Long-lived states in an intrinsically disordered protein domain

L. Fernandes, C. Guerniou, I. Marín Montesinos, M. Pons, F. Kateb, P. R. Vasos

Long-lived states (LLS) are relaxation-favoured eigenstates of J-coupled magnetic nuclei. LLS were measured, along with classical $^1$H and $^{15}$N relaxation rate constants, in aminoacids of the N-terminal Unique domain of the c-Src kinase (USrc), which is disordered in vitro under physiological conditions. The relaxation rates of LLS are a probe for motions and interactions in biomolecules. LLS of the aliphatic protons of glycines, with lifetimes ca. four times longer than their spin-lattice relaxation times, are reported for the first time in an intrinsically disordered protein domain (IDP). LLS relaxation experiments were integrated with 2D spectroscopy methods, further adapting them for studies on proteins.

[a] L. Fernandes, C. Guerniou, F. Kateb, P. R. Vasos
Equipe de RMN
Laboratoire de Chimie et Biochimie Toxicales et Pharmaco logiques, UMR8601.
Université Paris Descartes - CNRS, PRES Paris Sorbonne Citi.
45, rue des Saints-Pères.
75006 Paris (France).
Tel: +33 1 42 86 33 82
Fax: +33 1 42 86 40 50
E-mail: fatih.kateb@parisdescartes.fr, paul.vasos@parisdescartes.fr

[b] I. Marín Montesinos, M. Pons.
Bio NMR Group
Faculty of Chemistry, Department of Organic Chemistry.
University of Barcelona (Spain).
Introduction

The N-terminal region of c-Src kinase (USrc), composed of an SH4 domain (14 residues) and an unique domain (71 residues), has been shown to be intrinsically disordered in vitro.\[^1\] The SH4 domain is responsible for binding to the membrane, which is critical for c-Src activity,\[^2\] whereas the unique domain is involved in the regulation of different processes, e.g., controlling mitosis through phosphorylation of different Ser and Thr residues.\[^3\] The high sequence variability of USrc compared to corresponding domains within the family of Src kinases, corroborated with the high degree to which its sequence is conserved between species, indicate that its amino-acid composition is critical for its function.\[^1\] None of the X-ray structures of c-Src domains solved so far contain the disordered N-terminal domain. For the investigation of unstructured protein domains that are resilient to crystallisation, NMR is well positioned to obtain atomic-resolution information. The motivation of this study is to introduce new NMR experiments designed for disordered proteins and to maximise the amount of information obtained.

Relaxation of spin order involving one or several coupled nuclei can provide a sensitive probe to detect local dynamics reflecting transient secondary structure elements as well as long-range contacts leading to locally compact conformations. Heteronuclear \(^1\)H-\(^{15}\)N relaxation experiments are a widespread and powerful approach to study molecular motions on timescales ranging from picoseconds to milliseconds.\[^4\] Transverse \(^{15}\)N relaxation rate constants (\(R_2\)), longitudinal \(^{15}\)N relaxation rates (\(R_1\)), and heteronuclear {\(^1\)H}-\(^{15}\)N NOE in individual amino acids are sensitive to motions occurring on a time scale range from picoseconds to milliseconds. Heteronuclear relaxation experiments on IDPs reflect the exchange between an ensemble of conformations and provide an estimate of the local autocorrelation time, which is significantly shorter than the one expected for a globular protein with the same number of amino acids.\[^5\]\[^6\]

