**, 1915 (2009).
41. P. Vallurupalli, D. F. Hansen, P. Lundstrom and L. E. Kay, *J Biomol NMR* **45**, 45 (2009).
42. P. Lundstrom, H. Lin and L. E. Kay, *J Biomol NMR* **44**, 139 (2009).
43. N. R. Skrynnikov, F. A. Mulder, B. Hon, F. W. Dahlquist and L. E. Kay, *J Am Chem Soc* **123**, 4556 (2001).
44. F. A. Mulder, N. R. Skrynnikov, B. Hon, F. W. Dahlquist and L. E. Kay, *J Am Chem Soc* **123**, 967 (2001).
45. G. Cornilescu, F. Delaglio and A. Bax, *Journal of Biomolecular Nmr* **13**, 289 (1999).
46. Y. Shen, F. Delaglio, G. Cornilescu and A. Bax, *J Biomol NMR* **44**, 213 (2009).
47. N. R. Skrynnikov, F. W. Dahlquist and L. E. Kay, *J Am Chem Soc* **124**, 12352 (2002).
48. V. Y. Orekhov, D. M. Korzhnev and L. E. Kay, *J Am Chem Soc* **126**, 1886 (2004).
49. T. I. Igumenova, U. Brath, M. Akke and A. G. Palmer, 3rd, *J Am Chem Soc* **129**, 13396 (2007).
50. D. F. Hansen, P. Vallurupalli and L. E. Kay, *J Am Chem Soc* **130**, 8397 (2008).
51. A. J. Baldwin, D. F. Hansen, P. Vallurupalli and L. E. Kay, *J Am Chem Soc* **131**, 11939 (2009).
52. M. R. Hansen, L. Mueller and A. Pardi, *Nat Struct Biol* **5**, 1065 (1998).
53. P. Vallurupalli, D. F. Hansen and L. E. Kay, *J Am Chem Soc* **130**, 2734 (2008).
54. G. Cornilescu and A. Bax, *Journal of the American Chemical Society* **122**, 10143 (2000).
55. P. Vallurupalli and L. E. Kay, *Proc Natl Acad Sci U S A* **103**, 11910 (2006).
56. E. Z. Eisenmesser, et al., *Nature* **438**, 117 (2005).
57. R. Ishima, D. I. Freedberg, Y. X. Wang, J. M. Louis and D. A. Torchia, *Structure* **7**, 1047 (1999).
58. D. M. Korzhnev and L. E. Kay, *Acc Chem Res* **41**, 442 (2008).
59. E. D. Watt, H. Shimada, E. L. Kovrigin and J. P. Loria, *Proc Natl Acad Sci U S A* **104**, 11981 (2007).
60. M. Ruckert and G. Otting, *J Am Chem Soc* **122**, 7793 (2000).
61. D. F. Hansen, P. Vallurupalli and L. E. Kay, *J Am Chem Soc* **131**, 12745 (2009).
62. D. F. Hansen, P. Neudecker, P. Vallurupalli, F. A. Mulder and L. E. Kay, *J Am Chem Soc.* **132**, 42 (2010).
63. L. E. Kay, D. A. Torchia and A. Bax, *Biochemistry* **28**, 8972 (1989).
64. G. Lipari and A. Szabo, *J Am Chem Soc* **104**, 4546 (1982).
65. R. Ishima and D. A. Torchia, *J Biomol NMR* **34**, 209 (2006).
66. L. E. Kay and D. A. Torchia, *J Magn Reson* **95**, 536 (1991).
67. L. E. Kay and T. E. Bull, *J Magn Reson* **99**, 615 (1992).
68. F. A. Mulder, *Chembiochem* **10**, 1477 (2009).
69. P. Neudecker, et al., *J Mol Biol* **363**, 958 (2006).
70. R. Ishima and D. A. Torchia, *J Biomol NMR* **25**, 243 (2003).
71. K. Teilum, U. Brath, P. Lundstrom and M. Akke, *J Am Chem Soc* **128**, 2506 (2006).
72. D. S. Wishart, et al., *Nucleic Acids Research* **36**, W496 (2008).
73. A. Cavalli, X. Salvatella, C. M. Dobson and M. Vendruscolo, *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9615 (2007).

74. Y. Shen, et al., *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4685 (2008).
75. M. V. Berjanskii and D. S. Wishart, *J Am Chem Soc* **127**, 14970 (2005).
76. J. E. Johnson, Jr. and C. G. Hoogstraten, *J Am Chem Soc* **130**, 16757 (2008).
77. B. Furtig, C. Richter, W. Bermel and H. Schwalbe, *J Biomol NMR* **28**, 69 (2004).
78. D. M. Korzhnev, T. L. Religa, P. Lundstrom, A. R. Fersht and L. E. Kay, *J Mol Biol* **372**, 497 (2007).
79. D. D. Boehr, D. McElheny, H. J. Dyson and P. E. Wright, *Science* **313**, 1638 (2006).

Pramodh Vallurupalli is interested in studying the structure and dynamics of biomolecules like RNA and proteins involved in various processes like enzymatic catalysis, molecular recognition, folding etc. Solution NMR spectroscopy is the main technique used to study the structure and dynamics of the molecules involved in these processes. Previously he has worked on the structure of ribosomal RNA using solution NMR in the lab of Prof. Peter B. Moore during his PhD and worked on the development of NMR pulse sequences to study RNA and protein dynamics in Prof. Lewis E. Kay's lab.