Nuclear singlet states, discovered in M. Levitt’s group in Southampton,\[^7\] are the prototype of eigenstates with slowly equilibrated populations, which lead to the existence of LLS in the Liouville space. LLS can be created within \(J\)-coupled groups of spins (e.g., protons within an amino acid)\[^7\] either by selective excitation of diastereotopic protons, when the chemical shifts are well resolved, or by using pulse sequences recently developed by M. Levitt and collaborators, in the case of magnetic nuclei with very close chemical shifts.\[^8\]\[^9\] LLS can be sustained by radio-frequency irradiation or shuttling between observation and zero magnetic
Relaxation time constants of long-lived states, $T_{\text{LLS}}$, longer than 40 times the longitudinal relaxation time constant were measured on diastereotopic Gly protons in an Ala-Gly dipeptide, and $T_{\text{LLS}}$ reaching almost half an hour have been found for the nitrogen-15 spins of nitrous oxide. Aliphatic protons in systems with more than two coupled spins, notably serines, were shown to sustain LLS in urea-unfolded proteins. Gly and Ser residues are abundant in IDPs. Ser residues are often functionally important sites of post-translational modifications, and a large number of phosphorylation processes involve serine residues in disordered regions.

While $^{15}$N relaxation typically report on local correlation time of the NH bond vectors, LLS are less affected by dipolar interactions between spins on which they are excited, i.e., in the case of Gly residues, the interpair dipolar coupling. Therefore, LLS open the possibility to sense additional relaxation mechanisms. In the case of long-lived states covering only two spins, relaxation rates are completely insensitive to the intra-pair dipolar relaxation. As a consequence, LLS lifetimes are exquisitely sensitive to structural and dynamic effects arising from electronic or nuclear spins found in their external environment. This makes LLS well-suited potential sensors of both intramolecular or intermolecular (proteins, protein-ligand) interactions. LLS could thus be particularly relevant in the study of IDPs that are intrinsically disordered in solution, but that may fold in cells upon binding to their partners. In addition, since LLS lifetimes reflect the presence of a symmetry forbidden transitions, LLS may reflect access to non-symmetrical environments. Thus, classical relaxation measurements and LLS lifetimes provide potentially complementary information.

We report here the first observation of LLS pertaining to nuclei within an intrinsically disordered protein domain and we introduce 2D experiments to detect LLS in poorly dispersed spectra typical of IDPs. USrc, the protein domain specifically discussed in this manuscript, has been recently shown to participate in interactions that constitute a new layer of regulation of c-Src, which is a known oncogene.

The information that enables function in rigid proteins is encoded in their 3D structure. In contrast, IDP functionality is encoded in the capacity of the unfolded chains to dynamically sample preferentially some regions of conformational space. Residual dipolar couplings and relaxation measurements are physical observables directly sensing protein dynamics. LLS life times provide non-redundant information with respect to the most widely used $^{15}$N relaxation measurements. Increasing the number of experimental observables helps to alleviate the ambiguities arising from dynamic averaging. LLS are sustained by amino acids,
like Gly and Ser, that are very abundant in IDPs and therefore, this study was intended as a feasibility and proof of concept project for the application of LLS to study IDPs. Results and Discussion

$^{15}$N relaxation

Relaxation data for 65 $^1$H-$^{15}$N pairs out of 95 residues could be measured outside spectral overlap regions (Fig. 1). In USrc, like in other IDPs, the $R_2/R_1$ ratio, as well as the $^1$H-$^{15}$N NOE values are typically lower than in globular proteins of comparable molecular weight, reflecting its high local mobility on the relevant timescale (Fig. 1). A rigid globular protein with the molecular weight of USrc would be expected to have a correlation time $\tau_c \sim 15$ ns at 275 K, which would yield an $R_2/R_1$ ratio of $\sim 15$ and an NOE $\sim 0.8$. The flexibility associated with internal motions in this IDP has a sizeable effect on $^{15}$N relaxation rates. The rates are thus similar to the one that would be observed for a globular protein undergoing isotropic rotation, with a $\tau_c \sim 3$ ns, thus confirming the high mobility of all protein segments of USrc.

Since the classical model-free approach developed by Lipari and Szabo is not appropriate for IDPs, NMR $^{15}$N relaxation measurements ($^{15}$N $R_1$ and $R_2$ and $^1$H-$^{15}$N heteronuclear NOE) were analyzed using reduced spectral density mapping. The results, given in Fig. 1 (bottom), show that $J(0)$ and $J(0.87\omega_H)$ are, respectively, lower and higher than expected in the case of the absence of internal motions. This analysis did not point out to particular motions impacting $^{15}$N $R_2/R_1$ ratios in any of the residues in a different manner than the rest, showing similar motional properties along the protein chain. A few residues showed heteronuclear NOE values larger than 0.7, suggesting some residual local order. Two of these residues are Gly65 and Gly66, located in a region previously shown to display RDC features sensitive to the presence of urea. Paramagnetic relaxation enhancement experiments showed transient long-range interactions involving this region and other parts of the molecule. The apparent rigidity of the N- and C-termini on the NOE timescale (Fig. 1) may also reflect the stabilizing effect of long-range interactions within USrc, as already observed by paramagnetic relaxation enhancement experiments.

$^1$H $T_1$ and $T_{\text{LLS}}$ recorded at 275 K

$^1$H Long-lived states have recently been shown to be a good probe for processes such as protein denaturation or ligand-protein interactions. 1D NMR experiments using our
previous method\cite{25} to measure LLS lifetimes were applied on this new system (Fig. 2). Application to IDPs poses new challenges, especially because of the limited range of chemical shifts of their aliphatic protons. Because of signal overlap, Hα assignment, and therefore the chemical shift difference, Δν(Hα1, Hα2) could be unambiguously measured only for Gly28 and Gly66. For most of the other Gly residues, resolution was insufficient to afford stating whether the Hα protons had distinct chemical shifts or were degenerate. Severe overlap of the β protons of Ser residues with each other, as well as with other protons (below 3.8 ppm) prevented the determination of relaxation rates for individual Ser protons, with the exception of Ser89, which is part of the Strep-tag used for purification. Mono-exponential recoveries of longitudinal 1H magnetisation were found at 275 K, with the same relaxation rate constant throughout the spectral region corresponding to Hα protons of Gly residues. There are potential added contributions to intensities observed in this region from Ser Hβ protons, notably around 3.8 ppm. A uniform decay was also observed for the envelope obtained after exciting LLS in the same spectral range. The relaxation time constant were $T_{\text{LLS}} = 1.79 \pm 0.04$ s (Fig. 2) and $T_1 = 0.65 \pm 0.01$ s, i.e., $T_{\text{LLS}}/T_1 \sim 2.7$.

1H LLS created on Ser89, observed free from overlap at the resonances of their beta protons, show similar lifetimes to the Gly protons at 275 K, $T_{\text{LLS}}(\text{Ser}) = 1.74 \pm 0.02$ s. No LLS measurement could be made specifically for other Ser residues due to spectral overlap. No spin states longer-lived than those of Gly could be detected in the chemical shift region including the majority of β-protons of Ser. These results show that the ratio between LLS relaxation and longitudinal recovery time constants at 275 K is promising in terms of sensitivity to local dynamics of the LLS. However, the LLS-derived information along the protein sequence is poorly resolved.

1H $T_1$ and $T_{\text{LLS}}$ recorded at 293 K

In our case, raising the temperature to 293 K led to an improved ratio $T_{\text{LLS}}/T_1 \sim 3.9$ for the Gly resonances, with a longitudinal 1H relaxation time constant at this temperature $T_1 = 0.57 \pm 0.01$ s, and the LLS time constant $T_{\text{LLS}} = 2.23 \pm 0.01$ s. Improved LLS relaxation times were expected, since LLS relaxation rates were previously shown to decrease with decreasing local correlation times in a similar manner to $R_2$ relaxation, whether their relaxation stems from chemical shift anisotropy differences CSA's\cite{26} or from dipolar interactions with a third spin.\cite{27} \cite{28}
Interestingly, the Ser89 LLS lifetime has a steeper temperature dependence than that of Gly residues. Raising the temperature to 293 K causes a large increase in the lifetime of LLS associated to Ser89, from $T_{\text{LLS}}^{275\text{K}}(\text{Ser}) = 1.74 \pm 0.02$ s to $T_{\text{LLS}}^{293\text{K}}(\text{Ser}) = 4.26 \pm 0.04$ s. At 293 K, the ratio $T_{\text{LLS}}(\text{Ser})/T_{\text{I}}(\text{Ser}) \sim 7.4$ is nearly twice as large compared to the one obtained at 275 K (Fig. 3).

Such differences point out that either the motional properties of Ser89 are particularly sensitive to the change in conformational propensities brought about by increasing the temperature, or that its LLS environment is affected by other temperature-dependent factors, such as proton exchange with the solvent. Exchange between hydroxyl and water in the case of Ser residues in a tetrapeptide, yields a $k_{ex}$ of the order of $10^7$ s$^{-1}$ at this pH and around this temperature.$^{29}$ Considering that the effect of OD exchange on the Ser LLS system to which it is coupled ($H^\alpha$, $H^\beta_1$, $H^\beta_2$) occurs via scalar relaxation of the first and second kind,$^{28}$ one would indeed expect an increase in the relaxation time constants, when the temperature is raised to 293 K. Solvent exchange at the amide sites could, in principle, also affect LLS relaxation, but raising the temperature from 275 K to 293 K has almost no effect on Gly $T_{\text{LLS}}$.

This could be due to i) the slow rate constant of the amide proton exchange and ii) the similarity of the two $J$-couplings between the $H^N$ and the two $H^\alpha$ protons.$^{28}$ Alternative explanations for the increased LLS lifetime of Ser89 spins are possible. The observation of distinct LLS lifetimes and their temperature dependency for individual residues highlight the sensitivity of these measurements to the local environment of selected residues. To exploit the potential of LLS lifetime measurements in IDPs in subsequent studies, first the spectral overlap problem has to be solved.

**LLS relaxation experiments detected in two dimensions**

One way to address the overlap issue is to add a second dimension to LLS relaxation experiments. A 2D pulse sequence referred to as LLS-TOCSY was employed, where LLS evolution was coupled to a classical TOCSY evolution in the indirect dimension (Fig. 4a).

This sequence initially excites LLS, which in the case of a three-proton spin system take the form of linear combinations of population differences between singlet and triplet states on each of the spin pairs in the system.$^{15}$ Following this step, a sustaining period, which gives the actual relaxation delay, is introduced. The decoding re-transfers LLS to single-quantum
magnetisation. Depending on the weight of the populations of various pairs of spins in the created LLS\textsuperscript{[15]} and, more importantly, on the bias of the imperfect decoding pulse sequence, the single quantum coherences reporting on LLS relaxation may predominantly belong to one or the other of the three protons involved. In our case, the H\textsubscript{β} protons carry most of the resulting magnetisation, because the shortest and most effective decoding sequence is tuned to the coupling $^{2}\text{j}(\text{H}\textsubscript{α}\textsuperscript{1}, \text{H}\textsubscript{α}\textsuperscript{2})$. The TOCSY transfer step and indirect evolution dimension that follow re-distribute the magnetisation among Ser protons. A strong signal at the H\textsubscript{α} frequency in the indirect frequency dimension and at the H\textsubscript{β} in the direct frequency dimension corresponds to the magnetisation decoded on H\textsubscript{α} protons, as described above, at the outcome of the LLS sequence, and transferred to H\textsubscript{β} prior to the direct detection period. The use of the sequence is to alleviate potential effects of signal overlap in the H\textsubscript{β} region among various amino acids, in case the spectral crowding is less severe in the H\textsubscript{α} region of the spectrum.

This transfer of magnetisation from the H\textsubscript{β} to the H\textsubscript{α} region can also be achieved by the use of 1D TOCSY variant with selective pulses, placing the TOCSY transfer delay after the LLS pulse sequence (XXX REF).

In the particular case of USrc, the signals of Ser\textsuperscript{89} (Fig. 4b) could be better isolated using this TOCSY approach. Whereas on the 1D spectrum, only peaks corresponding to H\textsubscript{β} chemical shifts could be observed, LLS-TOCSY allows for the identification of correlation peaks between Ser H\textsubscript{α} and H\textsubscript{β} that could not be observed in the 1D experiment, thus helping in the assignment. Lack of symmetry with respect to the diagonal reflects unsymmetrical projection of LLS onto (H\textsubscript{α}, H\textsubscript{β}\textsuperscript{1}), (H\textsubscript{α}, H\textsubscript{β}\textsuperscript{2}) pairs on one side, and the (H\textsubscript{β}\textsuperscript{1}, H\textsubscript{β}\textsuperscript{2}) on the other side, as expected.\textsuperscript{[15]}

**Conclusion**

We report the measurement of relaxation time constants of symmetry-adapted spin states with long lifetimes (LLS) in a natural IDP. LLS have the potential to probe intra and intermolecular interactions in these proteins. To overcome limitations stemming from resonance overlap, a 2D version of the LLS experiments is suggested, increasing the applicability of these relaxation measurements to proteins. The temperature dependence of LLS relaxation times provides way to identify residues displaying singular dynamics within IDPs. In light of the long lifetimes that may be observed for Ser residues and the biological relevance of Ser phosphorylation in IDPs, LLS of phosphorylated serines within USrc or other IDPs may
represent a promising tool in the study of the structural and dynamic basis of regulation of eukaryotic signalling proteins through intrinsically disordered domains, provided overlap problems can be overcome. LLS could also be used as a new approach for in-cell studies of known IDPs in vitro, in the absence of their binding partners, to test whether they become structured in cells.

**Experimental Section**

The sequence of the protein domain under study is:

\[\text{MASNKSMPKDAQR} \text{RSLPEAVHGAGGGAFPASQTPSKPSA}\]
\[\text{DGHRGPSAAFAPAAEPKLGGSSDVTSPQRAGPLAGGSAWSHPQFEK}\]

where residues 1-14 (bold italics) belong to the SH4 domain, residues 15-85 to the UD, and the last 10 residues (italics) were introduced for protein purification.

A sample in 80 % D\(_2\)O and 20 % H\(_2\)O was prepared by rehydrating a lyophilized 0.4 mM solution of uniformly \(^{15}\text{N}\)-labelled protein in 50 mM phosphate buffer pH = 7.0. Measurements were carried out in a Shigemi tube at 275 K. The level of deuteration was chosen to optimize the observation of amide-protons while minimising external sources of relaxation for LLS experiments.\(^{10}\) \(^{25}\) This sample was used for resonance assignment and \(^{15}\text{N}\)-relaxation studies. A second sample in pure D\(_2\)O as a solvent but otherwise identical to the first one was used for measurements of LLS lifetimes, which are longer in deuterated solvents.

NMR spectroscopy

All NMR spectra were recorded on a Bruker spectrometer operating at a proton frequency of 500 MHz, equipped with a 5 mm TXI 1H/13C/15N probe, except if mentioned otherwise, as well as on a 950 MHz spectrometer equipped with a cryoprobe.

The \(\text{H}^a\), \(\text{H}^b\) chemical shifts were assigned using standard 3D and 2D TOCSY experiments acquired at 275 K in 80% D2O / 20% H2O. Spectra were processed using XWinNMR / TOPSPIN package.
Proton longitudinal relaxation rates at selected frequencies of Gly-Hα protons at 11.74 T were measured at 275 K and 293 K using a standard inversion-recovery sequence. Eleven time points were recorded with a recycle delay $\delta_1 = 1$ s. A three-parameter exponential was fitted to the peak intensities using a Scilab script.

$^1$H LLS experiments were performed at 11.74 T as described with $\tau_1 = 1/(4 \times 15) = 17$ ms, $\tau_2 = 1/(2 \times 55) = 9$ ms, an irradiation amplitude $\nu_1 = 1$ kHz during the variable evolution delays $\tau_M$, and a relaxation recovery delay $\delta_1 = 5$ s. Six different time-points $\tau_m$ were recorded at 275 K and 293 K.

$^{15}$N NMR relaxation measurements, including longitudinal ($R_1$), transverse ($R_2$) relaxation rates, and the $^{1}$(H)$-^{15}$N heteronuclear Overhauser effect (NOE) measurements were carried out at 275 K and 11.74 T using standard methods. For $R_1$ and $R_2$ relaxation rates, 9 and 7 time-points were recorded respectively in an interleaved manner. A two-parameter exponential decay was fitted to peak intensities using NMRViewJ. In all relaxation experiments, errors were estimated from the intensities of duplicate points.

2D LLS-TOCSY experiments were recorded according to the pulse sequence scheme described in Figure 4a, using 2048 and 144 points in the indirect and direct dimensions respectively. A DIPSI-2 TOCSY sequence was used as a TOCSY spin lock, with a total duration $\tau'_m = 60$ ms. Delays are the same as the one used in 1D experiments (see above). The phase cycle was $\phi_1 = 2(y)$, $2(-y)$, $\phi_2 = x$, $-x$, $\phi_3 = 2(x)$, $2(-x)$, $\phi_4 = 2(-x)$, $2(x)$, $\phi_{rec} = x$, $-x$, $-x$, $x$. Spectra were processed using the package XWinNMR.

**Acknowledgements**

The authors are very thankful to Mariano Maffei for protein expression and purification. We thank Aude Sadet, Nisrine El Hazzat and Gildas Bertho for valuable assistance with these experiments. We are grateful to the TGIR-RMN-THC Fr3050 CNRS and to the EPFL for NMR measurement time. This research was supported by a DIM Analytics / Ile de France contract via a PhD grant to LF, by the Univ. Paris Descartes, the French CNRS, The Spanish MINECO (Bio2010-15683), Generalitat de Catalunya (2009SGR13529 and EU BioNMR (contract 261863).
Keywords: NMR Spectroscopy . $^1$H spin. $^{15}$N spin. Relaxation. Long-lived states. Intrinsically disordered proteins . Kinases.
References

Figure 1. (a) $^{15}$N $R_2/R_1$ relaxation rates ratio (top) and $^1$H-$^{15}$N heteronuclear NOE (bottom) recorded at 500 MHz and 275 K, plotted versus USrc residue numbers. (b) Derived spectral density mapping at the frequencies of $^1$H, $J(0.87\omega_H)$ (top), $^{15}$N, $J(\omega_N)$ (middle), and $J(0)$ (bottom). Values are given in ns.rad$^{-1}$. The dashed lines represent expected values for a rigid, globular protein of the size of USrc.
Figure 2. (Left) 1D spectrum obtained following the excitation and locking of long-lived states during 2 s in Gly residues. (Right) Intensities from inversion-recovery experiments (‘+’) sampled at a frequency corresponding to the region of Gly Hα resonances, and LLS experimental decays corresponding to Gly Hα resonances (black filled circles). Errors are within the symbol size. The corresponding fitting curves are shown as full lines.

Figure 3. LLS decays of Gly residues (circles) and Ser89 (triangles) recorded at 293 K (black filled symbols) and 275 K (black open symbols). Fitting curves are shown as full lines for Gly and dashed lines for Ser residues.
Figure 4. (a) Pulse sequence scheme used for the LLS-TOCSY experiment. The delays used in the LLS build-up block (LLS encoding, sustain time and decoding) are given in the material and methods section (See Sarkar and co-workers\textsuperscript{[24]} for more details) (b) Section of the 2D LLS-TOCSY spectrum corresponding to the region of serine resonances.